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

the PRIMER
Winter 2013
Volume 10 Issue 1

Sequencing a fungal class to help prevent crop losses

A major DOE JGI Fungal Program Initiative is the Genomic Encyclopedia of Fungi, which employs a comparative genomics approach across a range of targets. Methodical exploration of fungal phylogenetic and ecological diversity by genome sequencing can help identify new metabolic pathways and enzyme activities. One critical application is “biorefinery” methods to convert biopolymers such as cellulose into simple sugars and then into biofuels.

Plant health is also critical for sustainable growth of biofuel feedstocks and can be influenced by symbionts, pathogens, and biocontrol agents. In 1970, for example, the Southern corn blight destroyed roughly 15 percent of the United States’ crop, with some states losing their entire harvest. As a result, corn prices rose by nearly 25 percent and prices for other grains rose as well due to increased demand. The economic losses were officially estimated at around $1 billion, roughly equivalent to $6 billion by today’s standards.

The pathogen behind the devastating crop disease was a fungus *(Cochliobolus heterotrophus)* that belongs to the *Dothideomycetes* class, which includes one of the largest groups of plant pathogens that infect nearly every major crop used for food, fiber or fuel. From a bioenergy perspective, plant pathogens could impact the nation’s goal of producing renewable fuels from sustainable biomass feedstocks.

Though outbreaks at the scale of the Southern corn blight have not been reported in recent years, many researchers at various genome centers have been sequencing fungal genomes from this class with the hope that their efforts could help mitigate such economic losses in the future. In 2008, several of these researchers converged at the DOE JGI to discuss these individual efforts, consolidate genomic data for comparative analysis, and develop larger-scale sequencing projects for the DOE JGI Community Sequencing Program.

The collaboration resulted in a large-scale comparative analysis of several *Dothideomycete* fungi. In a series of five papers that appeared in *PLoS Pathogens, PLoS Genetics* and *G3,* an international consortium of researchers described the genomes and biology of several *Dothideomycetes* including plant pathogens of switchgrass and other candidate feedstock crops.

The central publication linking all of the articles appeared in the December 6, 2012 issue of *PLoS Pathogens.* In this article, an international consortium led by DOE

(continued on page 6)

also in this issue

Unraveling clues for cotton fiber improvements

True to the advertising slogan, the cotton plant has woven itself into people’s lifestyles in various colors and textures. In the United States alone, more than 200,000 domestic jobs are related to cotton production and processing, with an aggregate influence of about $35 billion on the annual U.S. gross domestic product. The cotton fiber grown is valued at about $6 billion per year, with cottonseed oil and meal byproducts worth nearly another $1 billion.

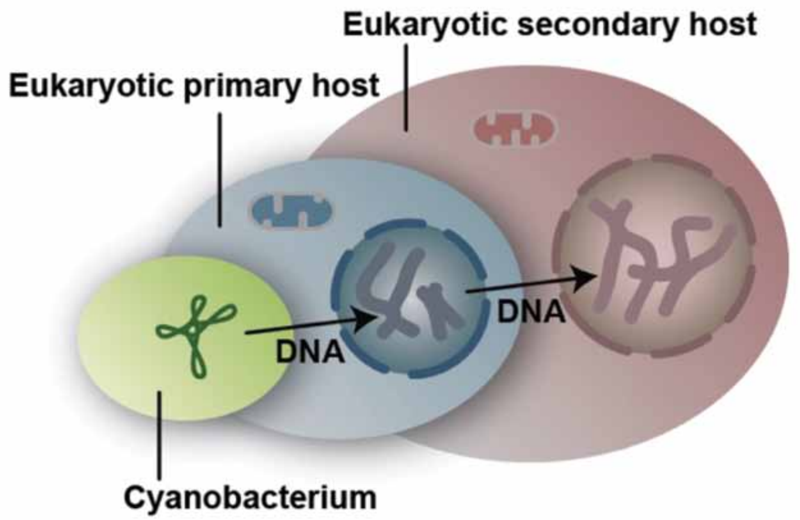
However to bioenergy researchers the most important property of cotton is its fiber composition. Each cotton strand is composed of more than two dozen coils of cellulose, a target biomass for next-generation biofuels.

