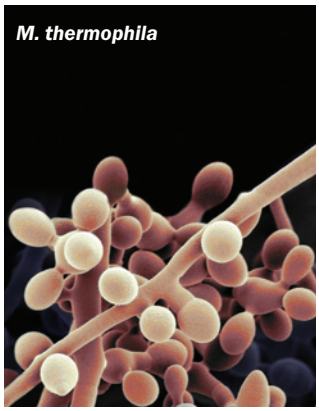
JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

the PRIMER

Fall 2011 Volume 8 Issue 4

also in this issue

Tagging Enzymes That Can Take the Heat



(Courtesy of Adrian Tsang, Concordia University)

"Thermophilic fungi represent excellent hosts for biorefineries where biomass is converted to biofuels."

Many of the enzymes currently being used in biofuel production are derived from species that thrive at temperatures of 20°C–35°C (68°F–95°F). Their inability to tolerate higher temperatures slows the conversion process, and allows contaminants to potentially reduce the final yield. To speed up the conversion process, researchers propose using enzymes that are stable above current working conditions. In a report published online October 2, 2011 in *Nature Biotechnology,* an international team of scientists including DOE JGI researchers compared the finished genomes of *Thielavia terrestris* and *Myceliophthora thermophila*, fungi that thrive in high-temperature environments above 45°C and whose enzymes, active ranging from 40 °C to 75 °C, may therefore be useful to improving the biofuel production process.

The 38.7-million base pair (Mbp) genome of *M. thermophila* and the 36.9 Mbp genome of *T. terrestris,* the first described for thermophilic eukaryotes, encode a multitude of enzymes that decompose biomass material. Their efficacies were tested on feedstocks representing the two major groups of flowering plants and the results suggest that compared to the other fungal cellulases, the enzymes in *M. thermophila* and *T. terrestris* have evolved to efficiently break down and convert biomass into simple sugars at a wide range of temperatures.

"These two thermophiles can be considered all-purpose decomposers with respect to their carbohydrate-active enzymes (CAZymes) and their ability to degrade plant polysaccharides," said study co-author and longtime DOE JGI *continued on page 5*

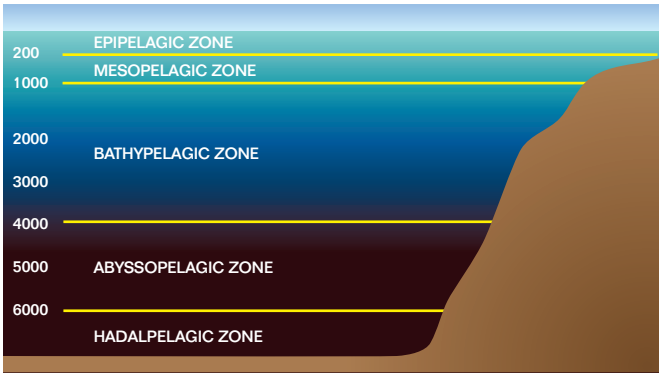
Carbon Fixation in the Dark Ocean

The oceans capture significant amounts of carbon, but the processes involved in their contribution to managing the global carbon budget are still not fully understood. One of the questions that remains only partially answered is how marine microbes process carbon without the aid of the sun. Many marine organisms rely on sunlight to produce the food they need to survive, but the light does not penetrate all the way to the bottom of the ocean.

In the September 2, 2011 issue of *Science*, DOE JGI scientists and longtime collaborators employed single cell sequencing techniques to identify a pathway by which microbes in the "twilight zone" that lies between 200 meters and 1,000 meters underwater capture carbon and derive energy without the presence of sunlight.

"This is the first application of a single-cell genomic approach to the deep ocean, one of the largest and least known biomes on the planet," commented marine scientist David Kirchman from the University of Delaware. "The paper radically changes our view about how microbes gain energy and flourish in the oceans."

