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

the
PRIMER
Summer 2009
Volume 6 Issue 1

also in this issue

Bugs, Big & Small, from Inside & Out, Join Sources of Green Genes to Fill CSP 2010 Pipeline

More than 70 projects were selected for the 2010 Community Sequencing Program (CSP) portfolio. They involve organisms from regions as far north as the Arctic and south to New Zealand. They include proposals to study microbial contaminants in alcohol that could impact biofuel production, microbial communities in the guts of insects from an area geography scholar Jared Diamond once described as "the nearest approach to life on another planet" and a novel bacterial isolate that could be used to remove heavy metal contaminants from freshwater streams.

"The information we generate from these projects promises to improve the clean, renewable energy pathways being developed now as well as lend researchers more insight into the global carbon cycle, options for bioremediation, and biogeochemical processes," said DOE JGI Director Eddy Rubin. "In translating DNA sequence data into biology, we generate valuable science that improves our understanding of the complex processes that support life on the planet, or imperil it."

Following along the lines of previous CSP-approved metagenomic projects involving the cow, termite and Amazonian stinkbird, Mike Taylor from New Zealand's University of Auckland proposed sequencing microbial communities inside

cont. on page 11

Moving Microbial Genomics Forward

BY MASSIE SANTOS BALLON

To get an idea of how many microbes there are on Earth, imagine taking just the smallest type of microbe, the viruses that infect other microbes, and lining them all up in the row. How far do you think that line would reach? From San Francisco to Paris? To the moon? To the sun? The answer: A thousand times the distance across the Milky Way.

Microbes are not only exceedingly numerous, but also extremely diverse, encompassing most of Earth's total biodiversity. This is one reason why two-thirds of the nearly 5,000 genome projects reported in the Genomes OnLine Database involve microbes. Another factor is the myriad potential applications of microbial research for targeted drug development, bioenergy, bioremediation, agriculture, and other human endeavors. But far more could be done with microbial genomics, according to DOE JGI Genome Biology head Nikos Kyrpides, if researchers would embrace the world of possibilities that extend beyond such human-centric studies.

In an article published in the July issue of the journal *Nature Biotechnology,* Kyrpides reflects on the role of microbial studies in the genomics revolution of the past decade, and considers the factors that have hindered the advancement of the field. Although nearly 1,000 microbial genomes have been sequenced over the past 15 years, he noted that the data obtained has been compromised by the lack of a common language and standardized procedures. Such standards, required for the exchange and integration of genomic data, still remain to be defined and codified. Due to biases in the selection of genomes for study, vast realms of biodiversity remain unexplored. Data release policies very early on led to what he terms "a violation of a longstanding precept in all scientific fields" when reviewers of submitted publications couldn't access the underlying genomic data.

Kyrpides offers numerous suggestions to meet these and other challenges that face genomics research in the decade ahead. For example,

cont. on page 11

Summer 2009 Vol. 6, Issue 1

Seeking Sequencing Standards in Santa Fe

BY MASSIE SANTOS BALLON

One thing became clear fairly quickly at the 4th Sequencing, Finishing, Analysis in the Future Meeting held in Santa Fe, NM from May 27-29, 2009: the name game genome researchers play needs new rules.

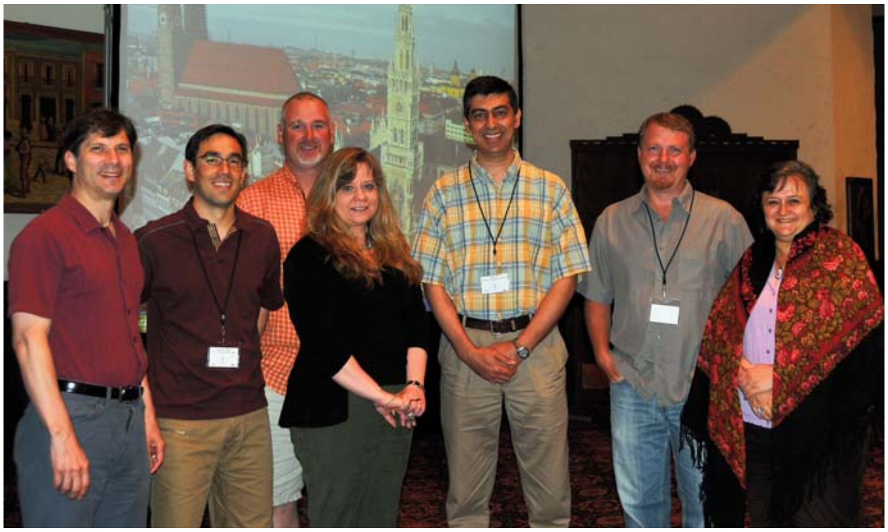
Speaking immediately after the opening keynote, Patrick Chain, head of Microbial Interactions program at DOE JGI, noted that there have traditionally been only two categories for sequenced genomes — “draft,” where the genetic data of an organism has been collected but not completely organized, or “finished,” in which the genetic code has been correctly put together and can be read like a published book from beginning to end. He then proposed adding four other labels so that researchers can determine the quality of the genomes they’re accessing.

Unfortunately, Chain said, pitching the idea on behalf of the International Genome Sequencing Standards Consortium, the advent of new sequencing technologies has increased the number of genomes being sequenced every year.

“Data generation is becoming cheaper, easier and faster,” he said, ”and this is only getting worse.” Based on the current output, Chain predicted that by 2012, there might be 12,000 incomplete genome sequences being made available to researchers who might not recognize that the datasets may lack the information they’re interested in using.

Chain said another part of the problem is that as new sequencing technologies become the machines of choice, they need to be evaluated so that researchers understand that the information that comes out of one company’s instrument may not be the same as the information provided by other equipment. The tags applied to the data may also vary based on the sequencing technology used.