In the December 20, 2012 edition of *Nature,* (http://bit.ly/NatCotton) an international consortium of researchers from 31 institutions including a team from the U.S. Department of Energy Joint Genome Institute (DOE JGI) presented a high-quality draft assembly of the simplest cotton *(Gossypium raimondii)* genome, a species indigenous to the Americas.

(continued on page 3)

Cotton field. (David Nance, USDA)

Proof of Plastids on “Permanent Loan”



Primary and secondary endosymbiosis is depicted in this diagram, which shows movement of DNA from the cyanobacterial progenitor of plastids to the primary host nucleus and, subsequently, to the nucleus of the secondary host. (John M. Archibald, Dalhousie University, Canada)

Many of the world’s most important photosynthetic eukaryotes such as plants did not develop the ability to combine water, solar energy and carbon dioxide. Instead, they gained the capability by indirectly stealing the chloroplasts or light-harnessing organelles from other organisms. Through a process called primary endosymbiosis, the microbe or archaeon containing these plastids were engulfed by a eukaryotic host that gained the photosynthetic ability. Through secondary endosymbiosis, a process thought to occur more frequently, this eukaryote was then engulfed by another. In an example of this event, the chloroplasts of red and green algae have subsequently come to reside within other, previously non-photosynthetic eukaryotes, transferring these properties on to their new host.

The processes of primary and secondary endosymbiosis have contributed to the global diversity of photosynthetic organisms that play a crucial role in regulating and maintaining the global carbon cycle. In most organisms that acquired photosynthesis by this mechanism, the nucleus from the ingested algal cell has disappeared, but in some cases it persists as a residual organelle known as a nucleomorph. Such organisms have four distinct genomes.

To better understand the process of secondary endosymbiosis and why nucleomorphs persist in some organisms, an international team of researchers at 27 institutions, including the DOE JGI, collaborated to sequence and analyze the genomes and transcriptomes or expressed genes of two tiny algae. The team led by John Archibald of Canada’s Dalhousie University published their findings on *Bigelowellia natans* and *Guillardia theta* on December 6, 2012 in *Nature.* (http://bit.ly/Sfjhcz)

The DOE JGI sequenced the genomes of *B. natans* (95 Mb) and *G. theta* (87 Mb) from single cell isolates provided by Bigelow Laboratory for Ocean Sciences as part of the 2007 Community Sequencing Program portfolio. In addition, the transcriptomes were sequenced separately by the National Center for Genome Resources in New Mexico. DOE JGI Fungal Genomics Program Head Igor Grigoriev, one of the study’s senior authors, called *B. natans* and *G. theta* “living fossils” because of the remnant nucleomorph. He added that algae are relevant to the DOE Office of Science research portfolio for their potential applications in the fields of bioenergy and environment, noting that the DOE JGI has published over 75 percent of the publicly available algal genomes.

“Iterations of endosymbiosis have led to a global diversity of these primary producers,” he said. “Sequencing these two algae, the first cryptophyte and the first chlorarachniophyte sequenced, helped us to fill in the gaps in the Eukaryotic Tree of Life, and obtain additional references for better understanding of eukaryotic evolution.”

Archibald noted that both algae have “a surprisingly complex suite of enzymes involved in carbon metabolism.” Having their genomes offers scientists additional resources for studying photosynthesis and biofuels development.