Understanding the flow and processing of carbon in the world's oceans, which cover 70 percent of *continued on page 5*



Fall 2011 Volume 8 Issue 4

Tringe makes the "Brilliant 10" List

DOE JGI Metagenome Program lead Susannah Green Tringe is one of 2011's "Brilliant 10." The annual list compiled by *Popular Science* recognizes scientists and researchers under the age of 40 who, according to Mark Jannot, editorial director of the Bonier Technology Group and Editor-in-Chief of Popular Science, "represent the best of what science can achieve and demonstrate American's continuing cutting-edge research." He said that nominations are solicited from hundreds of scientists and the list is cut down to those "whose work really blows the tops of our heads off."

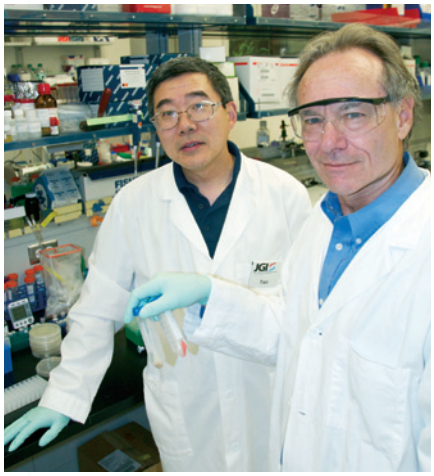
Popular Science recognized Tringe's $2.5 million award from the DOE Early Career Research Program to study wetlands in the San Francisco Bay-Delta region of California in naming her to the list. Her proposal to study the role of microbial communities in restored wetlands and their impact on long-term carbon sequestration over the next five years was one of just 65 accepted from more than 1,100 proposals.

Working with the U.S. Geological Survey, Tringe has already started sampling at Twitchell Island, where, she said, the carbon sequestration rates are estimated to be "comparable to, if not greater than" the carbon sequestration rates achieved from tropical reforestation. Tringe has been involved in several metagenomic studies, including the study of microbial communities in the termite hindgut, the leaf cutter ant and the cow rumen, as well as the microbial communities in ocean dead zones.

photo by David Gilbert, DOE JGI

Tringe wades into the "Brilliant 10" list.

Improving Ionic Liquid Tolerance in Biofuel Production



Tao Zhang (left), and Eddy Rubin (right)

The process of converting cellulosic biomass into biofuels involves a pretreatment process based on the ones used in the paper and pulp industry. Acids and bases break down the biomass and then enzymes make the plant sugars accessible for conversion to fuel.

Ionic liquids, a new class of solvents, have been reported to be much more efficient in treating the biomass and enhancing the yield of sugars liberated from it. However, while ionic liquids are useful for breaking down biomass, they may inhibit the ability of enzymes used in pretreatments to produce sugars for the next step of the biofuel production process. To help identify new cellulases that are tolerant of ionic liquids, a team of researchers led by DOE JGI Director Eddy Rubin and the Vice-President of the Deconstruction Division at the Joint BioEnergy Institute Blake Simmons analyzed the complete genome sequences of halophilic organisms. Their work was reported in the June 30, 2011 issue of *Green Chemistry.*

One of the organisms the researchers focused on was *Halorhabdus utahensis,* an archaeon isolated from the Great Salt Lake and sequenced at the DOE JGI as part of the Genomic Encyclopedia of Bacteria and *continued on page 8*

the PRIMER / 3
Fall 2011 Volume 8 Issue 4

Linking Fungi to Boreal Forests

Steve Kallick, courtesy of the Pew Environment Group

Boreal forest

"It's one of those fungi that everybody knows. It has such an aggressive form of cellulose breakdown," said DOE JGI collaborator Dan Eastwood of Swansea University in the United Kingdom. He was describing *Serpula lacrymans,* the fungus known to homeowners as dry rot and one selected for sequencing in 2007 by the Institute as a potential source of cellulases for commercialized biofuel production.

