

Drafting Maps for Plant Geneticists

Over centuries of exploration, surveyors and travelers from different countries would improve upon maps first drafted by those who went on the earliest voyages. The results reflected these additions and revisions, with multiple languages appearing on a single chart.

The plant research community has followed a path similar to those early explorers, developing a robust reference model in *Arabidopsis*. Researchers at the DOE Joint Genome Institute (JGI), however, want to expand the available model species by developing a series of reference standards for a number of DOE JGI Plant Flagship genomes sequenced by the Institute.

The project, now in its pilot phase, is called the Gene Atlas. “The word ‘atlas’ conveys the idea of travelling around the world, and a ‘gene atlas’ is meant to convey a similar sense of travelling in the gene space of various plant species,” said Jerry Tuskan, the Oak Ridge National Laboratory scientist who led the DOE JGI’s poplar genome project. He said that one of the goals of the Gene Atlas is to provide researchers with a way to confirm the validity of their results by presenting the gene

information gathered under several standard experimental conditions involving factors such as lighting and growth.

DOE JGI Plant Program head Jeremy Schmutz at the HudsonAlpha Institute for Biotechnology said the Flagship genomes for which references will be developed in the pilot phase of the study are the alga *Chlamydomonas*, the moss *Physcomitrella*, soybean, poplar and the grass *Setaria*. The

(continued on page 4)

also in this issue

Microbial activity in the Arctic	2
Focus on data analysis assets	3
DOE JGI at PAG XX.	3
A sequence for pest-control strategies	4
In the news	6-7

Terabase-Scale Projects Dominate 2012 Sequencing Portfolio

The 41 projects selected for the DOE JGI’s 2012 Community Sequencing Program (CSP) portfolio have been allocated more than 30 Terabases, a 100-fold increase compared with just two years ago, when just a third of a terabase was allocated to more than 70 projects. This amounts to the equivalent of at least 10,000 human genomes in data.

“These selections truly take advantage of the DOE JGI’s massive-scale sequencing and data analysis capabilities,” said Eddy Rubin, DOE JGI Director. “The projects span the globe and the unexplored branches of the tree of life, and promise to yield a better understanding of the interplay

(continued on page 5)



The green foxtail *Setaria viridis*, a model system for studying C4 bioenergy grasses, is a sequencing project led by the Danforth Plant Science Center’s Thomas Brutnell. (Kropsoq/Wikimedia Commons)

Arctic Microbes' Response to a Warming World

Trapped within the frozen soils of the Arctic is a volume of carbon estimated to be more than 250 times the amount of greenhouse gas emissions attributed to the United States in the year 2009. With rising global temperatures, scientists worry about the potential impact that the thawing permafrost and ensuing release of trapped carbon might have on the global carbon cycle.

To answer this question, researchers from the DOE JGI, Lawrence Berkeley National Laboratory (LBNL), and the U.S. Geological Survey (USGS) collaborated to understand how the microorganisms that reside in the permafrost are responding to their warming environment. From the metagenomic analysis that appeared in the November 6, 2011 issue of *Nature*, the team was able to generate the draft genome of a novel microbe that produces methane, a far more potent greenhouse gas than carbon dioxide.

"The permafrost is poised to become a major source of greenhouse gases as the

temperature in the Arctic is expected to increase dramatically compared to the expected temperature increase in many other regions of the world," said LBNL scientist and study senior author Janet Jansson. "By applying metagenomics to study microbial community composition and function, we can help to answer questions about how the currently uncultivated and unstudied microbial species residing in permafrost cycle organic carbon and release greenhouse gases during thaw. This will provide valuable information that could lead to improved carbon cycle models and eventual mitigation strategies."

Study co-author Mark Waldrop, a USGS researcher who has a 2010 DOE JGI Community Sequencing Program project on comparing the short-term microbial response of the thawing permafrost to the longer-term processes, and his USGS colleagues cored meter-deep permafrost samples from a peaty black spruce forest along Hess Creek, Alaska. The samples were sent to the Jansson lab for DNA extraction and analyses and at the DOE JGI, the team led by Director Eddy Rubin and his postdoctoral fellow Rachel Mackelprang (now at California State University-Northridge) sequenced microbial DNA from the samples.

