Form to Function Defines CSP 2013 DOE JGI Portfolio

In architecture, the term “form to function” refers to designing a building to best serve its intended purpose. Genomics researchers could well adopt the term to refer to the ongoing transition from studying only the genetic code of an organism, to also understanding what roles those genes play in the biology of the organism that encodes them.

Many of the 29 projects selected for the 2013 Community Sequencing Program (CSP) portfolio out of 108 submissions to the U.S. Department of Energy Joint Genome Institute (DOE JGI) combine sequence data generation with large-scale experimental and computational capabilities to enable fuller functional genome annotation. The projects will allow the DOE JGI to study in greater depth the genomes and activities of organisms that have potential applications in energy and environment. One approach focuses on RNA transcripts (RNASeq) that represent genes “turned on” under different circumstances, and examining the circumstances under which they are required, giving hints about what they do. Another involves transposon mutagenesis and sequencing (TnSeq), a method of generating large sets of random mutations in target DNA.

The organisms to be studied in the CSP 2013 portfolio are found in environments that run the gamut from a microbe found nearly two miles below the earth’s surface in a South African gold mine, to bacteria in the coastal waters (continued on page 4)

Carving out a niche in forest carbon

As the world’s most cultivated mushroom, the button mushroom is a popular item in grocery stores. In nature, unlike brown-rot and white-rot fungi, Agaricus bisporus is a known decayer of partially decomposed litter on forest floors and grassland soils rich in humus, a mix of soil and compost. Humus contributes the chemicals that drive the decomposition process, adding organic matter to deficient soils and contributing to overall plant health to foster root vitality and stimulate the growth of beneficial microbial communities in the soil.

“Compared to genomes of these fungi, that we previously characterized, Agaricus fits neither brown-rot nor white-rot classifications and its adaptation to (continued on page 5)
DNA Synthesis: The Write Stuff

STEPHEN TUNG

In a way, DNA synthesis works like sequencing in reverse. The process starts with a target nucleotide sequence from a database that is then synthesized from commercially available oligonucleotides (short DNA fragments). Scientists combine these custom fragments of DNA together until a full physical genetic sequence is realized.

Sam Deutsch, head of the DNA Synthesis Group at the DOE JGI, says his team harnesses the genomic information in publicly-accessible databases. DNA synthesis is especially useful for identifying the function of genes when the amount of genetic material is scarce or limited, as in the case of samples provided in metagenomics and single-cell genomics, Deutsch says.

"Only a handful of people might have access to certain metagenomics samples, for example from a deep-sea hyperthermal vent, or a permafrost core," he says. "Without the original sample, although the sequence may be available in databases, there’s no guarantee that you’re actually going to be able to identify and study the functions of particular genes."

In those cases, being able to recreate a physical version of a genetic sequence from a database of genomic information is critical for scientists investigating the function of genes deemed of interest for bioenergy and environmental applications. Deutsch’s team can also exhaustively generate different combinations that would be tedious to make individually, allowing users to explore how a suite of genes and their orthologs can work together.

Deutsch’s group can also optimize the sequence of DNA before it is generated to make it more compatible with the vector and host, thus improving both the chances of successful synthesis as well as proper function of the sequence in its new context. As such, Deutsch says, the DOE JGI offers a unique one-stop-shop from sequence annotation to synthesis to function that’s not found in other academic or commercial labs.

Nevertheless, Deutsch’s group is not in competition with companies.

"From the CSP proposals submitted annually, we’re trying to identify projects that wouldn’t be easily accomplished by companies," he says. “We want to do things that add an additional dimension.”

“In some experiments, we generate 100,000 combinations, for example, from a pool of 50 genes," says Deutsch. “This provides researchers with a powerful tool to find the optimum combination of genes for a specific function such as breaking down biomass at a specific temperature, or fix carbon or nitrogen.”

Deutsch’s group is trying to push the capabilities of DNA synthesis by making longer and longer sequences. This fiscal year, they made a 30-kilobase (Kb) molecule as a proof-of-concept. They’re currently working on two 50-kilobase molecules to serve as vectors for making multi-trait transgenic plants.

Germany’s seven-person group accepts proposals from researchers at the Bioenergy Research Centers anytime, and from the DOE JGI’s annual call for CSP projects (see page 1 for the recently-announced CSP 2013 portfolio). “Users will provide their sequences before the start of the fiscal quarter, and we should be able to complete them before the end of the quarter.”