In the July 14, 2011 issue of *Science Express,* Eastwood led an international team of scientists including DOE JGI researchers in comparing the 48.2 Megabase pair genome of *S. lacrymans*, the second brown rot fungus to have its genome sequenced, against 10 other published fungal genomes. Seven of the genomes used in the comparison were sequenced at the DOE JGI, among them *Postia placenta,* the first brown rot fungus sequenced. The analysis not only allowed researchers to understand the chemical reactions involved in the mechanism by which *Serpula* breaks down cellulose, it also sheds light on the role of brown rot fungi in the development of the largest terrestrial ecosystem — the subarctic cool climate boreal forest — and therefore the fungi's role in the global carbon cycle.

Eastwood said that the ability of wood-decaying fungi in general to break down lignocellulose is linked to the co-evolution of boreal forests and fungi. "If you go back far enough in time to the period when trees were developing," he said "there was no way to break lignocellulose down, which led to the coal seams we tap today. When the fungi figured out how to break down lignocellulose, the coevolution of the fungi and trees kick-started the carbon cycle again. The ancestor of all wood decay fungi we have today was a white rot and it was fungi derived from this ancestor which kick started the carbon cycle and lignocellulose decomposition. The brown rots evolved later from a white rot ancestry and, because they circumvent the lignin and go straight for the hemicellulose and cellulose, they are considered more efficient and is probably why they have been able to dominate boreal forests in more recent times."

The study allowed researchers to compare the gene families involved in the mechanisms by which brown rot break down cellulose and white rot fungi break down both cellulose and lignin. Study senior author Sarah Watkinson of the University of Oxford, however, emphasized the role of brown rot fungi in the global carbon cycle, noting that a third of the carbon sequestered in the soil of boreal forests are composed of the wood residues after the fungi break down the cellulose.

"When it grows in forests, it decays the wood and leaves behind the lignin," she said. "The residues of brown rot fungi contribute up to 30 percent of carbon in conifer forest soils, and conifer forests are a very large biome in the world and one of the largest carbon sinks on land. The activity of brown rot is very significant in global carbon cycling and I think that hasn't really been appreciated before."

Co-senior author Francis Martin from the French National Institute of Agriculture Research (INRA), an integral driver in many of the DOE JGI's fungal genome projects, also stressed the importance of this work in understanding the evolution of forest fungi, calling the dry rot fungus the "'missing link' along the 'saprotrophism-mutualism continuum.'"

DOE JGI Fungal Genomics head and study co-author Igor Grigoriev pointed out that there are now two white rot and two brown rot genomes available, and a dozen more are in queue. He added there are two variants of *Serpula*, the one that causes dry rot and one found in coniferous forests, and the DOE JGI has sequenced both. In fact, the Institute's sequencing efforts account for 40 percent of all fungal genomes deposited in the public databases. "Overall," he said, "we're getting to the point where we can do comparisons across the three fungal lifestyles: saprotrophs, which include wood-degraders, symbionts and pathogens, and we are entering the stage where we find that there are no clear, black and white associations with each category."

Fall 2011 Volume 8 Issue 4

DOE JGI Featured in NYTimes

Set against the backdrop of the nation's push toward significantly boosting renewable fuel supplies by 2022, an article in the July 14, 2011 issue of the *New York Times* focused on the cellulosic biofuels work being done in Northern California. In particular, the piece focused on various projects at the DOE JGI, the Joint BioEnergy Institute (JBEI) in Emeryville and the Energy Biosciences Institute in Berkeley and their industrial applications.

One of the projects prominently featured in the article was the massive-scale DNA sequencing project involving the cow rumen, led by former DOE JGI postdoctoral researcher Matthias Hess, now at Washington State University.

"I set out to really show that you can use sequencing to find something that can be applied," Hess said in the story. "And that's exactly what we did." The project took advantage of the advances in sequencing technology that allowed Hess and his colleagues to generate a quarter-trillion bases of sequence and then analyze the data to identify 30,000 novel, potential enzymes that could be used to break down cellulosic biomass for biofuel production.

