

Guts 'R Us?: CSP 2009 Selections Announced

Q: What do boat-boring bivalves and stinkbirds have in common?

A: Their guts are just two of 44 new CSP 2009 targets.

In the continuing effort to tap the vast, unexplored reaches of the Earth's microbial and plant domains for bioenergy and environmental applications, the DOE JGI has announced its latest portfolio of DNA sequencing projects for the coming year. The 44 projects, culled from nearly 150 proposals received through the Community Sequencing Program (CSP), will collectively generate more than 60 billion nucleotides of data.

"The scientific and technological advances enabled by

the information that we generate from these selections promise to take us faster and further down the path toward clean, renewable transportation fuels while affording us a more comprehensive understanding of the global carbon cycle," says Eddy Rubin, DOE JGI Director. "The range of projects spans important terrestrial contributors to biomass production in the Loblolly pine—the cornerstone of the U.S. forest products industry—to phytoplankton, barely visible to the naked eye, but no less important to the massive generation of fixed carbon in our marine ecosystems."

With new *cont. on page 7*



Among the CSP 2009 targets is the contents of the enlarged crop of the Hoatzin, *Opisthocomus hoazin*, or Amazonian stinkbird, which is unique among known avian species by virtue of its fermentative foregut.

also in this issue

JGI News	2	Genome of Simplest Animal Reveals Ancient Lineage	10
User Community Faces: Alexandra Worden	3	Using Metagenomics on Lake Washington Microbes	12
Super Bacteria for Super Alfalfa	4	JGI Announcements	16
CSP2009 Project at Contaminated Hanford Site	6		

Director Rubin Reviews Genomics of Biofuels in *Nature*

Genomics is accelerating improvements for converting plant biomass into biofuel, thus bringing closer to reality widespread use of an alternative to fossil fuel for the world's transportation needs, reports Eddy Rubin, DOE JGI Director, in the August 14 edition of the journal *Nature*. In "Genomics of cellulosic biofuels," Rubin lays out a path for how emerging genomic technologies will contribute to a biofuels future that is substantially different from the present corn-based ethanol industry—and in part mitigate the food-versus-fuel debate.

"The Apollo moonshot and the Human Genome Project rallied support for massive R&D efforts that created the capabilities to overcome obstacles that were not contemplated at the outset of these initiatives," says Rubin.



"Similarly, today's barriers to improving biofuels are significant, but genetics and genomics can catalyze progress towards delivering, in the not-too-distant future, economically viable and more socially acceptable biofuels based on lignocellulose."

While Rubin acknowledges that this strategy is in its infancy, rapid progress is being made.

"Over the past 10,000 years, wild plant species were selected *cont. on page 13*

JGI-Stanford Moves to HudsonAlpha Institute for Biotechnology

In Summer 2008, the DOE JGI's partner at the Stanford Human Genome Center (JGI-SHGC) moved to a new home, the recently opened HudsonAlpha Institute for Biotechnology (www.hudsonalpha.org) in Huntsville, Alabama. The newly formed research group, the HudsonAlpha Genome Sequencing Center, will expand its operations into next-generation genome sequencing and add capabilities to collect data to

augment the use of DOE JGI genomes in scientific discovery, bioenergy, and other directed breeding applications.

HudsonAlpha, a non-profit research institute, will continue to work as a DOE JGI partner with a focus on plant and eukaryotic genomics. It will also continue to assist with genome analysis for collaborators as well as form the cornerstone collaborative group for sequencing and genomic projects at HudsonAlpha. Director Rick Myers, Jeremy Schmutz, and Jane Grimwood are some of the Stanford Human Genome Center researchers who made the move to HudsonAlpha.

HudsonAlpha's mission is to use biotechnology to improve human health and quality of life, stimulate economic development, and inspire Alabama's youth to seek careers in science. HudsonAlpha believes that education supports innovation, giving rise to increased professional and economic opportunities. Investments in

science and technology education, including hands-on, research-based programs for students and educators, are critical to enabling discoveries that will improve health and keep America competitive in biotechnology. To push the boundaries of scientific and technological capabilities, education specialists at HudsonAlpha are supporting, creating and implementing programs toward making Alabama's students among the nation's best and brightest while augmenting the enterprising environment that will keep them home.

The HudsonAlpha Institute is the cornerstone of the 150-acre Cummings Research Park Biotech Campus. The campus embodies a synergistic cluster of talented scientists and business professionals that promises collaborative innovation to turn knowledge and ideas into commercial products and services for improving human health and strengthening Alabama's progressively diverse economy. ◉



The Stanford Human Genome Center has found a new home at the HudsonAlpha Institute in Huntsville, Alabama.

Photo: Cooper Carry

IMG, IMG/M Systems Upgraded, Education Site Launched

The DOE JGI has extended the capabilities of the Integrated Microbial Genomes (IMG) data management system, updated the content of the IMG/M metagenome data management and analysis system, and launched its educational companion site, IMG/EDU.

Version 2.6 of IMG includes new microbial genomes from the Version 29 release of the National Center for Biotechnology Information's (NCBI) Reference Sequence (RefSeq) collection. IMG 2.6 contains a total of 4,207 genomes consisting of 1,078 bacteria, 56 archaea, 40 eukaryotes, 2,230 bacteriophages, and 803 plasmids that did not come from a specific microbial genome sequencing project. Among these genomes, 3,737 are finished and 470 are draft genomes. IMG 2.6 contains 308 microbial genomes

sequenced at DOE JGI, consisting of 237 finished and 71 draft genomes.

IMG 2.6 includes the first version of chromosomal gene cassettes (positional gene clusters) together with estimates of their conservation across IMG genomes. The User Interface has been improved with tools for exploring gene cassettes conservation and includes additional graphical viewers for examining the functional annotation of genomes. More details on IMG 2.6 are available at: img.jgi.doe.gov/w/doc/releaseNotes.pdf.

The content of IMG/M, which provides tools for analyzing the functional capability of microbial communities based on their metagenome sequence, has been updated to include three additional metagenome datasets generated from microbial community samples that

cont. on page 9



During the week of September 15, more than 40 participants ventured to JGI's Production Genomics Facility from as far away as Australia, Greece, Hong Kong, Israel, and Russia for the Microbial Genomics & Metagenomics (MGM) Workshop. The full agenda can be found online at www.jgi.doe.gov/meetings/mgm/agenda.html.

A Conversation with Alexandra Z. Worden

BY MADOLYN BOWMAN ROGERS

Alexandra Z. Worden, microbial ecologist at Monterey Bay Aquarium Research Institute, studies picoeukaryotes, tiny marine algae less than 2 microns in diameter (about 1/50 the width of a human hair). These organisms are part of the photosynthetic community that forms the foundation of marine food webs. Because these photosynthetic organisms remove carbon dioxide from the atmosphere and “fix” it into organic molecules, they may play a significant role in global carbon cycling. Worden wants to know what regulates the growth of these microbes, what happens to the carbon they fix, and how this might play a role in climate change.

To answer these questions, Worden has turned to the power of genomic analysis. She has been a frequent collaborator with DOE JGI, serving as champion on marine microbiological sequencing projects in 2005 and 2007. In one of the 2009 Community Sequencing Program (CSP) projects, Worden heads an international collaboration that includes researchers from Austria, Wales, the University of Southern California, and the J. Craig Venter Institute in San Diego. The goal is to produce genome sequences of a ubiquitous group of predatory

protists, single-celled marine organisms that feed on picoeukaryotes and other small plankton.

