THE PRIMER

JGI
JOINT GENOME INSTITUTE
DEPARTMENT OF ENERGY

Winter 2018
volume 15
issue 1

in this issue

JGI gauged the size of a plant pan-genome using *Brachypodium distachyon*, growing here under Estelle Schaefer's gaze. (Marilyn Chung, Berkeley Lab)

Cataloging Candidate Genes for Plant Microbiome Studies

To help meet the food and energy needs of a rapidly growing global population, researchers are focusing on understanding and promoting beneficial plant-microbe relationships that may enhance crop breeding strategies.

In a study published December 18, 2017 in *Nature Genetics*, a team led by researchers at the U.S. Department of Energy (DOE) Joint Genome Institute (JGI), a DOE Office of Science User Facility, and the Howard Hughes Medical Institute (HHMI) at the University of North Carolina at Chapel Hill (UNC) has assembled and exploited a catalog of bacterial genomes to identify and characterize candidate genes that aid bacteria in adapting to plant environments. They were specifically interested in genes involved in bacterial root colonization, because most of the interaction between microbes and plants occurs at the interface between the roots and soil.

"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

continued on page 3

Unlocking Natural Diversity with Pan-genomes

GRETA LORGE

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. Having pan-genomes for plant crops that are important for fuel and food applications would enable breeders to harness natural diversity to improve traits such as yield, disease resistance, and tolerance of marginal growing conditions.

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. 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 and one of the JGI's Flagship Genomes.

"There are a vast

continued on page 3

Genus-Wide Fungal Study Highlights Functional Annotation Approaches

Aspergillus fungi are pathogens, decomposers, and important sources of biotechnologically-important enzymes, including more than 250 carbohydrate active enzymes that break down plant cell walls. The latter are of particular interest to DOE researchers working on the industrial production of sustainable alternative fuels using bioenergy feedstock crops.

In the January 23, 2018 issue of the *Proceedings of the National Academy of Sciences*, a team of researchers at the Technical University of Denmark (DTU), the JGI, and the DOE's Joint BioEnergy Institute (JBEI) at Berkeley Lab, report the first results of a long-term plan to sequence, annotate, and analyze the genomes of 300 *Aspergillus* fungi.

"With the JGI's strategic shift towards functional genomics, this study illustrates several new approaches for functional annotation of genes," said study co-author Igor Grigoriev, head of the JGI Fungal Program. "Many approaches rely on experiments and go gene by gene through individual genomes. Using *Aspergillus*, we're sequencing a lot of closely-related genomes to highlight and compare the differences between genomes."

In the study, the team sequenced and annotated six diverse *Aspergillus* species; four were sequenced using the Pacific Biosciences platform, producing very high quality genome assemblies that can serve as reference strains for future comparative genomics analyses.

"We picked four [species] that were distantly related," said study senior author Mikael R. Andersen of the DTU. "With that diversity comes also chemical diversity, so we were able to find candidate genes for some very diverse types of compounds." Andersen spoke about the project at the 2017 JGI Genomics of Energy & Environment Meeting. Watch his talk at bit./ly/JGI2017Andersen.

—GL

Full story at jgi.doe.gov/aspergillus-all-in-the-family-focused-genomic-comparisons/.

Comparative Genomics Traces Evolution of Armillaria

Clusters of fruiting bodies emerge on and around trees in *Armillaria*-infected areas in the fall. (Virág Tomity)

Often called the humongous fungus, *Armillaria* form some of the planet's largest living organisms. They are also among the most devastating fungal pathogens, capable of infecting more than 500 plant species found in forests, parks and vineyards, and breaking down all of the components of a host plant's cell walls. As part of an international team led by László G. Nagy of the Biological Research Center at the Hungarian Academy of Sciences, JGI researchers helped sequence and analyze four fungal genomes of the genus *Armillaria*. The study published October 30, 2017 in *Nature Ecology & Evolution*.