All of these changes, Chain said, have led to the Consortium’s proposal of four genome quality categories, as well as some suggested changes to the definitions of the traditional terms, none of which are pegged to a particular sequencing technology. A standard draft, for example, might be something produced in a lab other than a large sequencing center such as DOE JGI, which typically generates what he called a “high quality draft,” which has little or no manual review.



Meeting organizers (left to right): Michael Fitzgerald, Broad Institute; Patrick Chain, DOE JGI; Bob Fulton, Washington University in St. Louis; Donna Muzny, Baylor College of Medicine; Johar Ali, Ontario Institute for Cancer Research; Chris Detter, DOE JGI; Alla Lapidus, DOE JGI

In an “improved high quality draft,” the data has been reviewed both by people and machines to some extent so most of the genetic data is assembled correctly, but some errors may still be present. An “annotation grade” sequence would present all the information in various gene regions as accurately as possible, and the term “non-contiguous finished” would apply to genomes that are practically finished except for “recalcitrant regions” that are proving problematic. Such quirks, Chain said, would be included in the comments when the genome sequence is made available.

It’s worth wondering after all this what a genome sequence has to do to be labeled “finished.” To achieve this pinnacle, Chain said, the sequence would have no more than one error in every 100,000 base pairs.

Chain noted that one of the hardest parts of coming up with categories of sequence status had been getting everyone in the consortium to agree on the terms and definitions. “I can barely get agreement in my own group on some of these issues so this was really a fantastic achievement,” he said. He concluded by noting that the Consortium plans to submit their proposal for publication at a later date, and intends to publicize the standards on the websites of the various institutions affiliated with the group.

A number of other speakers referenced the need for a standardized terminology that could describe a genome sequence and its various components in their talks over the three days of the conference. Among them was Deanna Church from the National Center of Biotechnology Information, who referred to the need for standard terms during her talk when she mentioned

cont. on page 9

Salt Lake's Briny Waters May Harbor Biological Treasures

Photo by Charles Uibel, GreatSaltLakePhotos.com

BY MADOLYN BOWMAN ROGERS

In one of the 2009 Community Sequencing Projects, the DOE JGI are partnering with a consortium of researchers to sequence the genomes of microbial communities in Utah's Great Salt Lake. Researchers believe the lake's harsh environment harbors a wealth of undiscovered biodiversity that may provide new solutions for carbon sequestration and bioremediation, and might even give clues to the nature of life on other planets.

"We believe we're going to find new enzymes and new metabolic routes that have never been seen before," said the project's lead investigator, Bart Weimer, a microbiologist at University of California, Davis. The collaboration also includes researchers at the Utah State University, the U.S. Geological Survey, and the Utah Division of Water Quality.

The Great Salt Lake is the third saltiest body of water on earth, and the saltiest to support life. The salt content varies widely across the lake, but on average the lake is four times saltier than the ocean, and in the North Arm up to 20 times saltier. The lake also contains petroleum seeps, mercury and other heavy metals, and high volatile sulfur concentrations that give the lake a distinctive rotten-eggs stink. Despite these compounds, the lake supports five million migrating birds each year and is home to brine shrimp and several kinds of algae including diatoms, as well as an untapped diversity of microbial life.

Because the microbes can survive in such an extreme environment, Weimer believes they may contain unique proteins that enable them to chemically fix sulfur and carbon and to detoxify pollutants. Their biochemistry may suggest new methods for sequestering carbon and reducing acid rain, Weimer said, as well as novel pathways for bioremediation of heavy metals, aromatic hydrocarbons, chlorinated compounds and methylmercury (Weimer et al., 2009). Microbial life in the Great Salt Lake might also provide a model for life on planets such as Mars, where sulfate and salt concentrations are also very high.

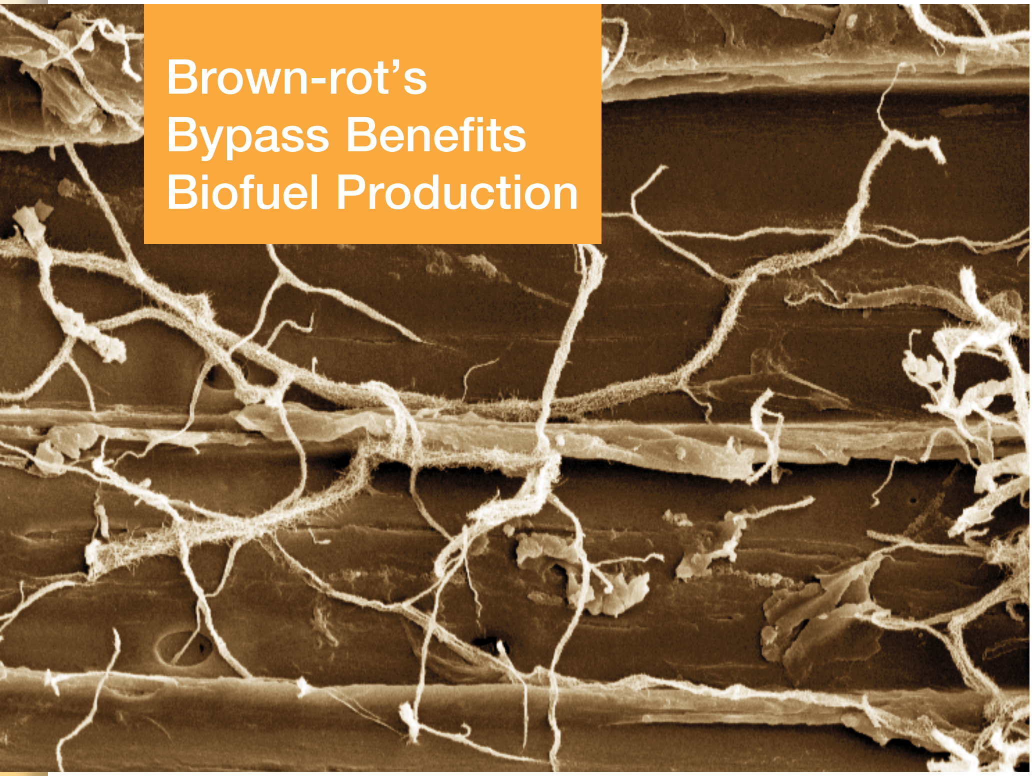
Since less than one percent of the microbes in the lake can be cultured, most of the microorganisms have never been studied before. The DOE JGI will sequence DNA from four sites in the lake to provide a baseline picture of microbial diversity. Researchers will combine the DNA data with measurements of microbial metabolites and environmental conditions such as salinity and oxygen to paint a complete picture of the lake's ecology.