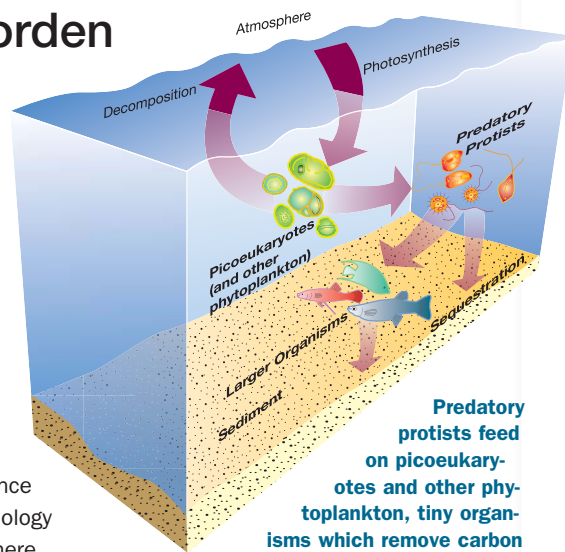
Worden has a doctoral degree in ecology from the University of Georgia and completed an NSF Postdoctoral fellowship at Scripps Institution of Oceanography in San Diego over a three-year period. After accepting a professorship at the University of Miami, she spent six months as a visiting scientist at the Station Biologique du Roscoff in France then built her group in the Marine Biology and Fisheries Department at UM, where she worked until Fall 2007. To view a video podcast about Worden and her research visit: www.youtube.com/user/JointGenomeInstitute

Q. What led you to look at the predatory protists?

A. It's critical that we know how predators select their prey, because that dictates the fate of the prey's photosynthetically fixed carbon. If one group of picoeukaryotes is actually eaten, while other groups of phytoplankton are not, it's a dramatically different role in carbon cycling and potential sinking of that material to the deep ocean.

Q. How does being eaten change the fate of the carbon?

A. When you're a micron [in size], and you're in seawater, it's like being in syrup. You're not able to move yourself through the water column. If you're one micron, and you die in the surface ocean, your carbon



Predatory protists feed on picoeukaryotes and other phytoplankton, tiny organisms which remove carbon dioxide from the atmosphere. Sequencing the predatory protists will improve our understanding of photosynthesis, as well as its links with carbon cycling and climate change.

is essentially released back to the environment. Larger organisms like these predators are better able to move through the water column—some sink—and they're also part of the food chain.

Q. What led you to start using genomics?

A. If you observe picoeukaryotes under a microscope, they show up as just a dot. So it's hard to get any handle on their physiology—what's controlling their growth. For me, genomics was a way to get at that. It's a window into the capabilities they might have. But it's really just a first step; then you need to find out what genes they're using, when and why. If we know what factors shift

cont. on page 14

“ I went into this wanting to help the environment, and I am daunted at this stage. Things are changing so fast, people are proposing scientifically unfounded quick fixes, and we're still at such a basic level of understanding. It is a true call to action. ”

—Alexandra Z. Worden

Nitrogen-Fixing Bacteria Hold the Key to Alfalfa as a Bioenergy Crop

BY MADOLYN BOWMAN ROGERS

By harnessing an ancient association between plants and bacteria, researchers may one day be able to engineer super-efficient crops for bioenergy production. To advance this goal, the DOE JGI is partnering with scientists at the University of Florence in Italy to sequence two new strains of nitrogen-fixing bacteria in one of the 2009 Community Sequencing Program (CSP) projects.

Sinorhizobium meliloti is the most well-studied and abundant species of rhizobia, a group of bacteria that infect the roots of legumes and form nodules in which they fix atmospheric nitrogen into plant matter. Rhizobia enable legumes such as peas, beans, clover, and alfalfa to grow in nitrogen-poor soils with no need for costly fertilizers. One strain of *S. meliloti* has been sequenced, but because different strains have adapted to diverse environments worldwide, *S. meliloti* strains exhibit tremendous genetic variability.

"We think we will find that 10 percent of the genome is completely new and different," says the project's lead investigator, Emanuele Biondi, a bacterial geneticist at the University of Florence. This could mean 500 to 800 novel genes per strain.

S. meliloti forms a symbiotic association with alfalfa, a crop traditionally used as animal feed but now under study by the USDA and the University of Minnesota as a promising feedstock for biofuel production. Alfalfa's helper bacteria give it a competitive advantage over current bioenergy crops such as corn, which require fertilizers. Because commercial fertilizer production is expensive and produces greenhouse gases, alfalfa would be a more environmentally sustainable biofuel crop than corn and other cereals. In addition, alfalfa can grow on marginal lands, so it would not use up prime cropland.

The two new strains of *S. meliloti* that the DOE JGI will sequence each contain desirable traits. Strain AK83 allows alfalfa to thrive in the desert-like North Aral Sea region of Kazakhstan, where extremely salty soils prevent many plants from growing. Strain BL225C comes from an agricultural region in the north of Italy and is highly efficient, producing plants with three to four times as much dry weight as strain AK83.

By comparing the genomes of these two strains with the previously sequenced reference strain, researchers may discover the genetic basis for traits such as salt resistance and nodulation efficiency, Biondi says. This would open the possibility of engineering a new strain by combining desirable traits already present in the species but not normally found together, a process Biondi refers to as "environmentally friendly genetic engineering" because it doesn't introduce foreign genes into the species.

"Our goal is to make a super-strain that can survive in extreme environments and also can nodulate very efficiently," Biondi says. Such a strain could produce high-yield alfalfa plants that can grow on land unsuitable for agriculture—an ideal bioenergy crop.

Madolyn Bowman Rogers, PhD, recently completed the Science Communication Program certificate at the University of California, Santa Cruz and spent the summer of 2008 as a public affairs intern at the U.S. Department of Energy Joint Genome Institute. ◉

Alfalfa (*Medicago sativa*) serves as the host plant for the nitrogen-fixing bacterium, *S. meliloti*.

Researcher Emanuele Biondi

Life at the Extreme: Microbes May Provide New Tools For Carbon Sequestration

BY MADOLYN BOWMAN ROGERS

Microbes that eke out their existence in a lightless, frigid, buried lake in Antarctica may be the most energy-efficient organisms on earth. To try to unlock the secrets of their remarkable abilities, the DOE JGI will partner with researchers at the University of Alberta and the University of Delaware in a 2009 Community Sequencing Program (CSP) project to sequence the genomes of this microbial community.

The microbes survive in the hostile environment of Lake Vostok, a freshwater lake about the size of Lake Ontario which has lain entombed under more than three kilometers of ice in the middle of Antarctica for perhaps 15 million years. The lake's near-freezing waters contain no light, no organic matter for food, and no geothermal activity or other energy sources. In addition, the lake's crushing pressure of nearly 400 atmospheres causes oxygen to become trapped in the waters. Lake Vostok is saturated with 50 times normal oxygen concentrations, which

would be toxic to most organisms.

"These are organisms really living at the extreme," says Phil Hugenholtz, head of the Microbial Ecology Program at the DOE JGI and lead investigator for the sequencing project. The collaboration also includes Brian Lanoil of the University of Alberta (formerly at UC Riverside), Craig Cary of the University of Delaware, and Victor Kunin of DOE JGI.

Although Lake Vostok itself has never been penetrated, the Vostok microbes were discovered in ice cores taken from just above the lake in 1998. Lanoil is leading the project to extract the microbes from the Vostok ice cores, which remain preserved at the National Ice Core Laboratory in Denver, Colorado.

Lanoil's team has developed stringent methods to insure the samples are not contaminated by even a single bacterium from our environment. The microbes will be handed off to the DOE JGI for sequencing, assembly, and analysis using DOE JGI's unique Integrated Microbial

Genomes with Microbiome Samples (IMG/M) data management system.