The team compared the genomes of *A. ostoyae*, *A. cepistipes*, *A. gallica*, and *A. solidipes* against 22 fungal genomes, many of which were previously sequenced and annotated by the JGI, and cataloged 20 gene families related to pathogenicity. They also identified plant cell wall degrading enzymes of potential interest to researchers looking for methods to convert plant biomass into alternative fuels or other bioproducts.

"*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, towards which understanding how they function in nature might be the first step," said study senior author Nagy.

Study co-author and JGI Fungal Program head Igor Grigoriev noted that these were among the first representatives of 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. The fungal genomes are all available on the JGI fungal genomics portal, MycoCosm, along with the fungal genome sequences used for comparison.

Full story at jgi.doe.gov/armillaria-white-rot-fungi-size-explained-gene-families/.

Plant Microbiome

continued from page 1

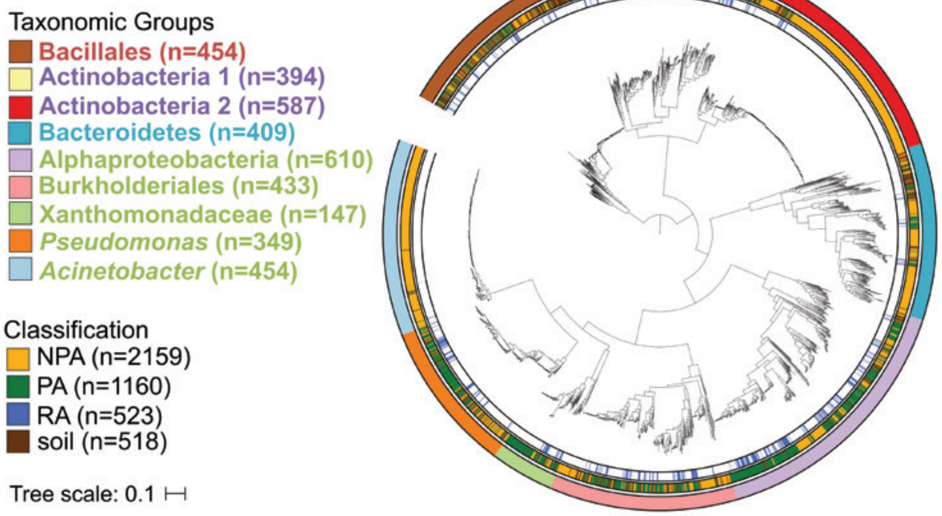
senior author Jeff Dangl, a HHMI investigator and the John N. Couch Professor of Biology at UNC.

Among the key insights gained from the study was that plant- and soil-associated genomes tend to be larger than control genomes from the same clade. This was found to be due in part to enrichment of genes involved in sugar metabolism and transport, likely an adaptation to photosynthesis-derived plant carbon, up to 20 percent of which is exuded through the roots as sugars to attract microbes, said JGI research scientist Asaf Levy, the study's co-first author.

The complete catalog of new genomes and plant-associated genes is available to the research community through a dedicated web portal at the Dangl lab's website called: Genomic Features of Bacterial Adaptation to Plants.

The database offers an unbiased way to identify potentially interesting bacterial genes involved in interaction with a plant—including many totally novel genes. "We are currently experimentally studying the function of many of these genes to gain a better functional understanding of the plant microbiome." Levy said.

—MSB and GL



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

Full story at jgi.doe.gov/functional-genomics-database-for-plant-microbiome-studies/.

Unlocking Natural Diversity

continued from page 1

number of genes that are not captured in a single reference genome," said 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." 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.

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

"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 added.

Additionally, genes found in only a subset of lines displayed faster rates of evolution lay closer to transposable elements 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 the project website, BrachyPan (brachypan.jgi.doe.gov/). The *Brachypodium distachyon* genome is available on the JGI Plant Portal Phytozome (phytozome.jgi.doe.gov/).

Full story at jgi.doe.gov/brachypodium-one-reference-genome-not-enough/.