"We've done a lot of diversity assessment of the lake in the last two years, and we're not even approaching the limit of microbial diversity," Weimer said. "The services that the DOE JGI will provide are going to eclipse anything that we could do in the next five to ten years."

FOR ADDITIONAL INFORMATION SEE:

Weimer, Bart C., Giovanni Rompato, Jacob Parnell, Reed Gann, Balasubramanian Ganesan, Cristian Navas, Martin Gonzalez, Mario Clavel, Steven Albee-Scott. Microbial biodiversity of Great Salt Lake, Utah. 2009. In Saline lakes around the world: unique systems with unique values. pp.15-22. Eds. Oren, A., D. L. Naftz, and W. A. Wurtsbaugh. S. J. and Jessie E. Quinney Natural Resources Research Library, Utah State University Press, Logan. (http://www.cnr.usu.edu/quinney/htm/publications/nrei)

Brown-rot's Bypass Benefits Biofuel Production



Lignin is to plants what a skeleton is for humans: an internal scaffold that keeps the organism upright and the softer insides protected. As a result, one of the problems with making biofuels from plants is getting past the lignin to break down the cellulosic biomass into sugars that can be fermented and distilled into liquid biofuel. Because the lignin prevents easy access to the cellulosic material, breaking down plant biomass has traditionally involved harsh chemicals and high heat, processes which add energy to the wrong side of the biofuel production equation.

The Department of Energy (DOE) and its Bioenergy Research Centers are approaching the problem from two different angles: developing plants bred specifically to become biofuel feedstocks; and identifying enzymes that can maximize the biomass being converted into fuel.

With the latter idea in mind, an international team led by researchers from the DOE JGI examined the complete genetic code of brown-rot fungi to find out how they can harness the fungi's ability to break down wood to make biofuel production more cost-effective.

"The microbial world represents a little explored yet bountiful resource for enzymes that can play a central role in the deconstruction of plant biomass—an early step in biofuel production," said Eddy Rubin, Director of the DOE JGI, where the brown-rot fungi's genome was sequenced. "The brown-rot *Postia placenta's* genome offers us a detailed inventory of the biomass-degrading enzymes that this and other fungi possess."

Brown-rot fungi affect softwood trees such as redwoods, pines, cedars and Douglas firs, removing practically all of the cellulose in the plant without removing the lignin as well. Interestingly, some researchers have noted that most brown-rot fungi can't work on cellulose if the lignin is missing. Approximately 10 percent of the decay noted annually in American timber is attributed to brown-rot fungi, which makes the wood look like a pile of cracked bricks. Researchers hope to identify the enzymes involved in attacking the wood and apply them toward other plant material.

"Nature offers some guidance here," said Dan Cullen, U.S. Department of Agriculture Forest Service, Forest Products Laboratory (FPL) scientist and one of the senior authors on the *PNAS* paper. "*Postia* has, over its evolution, shed the conventional enzymatic machinery for attacking plant material. Instead, the evidence suggests that it utilizes an arsenal of small oxidizing agents that blast through plant cell walls to depolymerize the cellulose. This biological process opens a door to more effective, less-energy intensive and more environmentally-sound strategies for more lignocellulose deconstruction."

Randy Berka, another of the study's senior authors and Director of Integrative Biology at Davis, Calif.-based Novozymes, Inc., noted that in sequencing the genome of brown-rot fungi, researchers have been able to compare the genetic blueprints of various types of wood-decaying fungi for the first time.

Brown-rot fungi is a type of basidiomycete fungi like white-rot fungi, which had been previously studied by the DOE JGI. White-rot fungi degrades both the lignin and the cellulose however, while brown-rot fungi only works on the latter.

"Such comparisons will increase our understanding of the diverse mechanisms and chemistries involved in lignocellulose degradation," Berka said. "This type of information may empower industrial biotechnologists to devise new strategies to enhance efficiencies and reduce costs associated with biomass conversion for renewable fuels and chemical intermediates."

The report by appeared in the February 4 online edition of the *Proceedings of the National Academy of Sciences (PNAS)* and involved over 50 scientists from Austria, Canada, Chile, the Czech Republic, France, Germany, Spain and the United States.

When One Cell is Enough

Studying microbes in a lab setting usually involves growing large quantities of them for genetic studies. Researchers at the DOE JGI demonstrated they could assemble a high quality draft of two marine microorganisms using just a single cell from each.

"As long as you can isolate a single cell, pick it from the environment, lyse it (blow it open), you can generate millions of copies of that genome and gain access to the information inside that organism," said DOE JGI researcher Tanja Woyke (pictured, top right).

Woyke and her colleagues at the DOE JGI sequenced genomes of two uncultured flavobacteria, marine microorganisms collected in Maine's Boothbay Harbor for Bigelow Laboratory collaborators Ramunas Stepanauskas and Michael Sieracki, who are particularly interested in genes encoding proteorhodopsins, which allow some cells to use energy from the sun without applying photosynthesis.