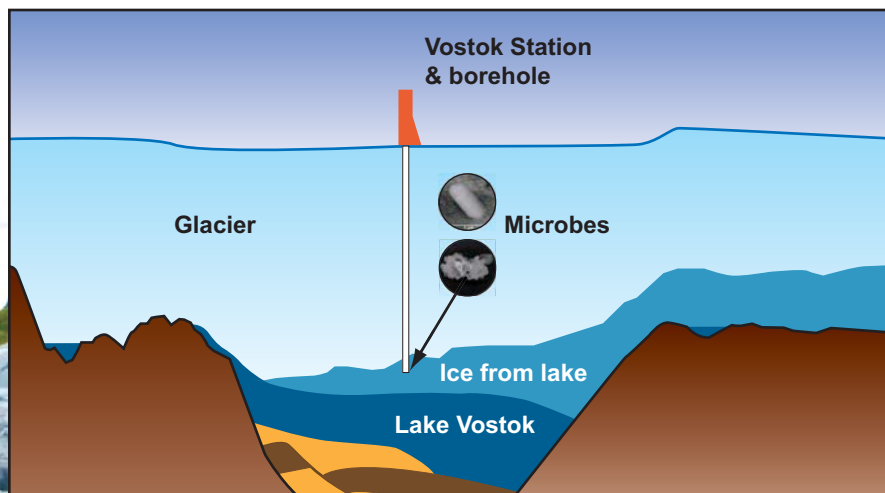
"This will be the first metagenomic analysis of this system," Hugenholtz says. "Metagenomics gives you insights into the functional capabilities of the organisms present—not just who's there, but what they're potentially capable of doing. And of course that's the really interesting thing, working out how they're adapted to that environment."

The microbes probably get their food by removing carbon dioxide from the atmosphere and fixing it into organic molecules, Hugenholtz says, a process that normally requires tremendous energy. Green plants also produce food this way, but they can power this process using the abundant energy of sunlight.

"Whatever the Lake Vostok microbes do, they must do it very efficiently, because they don't have much to eat and they don't have much to survive on," co-PI Kunin says, adding that high oxygen concentrations also interfere with carbon fixation. Therefore, the microbes have "probably the most efficient carbon sequestration you can get."

Since excess carbon dioxide in the atmosphere contributes to global warming, efficient carbon sequestration methods are of great interest to the DOE. Analysis of the bugs' genomes is expected to reveal novel enzymes and pathways for carbon sequestration, as well as other unusual adaptations to extreme cold, high pressure, and high oxygen. Because these conditions resemble those found on some outer planets and moons—for example, Jupiter's moon Europa—the microbes' genomes could even provide clues to the nature of extraterrestrial life.

A video podcast about the Lake Vostok CSP project can be viewed at: www.youtube.com/user/JointGenomeInstitute



Microbes were discovered in Lake Vostok ice cores taken in 1998 and are now being extracted and sequenced in hopes of discovering the secret to their survival in extreme conditions.

Uranium Plume Microbes May Inform Stewardship Efforts

BY MADOLYN BOWMAN ROGERS

Microorganisms in the soil may represent an untapped resource for cleaning up radioactive waste. In a 2009 Community Sequencing Project spearheaded by scientists at Pacific Northwest National Laboratory, the DOE JGI will sequence microbes in subsurface soil samples from a contaminated DOE site to glean clues about the microbes' biogeochemical abilities.

"This study might reveal organisms with novel properties and capabilities that are relevant to the fate and transport of subsurface contaminants," says the project's lead investigator, Allan Konopka, microbiologist at PNNL. "This fundamental scientific information would inform management decisions regarding what strategies would be best for long term stewardship at Hanford."

The sequencing is part of a research program funded by the DOE Environmental Remediation Sciences Program and carried out by scientists at PNNL and at several universities and other national labs. In the spirit of DOE team science, the project brings together microbiologists, ecologists, geologists, geochemists, hydrologists, and mathematical modelers to study how subsurface conditions and microbes in Area 300 of the Hanford Site interact to affect what happens to uranium

and technetium in subsurface soils.

The Hanford Site occupies 560 square miles along the Columbia River in southern Washington and was a major plutonium production facility for the Manhattan Project and during the Cold War, until it ceased operations in 1990. During cleanup operations, DOE removed ponds of uranium wastewater and contaminated sediment from

you're looking at something new, you don't want to be second-guessing yourself about the quality of the data," Konopka says.

By combining sequencing data with analysis of the physical properties, geochemistry, and mineralogy of the core samples, and using sophisticated data mining techniques, researchers hope to discover how environmental conditions

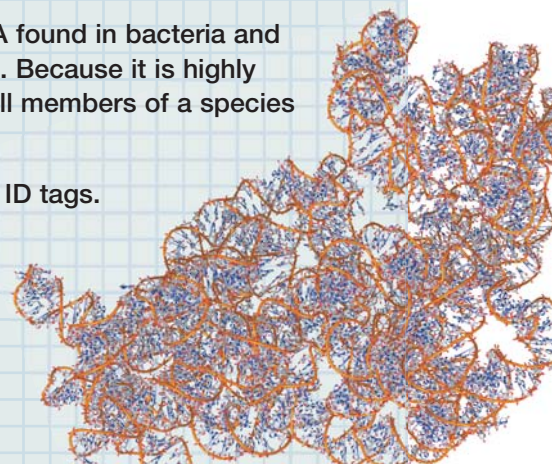
What is 16S rRNA?

16S rRNA is a type of ribosomal RNA found in bacteria and archaea—but not in higher life forms. Because it is highly conserved (resistant to mutations), all members of a species have the same 16S rRNA sequence.

16S rRNA sequences act as species ID tags.

16S rRNA sequencing can:

- Identify bacterial species
- Determine relatedness
- Construct evolutionary trees



Area 300, but a plume of low levels of radionuclides have persisted below the surface, contaminating groundwater above the drinking water standard.

During Summer 2008, project scientists will drill a borehole about 180 feet deep in Area 300 and extract genetic material from 20 sediment samples in five key regions, according to Konopka. The DOE JGI will sequence 16S rRNA [see box] to provide a rough picture of the microbes' species diversity and phylogenetic relationships.

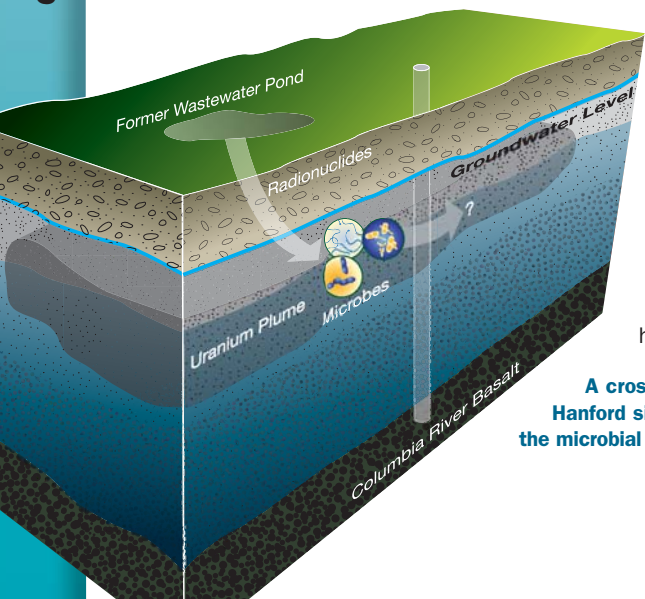
PNNL scientists chose the DOE JGI based on its reputation for providing high-quality sequence data. "When

and microbial communities influence each other, and how this in turn affects the transport and fate of radionuclides in the subsurface.

Research at other sites suggests that subsurface microbes can reduce the mobility of contaminants, preventing their transport off site. About one-third of the world's microorganisms have been estimated to exist in the subsurface, but little is known about their ecology. The project should enable researchers to build better models of subsurface biogeochemistry, Konopka says, and provide valuable information for decision makers regarding when and where to apply remediation techniques.

"What we learn here about the subsurface microbial ecology," he says, "will be applicable to a broad array of DOE sites in the arid western United States." ◉

A cross-section of the study plot at the Hanford site describes the strategy for sampling the microbial communities resident in the subsurface.



CSP 2009 Selections

cont. from page 1

Loblolly pine
(*Pinus taeda*)



sequencing strategies coming on-line at DOE JGI's Production Genomics Facility, Rubin says that the once daunting genome size of the Loblolly pine (*Pinus taeda*)—more than 21 billion bases—is now becoming tractable. Loblolly pine is the most commonly planted tree species in America, accounting for about 75 percent of all seedlings planted each year.