Trout Bog Lake (Image courtesy of the McMahon Lab)

Virophage Database Doubles from Mining Lake Datasets

Using metagenome data sets collected over several years in northern freshwater lakes, a team led by researchers at Ohio State University and the JGI uncovered 25 novel sequences of virophages, small viruses that live in giant viruses and use their machinery to replicate and spread. 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.

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' 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.

Sequencing and analyzing lake metagenomes from 3-year and 5-year time-series is allowing researchers to identify the community members, determine their metabolic pathways, and follow changes in communities over several years. "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."

Aside from doubling the number of virophages in public databases, the time series allowed Roux and his colleagues to determine whether factors such as the seasons or abundance of particular microbes influenced the viruses' own presence. 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.

To learn more about each of these stories, go to the JGI's News Releases page: jgi.doe.gov/category/news-releases/.

CAMI Challenge Benchmarks 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 high-throughput genomic sequencing of environmental communities ("metagenomics") and the development of computational tools to analyze these data.
To tackle assembling metagenomes into a set of DNA segments that together represent a consensus region of DNA or contigs, 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. Until now, however, there was a lack of consensus on how to evaluate their performance.

In an article published October 2, 2017 in *Nature Methods*, 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, formerly a JGI postdoctoral fellow and now head of the Computational Metagenomics group at Bielefeld University, 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."

The CAMI organizers evaluated computational tools in 3 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 4 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 benchmarking results are available on data.cami-challenge.org/results.

"JGI is very vested in not only benchmarking of lab protocols, but also computational workflows," said JGI Microbial Program head Tanja Woyke. "This makes our participation in critical community efforts such as CAMI so important."

Liverwort Lessons on How Plants Learned to Thrive on Land

Despite their unprepossessing appearance and lack of roots or vascular tissues for nutrient transport, liverworts 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 part of an international team led by researchers at Monash University in Australia, and at Kyoto and Kindai Universities in Japan, JGI scientists analyzed the genome sequence of the common liverwort (*Marchantia polymorpha*). The team identified genes and gene families crucial to plant evolution which have been conserved across plant lineages. The results were reported in the October 5, 2017 issue of *Cell*.

"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 freshwater," said JGI Plant Program head Jeremy Schmutz. "Land plants began with same parts present in *Marchantia* today, so the changes are all due to factors such as evolution, polyploidy, gene exchange and rounds of selection."

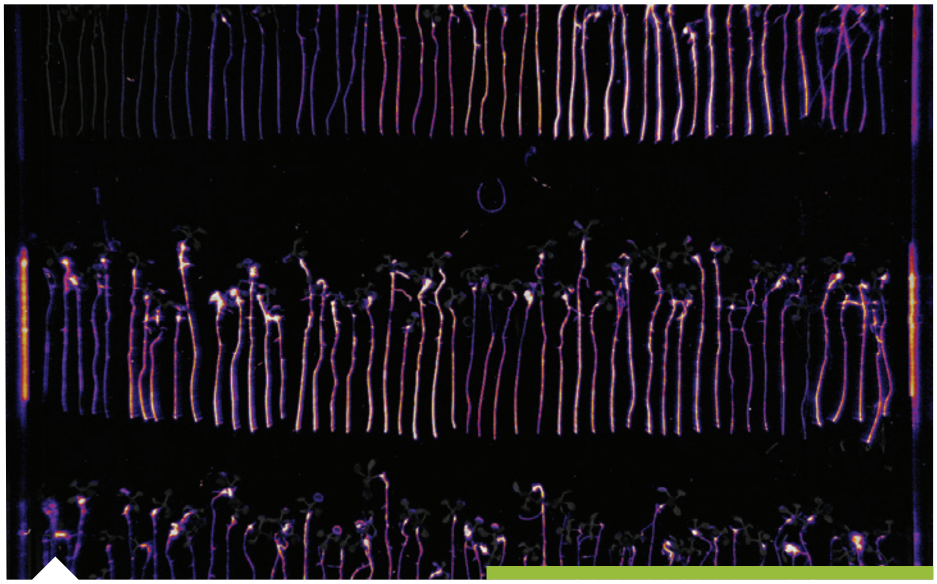
The liverwort's genome sequencing and annotation was done through the JGI's Community Science Program, and allows for genomic comparisons with other early plant lineages sequenced and analyzed by the JGI: the spikemoss *Selaginella moellendorffi* and the moss *Physcomitrella patens*.