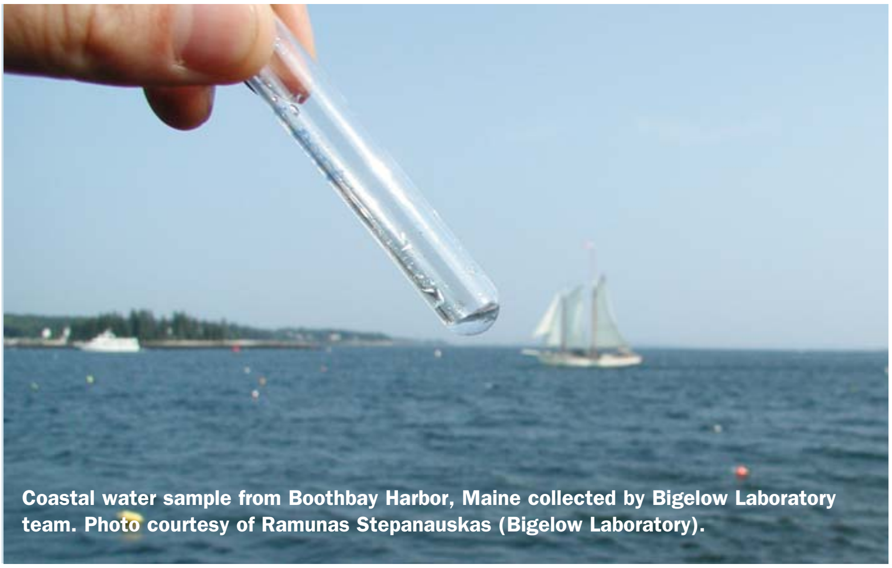
After the Bigelow scientists had picked out individual bacterial cells from their bacterial samples, Woyke and her team went to work. The single cells were then blown open and a process called multiple displacement amplification was applied to make millions of copies of the bacterial genomes for sequencing.

The resulting flavobacterial genome sequences are approximately 80 to 90 percent complete. Woyke credited DOE JGI's Cliff Han and his team at Los Alamos National Laboratory (LANL), which worked on closing gaps in the assembly.

"Even without completed genome assemblies, single cell sequencing offers radically new opportunities for the basic research and biotechnology applications of the microbial 'uncultured majority'," said DOE JGI collaborator Stepanauskas.

More importantly, the single cell sequencing approach means researchers don't have to culture organisms in order to sequence them. This could prove useful for studying organisms that can't exist in a lab environment.

"We estimate that roughly 99.9 percent of the microbes that exist on this planet currently elude standard culturing methods, denying us access to their genetic material, so we have to explore other methods to characterize them," noted DOE JGI Director Eddy Rubin. As a national user facility, he said, the DOE JGI is dedicated to helping researchers access genomes of interest that are relevant to the DOE's mission.


Coastal water sample from Boothbay Harbor, Maine collected by Bigelow Laboratory team. Photo courtesy of Ramunas Stepanauskas (Bigelow Laboratory).

"The power of single cell genomics is that it offers us the ability to sort out one cell from a complex environmental sample, liberate the DNA from that cell, and enzymatically produce millions of copies of that genome so that we have enough DNA to sequence it and characterize its metabolic potential," he said.

The technique doesn't just work on marine microorganisms either; Woyke said they are helping other DOE JGI collaborators who study organisms from various environments but can't provide enough sample to apply other test methods.

Woyke and her colleagues aren't the first to work on the single cell genome sequencing technique, but they've been able to provide a higher quality result compared to previous efforts. Still, the method needs more work. "One of the key issues that still needs refining is the lysis step, since many microbes will not lyse with alkaline solutions, the most common agent for the job. But we are actively working on that," she said.

The study was published in the April 23 edition of the journal *PLoS One.* Tanja Woyke can be heard discussing single cell genome sequencing as it applies to this particular project and other points of interest at http://www.jgi.doe.gov/News/podcast090421_woyke.mp3.

JGI
DOE JOINT GENOME INSTITUTE
4
GENOMICS OF
ENERGY &
ENVIRONMENT
2009
DOE JGI USER MEETING

JGI Annual User Meeting #4—Briefly In Review

A record 500 people attended the DOE JGI Annual Genomics of Energy and Environment User Meeting held March 25-27, 2009 at the Marriott Hotel in Walnut Creek, Calif.

This year the keynote speakers were Energy Bioscience Institute director Chris Somerville, Harvard University's George Church and founder of the eponymous genomic research institute J. Craig Venter. All three speakers focused on a number of scientific and technological techniques to move genomic sequencing forward in ways relevant to the DOE JGI's mission of clean energy generation and bioremediation.

Somerville's talk (http://www.scivee.tv/node/10674) focused on the development of cellulosic biofuels, which is being heavily promoted by the Obama administration and last year's keynote speaker-turned-energy secretary Steven Chu. He spoke about the challenges involved in harnessing nature, particularly the issue of carbon emissions from indirect land use, which has dogged bioenergy researchers since the concern was raised in a *Science* paper (see sidebar and Searchinger et al, 2008; 319: 1238-1240.) published last year. And he painted a sobering picture of the timelines involved in modifying plants whose genomes were being sequenced to become better feedstocks, which he considers a long-term solution, as well as when producing a billion gallons of cellulosic ethanol annually would likely happen.

> “My favorite among these [energy] crops is *Miscanthus*. In an average car you could drive around the world twice from fuel produced by one acre of biomass.”
>
> —Chris Somerville

During the talk, Somerville reminded people to think about the energy problem on a global scale. He also noted that thinking outside the box can lead to surprising collaborations and solutions being proposed.

“We learned it can be very productive to educate people who don't know about a topic about what the problems are,” he said.

Church's keynote (http://www.scivee.tv/node/10578) the following night focused on the technologies needed to move genomics research forward. He began by noting that the technology should be accessible enough that users feel comfortable not just using the equipment but modifying it to suit their needs.

“Think of the platform that reads sequences as an automated microscope. That's really what it is.” Adding a micromirror display, he said, would give the machine added functionality, allowing researchers to hold or release cells or sequences of interest.

The Harvard researcher also described a program he called Multiplex Automated Genome Engineering or MAGE that would allow researchers to study and design sequences “without any idea of where you are or where you're going at all.” As an example of how the system could be used, he talked about an agricultural study in which lycopene was added to a strain of tomato that lacked the antioxidant.

