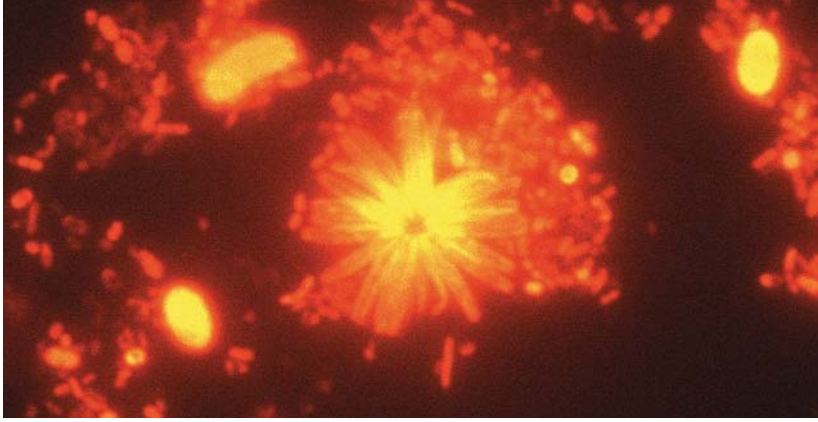


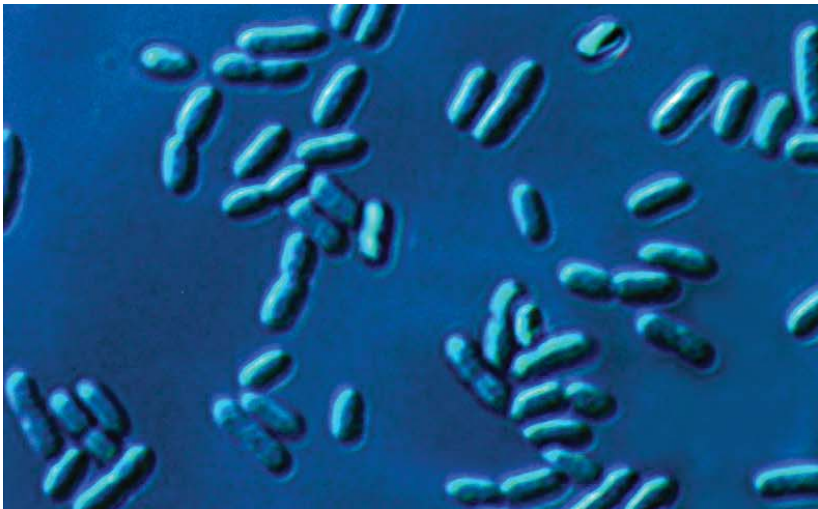
U.S Department of Energy

Joint Genome Institute

Progress Report 2009

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

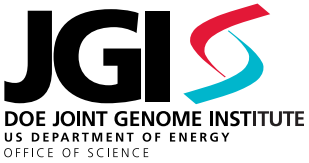
U.S. DEPARTMENT OF ENERGY
Office of Science



Representatives of the four DOE JGI genome programs (plant, fungal, metagenome, and microbe) grace the cover of this annual report. Three of the organisms were among the 81 projects selected in 2009 for the 2010 Community Sequencing Program portfolio. Photo credits, clockwise from bottom left: *Brachypodium distachyon* by Roy Kaltschmidt, LBNL; *Amanita thiersii* fungus by Joe McFarland; Cow rumen metagenome by Gemma Henderson, AgResearch; Soybeans by Roy Kaltschmidt, LBNL; *Zymomonas mobilis* Z4 by Katherine Pappas, University of Athens

U.S Department of Energy
Joint Genome Institute

Progress Report 2009



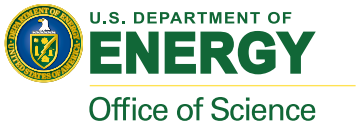


Table of Contents

Photo by Roy Kaltschmidt, LBNL

DOE JGI Mission

The mission of the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) is to serve the diverse scientific community as a user facility enabling the application of large-scale genomics and analysis of plants and microbes to address the DOE mission goals in bioenergy and the environment.

2009 DOE Joint Genome Institute Director's Perspective

Eddy Rubin, DOE JGI Director

One trillion bases and well beyond

The year 2009 was a successful year for the DOE Joint Genome Institute (DOE JGI) from the perspective of genome sequence data generation, project completion, and translating those data into meaningful and relevant science.

This last year, we surpassed the 1 trillion base milestone—which represents an eight fold increase over 2008. But to really get a sense for how far we have come, we produced some 50,000 times more data in the past year than we did in the Institute's first year of operation, just over a decade ago. This dramatically exceeds the famous "Moore's Law" of computer chip growth.

"Stupendous" science

Where does all this information go? It is made freely available to government, academic, and industrial collaborators around the world. We can point to a quote from a recent DOE Biological and Environmental Research review committee tasked to evaluate the science and operations of the DOE JGI, which assessed the science we generate as "nothing short of stupendous."

Over the past year, there has been a steady stream of nearly 130 original scientific publications, based on genomic data and its analysis by the DOE JGI, including 15 published in the journals *Nature* and *Science*. Below is a glimpse into the diversity of significant research contributed to the scientific literature by the DOE JGI:

- Soybean Genome (published in *Nature*, January 2010). One of the most important global sources of protein and oil, the soybean is now the first legume species with a published complete draft genome sequence. This affords scientists a better understanding of the plant's capacity to turn sunlight, carbon dioxide, nitrogen, and water into concentrated energy, protein, and nutrients for human and animal use.
- Genomic Encyclopedia of Bacteria and Archaea (published in *Nature*, December 2009). The analysis of more than 50 microbial genomes representing "phylogenetic dark matter of the biological universe" will shed light on the diversity of gene families and improve the understanding of how microbes acquire new functions.
- Sorghum Genome (published in *Nature*, January 2009). A major food and fodder plant with high potential as a bioenergy crop, the sorghum genome analysis is an important step toward the development of cost-effective biofuels made from nonfood plant fiber.
- *Postia placenta* Genome (published in the *Proceedings of the National Academy of Sciences (PNAS)*, February 2009). The analysis of this brown-rot genome offers us a detailed inventory of the biomass-degrading enzymes that this and other fungi possess—to catalyze important early steps of biofuel production.

