DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

the PRIMER

Spring 2012 Volume 9 Issue 2

Genomics Moving Beyond One Size Fits All

Notes from the 7th Annual DOE JGI Meeting

also in this issue

With a record 488 genomics researchers and bioinformaticians sitting in the Marriott Walnut Creek ballroom, *New York Times* science writer Carl Zimmer opened the DOE Joint Genome Institute's 7th Annual Genomics of Energy & Environment Meeting on March 20, 2012.

Sharing his thoughts about being "On the Genome Beat," he informed the audience that he was worn down by seeing news about scientists successfully sequencing yet another genome sequence, and that "maybe some genomes shouldn't be written about."

Zimmer's words engaged the audience in a discussion that ranged over the entire Meeting, with an "informal response" delivered in the closing keynote by Steven Benner from the Foundation for Applied Molecular Evolution (more details on the keynote addresses on page 2). Certainly the researchers that presented over the next two days on a slew of genomic projects worked to convince him and other sufferers of Yet Another Genome Syndrome (YAGS) of the value of their work, a question that Zimmer noted was often left unaddressed.

The presentations represented the four program areas of the DOE JGI. Collaborator Jody Banks of Purdue University discussed the *Selaginella moellendorffii* (spikemoss) genome, which had been selected as a DOE JGI Community Sequencing Program project in 2005 and was published in *Science* last year. One of the benefits of having the sequence, she said, is that bioenergy researchers can use the information to learn more about lignin development in plants. As the second most abundant polymer on Earth, breaking down lignin in plants is an obstacle in commercially producing biofuels from plant biomass.

Loren Rieseberg from the University of British Columbia talked about the challenges of sequencing the sunflower genome, noting that it is the only major crop domesticated in the United States. While the sunflower genome is roughly 3.6 Gigabase pairs in size and highly repetitive, he said the appeal in sequencing it lies in its potential as a biofuels feedstock.

(continued on page 4)

Realizing the Promise of Genomics

DOE JGI Director Eddy Rubin (left) and Carl Zimmer (right)

Closing keynote speaker Steven Benner

Delivering the opening keynote before an audience composed primarily of the researchers whose work he covers while "on the genome beat," *New York Times* science writer Carl Zimmer talked about the first time he realized the power of genomics.

He'd been working on the proofs of his 1999 book *"At the Water's Edge,"* which involved a story on how whales had evolved when molecular biologists used DNA studies to show that the whale's closest land-based relative was the hippo, which was completely different from what paleontologists had argued for years.

"What's amazing to me is how little DNA it took to just change the way people thought about this question," he said. "We're at this amazing point now where there have been an awful lot of genomes sequenced."

He then added that the sheer number of genomes published over the past decade — more than 3,000 projects from the DOE JGI alone — has led to his self-diagnosis as someone with the condition known as "YAGS" or "Yet Another Genome Syndrome."

YAGS, he said, resulted from the deluge of genome announcements in the news, with very little detail as to why these projects matter. This is not to say, he added, that genome studies have no benefit. He shared the anecdote of covering a story in Germany where the genome of the microbe involved had been sequenced in the few days it took to set up interviews with those involved.

"It's been fascinating to me watching genomes mature in terms of medicine but with maturity comes humility." There's an impression, he said, that "the sequence of a particular genome in addition to being a huge accomplishment, which it is, is also a game-changing, headline-making story." He suggested that scientists take into consideration that perhaps not all of their work is newsworthy.

Biochemist Steven Benner from the Foundation for Applied Molecular Evolution presented his closing keynote with a nod to Zimmer's comments. He recognized the "anti-genome sentiment in biology" but also noted the difference between the scientist's perspective and the public's perspective of genomic studies.

Unlike Zimmer and his colleagues, he said, scientists aren't trained to find that hook to get the public's attention or to speculate on possible benefits. "Look at genome sequencing from the chemistry perspective," he added. "German chemists of the 19th century did the chemistry of natural biostructures very well without knowing there would be a sequenced human genome."