"Its ability to efficiently convert CO₂ into biomass and its widespread use as a plantation tree have also made Loblolly a cost-effective feedstock for cellulosic biofuel production and a promising tool in efforts to curb greenhouse gas levels through carbon sequestration," says Rubin. Because of the pine's enormous genome, the project will begin with a targeted effort to understand the genome structure. Led by Daniel Peterson of Mississippi State University, the project is intended to zero in on genes that can be used for molecular breeding programs to improve Loblolly's use as a biomass feedstock, a carbon sequestration tool, and a source of renewable, high-quality raw materials for lumber and pulp fiber.

The Small Yet Mighty Duckweed

The CSP selections range from these towering pines to not-so-sizable aquatic plants such as duckweed—the smallest,

fastest-growing, and simplest of flowering plants. Greater Duckweed, *Spirodela polyrhiza*, is still relatively small, at less than 10 millimeters tall. Nevertheless, its utility is manifold: as a biotech protein factory, toxicity-testing organism, wastewater remediator, high-protein animal feed, carbon cycling player, as well as basic research and evolutionary model system.

"These plants produce biomass faster than any other flowering plant, and their carbohydrate content is readily converted to fermentable sugars by using commercially available enzymes developed for corn-based ethanol production," says Rubin. "Moreover, duckweed relates to all three of DOE JGI's mission areas: bioenergy, bioremediation, and global carbon cycling."

Spirodela species, which thrives on agricultural and municipal wastewater, efficiently extract excess nitrogen and phosphate pollutants. Duckweed growth on ponds effectively reduces algal growth (by shading), coliform bacteria counts, suspended solids, evaporation, biological oxygen demand, and mosquito larvae while maintaining pH, concentrating heavy metals, sequestering or degrading halogenated organic and phenolic compounds, and encouraging the growth of aquatic animals such as frogs and fowl. This project, submitted by Todd Michael of the Waksman Institute of Microbiology at Rutgers, The State University of New Jersey, unites the efforts of six institutions.

Termites of the Sea and other Aquatic Wonders

The DOE JGI has selected several metagenomes to sequence—complex microbial communities that are isolated directly from the environment or reside inside of a larger organism. These leverage DOE JGI's pioneering expertise honed from previous studies of acid mine drainage and the termite hindgut, where samples yielded scores of different microbes, pro-

ducing hundreds of enzymes with potentially useful industrial applications.

One such metagenome lurks inside of *Bankia setacea*, the giant Pacific shipworm. Shipworms, wood-boring marine bivalves, have been nicknamed "termites of the sea." These animals are capable of feeding solely on wood, utilizing a highly efficient system of symbiotic lignocellulose degradation that is biologically, functionally, and evolutionarily distinct from those found in termites, ruminants, and all other cellulose-consuming animals. Like termites, the ability of shipworms to consume wood depends on symbiotic bacteria that provide enzymes, including cellulases and other hydrolases critical for digestion of wood by the host. The enzymes are also potentially valuable for commercial bioconversion of lignocellulose to ethanol.

Analysis of the shipworm symbiont community metagenome will provide important insights into the composition and function of this unique lignocellulose-degrading bacterial community and will allow valuable comparisons to the recently sequenced termite symbiont metagenome. Unlike termites, shipworms accomplish the complete degradation of lignocellulose with a simple intracellular consortium of just a few related types of microbes. The project was proposed by Daniel Distel of the Ocean Genome Legacy Foundation.

Another marine organism, *Botryococcus braunii*, is a colony-forming green microalga, less than 10 micrometers in size, which synthesizes long-chain liquid hydrocarbon compounds and sequesters them in the extracellular matrix of the colony to afford buoyancy. A type of *B. braunii* produces a family of compounds termed botryococcenes, which hold promise as an alternative energy source. Botryococcenes have already been converted to fuel suitable for internal combustion engines. Geochemical analysis has shown that botryococcenes, presumably from cont. on page 13

Greater Duckweed
(*Spirodela polyrhiza*)



After the First Decade of Metagenomics, Adolescent Growth Spurt Anticipated

Mostly hidden from the scrutiny of the naked eye, microbes have been said to run the world. The challenge is how best to characterize them given that less than one percent of the estimated hundreds of millions of microbial species can be cultured in the laboratory. The answer is metagenomics—an increasingly popular approach for extracting the genomes of uncultured microorganisms and discerning their specific metabolic capabilities directly from environmental samples.

Now, some ten years after the term was coined, metagenomics is going mainstream and already paying provocative dividends according to a Q&A by the DOE JGI Microbial Ecology Program head Philip Hugenholtz and MIT researcher Gene Tyson, published in the September 25 edition of the journal *Nature*.

“By employing the techniques of metagenomics we can go beyond the identification of specific players to creating an inventory of the genes in that environment,” says Hugenholtz. “We find that genes occurring more frequently in a particular community seem to confer attributes beneficial for maintenance of the function of that particular ecological niche.”

Hugenholtz and Tyson were part of the team assembled by University of California, Berkeley geochemist Jillian Banfield to investigate microbial communities associated with the acid mine drainage of Iron Mountain in far Northern

California in 2004. In the dank recesses of the mine, protected by moon suits from the highly acidic effluent, the researchers scooped up pink biofilm growing on the surface of acid mine drainage streams. Extracting the nucleic acid from the sample and directing DOE JGI’s powerful DNA sequencing resource on them, the Banfield team was able to reconstruct the metabolic profiles of the organisms living under such inhospitable conditions—like putting many Humpty-Dumpties back together again. Their findings, published in *Nature* (428:37-43 February 1, 2004), showed that reconstructing the genomes of dominant populations from the environment was feasible and that the imprints of evolutionary selection could be discerned in these genomes.

Since this pioneering work, DOE JGI has gone on to characterize many other metagenomes with other newly selected targets in the sequencing queue at the Walnut Creek, California Production Genomics Facility. These range from the hindguts of termites, to plumb for microbes producing cellulose-degrading enzymes, likewise to microbial communities in the cow rumen, foregut of the Tammar Wallaby, and the crop of the Hoatzin, the Amazon stinkbird. An audio podcast, which includes an interview with Hugenholtz, can be found at: nature.edgeboss.net/download/nature/nature/podcast/v455/n7212/nature-2008-09-25.mp3?ewk13=1.

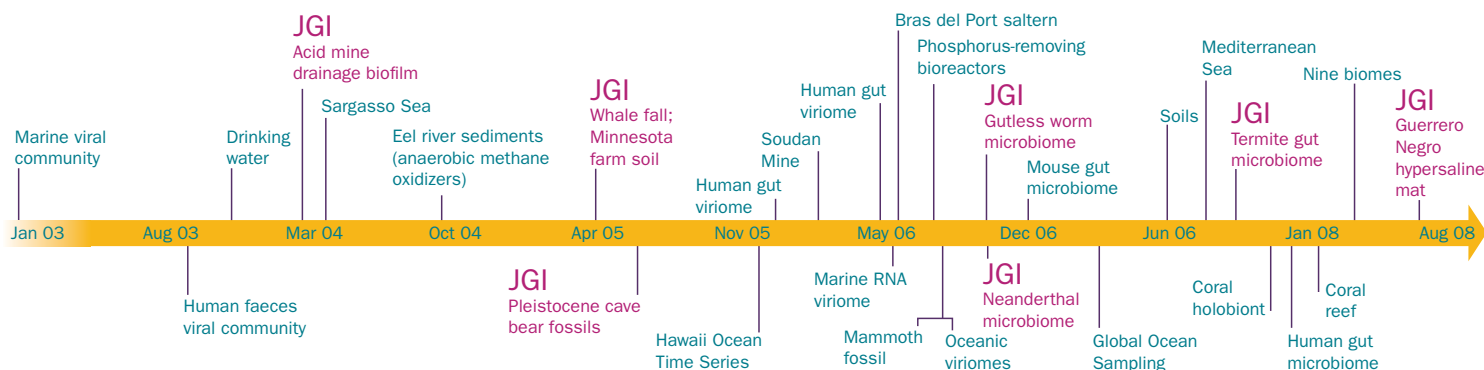
Responding to the steadily increasing

need to manage and interpret the terabases and terabytes of metagenomic data now bubbling up into the public domain, DOE JGI launched the Integrated Microbial Genomes with Microbiome Samples data management and analysis system (see page 2). IMG/M provides tools for analyzing the functional capability of microbial communities based on the DNA sequence of the metagenome in question.