The researchers identified many genes involved in carbon and nitrogen cycling in the metagenomic data, and found that their levels of abundance shifted in response to their thawing habitat. "These detailed analyses reveal for the first time the rapid and dynamic response of permafrost microbial communities to thaw," they concluded. "The thaw-induced shifts that we detected directly support conceptual models of carbon and nitrogen cycling in arctic soils, in which microbes play a central role in greenhouse gas emissions and



Researchers took core samples of the Arctic permafrost to assess their microbial composition and the impact on these populations by thawing conditions. (Courtesy of USGS Soil Carbon Research)

destabilization of stored permafrost carbon."

At the DOE JGI, Mackelprang and her colleagues generated nearly 40 billion bases of raw DNA sequence, necessary due to the high microbial diversity of the soil. They identified several microbes that produced methane as a byproduct, and were able to assemble a draft genome of a novel methanogen.

"This is the first example of a successful assembly of a draft genome from a highly complex soil metagenome," the authors wrote. "The abundance of this novel methanogen suggests that it may be an important player in methane production under frozen conditions."



Jansson discussed the research at LBNL's Science at the Theater panel on "Secrets of the Soil," held the day after the article was published. The video of the talk can be viewed on the Lab's YouTube channel at <http://bit.ly/LBLSoil>.

The 2012 Department of Energy Joint Genome Institute (DOE JGI)

SEVENTH ANNUAL

Genomics of Energy & Environment Meeting

March 20-22, 2012 | Walnut Creek, California

DETAILS AT
http://j.mp/JGI_UM

DOE JGI Data Analysis Assets Featured in *Nucleic Acids Research Database Issue*

As the global leader in generating genome sequences of plants, fungi, microbes and metagenomes, it was apt to see the DOE JGI's genome and metagenome comparative analysis systems featured in the Database Issue of *Nucleic Acids Research (NAR)*. These systems allow users to analyze and improve the functional characterization of a vast number of publicly available genome and metagenomes.

One of the Issue's "featured" articles (deemed by editors to be in the top five percent in terms of originality, significance and scientific excellence) covered the Genome Portal, through which DOE JGI's nearly 4,000 publicly available genomes and metagenomes can be accessed via a "Tree of Life" graphical catalog. "Users can also use it to navigate to the Institute's domain-specific comparative analysis systems, MycoCosm (fungal genomes), Phytozome (plant genomes), IMG (microbial genomes), and IMG/M (metagenomes)," added Inna Dubchak, the lead author of the paper.

Highlighted in the same paper, MycoCosm was initially released two years ago in response to requests for a central portal to access both fungal genomes and the tools for analyzing them. "The DOE JGI Fungal Genomics program alone aims to double sequencing and analysis throughput every year," wrote the team led by Fungal Genomes program head Igor Grigoriev in the report published online November 22, 2011. "This requires new analytical tools, further scalability in data storage and better integration for the DOE JGI to continue to enable science and serve as a central hub for user communities." Unlike tools developed by other sequencing centers, MycoCosm features community annotation, a model which allows registered users to validate and improve predicted gene models and annotations.

Another paper in the Issue covered

Phytozome, the comparative analysis platform for green plants featuring many genomes sequenced, assembled and annotated at the DOE JGI. Since 2008, the database has been updated 8 times, most recently in January 2012. As reported in the paper published online November 22, 2011, the version 8 release includes updates to flagship genomes *Brachypodium*, maize and *Medicago* and links to the Gene Atlas pilot project (see story on page 1). The updates also include newly-released DOE JGI genomes for the common bean phaseolus as well as the *Arabidopsis* comparator *Capsella rubella*, and externally-generated genomes for the apple and strawberry.