Benner also responded to the point Zimmer had raised about genome publications that had no discussion of their significance. He said that to truly understand the importance of the plethora of genome sequences, they should be considered from the perspective of planetary biology. Don't look at how long it is taking for the human genome project to realize the promise of personalized medicine, he said. Rather, look at the bigger picture and consider that human and global histories can be read in the genomes of various species, from dating the first beer ever brewed to charting evidence of climate change.

In the first example of his self-described long answer to Zimmer's keynote, Benner tracked the emergence and duplication of genes involved in fermentation in the yeast genome, then correlated the data to the rise of genes associated with alcohol tolerance in humans. His second anecdote involved correlating climate change with enzyme behaviors

(continued on page 8)

Spring 2012 Volume 9 Issue 2

Building Communities to Handle Big Data

During his presentation at the Genomics of Energy & Environment Meeting, Cold Springs Harbor Laboratory's Michael Schatz took a few minutes to remind the audience of the dramatic advances in sequencing technology that have led to the rise of several large-scale sequencing projects as researchers take advantage of the data deluge.

In separate presentations, representatives from two such large-scale projects — Adam Arkin from Berkeley Lab and Stan Wullschleger from Oak Ridge National Laboratory (ORNL) — described projects funded by the Department of Energy and invited the Meeting attendees to be active members of these research communities.

Project lead Wullschleger talked about the Next-Generation Ecosystem Experiments (NGEE), a series of experiments being conducted in the field and in the lab by several national labs and other institutions to better understand the response of the Arctic permafrost to the changing climate as seen from a variety of perspectives including hydrology, vegetation and energy. Armed with this information, he said, scientists could concentrate on developing improved simulations for climate change models.

One of the experiments Wullschleger described was a collaboration between researchers at DOE JGI, Berkeley Lab and the U.S. Geological Survey. As detailed in a report published in the journal *Nature* last year, the work focused on the microbial response to thawing permafrost.

Despite talk of improving computer models, the NGEE project is unarguably focused on data generation, with a significant amount through sequencing. As Schatz noted in his talk, however, there is still a bottleneck as computational tools needed to make sense of the data have not scaled-up with the sequencing technologies.

"What's required to turn those data into information are pure computational algorithms," he said.

These algorithms, he added, are important not just for *de novo* genome assembly, but to anchor variations of a reference genome, and in order to conduct studies in phylogeny and evolution among others. "The foundations of genomics will still be observation, experimentation and interpretation," Schatz said, "and technology will still push the frontier." He added that these technologies would lead to a corresponding rise in quantitative and computational demands.

Calling the Department of Energy the "stewards of large science," Berkeley Lab's director of the Physical Biosciences Division Adam Arkin talked about the plan to develop a DOE Systems Biology Knowledgebase (Kbase), another collaboration between several national labs and other institutions, and its potential role as for a community comprised of professional computational biologists, data generators and basic analysts, knowledge seekers and knowledge generators.

"Think of us as kernel developers," he told the audience. "It's your data and your tools we want to see in this framework." The system, he added, would leverage both ESNet and Magellan cloud infrastructure to access and store large-scale data.

Arkin said that with expertise in large-scale coordination of resources and scientific programs, scalable computational analysis and coordination with individual genome projects the DOE JGI "could become the central organizer of large-scale sequencing projects." One potential task of the DOE JGI, he added, could be to integrate and standardize omics data from multiple data sources, and everyone in the audience would be part of the user group improving the models.

Stan Wullschleger talking about the NGEE project

Adam Arkin discusses Kbase

To learn more about the NGEE, go to http://ngee.ornl.gov/. Wullschleger's talk is available online at http://bit.ly/JGI7NGEE.

More details about Kbase can be found at http://genomicscience.energy.gov/compbio/. Arkin's talk is online at http://bit.ly/JGI7Arkin.

The 7th Annual DOE JGI Meeting

(continued from page 1)

Shawn Kaeppler from the University of Wisconsin-Madison discussed work being done at the Great Lakes Bioenergy Research Center on maize and switchgrass and how their yields might be improved for bioenergy projects by concentrating on factors such as seed weight.