Other DOE JGI projects referenced in the article included the genome of fungus and primary industrial source of cellulose-degrading enzymes *Trichoderma reesei* and the termite hindgut metagenome. The shipworm and Tammar wallaby gut metagenome projects were also mentioned as examples of other organisms being studied for cellulases and other, similar enzymes.

The article then went on to discuss how the genome sequences and analyses generated at the DOE JGI can be used farther down the pipeline by researchers and collaborators at EBI and JBEI. One anecdote highlighted the collaboration between the DOE JGI and JBEI that resulted in the identification of enzymes in a bacterium from the Great Salt Lake that can tolerate the hypersaline conditions of biofuel production (further detailed on page 2).

Read the full article at http://nyti.ms/pGqCGc.

Bark beetle damage in Montana's Helena National Forest

U.S. Forest Service – Northern Region/Flickr

Chewing on Xylose Degradation

To chewing gum manufacturers, xylose is merely an ingredient for their products. For biofuel producers however, xylose is a five-carbon sugar that represents nearly half of the available plant sugars. In a paper published online July 25, 2011 in *Proceedings of the National Academy of Sciences,* researchers from the DOE JGI and the Great Lakes Bioenergy Research Center identified genes in yeast that could be used to break down xylose, increasing the conversion of cellulosic biomass to biofuels.

"Strains of yeast that are currently used for biofuel production convert xylose to ethanol slowly and inefficiently, and only do so after all the glucose is exhausted," said University of Wisconsin-Madison postdoctoral researcher and study lead author Dana Wohlbach. "For industrial purposes, the faster a yeast can consume the sugars, the better, since more sugar consumption means more ethanol."

A team of DOE JGI researchers led by Fungal *continued on page 8*

the PRIMER / 5
Fall 2011 Volume 8 Issue 4

Schmutz named Plant Program Lead

Jeremy Schmutz

Jeremy Schmutz at HudsonAlpha Institute for Biotechnology has been tapped to head the DOE JGI Plant Program. The move follows a reorganization at the DOE JGI, where the Plant and Fungal Programs are now under the Eukaryote Program headed by Dan Rokhsar, while the Microbe and Metagenome Programs fall under the Prokaryote Program headed by Nikos Kyrpides.

Schmutz has been involved in several high-profile plant genome sequencing projects at the DOE JGI, including the candidate bioenergy feedstock sorghum, the predominant source of biodiesel soybean and the first wild grass species to be sequenced *Brachypodium distachyon.* His involvement with the Institute dates back to the Human Genome Project, during which time Schmutz and his group finished and assembled the human sequence of chromosomes 5, 16 and 19. He currently leads the informatics and production sequencing groups at the HudsonAlpha Genome Sequencing Center.

Recently, Schmutz and HudsonAlpha colleague (and wife) Jane Grimwood completed their second and third biological collaboration. The first, Mollie, is flanked by new sisters Ella and Rosie, born June 14, 2011 in the image below.

Schmutz's daughters

Taking Heat *continued from page 1*

collaborator Randy Berka of Novozymes in Davis, Calif.

One of the team's findings during the comparative analysis involved the size of one glycoside hydrolase family in the two thermophilic fungi compared with the industry workhorse *T. reesei.* As hydrolysis is the chemical reaction by which simple, fermentable sugars are released from a polysaccharide, the researchers suggest this expansion of the gene family "may have evolved as a modified strategy for deconstruction of biomass polysaccharides compared to that of other species such as *T. reesei.*" In light of these results, the researchers concluded, *M. thermophila* and *T. terrestris* would be very useful to large-scale biofuel production.

The study results are applicable to the United States' Renewable Fuels Standard, which calls for the annual production of 36 billion gallons of biofuel by 2022. One of the biggest hurdles to achieving this goal lies in optimizing the multistep process involved in breaking down plant biomass and then converting it into fermentable sugars that can be refined into fuel.