Microbial and metagenomic genomes sequenced at the DOE JGI rely on the Integrated Microbial Genomes (IMG) system and are included into IMG Expert Review (IMG/ER) system for annotation review and comparative analysis. Since IMG was first released in March 2005, the catalog has grown to nearly 7,000 genomes with more than 11 million genes. All publicly available microbial genomes and metagenomes are included into IMG/MER, and those datasets not sequenced on site can also be submitted for inclusion with their original annotations or with IMG specific annotations. The names of metagenome datasets submitted for inclusion into IMG/MER are curated in the Genomes OnLine Database (GOLD) following a five-tiered classification similar to the phylogenetic classification of isolate genomes.

Since January 2009, over 930 metagenome sample datasets have been annotated using the IMG/MER annotation pipeline, with about 60 percent of these samples sequenced at DOE JGI.

The IMG/M system was the focus of a second featured article in the Database Issue, and also selected by editors in the top five percent in (continued on page 7)

PAG XX JGI Highlights (January 14-18, 2012)

Stephane Rombauts: Genome of *Tetranychus urticae* reveals herbivorous pest adaptations (see Primer story on page 4)

Stephen Wright: Comparative and population genomics in the Brassicaceae

Sook Jung: Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceae subfamilies

Fred Gmitter: New Citrus Genome Sequence Resources

Jeremy Schmutz: An improved reference genome for *Physcomitrella*

Larry Smart: Next Generation Sequencing of the *Salix purpurea* L. Genome and Transcriptome: Tools for the Genetic Improvement of Willow Biomass Crops

Jerry Tuskan: Eucalyptus Genome Sequence and Evolution

Alexander Myburg: Sequencing and Analysis of the *Eucalyptus grandis* Genome

Zining Wang: Development of a Whole Genome Marker Map Based on the Cotton D-genome Sequence of *Gossypium raimondii*

Robin Ohm: Comparative genomics of Dothideomycetes plant pathogens

Viola Manning: Sequencing and resequencing the necrotrophic Dothideomycete, *Pyrenophora tritici-repentis*, provides insights into the biology of an increasingly significant wheat pathogen

Katrien Devos: Structure and Evolution of the Foxtail Millet Genome

Brachypodium Community Meeting, Windsor Room

Sean Gordon: Genome Diversity in *Brachypodium distachyon*: Deep Sequencing of Highly Diverse Natural Accessions

Manoj Sharma: Insights into Switchgrass Genome Structure and Organization

Dan Rokhsar: General Annotation Methods

The full agenda is online at <http://pag.confex.com/pag/xx/webprogrampreliminary/start.html>

WEB-SPINNER'S SEQUENCE COULD CUT CHEMICAL PEST CONTROL

For a tiny pest that isn't quite the size of a comma on a keyboard the two-spotted spider mite has a big appetite. These web-spinners extract the nutrients they need from leaves of more than a thousand different plant species, including bioenergy feedstocks. Combating the mite's agricultural attacks through chemical means can cost around \$1 billion annually.

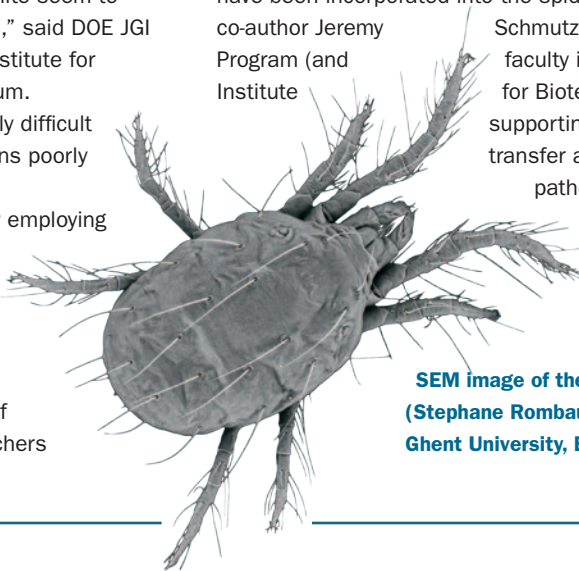
"Many aspects of the biology of the spider mite seem to facilitate rapid evolution of pesticide resistance," said DOE JGI collaborator Yves Van de Peer of the Flemish Institute for Biotechnology (VIB) and Ghent University, Belgium. "Control of these mites has become increasingly difficult and the genetic basis of such resistance remains poorly understood."