- *Micromonas* Comparative Analysis (published in *Science*, April 2009). A comparison of two geographically diverse isolates of this photosynthetic algal genus highlights the genes enabling them to capture carbon and maintain its delicate balance in the oceans.
- *Trichoderma reesei* Strain Comparative Analysis (published in *PNAS*, August 2009). Comparing strains of this important fungus used to produce industrial cellulases will lead to the development of more robust variants with higher productivities for such demanding applications as cellulosic biofuels production.
- Tracking Microbial Diversity without Cultivation (published in *PNAS*, September 2009). With a new cultivation-independent approach to characterizing marine microbial habitats by analyzing and comparing the strategies of the dominant organisms, scientists can better understand the carbon flux going through the environment, which in turn will inform how this process may be affected by global warming.
- Oxygen Minimum Zone (OMZ) Microbes (published in *Science*, October 2009). This study of the genomes of the uncultivated microbes found in OMZs will enable scientists to better understand how global geochemical cycles such as the carbon and nitrogen cycles result from microbial activities.
- Diatom Species Comparative Analysis (published in *Nature*, October 2009). Diatoms have a profound influence on climate, and this work will highlight their evolutionary origins and how they acquired advantageous genes from bacterial, animal, and plant ancestors, enabling them to thrive in today's oceans.
- Extreme Single Microbe Ecosystem (published in *Science*, October 2008). This remarkable microbe resides nearly 2 miles beneath the surface of the earth in a South African gold mine in complete isolation, getting its energy not from the sun but from hydrogen and sulfate produced by the radioactive decay of uranium.

The DOE Joint Genome Institute workforce gathers in the courtyard adjoining the Genomics Garden and helix sculpture at the Institute's headquarters in Walnut Creek. Photo by Roy Kaltschmidt, LBNL

In addition to the sequence information that the DOE JGI generates, we are playing a major role in responding to the needs of the research community by developing the computational and bioinformatic resources and standards to manage and analyze these data—converting this information into useful knowledge. Recent contributions include an enhanced Phytozome.net, a central "hub" for Web access to a rapidly growing number of plant genomes and the Expert Review version of the Integrated Microbial Genomes system to provide scientists with curation tools that improve the annotations of microbial genomes.

What lies ahead…

For 2010, the DOE JGI selected 81 new genome projects through its Community Sequencing Program. These offer a targeted portfolio of the planet's biodiversity, involving organisms that hail from Northern Arctic regions to New Zealand's South Island.

With its steadily increasing sequencing capacity, the DOE JGI will continue

to be active in advancing the field of metagenomics. The scale of these projects will increase from gigabases to terabases of data from sequence-based analyses of environmental microbial "communities." This addresses a key recommendation of the National Academy of Science report on *The New Science of Metagenomics* to:

"… characterize a few microbial habitats in great depth, using large multidisciplinary and multinational teams to address challenges in metagenomics that require massive datasets or highly diversified scientific approaches and engaging more investigators than would typically participate in a medium-sized center."

These projects will transform the current limited perspective from a genetic snapshot of a particular environment to the equivalent of genetic moving pictures. One such project to be launched in the coming year will be to explore the Great Prairie Soil Metagenome. The Great Prairie sequesters the most carbon of any soil system in the United States. It also yields large amounts of biomass annually, which may factor into future cellulosic biofuels strategies. The project will entail a comparison of samples from pristine, never-tilled prairie soil with Iowa cornfields that have been in continuous cultivation for over 100 years. The goal is to generate the information that will enable improved soil management, carbon sequestration, ecosystem services, and productivity.

In addressing the needs of our largest set of "customers" — the three DOE Bioenergy Research Centers — the DOE JGI's capabilities will remain essential for accelerating the progress of their research. These Centers are specifically focused on the development of next-generation cellulosic biofuels through efforts to improve biomass degradation and novel strategies for fuel production.

In collaboration with researchers at the Joint BioEnergy Institute, the DOE JGI sequenced the metagenome of a compost-inoculated switchgrass to identify enzymes that can break down the lignocellulose in this bioenergy feedstock candidate.

Working with researchers at the BioEnergy Science Center, DOE JGI generated the first sequenced genome of a marine bacterium that can degrade plant cell walls.

Finally, as reported in the November 2009 issue of *Science*, researchers from the Great Lakes Bioenergy Research Center described a symbiotic relationship between fungus-growing ants and bacteria, and proposed what might be the primary source of terrestrial nitrogen in the tropics.

These examples illustrate how the tools and infrastructure for the generation and analysis of sequence data are helping the DOE JGI and its partners advance DOE mission-relevant science. This, coupled with our commitment to continuously build upon these capabilities, will enable us to offer a unique breadth of genomics resources to our ever-expanding user community so that they are well-positioned to accomplish the science necessary for responding to present and future energy and environmental challenges.

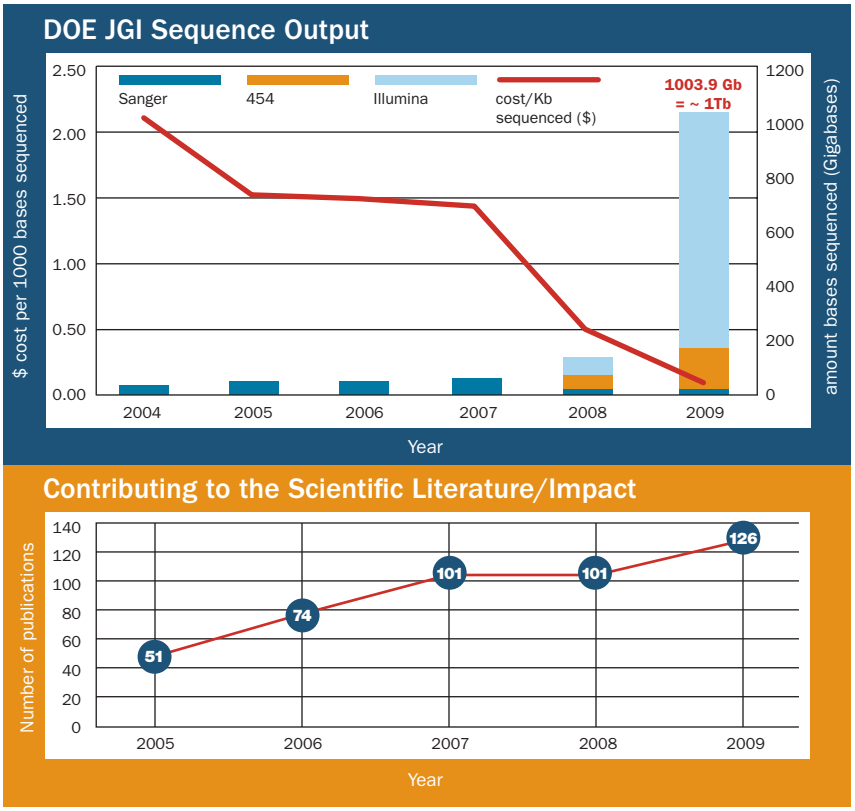
Edward M. Rubin, MD, PhD
Director
DOE Joint Genome Institute