"These thermophilic fungi represent excellent hosts for biorefineries where biomass is converted to biofuels as an alternative to modern oil refineries," said DOE JGI Fungal Genomics head and study co-author Igor Grigoriev.

"Thermostable enzymes and thermophilic cell factories may afford economic advantages in the production of many chemicals and biomass-based fuels," wrote the team including Berka and Grigoriev, and led by DOE JGI collaborator Adrian Tsang of Concordia University in Canada.

Dark Ocean *continued from page 1*

Earth's surface, is central to understanding global climate cycles. Carbon fixation in the dark ocean where photosynthesis cannot take place has so far been attributed primarily to archaea, but DOE JGI collaborator Ramunas Stepanuaskas, director of the Bigelow Laboratory's Single Cell Genomics Center, said these microbes alone cannot account for the carbon processes in the ocean depths.

"Previous oceanographic models suggested that archaea do not adequately account for the amount of carbon that is being fixed in the dark ocean," he said. "Our study discovered specific types of bacteria rather than archaea, and their likely energy sources that may be responsible for this major, unaccounted component of the dark ocean carbon cycle."

Stepanauskas and his colleagues isolated and identified microbes from water samples collected in the waters of South Atlantic and North Pacific. At DOE JGI, Microbial Program lead Tanja Woyke and her team then sequenced the genomes from the microbial single cells.

"This study represents an excellent example for the use of single cell genome sequencing to decipher the metabolic capabilities of uncultured natural microbial consortia, providing a powerful complement to metagenomics," Woyke said.

The results allowed the team to pinpoint the predominant bacterial lineages capable of trapping carbon in this deep underwater region. The information then led to the characterization of a microbial metabolic pathway that helps solve the mystery of how certain bacteria do this in the dark ocean.

"These previously unrecognized metabolic types of dark ocean bacteria may play an important role in global biogeochemical cycles, and their activities may in part reconcile current discrepancies in the dark ocean's carbon budget," they wrote in their paper.

Fall 2011 Volume 8 Issue 4

JGI IN THE NEWS

A "meraculous" algorithm for whole-genome assemblies

With the advent of sequencing technologies that allow researchers to generate the equivalent of a single human genome in days rather than the decades it took multiple organizations to complete a single one, the bottleneck in genomics research has shifted from sequence production to sequence assembly. For example, the Sanger platform routinely produced reads 700 basepairs long while the Illumina platform generates reads 35-100 basepairs in length, making the assembly process challenging.

In a paper that was published August 18, 2011 in *PLoS ONE,* DOE JGI researchers led by Eukaryote Program head Dan Rokhsar and Jarrod Chapman, have developed an efficient way to do short-read assemblies of eukaryotic genomes using a computer algorithm referred to as meraculous.

As a test case, Chapman and his colleagues used meraculous to assemble 75-bp Illumina reads of the yeast *Pichia stipitis,* previously sequenced by the DOE JGI in 2007 to help pinpoint enzymes in its genome that can help ferment the five-carbon sugar xylose for biofuels production.

"The meraculous assembly reconstructs 95% of the *Pichia* genome in long contigs and scaffolds without any errors," the team wrote in their paper, which includes a link from which the software can be downloaded. "Many stages of the meraculous algorithm are parallelized, and to document their scalability we describe an assembly of simulated data for the ~120 Mbp *Arabidopsis thaliana* genome, and show that for mammalian genomes the limiting memory structure requires less than 10 Gb of RAM."

Arabidopsis thaliana

(HermannFalkner/Flickr)

Mitigating Methane Emissions

Before the Tammar wallaby's complete genome was determined, researchers at the DOE JGI and Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO) collaborated to study the microbial communities in its gut. A cousin of the kangaroo, the wallaby's digestive system has been compared to that of ruminants such as sheep and cows. One bacterium discovered as a result of this work may prove useful in reducing methane emissions of livestock, and therefore of greenhouse gases on a global scale.

