

#### Extraordinary Month for JGI Science

In a single week during the month of June, three plant genome publications by researchers at the U.S. Department of Energy Joint Genome Institute were published in **Nature, Nature Biotechnology**, and **Nature Genetics**. These publications are featured on pages 1-3 of this issue.

#### **Lessons From Comparing Citrus Genomes**



Citrus is the world's most widely cultivated fruit crop and was originally domesticated in Southeast Asia thousands of years ago before spreading throughout Asia, Europe, and the Americas via trade. In the United States, the citrus crop was valued at over \$3.1 billion in 2013.

Citrus varieties are reproduced asexually by vegetative propagation, so trees producing a specific type of fruit are typically genetically identical. This breeding approach makes the fruit trees much more susceptible to disease. Researchers worldwide are mobilizing to apply genomic tools and approaches to understand how

citrus varieties arose and how they respond to disease and other stresses.

Initial support for the citrus genomics effort arose 10 years ago under the auspices of the inaugural round of the U.S. Department of Energy Joint Genome Institute's (DOE JGI) Community Science Program (formerly the Community Sequencing Program), which seeks to build scientific communities around cornerstone species of relevance to DOE missions in bioenergy, carbon cycling and biogeochemistry. By understanding the relationships between the various cultivated species with what they describe as "very narrow genetic diversity," the citrus consortium hopes to enable sequence-directed improvement, which could lead to crops that are more resistant to disease and stresses such as environmental changes.

In a study published in the July 2014 edition of Nature Biotechnology, an international consortium of researchers led by Fred Gmitter of the University of Florida Citrus Research and Education Center analyzed and compared the genome sequences of ten diverse citrus varieties. They found that these fruits are derived from two wild citrus species that diverged in Southeast Asia over five million years ago. By inferring the past hybridization events that gave rise to these common citrus varieties, the team hopes to enable strategies for improving citrus. continued on page 3

#### in this issue

Tree of Prodigious Potential	2
What Stop Codon?	4
Dead Sea Lessons	5
Diverse Cyanobacteria	5
Shearing Methane Emissions	6
A Spectrum of Fungal Rot	7
In Memoriam	8

# Common Bean Genome for Future Crop Improvements

All plants require nitrogen to thrive, and nitrogen fixation is the process by which atmospheric nitrogen is converted into ammonia. As many agricultural lands are deficient in nitrogen, farmers rely on fertilizers to supply this needed nutrient for their crops. Increasing crop yields is desired both for fuel and food production. According to the U.S. Department of Agriculture, the United States imports more than half of the nitrogen used as fertilizer, a total of nearly 11 million tons in 2012.

For the DOE Office of Science, sequencing the common bean matters because legumes such as the common bean and soybean can form symbiotic relationships with nitrogenfixing bacteria. To this end, a team of researchers led by Scott Jackson of the University of Georgia, Dan Rokhsar of the DOE JGI. Jeremy Schmutz of the DOE JGI and the HudsonAlpha Institute for Biotechnology, and Phil McClean of North Dakota State University sequenced and analyzed the genome of the common bean, Phaseolus vulgaris. The work was published in the July 2014 issue of Nature Genetics.

"Unlocking the genetic make-up of the common bean is a tremendous achievement that continued on page 3

#### **Eucalyptus**

#### A Genome with Prodigious Biofuel Potential

While native to Australia, eucalyptus trees are planted worldwide mostly for the value of its wood. Because of its wide adaptability, extremely fast growth rate and excellent wood and fiber properties, Eucalyptus trees are grown in 100 countries across six continents and account for over 40 million acres.

The tree's prodigious growth habit has caught the eyes of researchers seeking to harness and improve upon eucalyptus' potential for enhancing sustainable biofuels and biomaterials production. For the DOE, the energyrich cellulosic biomass of eucalyptus makes it one of the principal candidate biomass energy crops.