> “There are trillions of dollars invested in vehicle infrastructure and to think we'll switch overnight to ethanol and hydrogen isn't necessarily in the cards.”
>
> —George Church

Venter's presentation (http://www.scivee.tv/node/10653) on the last day of the User Meeting combined Somerville's focus on genome sequencing to design and develop better bioenergy sources with Church's emphasis on the technological upgrades needed to move the science forward. He spoke about what he termed “digitizing life,” moving toward reading the genetic code on the computer and writing it. He discussed the challenges of building a bacterial genome from the ground up, which led to the development of computer software that allows his researchers to act as digital

> “If you're looking for new mammalian genes, stop. Basically they've all been found.”
>
> —J. Craig Venter

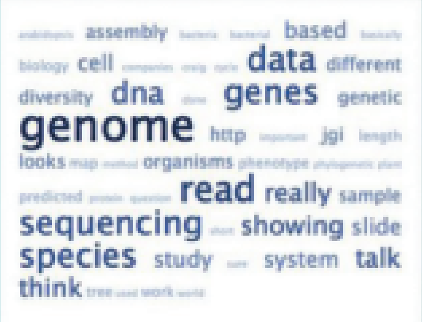
DNA designers to create the species needed to meet the ever-increasing demand for energy and the need to protect the environment.

"We have discovered 20 million genes and we're using them as design components of the future," Venter said. When asked at the end about making the software or the species being designed available, he quipped, "We are not ready to release our software or our bugs because the software has bugs."

Lessons Learned at the JGI Users Meeting

(excerpts from Jonathan Eisen: http://phylogenomics.blogspot.com/2009/03/lessons-learned-at-jgi-users-meeting.html)

"…here are my top lessons I got out of the meeting…"



- NextGen sequencing continues to open up new windows into biology.
- Ecological and population genomics are truly the next big thing.
- Related to the above point, one of the next revolutions is going to be in high throughput phenotyping — after all, we cannot solve the genotype-phenotype problem when we only know the genotype.
- Model / reference organisms are still in, but every single organism on the planet is now in play.
- NextGen sequencing has completely outrun the ability of even good bioinformatics people to keep up with the data and to use it well.
- Related to the above point, the NextNextGen (e.g., Pacific Biosciences) seems to be barreling along and almost ready for prime time. What are we going to do in terms of informatics then?
- Following up on the above point- we desperately need a MASSIVE effort in the development of tools for "normal" biologists to make better use of massive sequence databases.
- I am happy to report that just about everyone seems to be trying to use an evolutionary perspective as part of their work - especially in the selection of organisms for sequencing
- I am sorry to report that many of the evolutionary "perspectives" are a bit off kilter.
- If you study a plant or an animal and are not studying the microbes that live with them, you are missing something.
- If you study ANY organism and are ignoring epigenetics you are behind the curve
- Reading DNA is being used in every which way imaginable.
- Sequencing is definitely not over - it is just getting started.
- Next up – writing DNA.

THE INDIRECT LAND USE DEBATE

In a February 2008 issue of *Science*, Tim Searchinger and his colleagues at Princeton University argued that most studies advocating biofuel use as a way to reduce greenhouse gas emissions did not factor in the emissions resulting from the conversion of land to grow the additional energy crops.

The researchers argued that carbon sinks such as forests and grasslands would become carbon sources when they were plowed up for use as feedstock farmland. "At the broadest level, the key insight of our paper is that when cropland-based feedstocks are used for biofuels, the source of any potential greenhouse gas benefit is the use of the land to take carbon out of the atmosphere," Searchinger reiterated in a letter published by *Science* six months after the study was released.

"Leaving out emissions from land use change for such biofuels assumes that land is a carbon-free asset. We would argue that any calculation that ignores these emissions, however challenging it is to predict them with certainty, is too incomplete to provide a basis for policy decisions."

One of the examples Searchinger and his team used involved corn ethanol. Biofuel advocates had previously calculated that the energy crop would reduce emissions by 62 percent. The Princeton researchers said that when land use emissions were factored in, corn ethanol would actually release double the emissions that were to have been sequestered.

The study drew criticisms from biofuels advocates and researchers such as Chris Somerville, who noted that the study was "based on naïve assumptions."

The Princeton study, they said, only targeted land use emissions from biofuel production, instead of also considering the emissions from oil production. Another point they raised was that the study assumed a worst-case scenario in which all the land being converted for the added crops was pristine land. "According to indirect land use," Somerville commented during his keynote speech, "by not growing corn on unfarmed land, we're deforesting the Amazon." Other factors, the critics said, could make land available for biofuels without destroying forests.

Searchinger and his colleagues have refuted the counter-arguments, but the debate continues to this day.

JGI Publishes Sorghum Drought-Tolerant Biofuels Feedstock

For some 500 million people around the world, sorghum is a staple grain. Approximately 60 million tons of sorghum, which is more drought-tolerant than corn and other cereal crops such as wheat and oats, is produced annually.

Sorghum also requires a third less water to produce the same amount of ethanol per bushel compared to corn. Because of its potential as a whole plant fiber based biofuel feedstock, the DOE JGI and several partner institutions worked on the genome sequence of sorghum. The grass' complete genome was published in the January 29 edition of the journal *Nature.*

"This is an important step on the road to the development of cost-effective biofuels made from nonfood plant fiber," said Anna C. Palmisano, DOE Associate Director of Science for Biological and Environmental Research.

"Sorghum is an excellent candidate for biofuels production, with its ability to withstand drought and prosper on more marginal land," she said. "The fully sequenced genome will be an indispensable tool for researchers seeking to develop plant variants that maximize these benefits."

Researchers plan to use the genetic data not only to develop better food and feedstocks with improved drought tolerance and increased grain yield, but to produce grass strains suited specifically for biofuel production. Sorghum's position in the grass family — it is closely related to maize and sugarcane, and more distantly to wheat and rice — is expected to give researchers helpful insights into other plant genomes that could be used as biofuels.