In the June 30, 2011 issue of *Science Express,* CSIRO researchers reported that they'd been able to construct a partial draft genome of the microbe WG-1 identified from DOE JGI's metagenome assembly in a previous publication. "Our initial analysis of the metagenomic dataset showed that there were key bacterial and enzyme-based differences between the microbiota present in Tammar wallabies and other herbivores," said study senior author Mark Morrison, a CSIRO Science Leader in Metagenomics who works in their Division of Livestock Industries. "The current Science paper involves our use of computational methods to produce the information needed to isolate [WG-1] and evaluate its potential role in feed digestion and reduced methane emissions," via further experimentation and complete genome sequencing.

Though carbon dioxide is the greenhouse gas most people are familiar with, methane is 20 times more potent. In 2009, the U.S. Environmental Protection Agency calculated that 20 percent of the nation's human-related methane emissions were attributable to livestock digestive processes. In Australia, livestock emissions account for 12 percent of the country's total greenhouse gas emissions.

"We hope that in the next few years, in addition to strategies inhibiting the abundance of methane-producing microbes in livestock, we will have identified how to augment the growth of other bacteria so that feed digestion and fermentation remain optimal and with reduced methane emissions," Morrison said. "We think the research with the Tammar wallabies

continued on page 8

Wallaby

Mehgan Murphy, Smithsonian's National Zoo

the PRIMER / 7
Fall 2011 Volume 8 Issue 4

JGI IN THE NEWS

Growing a mitochondrial genome

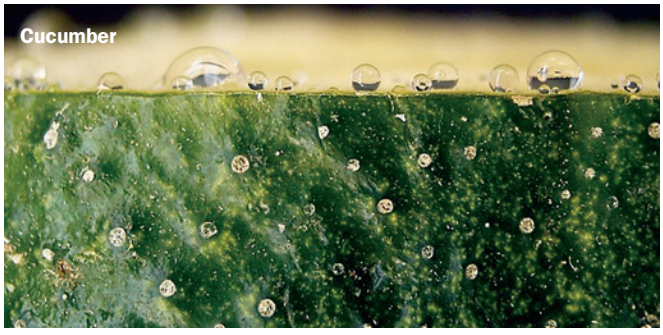
The mitochondria is the powerhouse of the cell and in plants the mitochondrial genome contains the genes that encode essential cell metabolism processes. The size of the mitochondrial genome varies widely among plants by as much as an order of magnitude.

To understand these variations, a team of researchers including DOE JGI's Kerrie Barry compared the mitochondrial genome of cucumber, one of the largest known mitochondrial genomes at nearly two million base pairs in size and comparable to the size of a bacterial genome, with the smaller mitochondrial genomes of watermelon (379 Mbp) and zucchini (983 Mbp).

"We help solve a 30-year mystery about the origins of its large size by showing that it mainly reflects the proliferation of dispersed repeats, expansions of existing introns, and the acquisition of sequences from diverse sources, including the cucumber nuclear and chloroplast genomes, viruses, and bacteria," wrote the team led by Jeffrey Palmer of Indiana University. "These data provide insights into the growth of plant mitochondrial genomes and novel perspectives on the pattern and process of intragenomic recombination in plant mitochondria." The results indicated that recombination dynamics among chromosomes help account for the range of mitochondrial genome sizes and why some repeat sequences appear asymmetrically compared to others.

Palmer was one of the researchers who originally wrote the proposal to have the cucumber genome sequenced as part of the DOE JGI's 2005 Community Sequencing Program portfolio. The project focused on understanding the unique properties of mitochondrial genomes, one of the last great frontiers of comparative sequencing, by studying several dozen seed plants including the cucumber, melon and mung bean.