Van de Peer and others researchers are now employing the 90-million nucleotide genome of the spider mite *Tetranychus urticae*, sequenced at the DOE JGI as part of the 2007 Community Sequencing Program portfolio, to develop and improve environmentally-friendly pest-control strategies. In the November 24, 2011 edition of *Nature*, he and an international team of researchers

reported on the use of spider mite genome to understand questions such as the pest's ability to rapidly develop resistance to pesticides, how they can serve more broadly as a model for pest-plant interactions and how they are likely to respond in a changing environment.

Among the arthropods, the two-spotted spider mite joins the water flea *Daphnia pulex* in the expanding portfolio of biologically-important model systems. The water flea, now officially recognized by the National Institutes of Health as an environmental sentinel, was the first crustacean to have its genome sequenced, and was the focus of a DOE JGI publication in the journal *Science* earlier this year (see <http://bit.ly/dLB6Ny>).

One of the discoveries Van de Peer and his colleagues noted in their report was the characterization of genes transferred between species. "It's exciting to identify microbial and fungal genes that have been incorporated into the spider mite genome," said study co-author Jeremy Schmutz, leader of the DOE JGI Plant Program (and Institute Schmutz, leader of the DOE JGI Plant Program (and Institute faculty investigator at HudsonAlpha for Biotechnology). "It adds evidence supporting the theory of lateral gene transfer as mechanism for plant pathogens to specialize on plants and increase the ability of their population's impact on our food sources."



SEM image of the spider mite *T. urticae*. (Stephane Rombauts and Wannes Dermauw, Ghent University, Belgium)

Gene Atlas *(continued from page 1)*

experiments for each plant are being overseen by several DOE JGI collaborators: University of California, Los Angeles' Sabeeha Merchant for *Chlamydomonas*, University of Freiburg's Stephan Rensing for *Physcomitrella*, the Danforth Plant Center's Tom Brutnell for foxtail millet, University of Missouri's Gary Stacey for soybean, and Tuskan for the poplar.

"The goal is to generate reproductive RNA sequencing data to inform gene expression levels," said Schmutz. "We hope to eventually generate data under 100 different conditions, but in this first

year we're focusing on 25." The team plans to do the studies on plants that are as close as possible to the sequenced genotype. For example, the foxtail millet plants used will come from the original cross and the soybean plants will be grown from seeds of the original plant, while the poplar team will use a clone of the original sequenced tree grown from a cutting.

By growing the various species under the same set of conditions, the team will be able to compile results that can be used by other researchers in their studies as a benchmark to determine how their

experimental results compare to the norm and what might be novel differences in gene function in their experiment.

Schmutz added that though the team will be determining under what conditions the organisms will be grown, they recognize that not all of the Flagship projects can be compared to one another. For example, unlike the other Flagship genomes, *Chlamydomonas* is a single celled organism and has no tissues, while *Physcomitrella*, an early plant, has something akin to apical meristematic cells, the tissue where growth takes place *(continued on page 7)*

CSP 2012 Projects *(continued from page 1)*

between climate, ecosystem and organism.”

The first genomic characterization of a microbial community resulted from a collaboration between the DOE JGI and Jill Banfield at the University of California, Berkeley and her colleagues and involved samples from a U.S. Environmental Protection Agency Iron Mountain Superfund site in Northern California. In her 2012 proposal, Banfield proposes to recover genomes from subsurface microbial communities at the DOE Integrated Field

Research Challenge bioremediation research site at Rifle, Colorado. She and her team intend to generate, initially, a terabase of sequence to identify novel rare microbes that might be useful for the environmental cleanup of metals and radionuclides, and which may also lend insight into subsurface carbon sequestration.