Producing large DNA molecules is particularly challenging. It works as a stepwise process in which 1 Kb building blocks are generated, these are then be assembled together into 10 kb fragments, which can be further combined to produce 50 Kb molecules or beyond. But in every assembly step they know that only about one third of the clones are correct. “Every time we

(continued on page 7)
White-rot wipes out coal-forming era

Coal, the fossilized remains of plants that lived 360 to 300 million years ago, generated nearly half of the electricity consumed in the United States in 2010, according to the U.S. Energy Information Administration. However, a seemingly innocuous change may have had a massive impact on the carbon cycle, bringing an end to the 60-million year Carboniferous period.

“We’re hoping this will get into the biology and geology textbooks,” said Clark University biologist David Hibbett, senior author of the June 29, 2012 study in *Science* that compared the complete genomes of dozens of species of fungi, many were sequenced at the DOE JGI. The evidence presented by Hibbett and his colleagues suggests that white rot fungi capable of breaking down lignin evolved 290 million years ago. This timeframe was calculated through molecular clock analysis, which assumes that genetic divergence is proportional to time, and is calibrated by fungal fossils. Prior to this period, there were brown-rot fungi that could only break down cellulose.

“The concept of the invention of an enzyme that can break down the ‘unbreakable’ is really great,” said Kenneth Nealson, Professor of Earth Sciences and Biological Sciences at the University of Southern California.

Study co-author Igor Grigoriev, head of the DOE JGI Fungal Genomics Program, said that this paper is the first product of the Genomic Encyclopedia of Fungi, the DOE JGI umbrella. “This paper is the first chapter in the Encyclopedia,” he said, “The data generated has produced the most comprehensive catalog of lignocellulolytic enzymes yet, which is of interest to industry. We’ve now got the blueprint of all genes across very diverse phylogenies, and we’ll get more. This is a huge step forward.”


The rhizosphere paper made the cover of the July 28, 2012 issue of *Nature*.

Scanning electron micrograph of wood being decayed by the white-rot fungus *P. strigos-zonata*. (Robert Blanchette, University of Minnesota)

The roots of a plant microbial network

The microbial communities in, on and around plant roots fight pests and manage carbon and other soil nutrients, ultimately contributing to plant health and growth. Despite this, much about processes by which the microbes act remain unknown.

Researchers from the DOE JGI and the University of North Carolina, Chapel Hill, led a team in digging to the root of plant-microbe interactions, identifying key microbial players surrounding and in the roots of *Arabidopsis thaliana*.

The study identified more than 750 operational taxonomic units (genetically distinct groups of microbes, similar to species) in soil and plant samples. Scientists were then able to extrapolate the metabolic functions of these groups of microbes.

“Understanding the rules that guide formation of the root microbiome are likely to contribute significantly to the success of agriculture and our understanding of the carbon cycle,” said study senior author Jeff Dangl, UNC’s John N. Couch Professor of Biology. “Science has long been fascinated with the spectrum of relationships between plant and microbe that span from pathogenic to mutually beneficial. With our results we are adding new details to this complex landscape.”

Although the microbial communities associated with a plant vary, this study offers a valuable data point of how plants and microbes work together. For instance, instead of spraying pesticides or fertilizer treatments, farmers could rely on a combination of microbial species to best improve productivity in a particular crop growing in a certain soil type.

“We can’t really know a plant genome’s full functional capacity until we also understand the functional capacity and the drivers governing assembly of its associated microbiome,” said co-author Susannah Tringe, head of JGI’s Metagenome Program.
“Form to Function” (continued from page 1)

off the west Antarctic Peninsula, to crops that could be grown in several regions across the United States and provide a renewable, sustainable source of biomass for biofuels, to fungal pathogens that pose a threat to energy crop yields.

Two DNA synthesis projects have been selected for the CSP portfolio. (Learn more about the DOE JGI’s DNA Synthesis program on page 2.) One of the two DNA synthesis proposals selected for the CSP 2013 portfolio came from Jef Boeke of Johns Hopkins University, who is heading the Synthetic Yeast genome project known as Sc 2.0. Baker’s yeast (Saccharomyces cerevisiae) is one of the most common microbial “platforms” for the production of fuels and other chemicals and is playing a key role in bioenergy research. The Sc 2.0 project intends to design, construct, and replace the native genome of baker’s yeast with a fully synthetic version. As part of Sc 2.0, the DOE JGI will synthesize chromosome IV, the largest of the 12-million base genome at half a million basepairs.

Included in the five plant projects selected is the Gene Atlas Pilot Project led by Gary Stacey, the Director of the University of Missouri Center for Sustainable Energy. The project aims to develop a comprehensive index of gene expression for several plant species deemed DOE JGI Flagship plants.