DOE JGI by the Numbers

The DOE JGI—a whole that is more than the sum of its parts

Part of the Lawrence Berkeley National Laboratory (LBNL), the DOE JGI in Walnut Creek, California, carries out production-scale sequencing and analysis, activities that are augmented by the specialized tasks drawn from the specific capabilities provided by its partner laboratories, Lawrence Livermore National Laboratory (LLNL), Los Alamos National Laboratory, Oak Ridge National Laboratory, Pacific Northwest National Laboratory, and the HudsonAlpha Institute for Biotechnology (in Huntsville, Alabama).

The DOE JGI Director, with input from the DOE JGI Joint Coordinating Committee (JCC), which is made up of two members from each of the partner laboratories, coordinates the vision for the overall DOE JGI as well as the activities carried out at each of the partner labs. The DOE JGI Director, representing the JCC, communicates the DOE JGI's technological and strategic vision to the DOE's Office of Biological and Environmental Research, and funding decisions are based upon these discussions. The DOE JGI 80,000-square-foot headquarters in Walnut Creek has a staff of 250, which draws primarily from LBNL and LLNL employees.



DOE JGI User Community

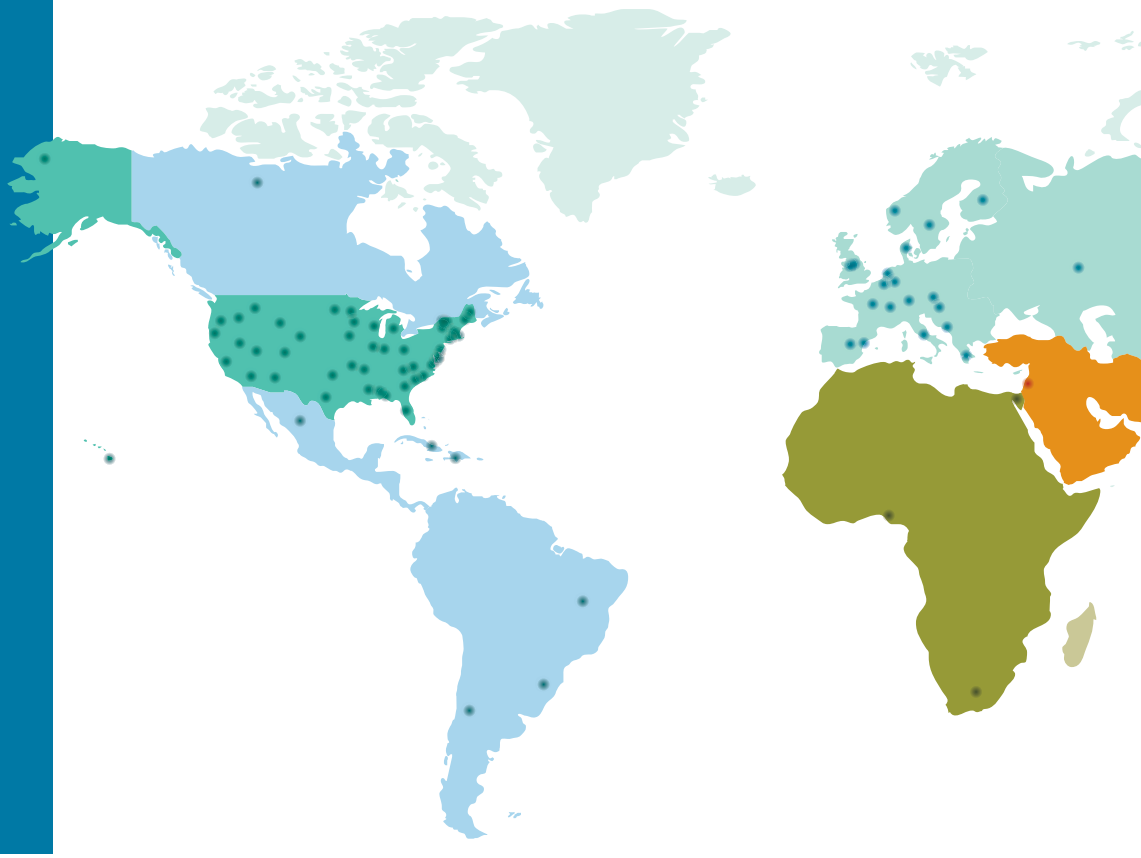
The DOE JGI's user facility approach is based on the concept that by focusing the most advanced sequencing and analysis resources on the best peer-reviewed proposals drawn from a diverse community of scientists, the DOE JGI will catalyze creative approaches to addressing DOE mission challenges. This strategy has clearly worked, only partially reflected in the fact that the DOE JGI has played a major role in more than 50 papers published in the prestigious journals *Nature* and *Science* alone over the past three years. The involvement of a large and engaged community of users working on important energy and environmental problems has helped maximize the impact of DOE JGI science.