Among the team’s findings is an answer to the question of why nucleomorphs still exist. “The reason for the persistence of nucleomorphs in both organisms appears to be surprisingly simple: they are no longer able to transfer their DNA to the host cell nucleus by the process of endosymbiotic gene transfer,” said Archibald. Unlike most other secondarily photosynthetic eukaryotes in which the endosymbiont’s genetic matter has completely migrated over to the host, in cryptophytes and chlorarachniophytes the nucleus and chloroplast from the engulfed algae remain partitioned off from the host cell. “As a consequence,” he added, “genetic and biochemical mosaicism is rampant in *G. theta* and *B. natans.”* Both researchers highlighted the unexpected finding of alternative splicing in *B. natans.* Grigoriev noted that the phenomenon is one typical for higher eukaryotes, and Archibald added that the levels "greatly exceed that seen in the model plant *Arabidopsis* and on par with the human cerebral cortex, unprecedented and truly remarkable for a unicellular organism. This challenges the paradigm that complex alternative splicing is a phenomenon limited to sophisticated multicellular organisms."

"The evolution of chloroplasts, the photosynthetic compartments of plants and algal cells, is complex but has had a profound effect on our planet," said Chris Howe, Professor of Plant and Microbial Biochemistry at Cambridge University in England. "As well as providing a goldmine of information on the general biology of these organisms, the paper shows us that the nucleomorph genomes have probably persisted simply because the mechanism for transfer of genes to the nucleus was closed off, rather than because nucleomorphs had to be retained as separate entities."

"Unraveling clues for cotton fiber improvements" *(continued from page 1)*

"This cotton data will help accelerate the study of gene function, particularly cellulose biosynthesis, the understanding of which is fundamental to improved biofuels production," said Jeremy Schmutz, head of the DOE JGI Plant Program and a faculty investigator at the HudsonAlpha Institute for Biotechnology, who led the effort to sequence and assemble the genome. He talks about the cotton genome project on YouTube at http://bit.ly/JGI-cotton-video. "In addition, the unique structure of the cotton fiber makes it useful in bioremediation, and accelerated cotton crop improvement also promises to improve water efficiency and reduce pesticide use."

Schmutz also said this is a good example of a project in which the DOE's genomic contributions were matched by resources from the research community.

These include genetic maps, RNA sequencing, additional genomic sequence and detailed genomic analysis, allowing a more detailed and meaningful interpretation of the results.

The cotton plants seen growing in typical fields in the U.S. are polyploid hybrids of two types of cotton (cotton A and cotton D). As one of the closest extant relatives of the tetraploid cotton genome, the diploid *G. raimondii* was selected for sequencing in part because it has a smaller genome and fewer repetitive elements than A-genome cotton and is much less complex than the polyploid cotton. Having data from multiple genomes for reference, such as those provided by USDA for this study, enabled the team to trace cotton's lineage, the evolution of hybrids and the gene duplications that allowed fiber development.

The DOE JGI's contribution of sequencing and assembling the 760-million basepair genome stems from a Community Sequencing Program proposal led by University of Georgia professor Andrew Paterson. "This study represents the first time that a polyploid plant was compared to its progenitors over the entire genome," he said. "This study reveals evolutionary processes salient to all plants and provides a strategy to better understand the genome of many other crops, such as canola, wheat, and peanut."

Learning more about the genetic contributions of the D- and A-genomes to the common cotton species can help researchers improve fiber traits. Don Jones, Director of Agricultural Research at Cotton Incorporated, said this *G. raimondii* gold standard genome will be the foundation for sequencing upland cotton, *G. hirsutum,* that makes up most of the worldwide field crop. Another species, *G. barbadense,* produces Pima cotton but accounts for less than two percent of the cotton crop. "This sequence is a cornerstone that will help advance our knowledge so we more thoroughly understand the biology that leads to enhanced yield, improved fiber quality, and better stress tolerance, all improvements that will benefit growers in the not-too-distant future."