“Now that we have generated these sequences at great expense, we need to add information to convert them into functional models,” the researchers noted in their proposal.

The expression responses of the alga Chlamydomonas reinhardtii, soybean (Glycine max), the moss Physcomitrella patens, poplar (Populus trichocarpa), and foxtail millet (Setaria italica) will be studied under a variety of environmental conditions, and the team also plans to study nitrogen metabolism with another five DOE JGI Flagship plants and three comparative model species.

From the seven fungal projects approved, one proposal aims to do for them what the ENCODE project is working toward for the Human Genome Project. Led by University of California, Berkeley researcher N. Louise Glass, the Fungal Nutritional ENCODE project aims to comprehensively map out the nutritional and metabolic regulatory networks of Neurospora crassa, which grows on decaying plant biomass, in order to identify and improve the productivity of cellulases that can be harnessed for industrial biofuel production.

Another project is the Mycorrhizal Genomics Initiative, led by Francis Martin of the French National Institute for Agricultural Research (INRA). The team plans to study the transcriptomes in several dozen species of fungi that form symbiotic relationships with plants to learn more about the mechanisms by which these interactions take place. Symbiotic fungi provide plants with nutrients and water and play key roles in managing carbon and nitrogen levels in the forest ecosystems.

Many of the 15 microbial and metagenome projects selected call for the use of single-cell genomics to study hard-to-culture microbes and metatranscriptomics to focus on the portion of the genome that encodes gene expression.

Berkeley Lab earth scientist Janet Jansson is the lead on a CSP project connected to DOE’s Next-Generation Ecosystem Experiment (NGEE), which aims to learn more about the response of the arctic ecosystem to the changing climate so that scientists can develop better simulations for climate change models. The CSP proposal focused on the genomics of microbial communities isolated from samples cored out of the permafrost in... (continued on page 5)
“Form to Function” (continued from page 4)

Barrow, Alaska. These frozen soils store a large amount of carbon, and rising global temperatures have also raised concerns about the potential outcomes when the permafrost thaws and the trapped carbon could be released into the atmosphere.

For more information about the NGEE project, watch ORNL's Stan Wullschleger discussing the project at the DOE JGI’s Annual Genomics of Energy and Environment Meeting at http://bit.ly/JGI7wNGEE.

During the pilot project known as the Genomic Encyclopedia of Bacteria and Archaea (GEBA), more than 250 uncultivated microbes from less explored branches of the Tree of Life were sequenced. In the CSP 2013 portfolio, the Functional Encyclopedia of Bacteria and Archaea (Feba) project led by Berkeley Lab’s Adam Deutschbauer plans to employ a high-throughput approach known as TnSeq to determine the gene functions in the genomes of 40 DOE-mission relevant microbes.

“While advances in high throughput DNA sequencing have made microbial genome sequencing easy, the functional annotation of genomes remains extremely hard,” Deutschbauer and his colleagues noted in their proposal. “There is therefore a great demand for novel high-throughput approaches to determine gene function from phylogenetically diverse microbes.”

The full list of projects that make up the CSP 2013 portfolio, supported by the DOE Office of Science, is available at http://www.jgi.doe.gov/sequencing/cspseqplans2013.html. The DOE JGI Community Sequencing Program also accepts proposals for smaller microbial and resequencing projects on a quarterly basis. For more information about the annual and quarterly CSP calls for proposals, go to http://www.jgi.doe.gov/CSP/index.html.

“Carving out a niche” (continued from page 1)

growing in a leaf-litter humic-rich environment is not typical of classic wood-degrading fungi,” said DOE JGI Fungal Program head Igor Grigoriev. The DOE JGI is among leading worldwide contributors of fungal genomes to the public databases, having sequenced over 150 fungal genomes, providing a vital computational infrastructure for such large-scale comparative analyses.

Working with Francis Martin, head of the ‘ARBRE’ Lab of Excellence at the French Institute INRA, and his colleagues at the DOE JGI, Grigoriev led an international collaboration of institutions in reporting the button mushroom’s genome, which first appeared online on October 8, 2012 in the Proceedings of the National Academy of Sciences.

“Our hypothesis was that metabolic strategies and niche adaptations of Agaricus might not be present in the white-rot and brown-rot wood-decomposing fungi,” study senior author Martin.

The researchers surveyed the genomes and the transcriptomes or subset of genes expressed under particular conditions of two A. bisporus lines, a commercial strain and related wild variety. The analysis revealed that several families of well-known sugar-degrading enzymes similar to the repertoire found in wood-decaying fungi are also found in Agaricus. However, high levels of enzymes such as heme-thiolate peroxidases and etherases in Agaricus predominate in the presence of humus-rich soil habitats, suggesting a higher ability to metabolize complex mixtures of derivatives of lignin and other polymers.