Another terabase-sized metagenome project focuses on the microbial communities in the McMurdo Dry Valley system of Antarctica. Considered to be among the

harshest and most extreme environments on the planet with low water and nutrient levels paired with high salt and ultraviolet radiation levels, the team led by Craig Cary at the University of Delaware and Charles Lee at the University of Waikato in New Zealand seeks to understand how the carbon cycle plays out in this ecosystem.

Among the large-scale and complex projects is a proposal from fungal researchers to fill in gaps in the Fungal Tree of Life by sequencing 1,000 fungal genomes over the next five years. Though more than 100 fungal genomes are now available, the team added, acknowledging the DOE JGI's role in these efforts, currently nearly three-quarters of these families have not yet been sequenced.

The single largest project comes from Jeff Dangl at the University of North Carolina and his colleagues and focuses on the rhizosphere, that narrow region where microbes in the soil colonize and interact with plant roots.

The team proposes to study the rhizosphere microbiomes of maize, *Arabidopsis* and a mustard relative known commonly as Drummond's rockcress, as well as potential biofuel crop *Miscanthus* and wild prairie grasses, to understand the plant genetics involved in determining the microbial communities associated with plant species.

The CSP call invited researchers to submit proposals for projects that advance capabilities in fields such as plant-microbe interactions, microbes involved in carbon capture and greenhouse gas emission, and metagenomics. The final list of projects was whittled down from 152 proposals submitted, culled in turn from the 188 letters of intent originally received. The projects were then reviewed and approved by an outside review panel before being vetted by the DOE.

For the complete list of CSP 2012 sequencing projects, see: <http://www.jgi.doe.gov/sequencing/cspseqplans2012.html>.

Backing Federal Funding for Basic Science

Against the continuing debate over the amount of federal funding that should be allocated to scientific research, io9 editor Annalee Newitz spoke with several DOE JGI researchers including Director Eddy Rubin. Her story ran in the November 4, 2011 edition of the *Washington Post*.

“Private industry isn't going to build supercomputers to study weather,” Newitz quoted Rubin as saying. “Or giant accelerators to probe the fundamental particles of nature. That's what the government needs to do.”

The article (<http://wapo.st/rCIBUD>) gave nods to DOE JGI projects such as the cow rumen metagenome, the microbial communities of the restored wetlands in the Sacramento/San Joaquin Delta and bioenergy feedstocks such as switchgrass.

“This kind of science is often called ‘basic,’” Newitz wrote, “though it's anything but.”

One of the people she interviewed was Eukaryote Super Program head Dan Rokhsar, whose math skills Newitz highlighted in an anecdote regarding the importance of the bioenergy-related research being done at the DOE JGI:

“Let's say you've got a goal that in 20 years, you want to have a much higher usage of biofuels,” Rokhsar said. He did some quick math: “We spend over \$1



Switchgrass flower (Jeremy Schmutz, HudsonAlpha Institute for Biotechnology/DOE JGI)

billion a day on foreign oil. So let's say that sequencing these genomes now allows some graduate student to clone a gene five years from now because they can look it up in our database. That's going to accelerate the research. Say that allows us to start using biofuels a month before our 20-year goal. You've saved \$30 billion from that month alone.’ The hope for that kind of dramatic breakthrough fuels the project.”

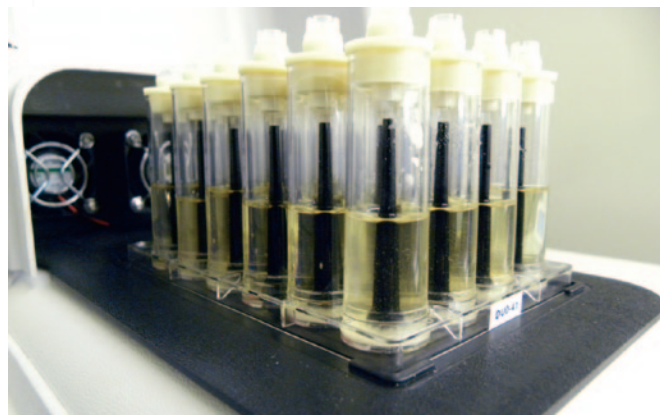
JGI IN THE NEWS**Engineering bacteria for biodiesel**