Genome sequencing is an essential diagnostic tool for understanding the basis of the eucalyptus tree's superior growth properties. As reported June 19, 2014 in the journal Nature, the international effort to sequence and analyze the 640 million base pair genome of Eucalyptus grandis engaged more than 80 researchers from 30 institutions, representing 18 countries. The project was led by Alexander Myburg of the University of Pretoria (South Africa); Dario Grattapaglia of the Brazilian Agricultural Research Corporation (EMBRAPA) and Catholic University of Brasilia; Gerald Tuskan of the Oak Ridge National Laboratory, the BioEnergy Science Center, and the DOE JGI; Dan Rokhsar of the DOE JGI; and Jeremy Schmutz of the DOE JGI and the HudsonAlpha Institute for Biotechnology.

Eucalyptus can be harvested from tropical and temperate zones and has over 700 species that are rich in genetic variation. Combing through the 36,000-plus genes found in eucalyptus, the researchers homed in on those that may influence the production of secondary cell wall

material that can be processed for pulp, paper, biomaterials and bioenergy applications. The extensive catalog of genes contributed by the team will allow breeders to adapt eucalyptus trees for sustainable energy production in regions, such as the U.S. Southeast, where it cannot currently be grown.

Approximately 80 percent of the woody biomass in a eucalyptus tree is made of cellulose and hemicellulose, both long chains of sugars, with the remaining biomass primarily comprised of lignin, the tough "glue" that holds it all together. The eucalyptus team identified genes encoding 18 final enzymatic steps for the production of cellulose and the hemicellulose xylan, both cell wall carbohydrates that can be used for biofuel production. "By tracing their evolutionary lineages and expression in woody tissues we defined a core set of genes as well as novel lignin-building candidates that are highly expressed in the development of xylem—the woody tissue that helps channel water throughout the plant—which serves to strengthen the tree," said Myburg.

An additional finding by the team was that among sequenced plants to date, eucalyptus showed the highest diversity of genes for specialized metabolites such as terpenes. These hydrocarbons serve as chemical self-defenses against pests, as well as providing the familiar aromatic essential oils used in both medicinal cough drops and for industrial processes.

"By having a library of these genes that control the synthesis of terpenes we are able to dissect which genes produce specific terpenes; then we can modify this biochemical pathway in the leaves so that we can develop the potential of eucalyptus as an alternative source feedstock for jet

fuel," noted ORNL's Tuskan.

A short interview with Jerry Tuskan on the implications of the team's Eucalyptus genome analysis can be viewed at http://bit.ly/eucalyptusTuskan.

The eucalyptus genome data are available publicly through the DOE JGI's comparative plant genomics portal known as Phytozome, now in its 10th revision (http://bit.ly/Phytozome-Eucalyptus).



# PRIMER

#### **Comparing Citrus Genomes**

continued from page 1

The genomes presented in the published study included pummelos, oranges and mandarins. DOE JGI researchers were involved in generating the high-quality reference genome of Clementine mandarin and the sweet orange genome.

The analyses revealed that while

pummelos represent a single citrus species, the same cannot be said of cultivated mandarins, even those long held as not having intermixed with other varieties. The "wild" Mangshan mandarin from China is an exception to the rule, as its genome revealed it was in fact a separate species from

other cultivated mandarins.

The Clementine mandarin and sweet orange genomes are on Phytozome at http://phytozome.jgi.doe.gov/. Gmitter's talk on citrus genomics at the DOE JGI 7th Annual 2012 Genomics of Energy & Environment Meeting is available at http://bit.ly/JGI7Gmitter.

#### Common Bean Genome

continued from page 1

will lead to future advances in feeding the world's growing population through improved crop production," said Sonny Ramaswamy, director of the USDA's National Institute of Food and Agriculture. "While we have much to learn about the application of genomics in agriculture, this study is ground-breaking. I applaud the work of this team of scientists and look forward to their continued work in this important area."

The team sequenced and assembled a 473-million basepair genome of the common bean. Though it is thought to have originated in Mexico, the common bean was domesticated

separately at two different geographic locations in Mesoamerica and the Andes, diverging from a common ancestral wild population more than 100,000 years ago. The team then compared sequences from pooled populations representing these regions, finding only a small fraction of shared genes. This indicated that different events had been involved in the domestication process at each location.