The report shows how the button mushroom’s genes are actually deployed not only in leaf decay but also wood decay and in the development of fruiting bodies (the above ground part of the mushroom harvested for food). The work also suggests how such processes have major implications for forest carbon management.

“The ability to use proteins prevalent in soil confers an advantage to Agaricus over other fungal scavengers,” said Martin. “To our knowledge, Agaricus had not been shown in nature to decompose wood. Yet, we now see how Agaricus has adapted to growing in this ecological niche. Our understanding of the carbon cycling role of Agaricus in ecosystems is a prerequisite to modeling and optimizing carbon management for sustainable forests.”

The comparative analysis also revealed a dozen other genes that are dialed up during mushroom formation. “Key master switches may be manipulated to control fruiting body formation—the mechanisms triggering the complex cascade that leads from undifferentiated mycelia, the mass of branching, thread-like fingers, to the button mushrooms most commonly consumed” said Martin.

The Agaricus genome was originally proposed by Mike Challen while at the University of Warwick and was sequenced under the auspices of the DOE JGI’s Community Sequencing Program. Challen is now at the Wellcome Trust Centre for Human Genetics, University of Oxford, UK.
Recap: EMSL’s Integration 2012

LOEL KATHMANN

The Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility at Pacific Northwest National Laboratory, held its annual user meeting in August 2012. With the theme “Integration 2012: Discovery at the Intersection of Biology, Energy, and Environment,” the meeting focused on advances in the biological sciences and brought together nearly 100 registrants from academia, national laboratories and user facilities, as well as industry.

“It was great to see our user community coming together at EMSL to share their knowledge and to strengthen our collaborations and partnerships. For EMSL, collaboration with researchers from across the scientific community is so important and our meeting is crucial to both forming and maintaining teams,” said Scott Baker, Lead for EMSL’s Biological Interactions and Dynamics Science Theme.

Keynote speaker Jay Keasling, CEO of the Joint BioEnergy Institute spoke to the intersection of biology, energy, and biotechnology research and about the technologies that have enabled them to be a world leader in plant genomics. Pennacchio also discussed the directions the DOE JGI is going; for example, its focus on single-cell studies of extremophiles: “By focusing on extremes, we think we can identify organisms that otherwise wouldn’t be culturable.” As extremophiles are tied to environmentally relevant roles such as bioremediation, the DOE JGI’s single-cell genomics studies, he said, will offer a better understanding of these microbes and the roles they play as well as help fill in the phylogenetic tree.

Attendees also heard from EMSL’s sponsor, represented by Todd Anderson of the DOE’s Biological and Environmental Research (BER) program. Anderson emphasized the power of the integrated, collaborative, and iterative core research model that brings together computational tools, namely KBase, and experimental tools, such as proteomics and genomics technologies at EMSL and the DOE JGI, to move forward research in bioenergy, carbon sequestration, environmental remediation, and synthetic biology.

Three breakout sessions offered users an in-depth and interactive perspective on state-of-the-art biology tools at EMSL that are enabling new scientific discoveries. The Cell Isolation and Systems Analysis session covered approaches for visualizing and analyzing transcriptomics data as well as EMSL’s super resolution fluorescence microscopy and fluorescence lifetime imaging tools. In the Advancements in Helium Ion Microscopy session, attendees were given a hands-on demonstration of EMSL’s helium ion microscope—the first HIM available to the user community. The Advancements in Mass Spectrometry session covered a broad range of mass spectrometry topics, including proteomics experimental design, the strength of top-down proteomics, and incorporating metabolomics into an EMSL user proposal.

Integration 2013 is right around the corner. Registration will open in Spring 2013 through www.emsl.pnl.gov.

Loel Kathmann is a communications specialist at EMSL.
“DNA Synthesis” (continued from page 2)

assemble fragments we expect that only one out of every three clones is going to be correct,” Deutsch says. “And every successive assembly round gets harder, the cloning efficiency drops, the sequencing verification gets harder, the handling of the molecules themselves gets harder. Everything gets harder.”

But despite its difficulty, his group has met with success. After only starting in March of last year, orders are growing substantially. “We just finished our first year of production... we had about 12 users,” Deutsch estimates. “This year we’ll probably go up to about 20.”

They’re looking forward to the next-next-generation sequence machines, which can read segments of DNA on the order of 10 Kb pairs at a time. Doing so would be a boon for the verification step. Another area they want to focus on is working with smaller quantities of samples and reagents. “Right now we work on the microliter scale; we want to reduce that to the nanoliter scale using nanofluidics,” says Deutsch. “That would substantially reduce costs and improve our efficiency.”