Biodiesel production typically starts with seeds from oil-rich energy crops such as soybean, palm or rapeseed, which are harvested and then converted into fatty acids for use as fuel. Growing enough of these crops for large-scale biodiesel production, enough to allow it to compete effectively with fossil fuel sources, is a challenge bioenergy researchers are working on.

One alternative source of biodiesel microbiologists are considering is bacteria. By synthesizing the fatty acids in bacteria and converting them into biodiesel, researchers say they can bypass several steps used in producing biodiesel from plants including the purification process, enabling them to reduce production costs and raise final yields.

Working toward this goal, the November 2011 issue of *Applied and Environmental Microbiology*, includes a report from Nikos Kyrpides and his former DOE JGI colleagues Parwez Nawabi and Thanos Lykidis, along with Stefan Bauer of the Energy Biosciences Institute, University of California, Berkeley, describing a process of engineering bacteria with the help of a novel fatty acid enzyme to produce biodiesel. The enzyme was identified and characterized from several bacterial sequences. They inserted the enzyme into the model microbe *E. coli* to prove that it was involved in fatty acids synthesis.

“The work described in this report provides an alternative route for the synthesis of biofuel molecules,” wrote the researchers. “The pathway described here is a first step in the generation of [biodiesel] and, with further optimization, may lead to the production of a cost-efficient next-generation biofuel.”



Laboratory small-scale optimization of biodiesel yields from *E. coli* — soon to be up-scaled for industrial production (Stefan Bauer/ Energy Biosciences Institute)

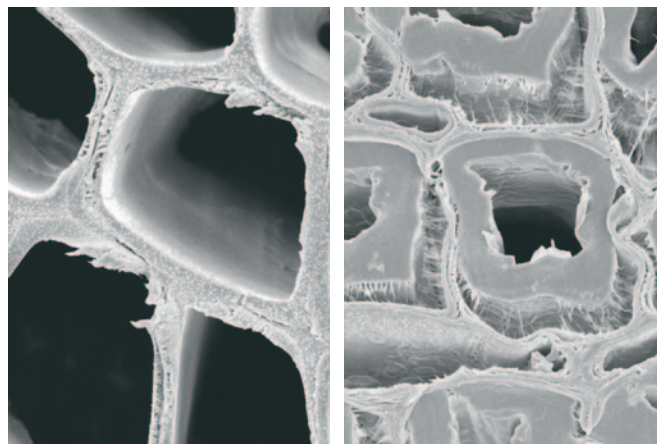
Tension wood helps tailor bioenergy feedstock

When the DOE JGI published the poplar genome sequence analysis in 2006, it bolstered several research efforts to fully develop that tree’s potential as a candidate bioenergy feedstock.

In a report published online October 4, 2011 in *Energy & Environmental Science*, DOE JGI’s Jerry Tuskan at Oak Ridge National Laboratory, the project lead on the poplar genome project, and colleagues at the BioEnergy Science Center at ORNL turned their attention on a characteristic of poplar trees that could be useful for increasing available sugars and ultimately, biofuels yield from trees.

To fully take advantage of the genomic information and target the genes associated with improving poplar’s potential as a bioenergy feedstock, the team wrote in their paper, researchers need to link phenotypic and genotypic data to the lignocellulosic challenges presented in converting poplar biomass to sugars for biofuels production.

In this particular study, they focused on tension wood, a natural response to stress from bending over in which cells are composed of 90 percent cellulose, twice as much as what is normally seen. The tension wood also has lower levels of lignin and higher levels of xylem and secondary cell wall material, all of which, the researchers wrote, “contribute to desirable characteristics of feedstock materials for bioethanol production. Understanding the underlying phenotypic, biochemical and morphological properties of model biomass such as *Populus TW* are central to informing the design of suitably tailored feedstocks.”