“Metagenomic tools are becoming more widely available and improving at a steady pace,” says Hugenholtz. “But, there are still computational and other bottlenecks to be addressed, such as the high percentage of uncharacterized genes emerging from metagenomic studies.”

In the *Nature* piece, Hugenholtz and Tyson go on to cite the emergence of next-generation sequencing technologies which are already creating a deluge of data that has outstripped the computational power available to cope with it.

“Nevertheless, it’s not necessary to compare all the data to glean useful biological insights,” Hugenholtz says. “What we can capture will help steer the direction toward a relevant data subset to investigate. At least with metagenomics, we have the environmental genetic blueprints awaiting our interpretation. We are still far from capturing and characterizing the dazzling diversity of the microbial life on earth—but at least we have hit upon the gold standard for scratching the surface.”



Lancelet Genome Shows How Genes Quadrupled During Vertebrate Evolution

BY ROBERT SANDERS, UC BERKELEY MEDIA RELATIONS

The newly sequenced genome of a dainty, quill-like sea creature called a lancelet provides the best evidence yet that vertebrates evolved over the past 550 million years through a four-fold duplication of the genes of more primitive ancestors.

The late geneticist Susumu Ohno argued in 1970 that gene duplication was the most important force in the evolution of higher organisms, and Ohno's theory was the basis for original estimates that the human genome must contain up to 100,000 distinct genes.

Instead, the Human Genome Project found that humans today have only 20,000 to 25,000 genes, which means that, if our ancestors' primitive genome doubled and redoubled, most of the duplicate copies of genes must have been lost. An analysis of the lancelet—or amphioxus—genome, being published in the June 19 issue of *Nature*, shows this to be the case.

"Amphioxus and humans had a common ancestor 550 million years ago, which allows us to use amphioxus as a surrogate for that ancestor in terms of understanding how vertebrate genomes evolved," says Daniel S. Rokhsar, a faculty member in the

University of California, Berkeley's Center for Integrative Genomics and program head for computational genomics at the DOE JGI.

Rokhsar and post-doctoral fellow Nicholas H. Putnam performed the genome-wide analyses of the amphioxus genome sequenced by JGI and are lead authors of the *Nature* paper.

For more information, visit: berkeley.edu/news/media/releases/2008/06/18_lancelet.shtml ◉



IMG Upgraded cont. from page 2

were the subject of recently published studies. These include the metagenomic and functional analysis of a microbial community from Lake Washington in Seattle (*Nature Biotechnology*, August 2008), a hypersaline microbial mat (*Molecular Systems Biology*, June 3, 2008), and the airborne metagenome in an indoor urban environment (*PLoS One*, April 2, 2008).

"The IMG/M system is the most comprehensive and constantly evolving resource for analyzing metagenomic data, offering an array of valuable analysis tools in conjunction with links to major public databases," says *Nature Biotechnology* paper author Ludmila Chistoserdova, a microbiologist at the University of Washington. "While very powerful, the system is user-friendly and intuitive, a true 'one-stop-shop' for metagenomics researchers."

More details on IMG/M are available at: img.jgi.doe.gov/img_mw_v241/doc/releaseNotes.pdf.

A special version of IMG, IMG/EDU, has been established to support DOE JGI's Education Program in Microbial Genome Annotation for teaching microbial genome

analysis and annotation using specific microbial genomes in the comparative context of all the genomes available in IMG (www.jgi.doe.gov/education/genomeannotation.html).

"IMG/EDU will serve as the core of a web-based portal that enables undergraduates to participate in microbial genome annotation," says Cheryl Kerfeld, head of DOE JGI's Education Program. "Currently, students at 12 schools nationally are using the portal in their molecular biology, genetics, microbiology, and biochemistry courses in which they examine gene calls and annotate genes and biochemical pathways. By helping to build curated genomes with researchers across the globe, undergraduates will discover the concepts and applications of bioinformatics using IMG/EDU."

IMG/EDU was developed and implemented by Nikos Kyrpides, Natalia Ivanova, Kostas Mavrommatis of the JGI Genome Biology Group and Victor Markowitz, Amy Chen, Ken Chu, Krishna Palaniappan, Ernest Szeto, Yuri Grechkin, and Anna Ratner of the Lawrence Berkeley National

Laboratory (LBNL) Biological Data Management and Technology Center (BDMTC).

An additional resource, IMG/ACT, provides support for managing student classes and assignments, as well as for sharing teaching materials and guiding students in their study of gene calls and functional annotations (jgi.doe.gov/education/annotation_tools.html).

IMG/ACT was developed and implemented by the JGI Scientific and Institutional Application Group's Anthony Kosky, Rene Perrier, David Hays, Joni Fazo, Leila Hornick, Kristen Taylor, Mike Brodhead, as well as the JGI Education Program's Kerfeld and Edwin Kim, with input from the 12 faculty members nationwide who are participating in the pilot.

IMG (img.jgi.doe.gov), IMG/M (img.jgi.doe.gov/m), and IMG/EDU (img.jgi.doe.gov/edu) are the result of a collaboration between the DOE JGI and BDMTC.

For more information on the Microbial Genomics & Metagenomics (MGM) Workshop, go to: www.jgi.doe.gov/meetings/mgm/index.html ◉

Genome of Simplest Animal Reveals Ancient Lineage, Confounding Array of Complex Capabilities

As Aesop said, appearances are deceiving—even in life's tiniest critters. From its discovery in the 1880s, clinging to the sides of an aquarium, to its recent characterization by the DOE JGI, *Trichoplax adhaerens*, a simple and primitive animal, appears to harbor a far more complex suite of capabilities than meets the eye. In a paper in the August 21 online edition of the journal *Nature*, a team of scientists establishes this organism as a branching point of animal evolution and identifies a set of its genes, or a "parts list," that has evolved along particular branches of the tree of life.

"Our whole genome analysis supports placing the placozoans after the sponge lineage branched from other animals," says Daniel Rokhsar, the paper's senior author. Rokhsar is head of DOE JGI's Computational Genomics Program and Professor of Genetics, Genomics, and Development at the University of California, Berkeley.

The analysis of the 98 million base-pair

genome of *Trichoplax* (literally "hairy-plate") illuminates its ancestral relationship to other animals. However, despite the finding, *Trichoplax* is the sole member of the placozoan ("tablet," or "flat" animal) phylum whose relationship to other animals, such as bilaterians (such as humans, flies, worms, and snails) and cnidarians (such as jellyfish, sea anemones, and corals), and sponges is contentious.

Earlier mitochondrial DNA studies suggested that this "mother of all metazoans," *Trichoplax*, was the earliest branch, before sponges diverged, but this remains debatable—even among collaborators.

A Simple Yet Complex Organism

Originally collected from the Red Sea, and cultured over the last 40 years in the laboratory, *Trichoplax* is a two-millimeter flat disk containing fluid sandwiched between two cell layers. It lacks organs and has only four or five cell types. Yet, despite its apparent simplicity, its genome encodes a panoply of signaling genes and transcription factors usually associated with more complex animals.

"*Trichoplax* has had just as much time to evolve as humans, but because of its morphological simplicity, it is tempting to think of it as a surrogate for an early animal," says Mansi Srivastava, the study's first author and a graduate student under the direction of Rokhsar at the Center for Integrative Genomics, U.C. Berkeley.

"*Trichoplax* is an ancient lineage—a good representation of the ancestral genome that is shedding light on the kinds of genes, the structures of genes, and even how these genes were arranged on the genome in the common ancestor 600 million years ago," says Srivastava. "It has retained a lot of primitive features relative to other living animals."