The Community Sequencing Program was created to provide the scientific community at large with access to high-throughput sequencing at the DOE JGI. Sequencing projects are chosen based on scientific merit—judged through inde-

1,771

JGI Users Worldwide in 2009

UNITED STATES	REST OF AMERICAS	EUROPE	MIDDLE EAST
1,129	88	433	7

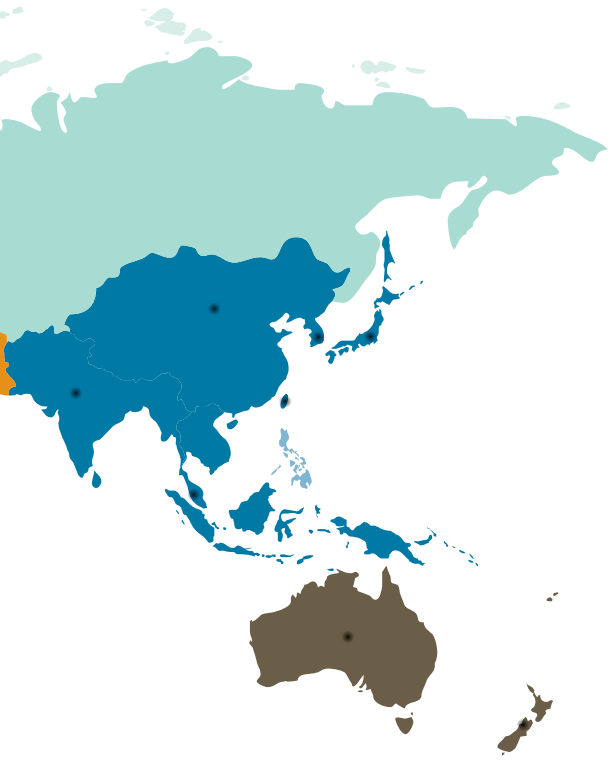


pendent peer review—and relevance to the DOE mission.

The DOE JGI's largest users are the DOE Bioenergy Research Centers (BRCs), which were launched in 2007 to accelerate basic research in the development of next-generation cellulosic and other biofuels through focused efforts on biomass improvement, biomass degradation, and strategies for fuels production. The three centers are the Joint BioEnergy Institute, led by LBNL and located in Emeryville, California; the BioEnergy Science Center at Oak Ridge National Laboratory; and the Great Lakes Bioenergy Research Center at the University of Wisconsin, Madison. By agreement with the DOE, the BRC projects are afforded top priority for sequencing and analysis at the DOE JGI.

After the BRCs, the DOE JGI user community draws heavily from academic research institutions, along with the national laboratories, federal agencies, and a small number of companies. Worldwide, there are nearly 1,800 unique collaborators on active projects. The broader user community is engaged in collaborations that are global in nature. In addition, the DOE JGI's influence is extensive, considering that hundreds more researchers every year tap sequence data posted by the DOE JGI on its numerous genome portals and at the National Center for Biotechnology Information's GenBank.

AFRICA 5
ASIA 57
AUSTRALASIA 52



Alabama	7	Ohio	8	Ireland	2
Alaska	1	Oklahoma	18	Italy	17
Arizona	10	Oregon	27	Netherlands	46
Arkansas	3	Pennsylvania	13	Norway	5
California	357	Rhode Island	3	Portugal	4
Colorado	16	South Carolina	16	Russian Federation	6
Connecticut	17	South Dakota	2	Spain	43
Delaware	7	Tennessee	27	Sweden	41
District of Columbia	2	Texas	31	Switzerland	11
Florida	29	Utah	13	Turkey	2
Georgia	40	Vermont	4	Ukraine	1
Hawaii	12	Virginia	5	United Kingdom	68
Idaho	8	Washington	38		
Illinois	34	West Virginia	1	Egypt	1
Indiana	19	Wisconsin	35	Israel	6
Iowa	11	Wyoming	1		
Kansas	4			Algeria	1
Kentucky	3	Puerto Rico	1	Nigeria	1
Louisiana	6	Canada	59	South Africa	3
Maine	8	Argentina	5		
Maryland	35	Brazil	10	China	8
Massachusetts	42	Chile	3	Hong Kong	5
Michigan	20	Colombia	2	India	4
Minnesota	15	Ecuador	2	Japan	17
Mississippi	14	Mexico	6	Malaysia	2
Missouri	19			South Korea	19
Montana	8	Austria	14	Taiwan	1
Nebraska	13	Belgium	8	Thailand	1
Nevada	6	Czech Republic	2		
New Hampshire	6	Denmark	13	Australia	38
New Jersey	13	Finland	14	New Zealand	14
New Mexico	25	France	52		
New York	35	Germany	72		
North Carolina	37	Greece	9		
North Dakota	5	Hungary	3		

Bringing the community together

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

Somerville's talk focused on the development of cellulosic biofuels and the challenges involved in harnessing nature, particularly the issue of carbon emissions from indirect land use, a persistent concern of bioenergy researchers. He also 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. Somerville reminded his listeners to think about the energy problem on a global scale.

Church's keynote 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 also modifying it to suit their needs.

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

Video recordings of these keynote talks as well as most other presentations from the User Meeting are available at: **http://www.scivee.tv/node/10579/video**

A list of the 2009 DOE JGI Annual Genomics of Energy and Environment User Meeting speakers can be found in Appendix E.

Keynote speakers, from left, Chris Somerville of the Energy Bioscience Institute, George Church of Harvard, and J. Craig Venter of the J. Craig Venter Institute address the 2009 DOE JGI Annual Genomics of Energy and Environment User Meeting. Photo by Roy Kaltschmidt, LBNL

2009 Published Highlights

An array of sorghum strains. Photo by Roy Kaltschmidt, LBNL

Sorghum's sequence boosts staple crop and feedstock production

For half a billion people around the world, sorghum is a dietary staple. Some 60 million tons of sorghum, which is more drought-tolerant that cereal crops such as wheat and oats, is produced annually.

As the technology to develop cellulosic biofuels matures, so does interest in using sorghum as a bioenergy feedstock because of its rapid growth. Sorghum was selected to be part of the DOE JGI Community Sequencing Program, in part because the crop is second only to corn as a biofuel feedstock in the United States, yet needs one-third less water to produce the same amount of ethanol.

Scientists at the DOE JGI and several partner institutions published the sequence and analysis of the complete genome of sorghum in a January 2009 issue of *Nature*. This is only the second grass genome to be completely sequenced; the first genome to be completed was that of rice. DOE JGI researcher Jeremy Schmutz from the HudsonAlpha Institute for Biotechnology said that sorghum's genome was 75 percent larger than that of rice, in part because plant genomes tend to have large sections of repetitive regions.