Richard Sayre from Los Alamos National Laboratory and the New Mexico Consortium talked about the poor conversion efficiencies of plants in converting solar energy, a hurdle that he said needs to be overcome in order to harness crops for biofuels production. His suggestion: consider microalgae as a candidate bioenergy feedstock given its higher oil yield and biomass production compared to land-based crops.

Siobhan Brady from the University of California (UC) Davis talked about the development of a root gene regulatory network based on studies using the model plant system *Arabidopsis*.

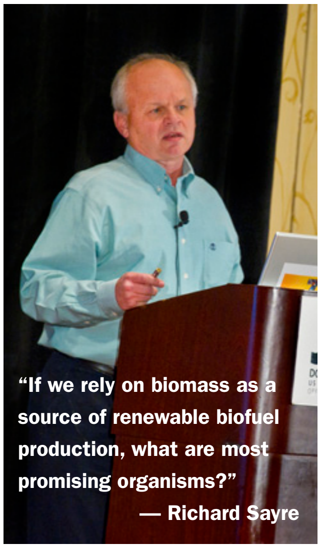
Several speakers were selected from the poster sessions. Wellington Muchero from Oak Ridge National Laboratory presented on genes involved in cell wall biosynthesis while Annika Mosier from UC Berkeley talked about enzymes that can break down complex carbohydrates in fungi and DOE JGI's Patrick Shih discussed the large-scale project to study cyanobacterial diversity known simply as Cyano-GEBA.

David Hibbett from Clark University talked about how mushrooms changed the world, though few people seem to realize it. "When I look at a picture of the forest, I think, where are the fungi?" he said. "It is a blind spot in our thinking about the world." He focused on wood-decaying fungi, both brown rot and white rot, and associated their development with the introduction of lignin-degrading organisms millions of years ago. He concluded with an image that prominently featured dinosaurs. "If I've been moderately successful," he said, "the next time you see dinosaurs from the Triassic period, you'll think, where are the mushrooms?"

N. Louise Glass, an Energy Biosciences Institute (EBI) affiliate at UC Berkeley, talked about the importance of identifying enzymes from filamentous fungi to break down plant biomass for bioenergy applications. Her talk focused in particular on the model fungus *Neurospora crassa* and the mechanisms by which its enzymes can sense and then break down plant cell walls.

Not all of the talks focused on genomic studies for bioenergy, however. For example, citrus breeder Fred Gmitter talked about improving citrus varieties through genomics, noting that several varieties are not in fact true species as most people might expect. "Citrus has incestuous genes. Nothing is pure," he quipped. He traced the origin of the grapefruit back to the pummelo, as well as referencing the first citrus genomes released last year on the DOE JGI's web portal Phytozome.

"Citrus has incestuous genes. Nothing is pure."
— Fred Gmitter



"If we rely on biomass as a source of renewable biofuel production, what are most promising organisms?"
— Richard Sayre

From UCLA, Robert Goldberg talked about the agricultural challenge of doubling the world's food supply on less land. Genomics, he said, allows researchers to find the genes to improve yield traits in crops, while also mitigating the effects of stresses caused by drought or diseases.

Eske Willerslev from the University of Copenhagen described how he and his team used next generation sequencing technologies and DNA from ancient human hair to trace the origin of people in the high arctic region, following the expansion of humans from Siberia into Alaska and Greenland.

With so many users in attendance, DOE Office of Biological and Environmental Research Program Manager Dan Drell led a session (see page 7) asking the audience to consider capabilities they would like to see the DOE JGI add on. The existing draft 10-year strategic vision document was discussed, with input from DOE JGI Director Eddy Rubin and user-led contributions from Sayre, Berkeley Lab's Adam Arkin and Stanford University's Christina Smolke.

Arkin followed up on his talking points from the Strategic

"The next time you see dinosaurs from the Triassic, you'll think, where are the mushrooms?"
— David Hibbett

Vision session with a longer presentation about the DOE's plan to develop a systems biology knowledgebase that is intended to be an open-source repository that researchers can use to access data sets and analysis tools, as well as share similar resources with the rest of the scientific community.