One of the 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*, another early plant lineage sequenced and analyzed by the JGI.

(Shohei Yamaoka, Kyoto University)

Establishing A Genome-Wide Map of Bacterial Genes

To better understand how microbes that contribute to plant health and growth colonize the root environment, researchers at the JGI and their collaborators at the Howard Hughes Medical Institute at the University of North Carolina applied a genome-wide transposon mutagenesis approach on the model plant growth-promoting bacterium *Pseudomonas simiae* using the model plant *Arabidopsis thaliana* as a host to generate a genome-wide map of bacterial genes that affect the efficacy of microbial colonization. Through randomly barcoded transposon sequencing (RB-TnSeq), the team identified 115 genes that, when mutated, cause reduced root colonization capabilities. These genes are involved in functions such as sugar metabolism, cell wall synthesis, and motility. The team also identified 243 genes that, when mutated, positively alter root colonization capabilities, many of them likely involved in amino acid transport and metabolism. The results were published October 4, 2017 in *PLOS Biology*.



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

How the Dry Rot Fungus Adapted to the Built Environment

Due to its aggressive capacity to damage the wood in homes, bioenergy researchers have been interested in harnessing the brown rot fungus *Serpula lacrymans* towards breaking down plant mass for conversion to sustainable, alternative biofuels and bioproducts. A team led by University of Oslo scientists and including JGI researchers compared the genomes of two strains, *S. lacrymans var. lacrymans* from Europe and *S. lacrymans var. shastensis* from North America, against a third *Serpula* species, *S. himantioides*, from Europe. They also pitted the three *Serpula* strains against three non-*Serpula* brown rot fungi in a head-to-head competition for resources on wood blocks of pine, fir, and spruce. The results, reported January 5, 2018 in The *ISME Journal*, reveal the evolutionary gains and losses that have allowed one strain in particular, *S. lacrymans var. lacrymans*, to adapt to its indoor home, essentially becoming an ecological specialist and, in so doing, losing its ability to harness woody substrates that other brown rots can access.

Dry rot attacking dresser. (Inger Skrede)

Insights Into Carbon Fixation in the Dark Ocean

Through single-cell genomics and community meta-omics, a team led by Bigelow Laboratory for Ocean Sciences researchers and including JGI scientists identified the most abundant and globally distributed nitrite-oxidizing bacteria in the planet's oceans. They determined that although these bacteria account for less than 5 percent of the populations in each of the various levels of the so-called "dark ocean"—depths below 1,000 meters where no sunlight penetrates—they capture, or "fix," close to half of the inorganic carbon in the waters. This study, the results of which were published in the November 24, 2017 issue of *Science*, indicates nitrite-oxidizing bacteria in the dark ocean may have a much greater role in the global carbon cycle than previously understood.

An Addiction Crucial to a Fungus' Reproduction

In heritable mutualisms, where both species of organism benefit from association, hosts pass on symbionts to subsequent generations. The origin of this relationship, however, is often antagonistic: the parasite first needs to secure its own transmission before working with the host. The bacterial endosymbiont *Burkholderia*, which lives within its host, the soil fungus *Rhizopus microsporus*, is recognized as a mutualist. A collaborative effort led by researchers at Cornell University and scientists at the JGI examined the *Rm-Burkholderia* relationship in order to understand how the antagonistic-to-mutualistic transition occurs. As reported in the November 29, 2017 issue of *Nature Communications*, the team found that the fungus is highly dependent on the bacteria to proliferate both sexually and asexually, consistent with the addiction model of mutualism evolution. 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.