"Sorghum will serve as a template genome to which the code of the other important biofuel feedstock grass genomes—switchgrass, *Miscanthus,* and sugarcane—will be compared," added Andrew Paterson, the publication's first author and Director of the Plant Genome Mapping Laboratory, University of Georgia.

Having the complete sorghum genome already benefits a number of current research projects involving the grain. Even before they knew sorghum's complete genome sequence, for example, the U.S. Department of Agriculture had been working on developing disease-resistant, low-lignin sorghum strains. USDA researchers based their studies on early sorghum genetics research that compared sorghum's genes to those of rice.

With less lignin in the plant, the sorghum would be easier to digest, making it a better livestock feed. It would also require less energy to be converted into ethanol biofuel. Similar research projects in other institutions would find the sorghum genome's availability equally useful.

The sorghum genome is only the second grass genome to be completely sequenced after rice and has turned out to be nearly 75 percent larger. Part of the reason for the difference in size, noted Jeremy Schmutz from the DOE JGI partner HudsonAlpha Institute for Biotechnology, is that plant genomes tend to have large sections of repetitive regions and sorghum is no exception. But the comparison between sorghum and rice genomes suggests that there may be a common grass genome structure as well.

"We found that over 10,000 proposed rice genes are actually just fragments," said DOE JGI's Dan Rokhsar. "We are confident now that rice's gene count is similar to sorghum's at 30,000, typical of grasses."

Summer 2009 Vol. 6, Issue 1

A Shortcut for Genome Sleuths

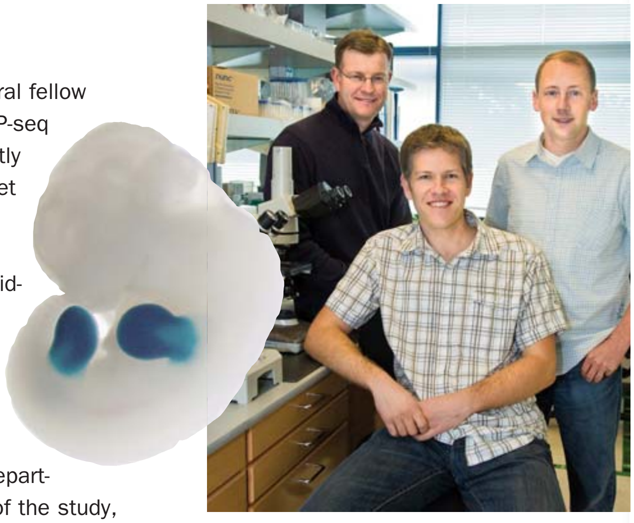
To identify potentially significant sequences in an organism's genetic code, such as those that control the genes responsible for an organism's development or key the development of better biofuel feedstocks, genome researchers typically compare the target genome with the genomes of other, similar species and look for sequences conserved in evolution.

DOE JGI researchers teamed with colleagues at Lawrence Berkeley National Laboratory and the University of California, San Diego to develop a tool that provides genome researchers with a shortcut for searching out locations on the genome where particular gene functions take place. The method is known as ChIP-seq and it can pinpoint the genetic switches by looking at how DNA and proteins interact.

As DOE JGI postdoctoral fellow Matthew Blow put it, ChIP-seq allows scientists to directly identify a genome-wide set of the genetic switches throughout the whole genome rather than individually tracking down and testing every gene for its possible functions.

Len Pennacchio, head of the DOE JGI's Genomic Technologies department and senior author of the study, added that ChIP-seq also provides researchers with an opportunity to learn more about noncoding sequences referred to as genomic "dark matter."

The research was described in the February 12 edition of the journal *Nature.*



The boxing gloves-like blue stain in this-mouse embryo indicates a specific gene function identified by researchers using a tool refined by (left-to-right) senior *Nature* paper author and JGI Genomic Technologies Program head Len Pennacchio and co-first authors Matt Blow and Axel Visel.

Coming to Terms in Santa Fe

cont. from page 2

that the same label might be applied to different things on different databases. And Owen White from University of Maryland suggested moving toward a common system that could pull the best information on bacterial genome sequences from a the various databases rather than relying on just a few during his keynote speech.

"Standards are going to be the key to understanding some of the difference between the things that are done well and the things that are sort of dubious," said Miriam Land, a DOE JGI researcher at Oak Ridge National Laboratory who annotates microbial genomes sequenced in DOE JGI headquarters in Walnut Creek, Calif.

Land was the final speaker at the conference and she addressed the need for standards in labeling the various components of a genome sequence. "I often run into people from the sequencing end who assume that what they see on the computer is right when they look at annotation," she said before explaining the process and the problems involved.

Land echoed Chain's comment from the first day of the conference on the steadily increasing list of genomes and sequences being made available. There's a movement to make the annotation process automated in order to speed up the work and make it more consistent. Unlike manual review, she said, computers will always interpret the data in the same way. And yet "consistent doesn't always mean better," Land noted. "Sometimes it's consistently wrong too."

There are a number of genome databases out there that serve as references for annotation work, Land said, and the information in them is based on observations made by other people. Each of these tools, however, was created with a specific purpose in mind, which is why the data provided by each might not always be useful to researchers. In that sense, she added, the quality of the annotation can become dependent on a person's point of view, which is in turn defined by their knowledge and experience.

All of these variations led to Land's call for annotation standards, which she didn't think was as impossible a process as Chain had made it out to be during his talk.

"We want a method of assessing and designation quality as currently all annotation is deemed uniform in quality," she said. "We want a standard vocabulary for when you're talking about the same thing." An "annotation quality sequence," she said in closing, deserves annotation of comparable quality.

Patrick Chain's talk can be viewed online at http://www.scivee.tv/node/11406. Miriam Land's talk can be viewed online at http://www.scivee.tv/node/11424. Other talks from the "Sequencing, Finishing, Analysis in the Future" conference can be found online at http://www.scivee.tv/node/11378.