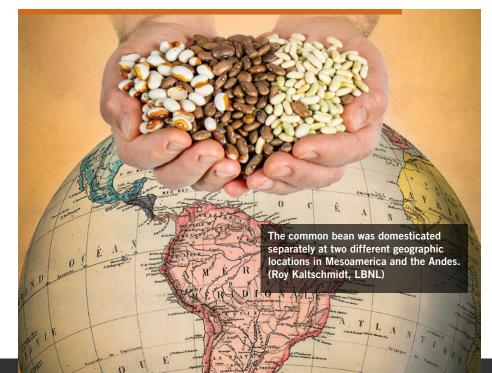
The team looked for regions associated with traits such as low diversity, flowering time, and nitrogen metabolism. Aside from finding dense clusters of genes related to disease

resistance within the chromosome, they also identified a handful of genes involved in moving nitrogen around. The team then compared the high quality common bean genome against the sequence of its most economically important relative, soybean. They found evidence of synteny, in which a gene in one species is present in another. They also noted that the common bean's genome had evolved more rapidly than soybean's since they diverged from the last common ancestor nearly 20 million years ago.

"Improvement of common bean will require a more fundamental understanding of the genetic basis of how it responds to biotic and abiotic stresses," the team concluded. "These findings provide information on regions of the genome that have been intensely selected either during domestication or early improvement and thus provide targets for future crop improvement efforts."

Phil McClean presented on the common bean genome project at the recent 9th Annual DOE JGI Genomics of Energy & Environment Meeting. Watch the video at http://bit.ly/JGIUM9McClean.

The *Phaseolus vulgaris* genome data are available publicly through the DOE JGI's comparative plant genomics portal Phytozome at http://bit.ly/Phytozome-commonbean.





#### Sometimes, "Stop" Codes for "Go"

The DNA sequence that genetically defines an organism includes 64 shorter, three-letter codons, nearly all of which code for the 20 amino acids. Three triplets, named amber, opal, and ochre, are known as stop codons, terminating the translation of RNA into protein. When an organism's machinery reads the instructions in the DNA, and builds a protein composed of amino acids, reaching amber, opal, or ochre signals the end of a protein.

It was long thought that there is only one "canonical" code, so each instruction means the same thing to every organism. While a few examples of organisms deviating from this canonical code had been discovered before, they were widely thought of as very rare anomalies. However, a DOE JGI study in the May 23, 2014 issue of *Science* shows that for some organisms, the instructions for *amber*, *opal*, and *ochre* mean anything but "stop."

"All along, we presumed that the code or vocabulary used by organisms was universal, applying to all branches of the tree of life, with vanishingly few exceptions," said DOE JGI Director Eddy Rubin, and senior author on the *Science* paper. "We have now confirmed that this just isn't so. There is a significant portion of life that uses different vocabularies where the same word means different things in different organisms."

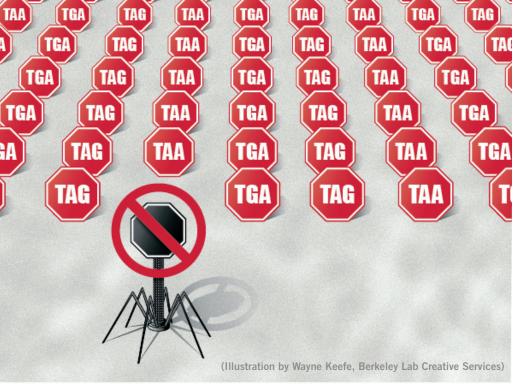
The team focused on uncultivated microbes whose genomes had been described through single-cell genomics and metagenomics, and on a collection of viral sequences. They analyzed nearly six terabases of sequence data from 1,776 samples collected from the human body and several sites around the world. "In this project, using metagenomics and single-cell genomics to explore uncultured microbes, we really had

the opportunity to see how the genetic code operates in the wild," Rubin said. "It is helping us get an unbiased view of how nature operates and how microbes manage our planet."