Further research is underway to defin-

itively place the organism's evolutionary position. "The latest and most complex analysis again suggests that placozoans populated the oceans long before sponges evolved," says Bernd Schierwater, director of the Institute of Animal Ecology & Cell Biology and head of the Center for Biodiversity at TiHo Hannover, Germany. Schierwater, a study co-author, joined Stephen Dellaporta and Leo Buss of Yale University in proposing the *Trichoplax* sequencing project in 2004 to DOE JGI's Community Sequencing Program.

"The outcome of the *Trichoplax adhaerens* genome sequencing is so exciting that we are now culturing another 13 placozoan species in order to identify the most basal placozoan lineage and genome," says Schierwater.

Trichoplax has no neurons, but has many genes that are associated with neural function in more complex animals. "It lacks a nervous system, but it still is able to respond to environmental stimuli," explains Rokhsar. "It has genes, such as ion channels and receptors, that we associate with neuronal functions, but no neurons have ever been reported."

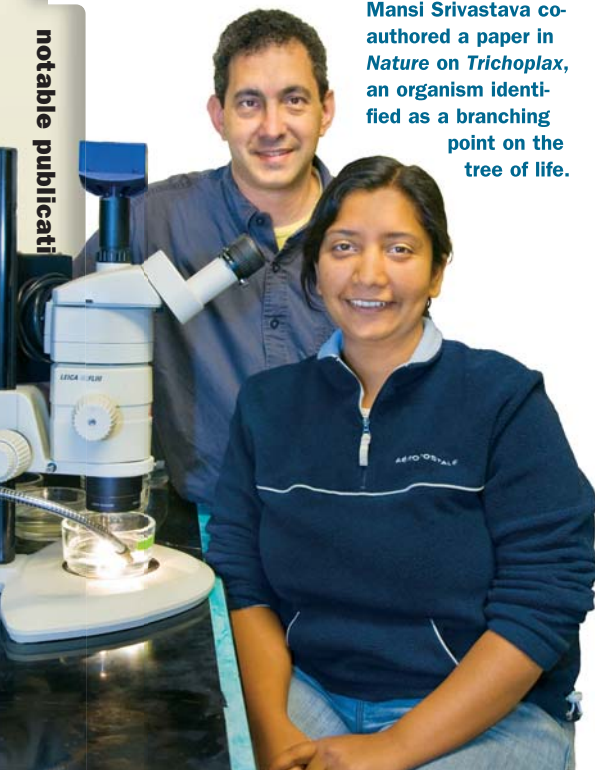
Not So Distant From the Human Genome

Of the 11,514 genes identified in the six chromosomes of *Trichoplax*, 80 percent are shared with cnidarians and bilaterians. *Trichoplax* also shares more than 80 percent of its introns—the regions within genes that are not translated into proteins—with humans. Even the arrangement of genes is conserved between the *Trichoplax* and human genomes. This stands in contrast to other model systems such as fruit flies and soil nematodes that have experienced a paring down of non-coding regions and a loss of the ancestral genome organizations.

cont. on page 14

Dan Rokhsar and Mansi Srivastava co-authored a paper in *Nature* on *Trichoplax*, an organism identified as a branching point on the tree of life.

notable publication

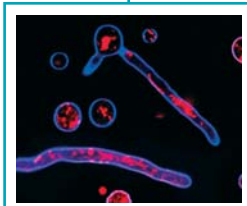


Other Notable Recent JGI Publications

Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)

Nature Biotechnology May 2008, Vol. 26 No. 5, pp. 553-60

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmöll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barbote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS.



Trichoderma reesei is the main industrial source of cellulases and hemicellulases used to depolymerize biomass to simple sugars that are converted to chemical intermediates and biofuels, such as ethanol. Unexpectedly, considering the

industrial utility and effectiveness of the carbohydrate-active enzymes of *T. reesei*, its genome encodes fewer cellulases and hemicellulases than any other sequenced fungus able to hydrolyze plant cell wall polysaccharides. Our analysis, coupled with the genome sequence data, provides a roadmap for constructing enhanced *T. reesei* strains for industrial applications such as biofuel production.

www.nature.com/nbt/journal/v26/n5/abs/nbt1403.html

Human-specific gain of function in a developmental enhancer

Science September 5, 2008, Vol. 321 No. 5894, pp. 1346-50

Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, Pennacchio LA, Rubin EM, Noonan JP.

Changes in gene regulation are thought to have contributed to the evolution of human development. However, in vivo evidence for uniquely human developmental regulatory function has remained elusive. In transgenic mice, a conserved noncoding sequence (HACNS1) that evolved extremely rapidly in humans acted as an enhancer of gene expression that has gained a strong limb expression domain relative to the orthologous elements from chimpanzee and rhesus macaque. www.sciencemag.org/cgi/content/abstract/321/5894/1346



Finishing in the Future

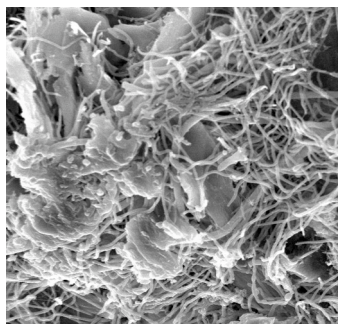
Keynote speaker and Nobel Prize Winner Sydney Brenner (right) regales JGI Deputy Director Jim Bristow (left) and JGI Microbial Interactions Program head Patrick Chain with tales of genomes he has known at the Third Annual Finishing in the Future meeting, which was held May 28-30, 2008 in Santa Fe, New Mexico.

The 2009 meeting dates are now set for May 27-29. For more information, see www.lanl.gov/finishinginthefuture or contact Chris Detter, cdetter@lanl.gov.

JGI in the News

Termite Bellies and Biofuels. Scientist Falk Warnecke's research into termite digestion may hold solutions to our energy crisis. (Published in *Smithsonian*, August 1, 2008, www.smithsonianmag.com/science-nature/termites-bellies-biofuels.html)

Gut Reactions. The termite's stomach, of all things, has become the focus of large-scale scientific investigations.



Could the same properties that make the termite such a costly pest help us solve global warming? (Published in *The Atlantic*, August 20, 2008, www.theatlantic.com/doc/200809/termites)

Electron-micrograph image of the termite's third gut, where food is turned into fuel.

Analysis of Lake Washington Microbes Shows the Power of Metagenomic Approaches

BY MADOLYN BOWMAN ROGERS

Today's powerful sequencing machines can rapidly read the genomes of entire communities of microbes, but the challenge is to extract meaningful information from the jumbled reams of data. In a paper published in *Nature Biotechnology* August 17, a collaboration headed by researchers at the University of Washington and the DOE JGI describes a novel approach for extracting single genomes and discerning specific microbial capabilities from mixed community ("metagenomic") sequence data.

For the first time, using an enrichment technique applied to microbial community samples, the research team explored the sediments in Lake Washington, located on Seattle's eastern edge, and characterized biochemical pathways associated with nitrogen cycling and methane utilization, important for understanding methane generation and consumption by microbes. Methane is both a greenhouse gas and a potential energy source.

"Even if you have lots of sequences, for complex communities it still doesn't tell you which organism is responsible for which function," says the paper's senior author, Ludmila Chistoserdova, a microbiologist at the University of Washington. "This publication presents an approach, via simplification and targeted metagenomic sequencing, of how you can go after the function in the environment."

Chistoserdova and colleagues study microbes that oxidize single-carbon compounds such as methane, methanol, and methylated amines, which are compounds contributing to the greenhouse effect and are part of the global carbon cycle.

"To utilize these single-carbon compounds, organisms employ very specialized metabolism," says Chistoserdova. "We suspect that in the environment, there are novel versions of this metabolism, and



The Methylamine-enriched microbe community of Lake Washington sediment features *Methylotenera* cells. Photo: D. Kunke; color, E. Latypova.

possibly completely novel pathways."