Smolke also backed up the call she'd made for more synthetic biology collaborations with the DOE JGI during her own, longer presentation on RNA control devices and their potential uses as, among other things, real-time metabolite sensors.

Michael Schatz from Cold Springs Harbor Laboratory focused on the developments in sequencing technology, and how the data deluge has led to a corresponding increase in quantitative and computational demands. (More details on the large-scale sequencing projects discussed at the Meeting on page 3.)

Echoing Schatz's talk on the promise and challenges offered in this age of large-scale sequencing, DOE JGI Deputy Director of Genome Technologies Len Pennacchio took a few minutes to introduce a new user program known as the Technology Development Pilot Program, meant to be a complement to the Institute's annual call for CSP proposals. (See page 7 for more details.)

Stan Wullschleger from Oak Ridge National Laboratory discussed another large-scale project to improve Arctic climate models through a multidisciplinary study of terrestrial ecosystems known as the Next-Generation Ecosystem Experiments (NGEE).

During the Meeting's final day, the DOE JGI showed a video produced by Berkeley Lab and featuring Metagenome Program lead Susannah Tringe and her DOE Early Career Research grant project to study microbial communities in the restored wetlands of California's San Joaquin Delta.

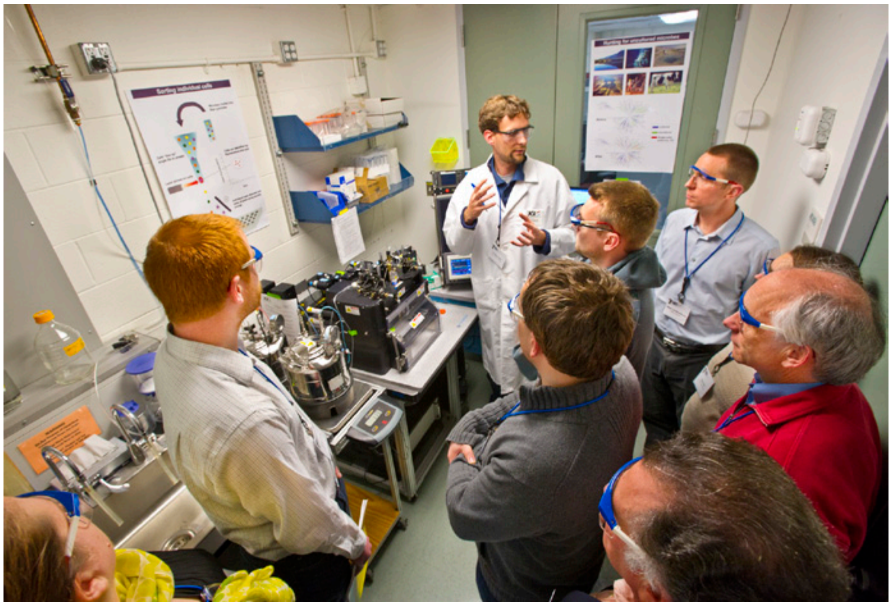
With so much of the microbial world still unknown, it was unsurprising that many talks from the Meeting focused on microbes and microbial communities. Justin Gallivan from Emory University's Center for Fundamental and Applied Molecular Evolution presented a talk on reprogramming bacteria to recognize new molecules, seek them out and possible destroy them. One example of a target molecule he said is atrazine, a herbicide so widely used that the amount sprayed on crops annually is at least five times greater than the mass of gold ever removed from Earth.

Many of the microbial talks focused on marine projects (see page 6). For example, Ramunas Stepanauskas of the Bigelow Laboratory for Ocean Science discussed his Institute's single-cell genomics pipeline as well as the recently published collaboration between Bigelow and the DOE JGI on microbes in the dark ocean and their role in the global carbon cycle. Stephen Giovannoni from Oregon State University talked about ocean bacteria that have streamlined their genomes, and what benefits might have afforded them. Finally, Matt Sullivan from the University of Arizona gave a talk about ocean viruses and their impact on global carbon cycling, referencing the two Community Sequencing Program projects he has with the DOE JGI.