The search for genetic vocabulary breakdowns began when the study's lead investigator came across bacteria with genes of only 200 base pairs in length instead of the expected 800-900 base pairs. They found that using the canonical codon table, *opal* gave the bacteria very short genes. However, when the team applied a different vocabulary in which *opal* was assumed to encode the amino acid glycine, the genes in the bacteria suddenly appeared to be of normal length.

Following this finding, the team looked for similar occurrences in enormous amounts of sequence data from uncultured microbes. They found that codons get reassigned more often that had been previously thought. "Reassignment of all three stop codons was found but with different preferences by domain and habitat," the team reported. "We observed distinct patterns of stop codon reassignment in the three domains of life, with bacteria showing only opal reassignments, ochre reassignments restricted to eukaryotes, and archaea devoid of codon reassignments. Among [DNA] viruses, we found both amber and opal reassignments." Another observation the researchers made was that beyond bacteria, these reassignments were also happening in phage, viruses that attack bacterial cells.

This work builds on a previous study in which DOE JGI researchers successfully employed single-cell genomics to shed insight on a plethora of microbes representing 29 "mostly uncharted" branches on the tree of life. Director Rubin spoke about this study at the recent DOE JGI Genomics of Energy & Environment Meeting: http://bit.ly/JGIUM9Rubin.



#### **Marine Microbes**

#### Fungal Lessons From the Lowest Point on Earth

Despite its name, the Dead Sea does support life, albeit a limited number of species. Some organisms thrive in the extremely salty environment at the lowest point on Earth by lying dormant when salt concentrations are very high. Other organisms need salt to grow. In the May 9, 2014 issue of *Nature Communications*, a team of researchers including the DOE JGI's Igor Grigoriev studied the genome of the filamentous fungus *Eurotium rubrum* to learn which survival strategy it uses.

"Understanding the long-term adaptation of cells and organisms to high salinity is of great importance in a world with increasing desertification and salinity," the team wrote. This work may also have biofuels applications as the DOE JGI and its partners are sourcing microbial and fungal enzymes for more effective biomass pretreatment with ionic liquids,

environmentally benign organic salts often substituted for volatile organic solvents.

The DOE JGI team sequenced, assembled and annotated the 26.2-million basepair genome of *E. rubrum*. They found that the fungus' proteins had higher aspartic and glutamic acid amino acid levels than expected. When the team compared *E. rubrum's* gene families against those in two other halophilic species, they found that high acidic residues were common in all three species, a general trait all salt-tolerant microbes share.

To learn more about *E. rubrum's* salt tolerance, collaborators at the University of Haifa then grew samples in liquid and solid media at salinities from zero up to 90 percent of Dead Sea water. For more details, read the full story at <a href="http://jgi.doe.gov/salt-needed-tolerance-lessons-from-a-dead-sea-fungus/">http://jgi.doe.gov/salt-needed-tolerance-lessons-from-a-dead-sea-fungus/</a>.



Dead Sea (Tami Kis-Papo, University of Haifa)

## **Exploring Genomic Diversity Within a Single Species**

By some estimates, a single liter of water can hold as many as 100 million cells of *Prochlorococcus*. These cyanobacteria are thought to be responsible for providing about 20 percent of the oxygen annually produced by the planet.

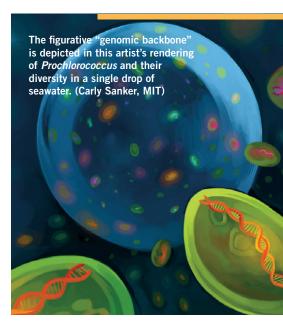
Despite being considered a single species, *Prochlorococcus* can be classified into several distinct major ecotypes, within which the cells display a wide range of genomic diversity.

In the April 25, 2014 issue of *Science*, MIT marine microbiologist Sallie Chisholm, a longtime DOE JGI collaborator, led a team that sequenced and assembled *Prochlorococcus* genomes from single cells collected from the northwestern Sargasso Sea. The team found cell clusters within known ecotypes that indicated that the relative abundance of cyanobacterial subpopulations shifted along with the

seasons, and that each of these socalled "genomic backbones" is comprised of highly conserved core gene alleles and a smaller distinct set of flexible genes.