Most of the microbes that oxidize single-carbon compounds are unculturable and therefore unknown, as are the vast majority of microbes on Earth. To find species of interest, the researchers sequenced microbial communities from Lake Washington sediment samples, Chistoserdova says, because lake sedi-

A Microbial First: Isotope Probing

To enrich the samples for the microbes of interest, the researchers adapted a technique called stable isotope probing. This is the first time the technique has been used on a microbial community, Chistoserdova says. The researchers used five different single-carbon compounds labeled with a heavy isotope of carbon, and fed each compound to a separate sediment sample. The microbes that could consume the compound incorporated the labeled carbon into their DNA, Chistoserdova says, while organisms that couldn't use the compound did not incorporate the label. The labeled DNA was then separated out and sequenced. In this way, microbial "subsamples" were produced that were highly enriched for organisms that could metabolize methane, methanol, methylated amines, formaldehyde, and formate.

The functionally enriched samples contained far fewer microbes than the total sample, Chistoserdova says. The sample that was fed methylated amines was simple enough that the research group was able to extract the entire genome of a novel microbe, *Methylotenera mobilis*,

"Even if you have lots of sequences, for complex communities it still doesn't tell you which organism is responsible for which function. This publication presents an approach, via simplification and targeted metagenomic sequencing, of how you can go after the function in the environment," says the paper's senior author, Ludmila Chistoserdova.

ment is known to be a site of high methane consumption. However, these sediment samples contained over 5,000 species of microbes performing a complex, interconnected array of biochemical tasks.

which normally comprises less than half a percent of the community but appears to be a first responder to methylated amines in the environment. The researchers were able to construct *cont. on page 15*

CSP 2009 Selections cont. from page 7

ancient *B. braunii* communities, also comprise a portion of the hydrocarbon masses in several modern-day petroleum and coal deposits.

While algae have been recognized for their role in carbon sequestration and for biofuels production, little information, either genetic or metabolic, has been reported for this particular alga. This project, led by Andrew Koppisch and colleagues from Los Alamos National Laboratory and five other institutions, will target the identification of specific metabolic pathways responsible for hydrocarbon synthesis to alleviate bottlenecks in biofuels production.

Other CSP 2009 Projects

One metagenome project entails a sampling of the foregut of *Opisthocomus hoazin*, or hoatzin, a leaf-eating Amazonian pheasant-like stinkbird. The unique fermentative organ of this prehistoric relic harbors an impressive array of novel microbes, like

that of cows and other ruminants. But instead of a rumen, stinkbirds possess a crop, an enlargement of the esophagus where the fermentation takes place—and is the source of the stink. The characterization of its contents will likely lead to the identification of novel microbial enzymes that degrade plant cell walls.

The most abundant source of carbon is plant biomass, composed primarily of cellulose, hemicellulose, and lignin. Many microorganisms are capable of utilizing cellulose and hemicellulose as carbon and energy sources, but a much smaller group of filamentous fungi has evolved with the ability to depolymerize lignin, the most recalcitrant component of plant cell walls. Collectively known as white rot fungi, they possess the unique ability to efficiently depolymerize lignin in order to gain access to cell wall carbohydrates for carbon and energy sources. *Ceriporiopsis subvermispora* rapidly depolymerizes lignin with relatively

little cellulose degradation. The annotated gene set of *C. subvermispora* and comparative analyses with the lignin degraders *P. chrysosporium* and *Pleurotus ostreatus* (both sequenced by DOE JGI) will advance the understanding of these complex oxidative mechanisms involved in lignocellulose conversions. This project was proposed by Dan Cullen from the University of Wisconsin–Madison.

Other CSP projects featured in this issue of *The Primer* are: predatory protists, led by Monterey Bay Aquarium Research Institute's Alexandra Worden (page 3), the Hanford Site's subsurface microbial communities, led by Pacific Northwest National Laboratory's Allan Konopka (page 6), nitrogen-fixing bacteria, with the University of Florence (page 4), and microbes frozen in Lake Vostok of Antarctica (page 5).

For the complete list of CSP 2009 sequencing projects, see: www.jgi.doe.gov/sequencing/cspseqplans2009.html.

Genomics of Biofuels cont. from page 1

for their desirable traits, resulting in today's highly productive food crops. But now, given the energy and climate challenges, we simply don't have thousands of years. So by applying the power of genomics to these problems, we are seeking to speed up both the domestication of energy crops and the technologies for converting them to suitable biofuels. The result will be a more carbon-neutral approach to meeting part of our transportation needs."

In the *Nature* review, Rubin describes the processes entailed in biofuel production from lignocellulose: the harvesting of biomass, pretreatment and saccharification, which results in the deconstruction of cell wall polymers into component sugars, and then the conversion of those sugars into biofuels through fermentation. Each step, he says, offers an opportunity

for genomics to play a significant role.

"With the data that we are generating from plant genomes, we can home in on relevant agronomic traits such as rapid growth, drought resistance, and pest tolerance, as well as those that define the basic building blocks of the plants cell wall—cellulose, hemicellulose, and lignin. Biofuels researchers are able to take this information and design strategies to optimize the plants themselves as biofuels feedstocks—altering, for example, branching habit, stem thickness, and cell wall chemistry, resulting in plants that are less rigid and more easily broken down."

For microbial biomass breakdown, Rubin says that many candidates have already been identified. These include *Clostridia* species, for their ability to degrade cellulose, and fungi that express

genes associated with the decomposition of the most recalcitrant features of the plant cell wall, lignin, the phenolic "glue" that imbues the plant with structural integrity and pest resistance. The white rot fungus *Phanerochaete chrysosporium* produces unique extracellular oxidative enzymes that effectively degrade lignin by gaining access through the protective matrix surrounding the cellulose microfibrils of plant cell walls.

Another fungus, the yeast *Pichia stipitis*, ferments the five-carbon "wood sugar" xylose abundant in hardwoods and agricultural harvest residue. Rubin says that *Pichia's* recently sequenced genome has revealed insights into the metabolic pathways responsible for this process, guiding efforts to optimize it in commercial production strains. Pathway engineering promises to produce cont. on page 15

A Conversation with Alexandra Z. Worden *cont. from page 3*

their growth or lead to one group being more abundant than another, we might be able to develop models that could predict a range of possible future scenarios, with, say, a 1° C temperature change in the system. This would help in appropriate decision making, which, given the current CO₂ situation, is really needed.

Q. What are some of the challenges of this project?

A. After sequencing the genome, we have this unassembled, fragmented blueprint, a list of parts. We don't know which of those parts are important, and for

some of them, we also don't know what they do. But they have the power to give us insights into mechanisms of control, and understanding the mechanisms is really the only way we can develop more predictive capabilities.

We're in a high level of urgency for figuring out fundamental aspects of community and carbon dynamics. I went into this wanting to help the environment, and I am daunted at this stage. Things are changing so fast, people are proposing scientifically unfounded quick fixes, and we're still at such a basic level of understanding. There

are some indications that in the north Pacific there have been major shifts in the photosynthetic community in the last 50 years. That's of huge concern, given that we don't know the drivers or consequences for marine life or climate regulation. Given that the oceans are 75 percent of the globe, and given that they drive what we breathe (both in a geological sense and in their role regulating the modern atmosphere), it's a pretty big hole in our knowledge.

Nevertheless I do feel hopeful. Look at where medicine was 100 years ago and where it is now. They did that without the tremendous computational power we have today. In the marine world, we start with that tremendous power. There's a lot we can figure out if we get enough minds thinking about the issues at hand.

Q. How has the collaboration with JGI helped advance your research?

A. It's an amazing resource. JGI is fabulous at sequencing and assembling genomes. They're also now tackling some of these much larger plant genomes and the untapped diversity of unicellular eukaryotic life. The challenge ahead lies in the data analysis—for example, modeling the genes accurately and annotation.

Q. What are the personal rewards of your work, and what is its value to society?

A. Genomes serve as tools for doing ecological studies, so getting these tools out to a broader community, in a high-quality, annotated format that enables the scientific community to go from there, is one reward. Bigger-picture reward, I would love to participate in developing predictive carbon cycling models.