In the future they hope to focus on “systems biology” approaches, whereby different data types (expression, transposon bombing, metabolic fluxes) are used to decide which synthetic DNA modifications are most likely to have an impact on cellular physiology. These studies will involve the synthesis of larger DNA molecules from single operons to genome wide modifications. However, he says the field is progressing rapidly. “These things are moving so fast,” Deutsch says. “In our second year we’re already doing a lot of the operon shuffling, and we’re starting a pilot project for modifying the entire sequence of an organism,” both years ahead of schedule.

“...The vast amounts of sequencing data being generated at JGI and other places allow you to make hypotheses at the gene, genome and environment levels but those cry out for validation,” Deutsch says. “A few years ago, there were only a few hundred thousand predicted genes in databases. Nowadays, a single project will generate a million new gene predictions. That’s such a rich potential to find new functions. Our synthesis capability will help position JGI to start testing these predictions.”

IMG Version 4 Released

The Integrated Microbial Genomes (IMG) system, a community resource for analysis and annotation of genome and metagenome datasets in a comprehensive comparative context, has been upgraded. IMG supports the DOE JGI’s Microbial Genomics and Metagenomics Programs and version 4.0 consolidates the individual databases so that microbial genomes and metagenomes will both share a single integrated database. Both Programs will retain a publicly accessible portal (IMG for microbial isolates and IMG/M for metagenomes) and a password-protected portal (IMG/ER and IMG/M-ER) for collaborators who are not yet ready to release their data.

The IMG system is a crucial piece of the international effort led by the DOE JGI to generate the Genomic Encyclopedia of Bacteria and Archaea (GEB) and is the means by which all the data produced are immediately released to the community, fully annotated. To fill in the large gaps regarding microbial diversity in the Tree of Life, in 2007, the DOE JGI embarked on the two-part pilot phase of the GEB project, completing 250 microbial genomes from branches about which very little or nothing at all were known. The next phase was begun under the DOE JGI 2012 Community Sequencing Program (CSP) project known as 1,000 Microbial Genomes or KMG, and the follow-up KMG II was approved as part of the 2013 CSP portfolio.

DOE JGI Prokaryote Super Program Head Nikos Kyrpides said that both GEB and KMG fall under the Microbial Earth Project (MEP), which he calls “the overall umbrella project that aims to generate a genome sequence for all type strains.”

Version 4 of the IMG system is dedicated to the memory of Dr. Iain Anderson (see page 8).
IN MEMORIAM

Dr. Iain J. Anderson

The DOE JGI community was deeply saddened to learn that Dr. Iain J. Anderson unexpectedly passed away on October 14, 2012. Iain had worked at the DOE JGI for over seven years and was a major contributor to the IMG and GEBA projects. Iain received his Ph.D. in biochemistry from the University of Wisconsin-Madison. He spent two years doing postdoctoral research on methanogens in Barry Whitman’s lab at the University of Georgia. Iain joined Integrated Genomics Inc. where he was responsible for the curation of transporters and the analysis of microbial genomes. He then moved to The Institute for Genomic Research analyzing protist and fungal pathogens. He joined the DOE JGI’s Genome Biology program in 2005 to lead the Archaeal diversity genomes project. He was also responsible for the curation of amino acid metabolism data. Iain was also a technical reviewer for the DOE JGI Community Sequencing Program for 2012 and served as an instructor at the Microbial Genomics and Metagenomics workshops. Iain was a great team member who worked tirelessly behind the scenes to keep the Microbial and Metagenome programs moving forward. While he was a quiet individual, Iain’s contributions spoke volumes. He is survived by Jean, his wife of 15 years, and their daughter Chloe.

Confirmed Speakers:
Eric Allen, UC San Diego
Rick Amasino, Univ. of Wisconsin
Greg Bell, ESnet, Berkeley Lab
Paul Blaimley, Broad Institute
Eoin Brodie, Berkeley Lab
Penny Chisholm, MIT
Joe DeRisi, UC San Francisco
Nicole Dubilier, MPI, Bremen
Tom Gilbert, Univ. of Copenhagen
Sam Hazen, Univ. of Massachusetts
Eric Karsenti, EMBO Heidelberg
Zander Myburg, Univ. of Pretoria
Jack Newman, Amyris
Wayne Reeve, Murdoch University
Bob Schmidt, SG Biofuels
Chris Voigt, MIT

For more information and to register:
http://www.jgi.doe.gov/meetings/usermeeting/