DOE JGI Plant Program Head Jeremy Schmutz, out standing in a cotton field, was one of the recipients of the 2012 Cotton Biotechnology Award, presented at the Plant and Animal Genome (PAG) Conference held January 12-17, 2013 in San Diego, Calif. This year's recipients formed the core group of scientists who, with Schmutz, led the effort that resulted in the publication of what has been recognized as the "gold standard" sequence of the simplest cotton-D genome in a December 2012 issue of *Nature.* The other scientists who received the award were: Andrew Paterson, Director of the Plant Genome Mapping Laboratory at the University of Georgia; Jonathan Wendel at Iowa State University; Dan Peterson, Director of the Mississippi State Institute for Genomics, Biocomputing and Biotechnology; and Daniel Rokhsar, DOE JGI Eukaryote Super Program Head and Chief Informatics Officer. (Holly Ralston, Director of Communications, HudsonAlpha)

Cassava brief: the problem and the genomics approach

MTAKAI NGARA

What keeps Mtakai Ngara and Teddy Amuge up at night? Thinking about cassava. These young, ambitious, researchers working at the International Institute for Tropical Africa (IITA) just outside Nairobi, Kenya are learning more about genomics to help breed more effective cassava to feed hungry mouths in their native Africa and further afield. To feed their passion for genomics, they made a trip to the DOE JGI and UC Berkeley last November for a two-week study visit.

Cassava *(Manihot esculenta),* also known as manioc, tapioca or yuca, a herbaceous root tuber, is an important source of calories for 800 million people within the tropics of Africa, Latin America and Asia. Due to its minimal nutrient requirements and flexibility in harvesting times, low-income smallholder farmers in Africa grow cassava. "With certain varieties, farmers can generate decent yields in drought and disease prone regions and infertile soils," Ngara pointed out, "hence earning it the epithet 'poor man's crop.'"

Yield improvement forms the centerpiece of a number of cassava studies at the moment. As a remarkable energy source, improving production will contribute towards meeting immediate food security and potentially future biofuels needs.

Currently, average cassava production is estimated at 15 metric tons per hectare (6.7 U.S. tons per acre), which is six times less than the projected potential of 90 metric tons per hectare) and one of the critical constraints to its yield is disease. "Cassava brown streak disease (CBSD) is caused by a monopartite RNA virus from *Potyviridae* family," said Amuge, whose Ph.D. focuses on understanding the virus, which she noted is one of the leading threats to cassava production in Africa.


(Teddy Amuge, IITA)

The existence of cassava varieties that possess tolerance to CBSD provide an invaluable opportunities for establishing resistant varieties through breeding and/or transformation using genomics-based approaches as well as understanding the disease biology and developing high-resolution markers for the trait.

Genetic hybridization between domesticated and wild varieties of cassava has been hypothesized to be responsible for the introduction of genes for CBSD tolerance from the wild relatives into some of the farmer-grown lines. Two wild cassava relatives and tree cassava genomes have been sequenced with the aim of identifying genomic regions that may be involved in this process. This is a research area Ngara would like to develop a Ph.D. thesis on, as he explores options for graduate school.

Albert and Namikonga are two of the established varieties for CBSD susceptibility and tolerance respectively. Based on a low-resolution genetic map, the approximate genomic regions (known as quantitative trait loci or QTLs) underlying CBSD tolerance have been anchored on the reference genome and annotated. A larger (high resolution) mapping population has recently been established with the intent to genotype the markers by sequencing reduced representation of the populations' genomes. "Understanding CBSD at the molecular level is key to its management," said Amuge, who is about halfway through her Ph.D. She has set up a time-course RNA sequencing experiment based on CBSD-infected Albert and Namikonga. She hopes this study will unravel the molecular interactions that underpin CBSD, potentially revealing the key to engineering resistant varieties.

In collaboration with a team from the DOE JGI, UC Berkeley and Dow AgroSciences, a total of 15 whole genome sequences have been generated from farmer-grown varieties using Illumina's paired-end sequencing. "We have identified a whole-genome repertoire of SNPs and INDELs," said Simon Prochnik, their DOE JGI host. "Genomic resources like these applied to modern breeding


DOE JGI's Simon Prochnik with a cassava in French Polynesia. (courtesy of Simon Prochnik)

methods such marker assisted and genomic selection provide hope for breeding tolerance to CBSD and other diseases as well as important agronomic traits. It's exciting that Teddy and Mtakai are using their newfound knowledge now that they are back in Africa. And I have a feeling they learned a lot more than we expected on their visit."