The Meeting closed with a different perspective from Zimmer's keynote, delivered by Steve Benner from The Foundation for Applied Molecular Evolution. While current genomics research does not provide enough information yet regarding its significance or merit, Benner said that the true value of the data would become clear when more genome projects are considered from a "planetary biology" perspective.

Many of the talks from the Genomics of Energy & Environment Meeting have been posted on the DOE JGI's YouTube channel. They can be viewed at http://bit.ly/JGIUM7videos.

Zooming in on the Small Stuff



DOE JGI's Christian Rinke shows off the flow cytometer used for studies in single-cell genomics.

Just as advances in sequencing technologies have allowed researchers to work on large-scale projects of increasing complexity, they have also made it possible to focus on the denizens of the microbial world, whose identities and role in regulating global processes remain largely unknown. One of the presenters at the Genomics of Energy & Environment Meeting who focused on marine microbes was Stephen Giovannoni of Oregon State University. He talked about *Pelagibacter*, a bacterial genus in the SAR11 clade, which includes some of the most abundant marine organisms.

Giovannoni said *Pelagibacter* species also rank among the smallest microbial genomes, and his talk focused on how genome expansion has been suppressed over time. One of the current theories suggests the *Pelagibacter* genome has been streamlined because these microbes have stricter food requirements compared with others. Another theory to explain the small genome is that the microbe exploits genes from marine organisms with which they have formed symbiotic relationships.

One technique that is becoming increasingly popular for studying uncultured microbes is single-cell genomics. At the DOE JGI, the number of projects that focus on sequencing and assembling genomes from a single microbial cell has grown more than 20-fold over the past three years. Many of these projects are done in collaboration with the Bigelow Laboratory for Ocean Sciences, and at the Genomics of Energy & Environment Meeting, Ramunas Stepanauskas, director of Bigelow's Single Cell Genomics Center, talked about a joint effort that was published in *Science* last year about understanding the roles of marine microbes in relation to the global carbon cycle.

Stepanauskas also talked about Bigelow Lab as a single-cell genomics facility, funded through user fees. "It's a waste of resources not to be available to community as a whole," he said. He described the differences between the genomes of uncultured microbes and their cultured relatives. He also reminded the audience that despite the progress made using this technique, single-cell genomics still needs to resolve problems with contamination, incomplete lysis and uneven amplification, but, he added these are solvable with right facility.

From the University of Arizona, Matt Sullivan did a lateral shift, opting to focus on marine viruses rather than marine microbes. "The problem with viral metagenomics," he said, "is that nearly everything is unknown. The challenge is to understand that unknown."

One method of identifying the viral hosts is to stain the viruses and use them as probes. He talked about work done involving *Prochloroccus* and *Synechococcus*, cyanobacteria sequenced by the DOE JGI, which showed that the viral genomes contained genes "stolen" from the bacteria.

Sullivan also pointed out that there are 10 times more viruses than cells in the ocean. "We have up to 100 million viruses per milliliter of seawater," he said, adding that these marine viruses cycle 150 gigatons of carbon each year. Sullivan then discussed two Community Sequencing Program projects he has with the DOE JGI involving ocean viruses sampled from oceaonographic transects in California and Canada. He noted in his presentation that while most of the reads generated for his CSP projects on viral metagenomics remain unknown, applying their protein clustering methods to the dataset has allowed them to double the amount of organized viral sequence space. He added that his team is using a combination of model systems, new and experimental methods and informatics to explore this unknown space.

Envisioning the Future of the DOE JGI

Taking advantage of the DOE JGI's users in attendance at the 7th annual Genomics of Energy & Environment Meeting, DOE Program Manager Dan Drell and Director Eddy Rubin asked the audience to review the draft 10-Year Strategic Vision document released in October 2011, and consider what capabilities they would like to see the DOE JGI add on to remain relevant. For example, Rubin said, users currently send in a tube of DNA and interactions are limited to sequencing and analysis. Additional capabilities would mean users could interact in more expanded ways.

To help the discussion along, Berkeley Lab's Adam Arkin, Los Alamos National Laboratory's Richard Sayre and Stanford University's Christina Smolke shared their own thoughts on how they saw the DOE JGI being of the most use to the scientific community.