Ocean Sentinels for Climate Change

Scientists studying climate change often look for organisms that can serve as the local version of the canary in a coal mine. Such sentinels help researchers identify warning signs of environmental changes that are likely to upset the balance of the ecosystem. In the Arctic region, the scientists watch the polar bears. In the oceans, however, they might be watching something much, much smaller.

An algal genus known as *Micromonas* thrives in waters ranging from the equator to the poles and plays a key role in the global carbon cycle. These microorganisms lend marine scientists insight on how climate change and rising ocean temperatures affect marine ecosystems and which organisms might become major players in the global carbon cycle.

Approximately a 50th of the width of a human hair, *Micromonas* could give Olympian Michael Phelps a run for his gold medals, being able to move through water at the rate of 50 body lengths per second. The tiny marine microorganism can also move toward the sunlight that powers it, much like sunflowers constantly keep their faces turned toward the light throughout the day.

The DOE JGI sequenced genomes of two *Micromonas* samples being studied by Alexandra Z. Worden of Monterey Bay Aquarium Research Institute (MBARI). Taken from opposite ends of the world — the South Pacific and the English Channel — approximately 10,000 genes were identified in each isolate.

DOE scientists plan to use the genetic information to develop algal biofuels, among other things. The algal sequences could also provide researchers with a look at the encoded evolutionary record of how plants moved from the water to land.

"Genome sequencing of *Micromonas* and the subsequent comparative analysis with other algae previously sequenced by DOE JGI and Genoscope [France], have proven immensely powerful for elucidating the basic 'toolkit' of genes integral not only to the effective carbon cycling capabilities of green algae, but to those they have in common with land plants," said DOE JGI Director Eddy Rubin.

But the genomic results also revealed that the algae weren't as similar to each other as the researchers had previously thought.

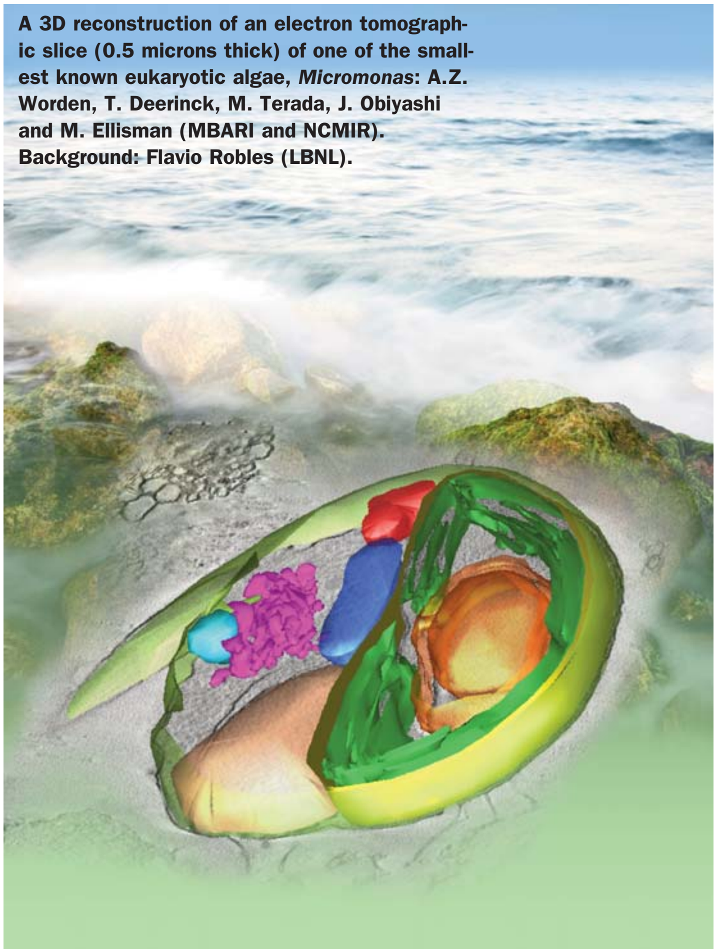
"These two picoeukaryotes, often considered to be the same species, only share about 90 percent of their genes," said Worden. To put this in perspective, humans and some primates have about 98 percent genes in common.

The information suggests the algae in each region may thus react differently to coming environmental changes.

"By understanding which genes a specific strain employs under certain conditions, we gain a view into the factors that influence the success of one group over another," Worden said. "We may then be able to develop models that could more effectively predict a range of possible future scenarios, that will result from current climate change."

Worden thinks, for example, that the algae would adapt more easily than other marine species because it has already shown that it can thrive in a variety of temperatures.

Only five other single celled marine microorganisms had had their genomes sequenced prior to the publication of the work, which appeared in the April 10 issue of the journal *Science.*



A 3D reconstruction of an electron tomographic slice (0.5 microns thick) of one of the smallest known eukaryotic algae, *Micromonas*: A.Z. Worden, T. Deerinck, M. Terada, J. Obiyashi and M. Ellisman (MBARI and NCMIR). Background: Flavio Robles (LBNL).

CSP 2010 Selections

cont. from page 1

gut samples taken from 11 insect species endemic to New Zealand and 3 termite species from Australia that potentially have novel wood-degrading enzymes.

Another metagenome project involves the desert locust *(Schistocerca gregaria),* swarms of which spread over up to 20 percent of the world's land mass and affect the livelihoods of up to 10 percent of the world's population just by consuming the equivalent of their body mass daily. DOE JGI researcher Falk Warnecke proposed sequencing the insect's gut wall and the microbial community inside, in part to find out if the locust or the microbes should get the credit for degrading lignocellulose during these meals.

Many sequencing projects focus on breaking down plant cell walls to get more energy out of biofuel feedstocks. Two involving fungi come from Daniel Cullen, U.S. Forest Service, Forest Products Laboratory, who proposed sequencing a lignin-degrading white rot fungus known as *Phlebiopsis gigantea,* and Harvard University's Anne Pringle, whose proposal focuses on the common *Amanita thiersii* which breaks down cellulose.