The group will play a critical role in the partnership known as the Next Generation Cassava Breeding project, a five-year, $25.2 million research effort supported by the Bill & Melinda Gates Foundation and the Department for International Development of the United Kingdom, coordinated by Cornell University.

Mtakai Ngara is an IITA bioinformaticist working with the cassava genome.

(Left to right) Mtakai Ngara, Teddy Amuge and Jessen Bredesen. (David Gilbert, DOE JGI)

Adding another Archaeal phylum

Perhaps the most enduring scientific legacy left by University of Illinois microbiologist Carl Woese when he died on December 30, 2012 is the discovery of what has been accepted as the third branch of the Tree of Life: *Archaea.* In a 1980 essay for *Scientific American,* Woese, the recipient of a National Medal of Science in 2000, laid out the line of reasoning that led to the seminal paper published in *Proceedings of the National Academy of Sciences* just a few years earlier arguing for a third kingdom aside from Bacteria and Eukarya:

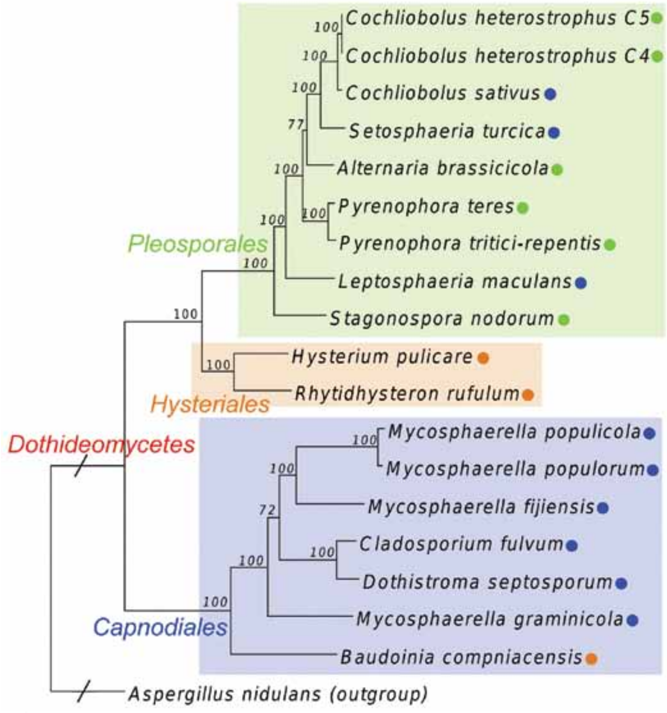
"Initially some biologists rejected out of hand the notion of a 'third form of life,'" he recalled in his article. "How could some thing that looked like a bacterium not be a bacterium and indeed not be related to bacteria? In time the simplicity of our argument and the accumulation of evidence prevailed."

In the three decades since Woese's paper was published, several more archaeal phyla have been recognized, many of them through microbial field studies conducted in extremophilic environments such as Yellowstone National Park in Wyoming.

A report published online November 15, 2012 in *The ISME Journal* (http://bit.ly/iJW12Geo) now suggests the presence of still another phylum, Geoarcheota. The work was led by longtime DOE JGI collaborator Bill Inskeep of Montana State University, and was based on metagenomic studies conducted on acidic iron-oxide microbial mats from the 100 Spring Plain at the Norris Geyser Basin in Yellowstone. The samples from the microbial mats were sequenced at the DOE JGI as part of a larger project under the aegis of the Community Sequencing Program focused on several sites in Yellowstone. The *de novo* assemblies revealed a novel archaeal group as a dominant member of the microbial mat community, one that did not fit into any existing *Archaea* phylum.