Nutrient Availability in Model Wetlands

Studying microbial communities in San Joaquin Delta rice fields, a model wetland system, JGI researchers linked microbial metabolism and nutrient availability to soil carbon cycling rates. Using a combination of metagenomic sequencing of soil samples, biogeochemical characterization and weekly greenhouse gas emission measurements led to the team's results, published in the December 2017 issue of The *ISME Journal*. Wetland ecosystems can trap as much as 30 percent of global soil carbon but contribute nearly 40 percent of global methane emissions, providing an opportunity to understand their roles as both carbon sinks and carbon sources. In showing how microbial metabolism is regulated by coupled nutrient cycling and soil carbon availability, researchers demonstrate how genomics studies of microbial communities can be scaled up to the ecosystems level, which will contribute to a deeper understanding of ecological processes and will aid the development of better global carbon cycling models.

To learn more about each of these stories, go to the JGI's Science Highlights pages: jgi.doe.gov/category/science-highlights.

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

Succulent Genes for Water Use Efficiency

In water-poor conditions, some plants collect and process carbon at night rather than during the day through crassulacean acid metabolism (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 the genome of the succulent *Kalanchoë fedtschenkoi* (lavender scallops) and compared it to the genomes of other plants that also use CAM photosynthesis. As the first CAM eudicot—a large group of flowering plants that includes sunflower, coffee, and the common bean—to have its genome sequenced, *K. fedtschenkoi* offers researchers an emerging model to trace the evolution of CAM photosynthesis through these plant lineages. Understanding how the CAM photosynthesis pathway evolved in plants would enable researchers to engineer a photosynthesis pathway to improve crop yields for fuel or food applications on arid or marginal lands.

K. fedtschenkoi (Forest and Kim Starr, Flickr CC BY 2.0)

At the International Plant and Animal Genome Conference (PAG XXVI) held January 13–17, 2018 in San Diego, California, JGI and the DOE Systems Biology Knowledgebase (KBase) hosted a joint workshop on January 15 and shared a booth. Besides JGI Plant Program and KBase updates, the workshop featured JGI collaborators who are applying genomic sequence information from the JGI to elucidate functions of plant systems. Danforth Center's Sarit Weissman (left) spoke about *B. distachyon's* response to drought, while the University of Georgia's Jim Leebens-Mack (center) described the Open Green Genomes Initiative, which lists 98 co-primary investigators with him. Right: Among those who stopped by the booth was fungal researcher Scott Baker (left) from the national user facility Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory, seen here with ORNL/KBase's Ben Allen (center left), Priya Ranjan (center right) and Meghan Drake (right).

Meet the JGI Users

Hear how partnering with the JGI and availing of the scientific knowledge, tools and technologies has benefited researchers. Check out the profiles of JGI users at jgi.doe.gov/category/featured-profiles/

Clockwise from top left: J. Chris Pires, University of Missouri; Colleen Hansel, Woods Hole Oceanographic Institute; Brian Hedlund, University of Nevada, Las Vegas; Cat Adams, University of California, Berkeley; Tobias Erb, Max Planck Institute for Terrestrial Microbiology. (Images of Brian Hedlund and Cat Adams by Kristine Wong; other images courtesy of the researchers themselves.)

Proposals due March 1, 2018 for CSP New Investigator Call

The call targets investigators and research initiatives new to the JGI, with an emphasis on providing pilot data to assess feasibility of a novel approach or form the foundation for a large-scale Community Science Program proposal submission. Projects must be independent of ongoing JGI proposals, and lead PIs cannot have been lead PI on any previously accepted JGI CSP or FICUS proposal.

bit.ly/JGI-CSPNewInvestigatorcall

FICUS JGI-EMSL Call

Letters of Intent (LOI) due March 21, 2018

bit.ly/FICUSJGI-EMSL

CSP Annual Call

FY19 Letters of Intent (LOI) due March 30, 2018

bit.ly/JGI-CSPannualcall

Contact The Primer

David Gilbert, Managing Editor
DEGilbert@lbl.gov
Massie Santos Ballon, Editor

18-JG-4720