"This work explains how the cucumber mitochondrial genome got so large and provides new tools for examining the complex dynamics of these cool genomes," noted *Plant Cell* science editor Jennifer Mach in an editorial feature on the research. Both pieces appeared in the July 2011 issue of *The Plant Cell.*



(Frank Vincentz/Wikimedia Commons)

An Inventory of Plant Proteins

In plants and algae, the plastid is the part of the cell where many metabolic processes such as light reaction and carbon fixation take place. To determine which proteins are involved in these functions, DOE JGI bioinformaticist Simon Prochnik and collaborators at UCLA and the Carnegie Institution at Stanford University employed phylogenomics on 20 plant and algal genomes — including *Volvox,* spikemoss, and many others sequenced by the DOE JGI — to compile an inventory of nearly 600 proteins conserved in plants and algae. The work appeared in the June 17, 2011 issue of *The Journal of Biological Chemistry.*

The result of the team's efforts is a resource known as GreenCut2. "We present this annotated inventory of 597 proteins as a resource for functional analyses of plant-specific biochemistry," they wrote.

Prochnik and his colleagues culled the list of proteins with the help of a computer algorithm also called GreenCut2 that can distinguish proteins involved in photosynthesis present in plant and green algal genomes, but not in non-photosynthetic organisms. They found that well-conserved proteins are more likely to have been studied in a reference organism, while other proteins may be associated with functions more specific to certain eukaryotes. Comparative analyses of the proteins in this inventory is opening windows for discoveries about the roles that these proteins play in photosynthetic cells, the evolution of chloroplasts, and how photosynthetic cells might be tailored for survival under different environmental conditions.

Jing-Ke Weng, Salk Institute

8 / the PRIMER
Fall 2011 Volume 8 Issue 4

DOE JGI hosted the *Guillardia theta* and *Bigelowiella natans* Jamboree held September 7-9, 2011. The algal event brought together more than two dozen collaborators to explore the genomics and biology of these microbial eukaryotes.

Ionic Liquids *continued from page 2*

Archaea (GEBA) project. The team cloned and expressed a gene from *H. utahensis* in another haloarchaeal microbe, identifying a salt-dependent enzyme that can tolerate high temperatures and is resistant to ionic liquids. "This is one of the only reports of salt-tolerant cellulases, and the only one that represents a true 'genome-to-function' relevant to ionic liquids from a halophilic environment," said Simmons. "This strategy enhances the possibility of identifying true obligatory halophilic enzymes." He added that the collaboration between the two Institutes "has established a very important link between genomic science and the realization of enzymes that can handle very demanding chemical environments, such as those present in a biorefinery."

Wallaby *continued from page 6*

have provided us another group of bacteria to target."

The Tammar wallaby's gut microbiome was sequenced by the DOE JGI under the 2007 Community Sequencing Program portfolio in part to compare how plant biomass is degraded by the marsupial gut microbiota compared to other gut microbiota. "This new work builds on the growing knowledge portfolio of DOE JGI gut metagenome projects targeting biomass-degrading organisms, such as the termite hindgut metagenome and the recent cow rumen metagenome [published in *Science* earlier this year]," said DOE JGI Metagenome Program lead and study co-author Susannah Tringe. (Read about Tringe's inclusion in this year's "Brilliant 10" list on page 2.) "The DOE JGI is uniquely positioned, working in concert with our community of collaborators, to apply the power of genomics to important societal issues."

Xylose *continued from page 4*

Genomics head Igor Grigoriev sequenced the genomes of two types of fungi that rely on xylose-rich bark in the habitats of bark beetles. The work, added Wolbach, offers a genomic toolset that can improve the conversion efficiency of yeast on cellulosic biomass to produce biofuels. The information allowed the scientists to identify the differences between non-xylose fermenting yeasts and xylose-fermenting yeasts.

"By comparing the genome sequences and expression patterns of many yeasts — rather than just looking at one — we were able to identify elements common to all xylose-fermenting yeasts and elements absent from non-xylose fermenting yeasts," she said. The team then introduced several genes into the yeast *S. cerevisiae,* allowing it to consume xylose. They found that the introduction of the gene CtAKR in particular significantly increased xylose consumption, enabling yeast to more efficiently metabolize xylose to biofuels.

Contact The Primer
David Gilbert, Editor / DEGilbert@lbl.gov

CSO 22314