Carl Woese (Wikimedia Commons)

Sequencing a fungal class to help prevent crop losses *(continued from page 1)*



Genome-based phylogenetic tree of 18 *Dothideomycetes* computed using 51 conserved protein families. Lifestyles and strategies of pathogenesis (green circle for necrotrophs, orange circle for saprotrophs and blue circle for [hemi]biotrophs) are indicated. *Aspergillus nidulans* was used as an outgroup and its branch on the tree is not drawn to scale. (Ohm RA et al. *PLoS Pathog.* 2012 Dec;8(12):e1003037.)

JGI Fungal Program Head Igor Grigoriev and U.S. Department of Agriculture scientist Steven Goodwin compared 18 fungal genomes, 14 of which had been sequenced during the project.

The team found that while the 18 genome sequences had a wide range of sizes due to repeats, they shared core genes and their lifestyles were reflected in the gene sets: “For example, sets of genes involved in carbohydrate degradation and secondary metabolism are expanded in necrotrophs,” wrote DOE JGI’s Robin Ohm, the first author of this study, and his colleagues. “Many genes involved in pathogenesis are located near repetitive sequences, which are believed to speed up their evolution. Blocks of genes with conserved gene order were identified across most of sequenced *Dothideomycetes* and expressed during infection of one of them suggesting their ancient origin and role in pathogenesis. In addition to this we deduced the mechanism for mesosynteny, a type of genome evolution, particular to *Dothideomycetes,* which reshuffles genes within boundaries of chromosomes.”

The central paper was followed by more focused studies with an emphasis on functional genomics. One of them, from the teams of Gillian Turgeon from Cornell University and Shaobin Zhong from North Dakota State University, reported a detailed analysis and evolution of secondary metabolism genes producing chemicals critical for plant infection within *Cochliobolus* species including those that caused the 1970 the Southern corn blight.

Another team led by University of Hawaii at Manoa’s Yangae Cho demonstrated the functional role of transcription factors regulating virulence genes in *Alternaria bracicicola* genome, explored in studies on model plant *Arabidopsis thaliana.*

In another article, researchers led by Rosie Bradshaw at New Zealand’s Massey University and Pierre de Wit from the Netherlands’ Wageningen University, compared the genomes of two fungal plant pathogens, one which causes leaf mold on a tomato and another that causes needle blight on pine species around the world. At first glance, the diseases appear unrelated but they are both caused by closely related *Dothideomycetes* species. The researchers found that both genomes share genes responsible for both diseases but adaptation to different hosts and lifestyles resulted in rearrangement and different regulation of these genes, highlighting the importance of functional genomics. The pine needle blight can infect at least 80 species of conifers around the world and cause devastating economic losses, nearly U.S. $20 million in losses annually, in New Zealand alone.

The pathogen *Pyrenophora tritici-repentis (Ptr)* causing tan spot of wheat was the focus of study by Linda Ciuffetti from Oregon State University. In 2012, when harsh weather conditions

Infected needles on Austrian pine. (Robert L. Anderson, USDA Forest Service, Bugwood.org)

threatened the wheat harvest in Kansas, the nation's single largest producer of the grain, wheat prices rose by 10 percent as a result. In North Dakota, which ranks second after Kansas in wheat production, tan spot caused by the *Dothideomycete* species and other diseases can reduce the crop by 10-15 percent annually. The direct cash value of North Dakota's wheat crop was $2.2 billion in 2010, and generated an additional $4.5 billion in indirect commercial activity. Comparing different *Ptr* strains revealed a highly-adaptable genome that allows the influx of transposable elements (TEs), segments of DNA that can change their locations in a genome and even move to other genomes, increasing the pathogenicity of *Ptr* and in part accounting for its increased impact in recent decades. "In fact, the ways in which TEs can affect the structure and evolution of individual genes appear to be almost unlimited," they wrote.