More broadly, for society, I think we could do preventative medicine in the oceans if we understood the system a little better. So this work has tremendous value for advancing a preventative approach for the ways humans interact with the environment both on land and in the sea. ○

Trichoplax, a Simple and Ancient Animal *cont. from page 10*



A bright-field high magnification image of *Trichoplax adhaerens* captured by Nature paper author Ana Signorovitch, Yale University.

With its pancake shape, gutless feeding, and genomic primitiveness, the rich array of metabolic capabilities begs additional consideration. While it has been observed to motor around via cilia, eat by mounting its prey, and reproduce by fis-

sion (pulling itself into pieces), it may in fact have a secret sex life.

"Some of our new placozoan species show frequent sexual reproduction while others never show any signs of sex," says Schierwater. "The genome data allow us to search for the genes responsible for sex and life cycle complexity."

Adds Rokhsar: "It's remarkable that we have the whole genome sequence, but we still know so little about this animal in the wild. Hopefully the genome sequence will stimulate more studies of this enigmatic creature."

Other authors include DOE JGI's Jarrod Chapman, Nicholas Putnam, Uffe Hellsten, Alan Kuo, Asaf Salamov, Harris Shapiro, and Igor Grigoriev; Jane Grimwood and Jeremy Schmutz of the Stanford Human Genome Center and DOE JGI; Emina Begovic, Therese Mitros, and Meredith Carpenter of U.C. Berkeley; Takeshi Kawashima of the Okinawa Institute of Science and Technology; Ana Signorovitch, and Maria Moreno, Leo Buss, and Stephen Dellaporta of Yale University; and Kai Kamm the University of Veterinary Medicine Hannover, Germany. ○

Lake Washington Microbes cont. from page 12

much of *M. mobilis*'s biochemistry and predict that it is also involved in nitrogen cycling, demonstrating the utility of metagenomic analysis.

A Pioneering New Technique for Metagenomic Assemblies

The DOE JGI performed the sequencing and assembly of these complex metagenomic data sets. The complexity of the community's sequence samples created new challenges for genome assembly. "It is very important for metagenomic assemblies to rely on high-quality reads," says Alla Lapidus, microbial geneticist at the DOE JGI and co-author of the paper. If some of the sequence is of low quality, she says, it can lead to errors in assembly and gene annotation.

Because of the need for higher quality control, Lapidus says, the DOE JGI devel-

oped a new quality-control approach that involves using a computer tool called LUCY to trim out low-quality sequences in combination with the Paracel Genome Assembler, which appeared to be more appropriate for metagenomic assemblies. This approach was pioneered on the Lake Washington project, Lapidus says, and due to its superior results it is now the standard metagenomic assembly method at the DOE JGI.

"The DOE JGI's unique Integrated Microbial Genomics with Microbiome Samples (IMG/M) [img.jgi.doe.gov/m] data-management system was used for detailed annotation, and was instrumental for efficient comparative analysis and metabolic reconstruction of the samples," Lapidus says.

Michael Galperin, a microbial geneticist at the National Center for Biotechnology Information at the National Institutes of

Health, who was not involved in the study, said in an email that the paper describes "an interesting novel approach" and the results "constitute a significant advance in the emerging discipline of metagenomics."

"I think other people can use the same approach in different environments, as long as they have an enrichment technique," Chistoserdova says. "For us this work is just the beginning because now we will be using this metagenomic sequence as a scaffold for downstream experiments in our lake."

Other authors include DOE JGI's Natalia Ivanova, Alex Copeland, Asaf Salamov, Igor Grigoriev, Susannah Tringe, David Bruce (Los Alamos National Laboratory), and Paul Richardson; and Ernest Szeto and Victor Markowitz of the Data Management and Technology Center at Lawrence Berkeley National Laboratory. ○

Genomics of Biofuels cont. from page 13

a wider variety of organisms able to ferment the full repertoire of sugars derived from cellulose and hemicellulose and tolerate higher ethanol concentrations to optimize fuel yields.

Rubin also touches on the emerging technology of metagenomics—characterizing, without the need for laboratory cul-

ture, the metabolic profile of organisms residing in an environmental sample—for the identification of enzymes suitable for industrial-scale biofuel production.

"Using this prospecting technique, we can survey the vast microbial biodiversity to gain a better picture of the metabolic potential of genes and how they can be enlisted for the enzymatic deconstruction of biomass and subsequent conversion to high-energy-value fuels," he says.

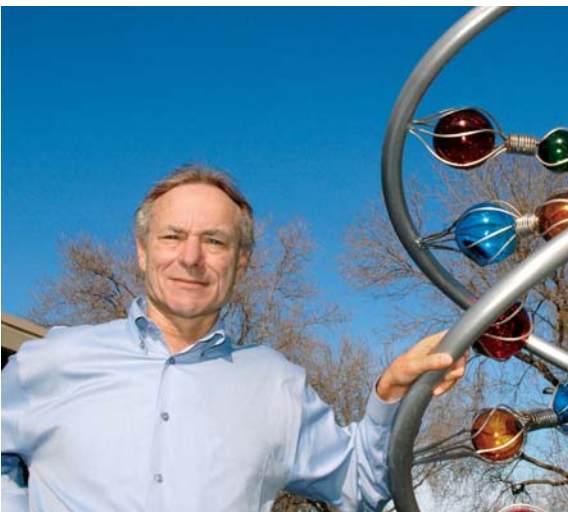
As an example, Ruben cites an analysis of the hindgut contents of nature's own bioreactor, the termite, which has yielded more than 500 genes related to the enzymatic deconstruction of cellulose and hemicellulose (published in *Nature* 450:560-565, November 22, 2007).

The *Nature* review goes on to list the feedstock genomes, microbial "biomass

degraders," and "fuel producers" completed or in progress. These include the first tree genome completed—that of the poplar *Populus trichocarpa*—and other plants in the sequencing queue, such as soybean, switchgrass, sorghum, eucalyptus, cassava, and foxtail millet. In addition, Rubin points to oil-producing algae as an alternative source for biodiesel production; the alga *Chlamydomonas reinhardtii* is just one of several algal species that has been characterized for their ability to efficiently capture and convert sunlight into energy.

"Given the daunting magnitude of fossil fuel used for transportation, we will likely have to draw from several different sources to make an appreciable impact with cellulosic biofuels, all of which will, in some significant way, be informed by genomics," says Rubin. ○

JGI Director Eddy Rubin is leading an effort, supported by the Energy Biosciences Institute, to identify enzymes produced by the abundant microbes in the rumen of forage-feeding animals that can potentially be adapted for cellulosic biomass conversion to biofuel.



Genomics Garden

JGI Operations Manager Ray Turner inspects the bountiful first harvest from the Production Genomics Facility's bioenergy feedstock garden. The bulk of the *Zea mays* (corn, shown here) crop was enjoyed by the indigenous ground squirrel population, while the nearby *Miscanthus* and switchgrass stands were left untouched.



JGI Planning Retreat



The DOE Joint Genome Institute convened a planning retreat on August 27-29, 2008 in Pacifica, California, to jumpstart a strategic plan that will guide JGI's path forward over the next five years. Evening presentations included those from Steven Beckwith, University of California Vice President for Research and Graduate Studies, and James Siegrist, Lawrence Berkeley National Laboratory Associate Laboratory Director for General Sciences (not pictured). Other participants (pictured) were from LANL, LLNL, NIH, ORNL, PNNL, HudsonAlpha, and DOE, and several other organizations.

Poplar Watch

In September, this poplar, or black cottonwood, measured here by JGI Instrumentation Engineer Steven Wilson, reached 20' 8" (630 cm), or nearly five times its height at planting. The tree was planted two years ago in commemoration of the publication in the journal *Science* of JGI-led research on the genome of the *Populus trichocarpa*, the first complete DNA sequence of a tree.



JGI User Meeting

SAVE THE DATE

march 25-27, 2009

CSO 15510

Contact The Primer

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