"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."

According to Andrew Paterson, the publication's first author and Director of the Plant Genome Mapping Laboratory at the University of Georgia, the information will be used not just to develop strains with increased grain yield, resistance to plant stressors, and biofuel-specific strains: "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."

Biomass breakdown with brown-rot genome

One of the challenges in commercially producing biofuels from cellulosic biomass is the cost involved in breaking down the fibrous plant material. In a February 2009 issue of the *Proceedings of the National Academy of Sciences (PNAS)*, researchers from the DOE JGI and several partner institutions around the world reported the genome sequence of *Postia placenta*, a brown-rot fungus with biomass-degrading enzymes that can easily break down wood. Approximately 10 percent of the decay noted annually in American timber is attributed to brown-rot fungi. Researchers hope to identify the enzymes involved and apply them toward other plant material.

Dan Cullen, a Forest Products Laboratory scientist at the U.S. Department of Agriculture's Forest Service, and one of the senior authors on the *PNAS* paper, suggested that *Postia* uses oxidizing agents to depolymerize the cellulose,

Photo courtesy of Jessie Micales (Northern Research Station)

"Nature offers some guidance here."

—Dan Cullen, USDA-FS, FPL

deconstructing the lignocellulose in a less energy-intensive manner than had been previously expected.

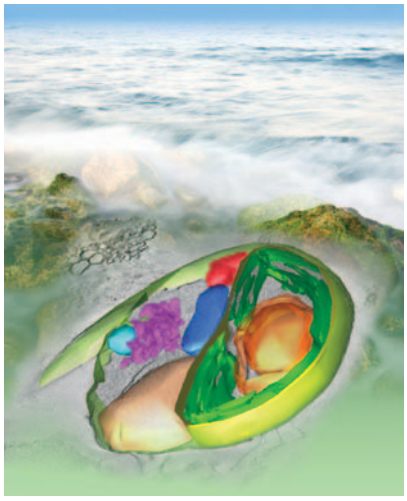
The fungal genome also allows researchers to compare various types of wood-decaying fungi. "For the first time we have been able to compare the genetic blueprints of brown-rot, white-rot, and soft-rot fungi, which play a major role in the carbon cycle of our planet," said Randy Berka, one of the study's senior authors and Director of Integrative Biology at Novozymes Inc. of Davis, California. "Such comparisons will increase our understanding of the diverse mechanisms and chemistries involved in lignocellulose degradation. 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."

Tiny *Micromonas* influence global carbon cycle

Trees and plants are closely related to unicellular algae, including a family known as picophytoplankton. In the green Tree of Life, land plants make up one main branch of photosynthetic eukaryotes while algae are part of the other branch.

Measuring 1/50th the width of a human hair, *Micromonas*, one of the smallest known algae, thrive in waters from the equator to the poles and are key players in several biogeochemical cycles, including the carbon cycle.

In April, the sequenced genomes of two *Micromonas* algae collected from the South Pacific and the English Channel



A 3-D reconstruction of *Micromonas*, one of the smallest known algae. A.Z. Worden, T. Deerinck, M. Terada, J. Obiyashi and M. Ellisman (MBARI and NCMIR); Background: Flavio Robles (LBNL).

were published in *Science*, joining a very short list of sequenced single-cell marine microorganisms. With 20 million base pairs and 10,000 genes in each genome, the information is expected to help researchers understand how photosynthesis contributes to the Earth's biogeochemical cycles and algae's particular role in the carbon cycle. The sequence is also going to be of use for researchers developing algal biofuels.

The two strains of *Micromonas* were found to share 90 percent of their genes, and the differences likely represent adaptations to their respective environments. These sequence changes allow researchers to suggest that *Micromonas* might thrive better in a changing climate than other algal species. "This also means that as the environment changes, these different populations will be subject to different effects, and we don't know whether they will respond in a similar fashion," said the study's first author, Alexandra Worden from the Monterey Bay Aquarium Research Institute.

Genome sequencing with a single cell sample

Researchers at the DOE JGI demonstrated that sequencing a high-quality draft genome is possible even if there is only a single cell from the organism. In April, DOE JGI researcher Tanja Woyke and her team picked out individual bacterial cells from coastal water samples provided by scientists at the Bigelow Laboratory and used a process known as multiple displacement amplification to make millions of copies of the genomes for sequencing. The resulting genomes of two uncultured flavobacteria published in *PLoS ONE* are 80 to 90 percent complete.

"As long as you can isolate a single cell, pick it from the environment, lyse it, you can generate millions of copies of that genome and gain access to the information inside that organism."

–Tanja Woyke

"We estimate that roughly 99.9 percent of the microbes that exist on this plant currently elude standard culturing methods, denying us access to their genetic material," said DOE JGI Director Eddy Rubin. "So we have to explore other methods to characterize them."

DOE JGI researcher Tanja Woyke. Photo by Massie Ballon, DOE JGI

Woyke said these results are already of higher quality than previous single-cell sequencing attempts, but there is room for improvement of the method. DOE JGI is using the technique to help other collaborators who study organisms from a variety of environments but cannot obtain enough genetic material to make use of more conventional sequencing methods.

"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 Bigelow researcher and study co-author Ramunas Stepanauskas.

Call for improved genome sequencing standards

As technological improvements speed up the process of sequencing DNA while simultaneously reducing the costs involved, the bottleneck shifts from production to data analysis. DOE JGI researchers are at the forefront of a movement to update sequencing standards as more centers around the world sequence organisms.

In July, Genome Biology Program head Nikos Kyrpides suggested that though 1,000 microbial genomes have been sequenced in the past 15 years, lack of procedural standards has compromised the collected data. In *Nature Biotechnology*, he called both for shared standards in genomic data collection and analysis, and urged researchers to sample a larger representative sample of the microbial diversity rather than focusing on the small fraction related to human health and activities. This can be done with the help of techniques such as single-cell sequencing and pan-genomic comparisons.