If each of the cyanobacterial backbone subpopulations were counted as distinct species, they noted, then *Prochlorococcus* consists of thousands of species. Additionally, each of these groups likely helps maintain the "dynamic stability of the *Prochlorococcus* 'collective' in the global oceans." Such a large set of coexisting subpopulations with distinct genomic backbones, the team concluded, may be characteristic feature of free-living bacterial species with huge populations in highly mixed habitats.

The vast diversity of these *Prochlo-rococcus* subpopulations was validated with help from the DOE JGI's Rex



Malmstrom. For more details, check out the full story at <a href="http://jgi.doe.gov/discovering-diversity-one-cell-time/">http://jgi.doe.gov/discovering-diversity-one-cell-time/</a>.



#### Mitigating Methane Emissions at the Rumen Microbiome

The U.S. Environmental Protection Agency (EPA) attributes one-fifth of methane emissions to livestock such as cattle, sheep and other ruminants. The number is of concern as the greenhouse gas is some 28 times more potent than carbon dioxide and has been steadily growing since the 18th century, currently exceeding 1,800 parts per billion in the atmosphere.

Though ruminant livestock are the largest single source of methane emissions, it turns out the levels they produce vary. To understand why this is so, a team of researchers led by the DOE JGI deployed high throughput DNA sequencing and specialized analysis techniques to explore the contents of sheep rumens. They collaborated with New Zealand's AgResearch Limited to see what role the microbes living in the rumen play in this process. The study appeared online June 6, 2014 in Genome Research.

"We wanted to understand why some sheep produce a lot and some produce little methane," said DOE JGI Director Eddy Rubin. "It's not so much the actual composition of the microbiome that determines emission—which conventional wisdom would suggest—but mostly transcriptional regulation within the existing microbes that makes the difference, which is a concept that is relatively new in metagenomic studies."

The researchers took advantage of a large sheep screening and breeding program in New Zealand that aims to breed low methane-emitting ruminants without impacting other traits such as reproduction and wool and meat quality. The methane yields from a cohort of 22 sheep were measured and based on the data, the team selected four sheep with the lowest methane emissions, four sheep with the highest emissions and two sheep with intermediate emission levels.

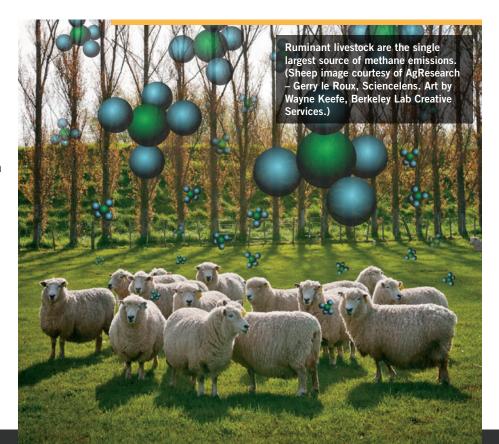
Rumen metagenome DNA samples collected on two occasions from the 10 sheep were sequenced at the DOE JGI, generating 50 billion bases of data each.

"The deep sequencing study contributes to this breeding program by defining the microbial contribution to the methane trait, which can be used in addition to methane measurements to assist in animal selection." said senior scientist Graeme Attwood of AgResearch Limited, a senior author on the paper.

In sheep with low methane emissions, the team found elevated levels of one particular species of methanogen while sheep with high methane emissions had elevated levels of another group of methanogens. Upon further exploration, they identified a methane-producing pathway and three variants of a gene encoding an important methane-forming reaction that were involved in elevated methane yields. While the overall changes to the methane-producing microbial

community structure and methanogen abundance across sheep were rather subtle, the team reported that the expression levels of genes involved in methane production varied more substantially across sheep, suggesting differential gene regulation, perhaps controlled by hydrogen concentration in the rumen or by variations in the dwell time of their feed. The team's findings suggest new possible targets for mitigating methane emissions at the microbiome level.