"It's hard to predict 10 years out when sequencing technologies are changing so quickly," noted Sayre, who spoke about tapping algal biomass for fuel development. One of the challenges in harnessing this potential feedstock, he said, is that though there are more than a dozen algal taxa with diverse photosynthetic systems, very little is known about their genomic diversity. Genomic studies would assist in understanding these photosynthetic processes to improve algal biomass accumulation for bioenergy applications.

"Framing genomics in grand challenges facing bioenergy and the environment is a way to addressing this concern," Sayre said. "It is important as we go forward that the public buy into this concept in a big way. Anything we can do to help send a message to show this has broader ramifications is important for continued investment in the DOE JGI."

Arkin spoke about the DOE JGI's potential role in "harnessing the universe of biological function" through the DOE Systems Biology Knowledgebase (see page 3). He said that filling in the knowledge gaps about plant and microbial ecosystems in order to tap their potential applications calls for a large-scale coordination of resources and scientific programs in order to make comparisons across species, across communities, across space and time. "The DOE JGI," he said, "is in a unique position to do this type of aggregation."

Smolke described the potential of combining synthetic biology and systems biology, harnessing DNA sequencing and DNA synthesis to more effectively make use of information gained from nature. However, she added, what's needed are tools such as high-throughput functional assays or screens that can extract codes from sequence information. In addition, she said, regulatory toolsets would be useful for truly understanding sequence/structure/function relationships.

Discussing the development of a 10-year strategic vision for the DOE JGI at the Meeting (left to right): DOE Program Manager Dan Drell, LANL's Richard Sayre, DOE JGI Director Eddy Rubin, LBNL's Adam Arkin and Stanford University's Christina Smolke.

DOE JGI Launches TDP

Designed to complement the Community Sequencing Program, the Technology Development Program (TDP) will provide users with early access to nascent technologies at the DOE JGI and in close collaboration with onsite scientists. Projects funded through the TDP are expected to use groundbreaking methods that are currently emerging (and not offered through the CSP program) to address questions of immediate DOE mission relevance. High-risk/high-reward projects exploring the very limits of current technology in the context of cutting-edge scientific questions are encouraged. More information is available at www.jgi.doe.gov/programs/TDP/index.html.

Representatives from several Berkeley Lab national user facilities including the DOE JGI were in Washington, DC on March 28-29, 2012 to meet with Congressional staff for California Members and with staff on the relevant congressional committees. From left to right: Seth Taylor from the Molecular Foundry, David Skinner from NERSC, Susan Bailey from the Advanced Light Source (ALS), Craig Tull and Chris Tracy from ESNet, Jim Ciston and Petra Specht from NCEM, Yves Idzerda from the ALS and Julian Borrill from NERSC. They also took part in an exhibition arranged by the National User Facility Organization (NUFO) in response to Congressional invitations to recognize exciting science and promote the overall benefits of user research. Not pictured: DOE JGI's David Gilbert, who took this photo.

Realizing the Promise of Genomics

(continued from page 2)

as seen through the emergence of ruminants and primate evolution.

"There are not that many sequences completed to planetary biology level yet," Benner admitted in closing. "But planetary analysis will indeed revolutionize biology."

Until that day comes, it seems scientists and science communicators will have to work together to keep people reaching for the stars.

Carl Zimmer's keynote presentation is available online at bit.ly/UM7Zimmer. Steven Benner's keynote presentation is online at http://bit.ly/UM7Benner.

EIGHTH ANNUAL Genomics of Energy & Environment Meeting

MARCH 25-29, 2013
WALNUT CREEK, CA

Presentations and poster sessions on DOE-relevant science including microbial ecology and bioprospecting; genomic analysis of biofuels crops; single-cell genomics; and synthetic biology. Also bioinformatics and new genome sequencing technology tutorials, tours and more. Be there!

http://1.usa.gov/JGI-Annual-Meeting

Photos by Roy Kaltschmidt, LBNL, unless otherwise noted

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

CSO 23839