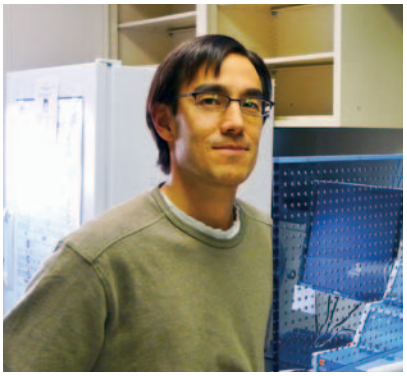
"The remarkable number of microbes—already estimated to be several orders of magnitude greater than the number of stars in the universe—urgently calls for a transition from random, anecdotal, and small-scale surveys towards a systematic and comprehensive exploration of our planet."

–Nikos Kyrpides

"The remarkable number of microbes—already estimated to be several orders of magnitude greater than the number of stars in the universe—urgently calls for a transition from random, anecdotal, and small-scale surveys towards a systematic and comprehensive exploration of our planet," Kyrpides said.

In October, an international consortium of researchers led by DOE JGI metagenomics scientists Patrick Chain and Chris Detter proposed modifying the

DOE JGI Genome Biology head Nikos Kyrpides. Photo by David Gilbert, DOE JGI



Patrick Chain, a DOE JGI metagenomics scientist at Los Alamos National Laboratory. Photo by Chris Detter, LANL

current sequencing standards from the existing terms "draft" or "finished" by adding several stages in between. With a conservative estimate of 12,000 draft genomes available to researchers, Chain said, further descriptions are necessary to describe the data's usefulness.

"In the past, we've been limited to two options, requiring us and the other centers to come up with internal definitions," said Chain, the *Science* paper's first author who does research at Los Alamos National Laboratory (LANL). "But these are not clear and they're not propagated to the databases to which we submit sequences. So when users try to download genomes, they get data of unknown quality with no information, or a complete genome that they assume has been checked for missing-data errors."

Chain's colleague Detter, head of LANL's Genome Sciences program, noted that the proposed six categories for sequencing standards are still a work in progress, and that the ensuing discussions will lead to the definition and implementation of sequencing standards that can be applied to as many projects as possible around the world.

Improving fungal genomes for industrial enzyme production

Shortly after the Second World War, mutations in strains of the fungus *Trichoderma reesei* changed its status from bane of the American Army quartermasters to key producer of industrial enzymes known as cellulases, which can break down biomass for biofuel production.

DOE JGI scientist Wendy Schackwitz pointed out that the mutagenic treatments applied to the fungi were random, and in August an international team led by Schackwitz and her fellow co-first author, Stéphane Le Crom, from the French institute École Normale Supérieure, developed the first mutation map of *T. reesei* to identify which genetic mutations are responsible for boosting cellulase production. The information could be used to develop a fungal strain that can lead to the commercial-scale production of cellulosic ethanol.

DOE JGI scientist Wendy Schackwitz. Photo by David Gilbert, DOE JGI

Scott Baker, a DOE JGI scientist at Pacific Northwest National Laboratory and a senior author of the study that appeared in *Proceedings of the National Academy of Sciences (PNAS)* added, "We now have a blueprint on which we can do future studies to see which genes are related to the enzymes. If you can produce more enzyme more efficiently, that makes your process — in this case the production of biofuel — more economical."

A way to predict microbial lifestyles

More than two-thirds of the Earth's surface is covered by water, and these oceans are home to two kinds of microorganisms: copiotrophs that thrive in nutrient-rich waters often associated with warmer regions; and oligotrophs that prefer nutrient-poor waters. In September, an international team of scientists led by the University of New South Wales in Australia and the DOE JGI described a method to study the oceans' microbial diversity without the need to cultivate samples in the laboratory.

To prove the technique's efficacy, the researchers compared the genome of a copiotroph sampled off the Australian coast against the genome of the oligotrophic bacterium *Sphingopyxis alaskensis*, which was collected from waters off the Alaskan coast and sequenced by the DOE JGI. The goal was to study the biochemical pathways in microbial genomes in order to use that information to help reduce greenhouse gas emissions.

The work reported in the *Proceedings of the National Academy of Sciences* also lent credence to a long-held theory that though more microbial genome projects involve copiotrophic bacteria, in part because they are easier to cultivate, oligotrophic bacteria are more representative in number of the microorganisms in the ecosystem.

"Despite the number of microbial genome projects being done, these organisms represent just a fraction of the microbial diversity on the planet," noted DOE JGI Genome Biology Program head and study co-author Nikos Kyrpides. "To sequence microbial genomes that are representative of the environments in which they were collected and for a more systematic and comprehensive sampling of the Tree of Life, researchers need to increasingly develop and rely on other techniques such as single-cell sequencing to isolate DNA samples from harder-to-cultivate microbes residing in environments where nutrients are scarce."

An oligotrophic bacterium collected off the Alaskan coast was sequenced by DOE JGI. USGS/Image by Bruce F. Molnia

"Despite the number of microbial genome projects being done, these organisms represent just a fraction of the microbial diversity on the planet."

—Nikos Kyrpides

Tracking global cycles in a Canadian "dead zone"

Known colloquially as "dead zones," oxygen minimum zones (OMZs) are spreading in oceans all over the world, affecting the delicate balance of life in these ecosystems.

One such OMZ is Saanich Inlet, a fjord off the coast of British Columbia, Canada. In October, a team of researchers from the University of British Columbia (UBC) and the DOE JGI described in the journal *Science* a composite genome for the most abundant organism, known as SUP05, collected in these waters over several seasons. "By studying the genomes of the uncultivated microbes found in OMZs, we can better understand how they participate in global geochemical cycles such as the carbon and nitrogen cycles," said DOE JGI scientist Susannah Tringe.

Study first author and UBC researcher David Walsh collecting ocean samples. Photo courtesy of the Hallam Lab

Study senior author and UBC professor Steven Hallam described SUP05 as a paradoxical organism, one that fixes carbon dioxide and removes toxic sulfides, but which might also be producing nitrous oxide, a more potent greenhouse gas than either carbon dioxide or methane. "Just as cyanobacteria play an essential role in producing atmospheric oxygen, in future oceans this could be one of those organisms that play similarly integral roles, albeit with different ecological outcomes," Hallam said. "Global warming is changing the chemistry of the oceans and one of the byproducts of change is that the ocean pH is becoming acidic. Blooming SUP05 populations have the potential to help offset rising carbon dioxide levels that ultimately lead to ocean acidification."