Normal wood (left) and tension wood (right). (Images by DOE’s BioEnergy Science Center/Georgia Institute of Technology)

JGI IN THE NEWS

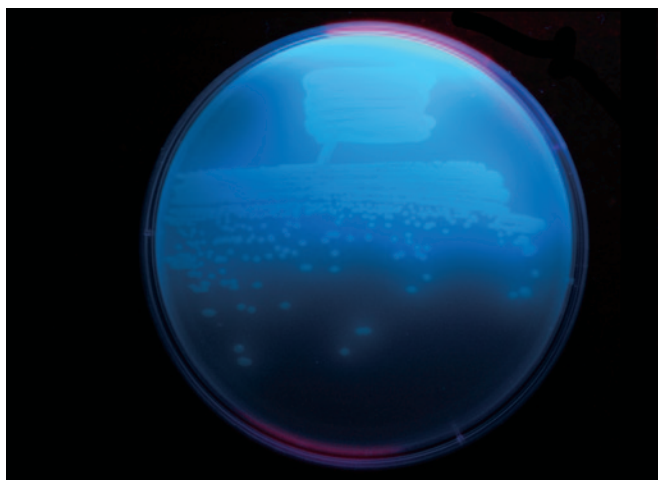
Treating contaminated MDA reagents

Single cell genomics is a method used to get around the research roadblock thrown up by the need for large amounts of microbial DNA that may be difficult to culture or is uncultured in order to do traditional sequencing techniques. In this process, a single cell isolated from an environmental sample is amplified to make millions of copies of the microbial genome for sequencing.

One challenge single cell genomics researchers face, however, is that the reagents used in the multiple displacement amplification (MDA) process along with the phi29 polymerase are contaminated with DNA that also ends up being amplified along with the target DNA, ultimately impacting the technique's effectiveness in characterizing unknown microbial genomes.

In a report published October 20, 2011 in *PLoS ONE*, DOE JGI Microbial Genomics Program head Tanja Woyke and Process Optimization head Jan-Fang Cheng led colleagues and longtime collaborators from the Bigelow Laboratory for Ocean Sciences in testing UV-irradiation of up to 90 minutes as a way to standardize UV treatment protocols for removing the contaminant DNA that comes in MDA reagents without affecting enzyme activity or introducing potential biases into the process.

"From the analysis of genomic sequence data of >100 *Escherichia coli* single cells, we demonstrate the optimal range of UV treatment of MDA reagents for efficiently removing contaminant DNA without a significant reduction of the Phi29 activity or introducing additional single cell genome coverage bias or artifacts," the team wrote in their paper.



P. fluorescens (Image by Ninjatacoshell/Wikimedia Commons)

Gene Atlas *(continued from page 4)*

on plants, but they are called protonemal apical cells.

Another goal of the Gene Atlas, Schmutz said, is to build up communities around these genomes. "We want to attract young scientists who want to work on cellulosic fuels and carbon cycling," he said. "The Plant Flagship genomes are among the most important feedstock species we've ever worked on. The Gene Atlas datasets will offer researchers validation downstream of the sequence and analysis, offering them a way of interpreting the data in light of what's 'normal' in the plant."

The project is also poised to move the DOE JGI beyond genome sequencing. "We're laying the foundation to do functional work in these important organisms," Schmutz said. "The data we're collecting will allow people access to an information architecture so they can predict gene pathway functions based on validated expression levels across tissues and conditions for the Flagship genomes."

Database Issue *(continued from page 3)*

terms of originality, significance and scientific excellence. The paper's lead authors are DOE JGI's Prokaryote Super Program Head Nikos Kyrpides and Chief Informatics Officer Victor Markowitz, also the head of the Biological Data Management and Technology Center at Lawrence Berkeley National Laboratory, and reported new data analysis tools available to users including the ability to estimate the phylogenetic composition of a metagenome sample, as well as the ability to compare up to five user-selected metagenomes based on their BLAST hits. There are also "Abundance Profile" tools that allow users to compare the functional capabilities of metagenomes and genomes.