Screening and breeding for lowmethane producing sheep is still underway, and importantly, low-methane lines then need to be tested for stability of the trait. Attwood noted that, "there needs to be an incentive for farmers to incorporate low methane animals into their flocks. If everything went well, you could expect introduction of the low methane trait to begin in three years, and for there to be slow but incremental changes to the sheep industry in subsequent years."



### **Fungal Rots**

#### **Exploring the Range of Fungal Rots**

If a fungus can break down all the components – cellulose, hemicellulose and lignin – of plant cell walls it is considered a white rot fungus. If a fungus can only break down cellulose and hemicellulose but not lignin, it is classified as a brown rot fungus. Known white rot fungi produce certain lignin-degrading enzymes called class II peroxidases or PODs, and a variety of enzymes that go after crystalline cellulose.

In a study published July 8, 2014 in the *Proceedings of the National Academy of Sciences*, a team led by DOE JGI fungal researchers suggests that categorizing wood-decaying fungi as either white rot or brown rot may not be as clear-cut as previously thought. Their findings both complicates and broadens the range of fungal decay strategies to be explored for commercializing the process of biofuels production.

"It is important to identify a whole range of enzymes sourced from nature that can be used to develop second-generation biofuels in terms of breaking down lignin and other components in plant cell walls," said DOE JGI Fungal Genomics head Igor Grigoriev.

Researchers analyzed 33 basidio-mycete fungal genomes, 22 of them wood decayers, and including four recently sequenced by the DOE JGI. Based on previously sequenced genomes, the team observed that two of the new fungi, *Botryobasidium botryosum* and *Jaapia argillacea*, had the cellulose-attacking enzymes characteristic of white rot fungi, but lacked PODs, making them similar to brown rot fungi.

Applying a statistical process called Principal Components Analysis to find similarities in fungi based on their plant biomass degrading genes, they found that the two new fungi grouped close to *Phanerochaete chrysosporium*, the first white rot species

sequenced. This was a curious finding because the new fungi were phylogenetically distant from *P. chrysosporium* and didn't have PODs.

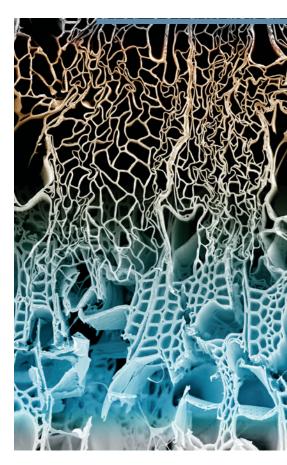
The team then grew isolates of *B. botryosum* and *J. argillacea* on pine and aspen wood. They found that the fungi superficially degraded the wood surfaces but in localized areas, went further and broke down the cell walls and removed cellulose, hemicellulose and lignin.

"[They] show similarities to white rot fungi in ... all predicted carbohydrate- and lignin-active enzymes and can degrade all components of wood, but they do so without the PODs that are a hallmark of white rot," the team reported in their paper. They also found a correlation between secondary metabolism genes, which are crucial for fungal survival, and brown rot fungi. These results, they added, suggest that the perceived dichotomy of white rot and brown rot is too simplistic and suggest that fungal wood decay capabilities be categorized instead on a continuum.

Dan Eastwood, a fungal researcher at Swansea University who was not involved in the study, pointed out that fungi don't have to follow rules to exhibit a decay form. "The manuscript is very timely and provides evidence for what many people in the field have suspected for some time – that simple descriptors of wood decomposition do not necessarily reflect the diversity in decay strategies exhibited by fungi," he said. "This is particularly the case when discussing the brown rot wood decay mechanism where distantly related species have evolved superficially similar decay mechanisms. This manuscript uses whole genome sequence information to outline the argument for advancing our understanding of wood decomposition away from a simplistic white versus brown rot dichotomy."