Releasing the first volume of a genomic encyclopedia

In December, a team of researchers led by DOE JGI Phylogenomics Program Head and University of California, Davis, professor Jonathan Eisen published the first "volume" of a genomic encyclopedia that represent little-studied branches of the Tree of Life.

"We've done a very poor job of sampling across the tree in microbial studies," said Eisen of the study, which appeared in *Nature*. "If you look at phylogenetic diversity in the bacterial kingdom, most of the available genomes come from just three of the 40 major phyla. The same trend holds for archaea, eukaryotes, and viruses.

Image by Roy Kaltschmidt, LBNL

"The solution is to use the tree to guide us, going through phylogenetic diversity to explicitly fill in missing branches of the tree with missing data. Building a more balanced catalog of the diversity of genomes present on the planet should facilitate searches for novel functions and our understanding of the complex processes of the biosphere."

For the pilot project, Eisen and his colleagues collaborated with researchers at the German Collection of Microorganisms and Cell Cultures. The findings indicate that though nearly 2,000 microbial genomes have been sequenced, the number is a tiny fraction of the estimated nonillion (10^{30}) individual microbes in, on, and around the planet. To even sample half of the known phylogenetic diversity, Eisen said, researchers would need to sequence another 10,000 genomes of what are still mostly uncultured organisms, and then understand the biology of the organisms.

"You might not care about the genomes sequenced in this study, but they provide the ability to study other genomes you might care about more."

—Jonathan Eisen

Soybean drafted, largest DOE JGI plant project to date

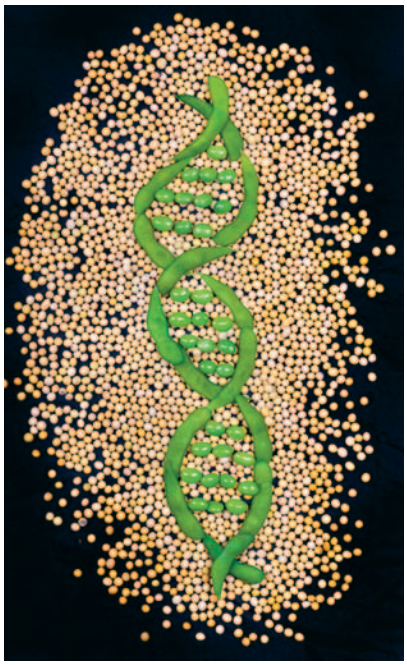
In January 2010, the soybean became the first legume to have a published complete draft genome sequence. "The soybean sequence project is the largest plant project done to date at the DOE Joint Genome Institute," said DOE JGI first author Jeremy Schmutz of the Alabama-based HudsonAlpha Institute for Biotechnology. "It also happens to be the largest whole-genome shotgun plant that's ever been sequenced. We took the approximately 1.2 gigabase genome, broke it apart, and reassembled it like a puzzle."

Detailed in *Nature*, the crop's sequence is expected to be used for both agricultural improvement and biodiesel production. "In recent years we've kind of tapped out on traditional soybean breeding and can't seem to increase the yield anymore. Using genomics allows us to breed specific genes; identify specific traits such as drought tolerance, pathogen resistance and more seed production; and breed them back into soybean lines," Schmutz said.

Another major application is expected to be biodiesel production. Schmutz noted that while the plant doesn't currently produce enough oil to compete with petroleum products, the soybean is the major source of biodiesel worldwide.

"The soybean genome's billion-plus nucleotides afford us a better understanding of the plant's capacity to turn sunlight, carbon dioxide, nitrogen, and water into concentrated energy, protein, and nutrients for human and animal use," said Anna Palmisano, Associate

Photo by Roy Kaltschmidt, LBNL



Director of the DOE Office of Science's Office of Biological and Environmental Research. "This opens the door to crop improvements that are sorely needed for energy production, sustainable human and animal food production, and a healthy environmental balance in agriculture worldwide."

Researchers from the DOE JGI, the U.S. Department of Agriculture-Agricultural Research Service (USDA-ARS), the National Science Foundation, the University of Missouri, Purdue University, and a dozen other institutions contributed to the work.

Brachypodium genome completes trio of representative grasses

Grasses are being considered for use as feedstocks in the commercial production of cellulosic biofuels but the long life cycles and complex genomes of crops such as wheat and barley have made it difficult for bioenergy researchers to study them. The complete sequence of the wild grass *Brachypodium distachyon*, also known as purple false brome, is useful in part because of its smaller genome compared with other reference grass genomes and because of its shorter life cycle. In February 2010, an international consortium of researchers known collectively as the International Brachypodium Initiative presented the complete sequence of the first representative grass from the third subfamily in *Nature*.

"*Brachypodium* has the traits required to serve as a functional model system to more rapidly gain the knowledge about basic grass biology necessary to develop superior grass crops," said the study's first author and DOE JGI collaborator John Vogel (USDA-ARS). The grains that make up food staples around the world belong to three grass subfamilies. Rice, sorghum, and maize have already been sequenced, providing reference genomes for their respective subfamilies.

Image by Roy Kaltschmidt, LBNL

B. distachyon's genome also allows researchers to compare grass genomes across all three subfamilies for the first time. The information is useful to researchers who want to track how wild grasses were domesticated to become staple crops, and to help them develop varieties better suited to human needs.

DOE JGI in the News

NewScientist

Gold-mine bug DNA may be key to alien life

Published in *New Scientist,* October 9, 2008

A bug discovered deep in a gold mine and nicknamed "the bold traveller" has astrobiologists buzzing with excitement. Its unique ability to live in complete isolation from any other living species suggests it could be the key to life on other planets.