The comparative analyses conducted on such a large scale affords researchers insights into fundamental questions regarding fungal lifestyles, evolution and adaptation to diverse ecological niches, information crucial toward applying fungal genomics to challenges in energy and environment such as plant pathogenicity and biomass conversion. "However, even with this unprecedented scale of genomic studies of plant pathogens we are only scratching the surface," Grigoriev concluded. "Broader and deeper genomic sampling of *Dothideomycetes,* found on every continent and in a wide range of environmental extremes, as well as other important groups of fungi will continue under the DOE JGI's 1000 Fungal Genomes project."

References:

- Ohm RA et al. doi: 10.1371/journal.ppat.1003037. (http://bit.ly/OhmDcp1)
- Cho Y et al. doi: 10.1371/journal.ppat.1002974. (http://bit.ly/ChoDc2)
- Condon et al. doi: 10.1371/journal.pgen.doi/1003233 (http://bit.ly/ConDcp3)
- de Wit PJ et al. doi: 10.1371/journal.pgen.1003088. (http://bit.ly/deWDcp4)
- Manning VA et al. doi:10.1534/g3.112.004044 (http://bit.ly/MannDc5)

Adding another Archaeal phylum *(continued from page 5)*

"Analysis … reveals a dominant archaeal population with novel strategies for energy conservation and oxygen respiration," Inskeep and his colleagues reported.

"Evidence is provided here that these organisms are members of one the earliest currently known lineages of the domain Archaea, and that members of this phylum- level group are limited primarily to thermophilic habitats."

Norris Geyser Basin in the Yellowstone area — algae on left, bacteria on right. (Daniel Mayer, Wikimedia Commons)

THE 2013 DEPARTMENT OF ENERGY
JOINT GENOME INSTITUTE (DOE JGI)

Genomics of Energy & Environment Meeting

March 26-28, 2013 / Walnut Creek, CA

Scientists interested in learning about state of the art genome sciences and associated technologies as well as their potential applications to challenges in bioenergy and environmental issues are invited to participate in the 8th Annual DOE Joint Genome Institute Genomics of Energy & Environment Meeting. This international gathering offers invited presentations, workshops and tutorials on sequence-based bioinformatics, data management systems and new genomic technologies, as well as poster sessions and facilities tours.

Short talks will be chosen from submitted abstracts.

Topics include:

- DNA Synthesis & Synthetic Biology
- Application of Single-Cell Genomics in Microbial Ecology & Bioprospecting
- Genomic Analysis of Biofuel Traits in Biomass Feedstocks
- High-Performance Computing for Next Generation Sequencing Applications
- Biodiversity Surveying Functional Metagenomics
- Metaproteomics of Marine Bacterial Symbionts
- Massive Scale Data Transfer
- Analyses of Root-Nodulating Bacteria
- and more.

Confirmed Speakers:

- Eric Allen, UC San Diego
- Rick Amasino, University of Wisconsin
- Greg Bell, ESnet, Berkeley Lab
- Eldredge Bermingham, Smithsonian Tropical Research Institute
- Paul Blainey, Broad Institute
- Eoin Brodie, Berkeley Lab
- Penny Chisholm, MIT
- Joe DeRisi, UC San Francisco
- Tom Gilbert, University of Copenhagen
- Sam Hazen, University of Massachusetts
- Eric Karsenti, EMBO Heidelberg
- Zander Myburg, University 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/

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

U.S. DEPARTMENT OF ENERGY Office of Science

CSO 24650

@ PAGXXI

Brian Davison, (left) the DOE Systems Biology Knowledgebase (KBase) Community Outreach Team Co-lead and Chief Scientist for Systems Biology and Biotechnology at Oak Ridge National Laboratory (ORNL), with RTI researcher Safiyh Taghavi (center) and Senior Director of the RTI Center for Agricultural and Environmental Biotechnology Niels van der Lelie (right) at the KBase booth during the recent Plant and Animal Genome (PAG) XXI Conference held January 12-16, 2013 in San Diego, CA. (Brian Davison, ORNL)