"Data sets from next generation sequencing technology platforms often result in million sequences rendering storing and accessing of data in the standard relational data bases inefficient," the team wrote in their report published online November 15, 2011. "As we expect an exponential growth of the size of metagenome data sets by these platforms, we are devising new data management and classification techniques for organizing metagenome data in support of effective analysis."

Among these techniques is one already implemented in the Genomes OnLine Database (GOLD), which was launched in 1997. While the DOE JGI's Genome Portal offers users a way of tracking only the DOE JGI's ongoing projects, GOLD allows users to monitor genome and metagenome projects worldwide. Dino Liolios, co-lead author of the GOLD paper published online December 1, 2011, said the main page of the GOLD site has recently begun making distinctions between these two types of *(continued on page 8)*

8 / the PRIMER

Winter 2012 Volume 9 Issue 1

Database Issue *(continued from page 7)*

projects being monitored.

With more than 11,000 projects documented, the GOLD database heavily relies on the use of genome standards and has been pioneering the development and implementation of several standards for various data types. One such example is the development of a standardized Metagenome classification scheme. Metagenome datasets are organized in three main ecosystem classes, environmental, host associated, and engineered, then further divided in subclasses characterized by ecosystem categories (e.g., aquatic, terrestrial, air for environmental metagenomes), ecosystem type (e.g. freshwater, marine), ecosystem subtype (e.g., ground-water, drinking water), and specific ecosystem (e.g., cave water, filtered water).

"[The scheme] represents the first and so far the only proposed classification system for metagenomes," Liolios and his colleagues noted in their report. With

ongoing grand-scale sequencing initiatives including thousands of metagenomic and uncultured samples, the team added, the classification system implemented will become even more necessary to track projects in coming years.

"JGI's comparative analysis systems have matured over the past years and are recognized as important resources for conducting genome and metagenome studies worldwide" said Markowitz. "Unlike the decreasing cost of generating 'raw' genomic sequence datasets, the biological interpretation of these datasets and their integration into comparative analysis systems are complex and require substantial computational resources that are not readily available to individual scientists or research groups. JGI's comparative analysis systems empower scientists around the world to conduct studies that otherwise would be very expensive or out of reach."



As of November 19th, 2011, DOE JGI's Integrated Microbial Genomes with Microbiome Samples (IMG/M) data integration and computational analysis system exceeded the one billion genes mark. The sample that helped break the data barrier was from the gut microbiome of Panamanian leaf-cutter ants.

System Stats:

Genome Portal

(<http://genome.jgi.doe.gov/>)

4,000 publicly available projects

Phytozome (www.phytozome.net/)

More than 24 plant genomes
Used for plant genome studies
published in over 210 scientific
papers

MycoCosm (<http://genome.jgi-psf.org/programs/fungi/index.jsf>)

More than 100 fungal genomes
Used for fungal genomic studies
published in over 369 scientific
papers

IMG (<http://img.jgi.doe.gov/w/>) and **IMG/ER** (<http://img.jgi.doe.gov/er/>) have been used for microbial genome studies published in over 410 scientific papers

IMG contains more than 3,000 bacterial, archaeal and eukaryotic genomes

IMG has nearly 3,000 viral genomes and more than 1,100 plasmids that did not come from a specific sequencing project

IMG/MER pipeline has been used to annotate 930 metagenome sample datasets

GOLD (<http://www.genomesonline.org>)

More than 11,000 genome projects registered, with more than 200 metadata fields.

References

Genome portals: http://1.usa.gov/NAR_Portal

Phytozome: <http://1.usa.gov/Phytozome>

IMG/M: http://1.usa.gov/NAR_IMG_M

GOLD: http://1.usa.gov/NAR_GOLD

IMG: http://bit.ly/NAR_IMG

Contact The Primer

David Gilbert, Editor / DEGilbert@lbl.gov