Grigoriev noted this wasn't the first time they'd seen a genome that appeared to blur the definitions between white rots and brown rots. "We thought we saw an anomaly with a previously sequenced white rot fungus Schizophyllum commune," he said. "Now we see a trend. This is the value of having multiple data points and so many fungal genome sequences. Now that it's clear that POD is not the only marker for white rot, we should broaden our search for enzymes that have bioenergy applications. This is the whole point of doing fungal genomics at scale."

For more information about the DOE JGI Fungal Program, view a short video at http://bit.ly/JGI-Fungal-vid.



Micrographs of transverse sections of *Populus* wood decayed by brown rot (top photo) and white rot (bottom photo) fungi. (Images by Benjamin Held, University of Minnesota. Artist interpretation by Zosia Rostomian, Berkeley Lab Creative Services)

#### DOE JGI, UC Merced Launch Graduate Internship Program

Travis Lawrence (third from the right) and Keedrian Olmstead (second from the left) are the first students participating in the DOE JGI/University of California, Merced Genomics Distinguished Graduate Internship Program. For eight weeks, Olmstead and Lawrence collaborated with Axel Visel (third from the left) and Zhong Wang (rightmost), both of whom have adjunct appointments in UC Merced's Quantitative and Systems Biology Graduate Group. Olmstead worked with Ben Cole (leftmost) on a project involving bacteria in Arabidopsis plant roots, while Lawrence worked with

Nicole Johnson (not pictured) on developing an algorithm for RNA-Seq assembly. The Program challenges UC Merced graduate students with individual projects that provide hands-on experience in cutting-edge genomes research and apply experimental and computation tools to solve research problems. Visel and Wang conceived the program and developed it with assistance from Eddy Rubin (second from right), DOE JGI Director, and Juan Meza. Dean of the UC Merced School of Natural Sciences. The Program's oversight committee is comprised of Visel and Wang and UC



Merced School of Natural Sciences assistant professor Suzanne Sindi.

#### In Memoriam: Falk Warnecke

Falk Warnecke was a postdoctoral fellow at the DOE JGI from March 2005 until June 2009. Among the publications that resulted from his work here, he was first author on the termite hindgut metagenome paper that appeared in *Nature*. (More information about the paper at <a href="http://jgi.doe.gov/news\_11\_21\_07/">http://jgi.doe.gov/news\_11\_21\_07/</a>.) Warnecke recently died at the age of 42. On July 16, 2014, Phil Hugenholtz, who previously headed the DOE JGI Microbial Ecology Program and is now Director



of the Australian Centre for Ecogenomics at the University of Queensland, reminisced about his former colleague:

"I heard the sad news of Falk Warnecke's death this morning while attending a conference on insect biology in Cairns, Australia. Falk was a founding member of the Microbial Ecology Program at JGI together with Victor Kunin and Hector Garcia Martin. Although, they did not see eye to eye initially (or speak the same science lingo), they made a formidable team once they had learned to communicate and respect each other's strengths. Falk was the microbial ecologist in the team, trained in the rigorous German tradition. He led the charge on our termite gut microbiome project together with Jared Leadbetter, culminating in the much-cited Warnecke et al paper published in Nature in 2007. (See and read about Falk tracking termites at http://bit.ly/Atlantermite08.) Just this morning in my session on termitemicrobe symbiosis, the Warnecke et al paper featured in three different talks attesting to its ongoing impact in the

Falk was struck down in his prime by a brain tumor shortly after leaving

JGI. What I had mistaken as his adopting the laid back Californian attitude in his final year at JGI was the tumor pressing on his frontal lobe and disinhibiting him. Despite the grim prognosis (5 year average survival), Falk started his own group at the University of Jena in Germany. I met him several times at conferences after that, and not once did he complain about his fate or show any self-pity. He threw himself bravely into his work and hobbies (notably hiking) and made the most of the time he had left. My condolences to his family and friends, and to the field of microbial ecology, which is poorer for his passing."

The termite hindgut project led to Hugenholtz and Warnecke's appearance on *U.S. News & World Report's* 2009 list of "scientists on the cutting edge of energy & environmental research." Learn more at http://bit.ly/09UNWRtermite.