Other approved projects focus on the fermentation process of alcohol production. David Mills from the University of California, Davis and Trevor Phister from North Carolina State University both proposed studying bacterial contaminants that destroy alcohols and reduce biofuel production efficiencies: the bacteria that turns wine into vinegar *(Acetobacter aceti)* and the bacteria responsible for the distinctive "horse-sweat taint" in contaminated wine *(Dekkera bruxellensis).*

The DOE JGI is also using DNA sequencing for projects related to carbon cycling and biogeochemistry. A project from Mark Waldrop of the U.S. Geological Survey looks at microbes trapped in the Alaskan permafrost, a carbon sink threatened by global warming. The samples to be sequenced come from the USGS' Yukon River Basin project, where geoscientists have been studying the region's response to climate change since 2001.

***P. gigantea* found fruiting on a red pine log in a northern Minnesota forest, where the fungus, an aggressive saprophyte, is commonly found on downed logs or cut timber. Photo by: Robert A. Blanchette, Professor, University of Minnesota**

Another project related to the carbon cycle comes from Ramunas Stepanauskas at the Bigelow Laboratory for Ocean Sciences in Maine and involves microbial genomes collected from the Atlantic Ocean near South America. Stepanauskas intends to use single cell genome sequencing techniques, building on previous collaborations with DOE JGI research scientist Tanja Woyke and her colleagues (see story on page 5).

One sequencing project is led by Volodymyr Dvornyk from the University of Hong Kong and involves cyanobacterial

cont. on page 12

Kyrpides on Microbial Genomics

cont. from page 1

the list of microbial genomes for potential sequencing has been limited to the approximately one percent of organisms that can be cultured in the lab and their DNA then extracted. He sees a way forward using single-cell genomics — a technique now being pursued in earnest by DOE JGI researcher Tanja Woyke and her colleagues — in partnership with environmental metagenomics to provide a more holistic understanding of a community and its individual members.

Kyrpides also suggests several innovative approaches for easing the data processing bottleneck accompanying the exponential increase in genomic data. All-versus-all gene comparisons — previously a common practice — will become infeasible. To reduce the size of the datasets, he proposes a proxy approach in which one protein from each protein family or one species from each genus represents the group. Taking this one step farther, all the genes from all the sequenced strains in a species — the pan-genome for that species — would constitute the genome representing that species for gene comparisons.

Echoing other researchers, most notably DOE JGI's Patrick Chain and Miriam Land during the recent "Sequencing, Finishing, Analysis in the Future" Conference, Kyrpides calls for the development of genome annotation standards and their adoption by sequencing centers around the world — a necessity for meaningful genome comparisons.

With new tools in hand and international initiatives for increased collaboration underway, the field of microbial genomics is poised for a decade of exciting advances. Sharing his vision for the future, Kyrpides observes: "To explore and seek to understand how the Earth breaths, grows, evolves, renews, and sustains life — all essentially the work of the microbial world — is the great adventure now beckoning to us. Microbial genomics paves the way forward."

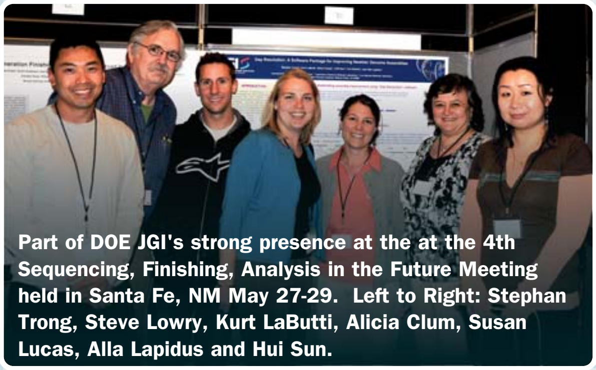
CSP 2010 Selections

cont. from page 11

samples of *Nostoc linckia* collected from opposite faces of a valley known as Evolution Canyon north of Tel Aviv, Israel. The south side of the valley receives 800 percent more solar radiation than the north side of the valley, so the samples offer researchers an opportunity to learn how long-term stress affects the evolution of the bacterial genome.

Among approved projects that focus on microbes that can help clean up the environment is one from Gillian Lewis at the University of Auckland in New Zealand. She proposed sequencing a novel isolate of a freshwater bacterium from the *Siderocapsaceae* family that can remove manganese contaminants.

For the complete list of CSP 2010 sequencing projects, see: http://www.jgi.doe.gov/sequencing/cspseqplans2010.html



Part of DOE JGI's strong presence at the at the 4th Sequencing, Finishing, Analysis in the Future Meeting held in Santa Fe, NM May 27-29. Left to Right: Stephan Trong, Steve Lowry, Kurt LaButti, Alicia Clum, Susan Lucas, Alla Lapidus and Hui Sun.

Matthias Hess, DOE JGI postdoctoral fellow presented the work on mining the metatranscriptome of the cow rumen microbiota for feedstock-targeted glycosyl hydrolases, at the 31st Annual Society of Industrial Microbiology's Symposium on Fuels and Chemicals May 4 in San Francisco.

Twenty-three feet, four and a quarter inches and counting...At a recent gathering of poplar researchers at the base of DOE JGI's sentinel *Populus trichocarpa* Jerry Tuskan (arms folded, above the helix), predicted at least another six-foot growth spurt by year's end.

U.S. House of Representatives Committee on Science and Technology staffers (left to right) Adam Rosenberg, Margaret Caravelli, Jetta Wong and Elizabeth Chapel, visited DOE JGI for a briefing on May 28.

JGI USER MEETING
5
2010
SAVE THE DATE!
MARCH 24-26
WALNUT CREEK, CA

CSO 16843

Contact The Primer
David Gilbert, Editor / DEGilbert@lbl.gov / (925) 296-5643