A community of the bacteria *Candidatus Desulforudis audaxviator* has been discovered 2.8 kilometers beneath the surface of the Earth in fluid-filled cracks of the Mponeng gold mine in South Africa. Its 60° C home is completely isolated from the rest of the world, and devoid of light and oxygen.

http://www.newscientist.com/article/dn14906-goldmine-bug-dna-may-be-key-to-alien-life.html

Ethanol PRODUCER MAGAZINE

Unraveling the mysteries in sorghum's 'simple' genome

Published in *Ethanol Producer* magazine, May 2009

Mapping out a genetic blueprint involves determining several short sections that would make up a gene, and sequencing those in bulk. Scientists then take all of those pieces and put them together like a puzzle—a difficult puzzle, according to Dan Rokhsar, the computational biology leader at the DOE JGI.

"There are parts of the genome that are very complicated to assemble," Rokhsar says. "Think of them as large stretches of blue sky in a puzzle, with wispy clouds throughout. So, you have to use that cloud information to put together those regions of the genome." As the puzzle starts to take shape, he says, the reconstruction of entire genes and eventually whole chromosomes takes place.

http://www.ethanolproducer.com/article.jsp?article_id=5585

DiscoveryNews

Community genome could produce biofuels

Published in *Discovery News* and MSNBC.com, July 6, 2009

The genomes of 17 different ants, fungi, and bacteria that eat through hundreds of pounds of leaf matter a year could ultimately lead to new techniques for making biofuels.

Scientists from the University of Wisconsin, the DOE JGI, and Emory University are sequencing the first-ever community genome, searching for clues to how what's essentially a 50 million-year-old bioreactor operates.

http://news.discovery.com/tech/biofuel-leaf-cutter-ants.html?FORM=ZZNR

http://www.msnbc.msn.com/id/31767229/ns/technology_and_science-science/

Forbes

Travels with a geek: Unusual spots for the scientific at heart

Published in *Forbes,* July, 2, 2009

At laboratories like the DOE JGI, the human genome was deciphered. The techniques used for the Human Genome Project are now applied to fish, animals, viruses, and plants, and by organizing a tour, it's possible to see the machinery used for DNA sequencing and get a detailed explanation of how it all works from someone who works there.

http://www.forbes.com/2009/07/02/geek-atlas-travel-technology-breakthroughs-oreilly.html

Wine Spectator

Taking the barnyard out of your wine

Published in *Wine Spectator,* August 7, 2009

Wineries have tried a number of chemical mixtures to ward off infection, but none has proved fully effective. Trevor Phister believes the genome will provide answers on how *Brettanomyces* survives the initial battle with *Saccharomyces*, how it spreads so fast and, ultimately, on how to stop it.

To decode the *Brettanomyces* genome, Phister will work with the DOE JGI. The seemingly incongruous pairing of the Energy Department and winemaking stems from problems with Brett contamination in biofuel production. Biofuel producers use *Saccharomyces* to convert organic waste into ethanol fuel. In 2008, *Brettanomyces* eradicated and replaced over 2,000 pounds of *Saccharomyces* in an ethanol production plant in just a few weeks.

http://www.winespectator.com/webfeature/show/id/40447

The New York Times

Scientists start a genomic catalog of earth's abundant microbes

Published in *The New York Times,* December 28, 2009

"Microbes represent the vast majority of organisms on earth," said Hans-Peter Klenk, a microbiologist for the German Collection of Micro-organisms and Cell Cultures, a government microbiology research center.

Yet scientists still know very little about our microbial planet. The genomes of only about 1,000 species of microbes have been sequenced. That leaves 99.99999 percent to go. Making matters worse, the genomes scientists have sequenced so far are clustered together in groups of closely related species, leaving vast stretches of the microbial Tree of Life virtually unexplored. It would be as if all we knew about the animal kingdom were based entirely on a stuffed ferret and a pickled tarantula.

To shed light on this enormous stretch of biological darkness, the DOE JGI has started what it calls a "genomic encyclopedia." It is filling the encyclopedia with the genomes of microbes from remote reaches of the Tree of Life.

http://www.nytimes.com/2009/12/29/science/29microbes.html?_r=2

The Scientist

A mossy renaissance

Published in *The Scientist,* February 2010

Physcomitrella holds a unique phylogenetic position; it evolved when aquatic plants transitioned to terrestrial living approximately 400–500 million years ago. Sequencing the genome could help scientists better understand plant evolution, specifically how aquatic plants adapted to deal with the stressful heat and dehydration that come with living on land.

http://www.the-scientist.com/article/display/57103/

CSP 2010 Portfolio

Since 2005, the DOE JGI has primarily focused on sequencing organisms selected for their relevance to the DOE science mission areas of bioenergy, carbon cycling, and biogeochemistry, which is the study of the global processes involved in making and keeping the Earth habitable. These organisms are selected on the basis of scientific merit — judged through independent peer review — through the Community Sequencing Program (CSP), which provides the scientific community at large with free access to high-throughput sequencing at DOE JGI.

For the CSP 2010 portfolio, a total of 81 organisms from regions as far north as the Arctic and south to New Zealand were selected. The DOE JGI's recent transition to new sequencing technologies has increased its sequencing throughput almost five times, from over 60 billion nucleotides allocated for CSP projects last year, to about a third of a trillion nucleotides for the CSP 2010.

Many of the projects are bioenergy-related. A key component to achieving cost-effective cellulosic biofuels is identifying more effective enzymes to break down plant fibers into sugars that can then be converted to fuels. An equally important step involves developing more effective ways to ferment plant-derived sugars into liquid fuels.

DOE JGI will employ a variety of sequencing methods, ranging from whole-genome shotgun sequencing that produces high-quality draft sequences, to next-generation technologies that can generate millions of sequence reads per run, to single-cell sequencing techniques that allow access to genomes when only minute quantities of DNA are available. Beyond sequencing, DOE JGI offers assembly, annotation, and genome analysis services for all approved CSP projects.

A *Posidonia oceanica* meadow off Formentera Island, Spain. The marine plant plays host to *Marinomonas mediterranea*, one of the CSP 2010 projects related to carbon cycling. Photo by Manu Sanfelix

Bioenergy

Two-thirds of the energy consumed in the United States is used by transportation and industry. As a result, one of the DOE's main goals is to contribute to the development of clean and sustainable alternative energy sources that can compete with petroleum-based fuels. A significant portion of the CSP projects are related to bioenergy and tend to focus on one of three categories: developing plant feedstocks; using microbes to break down cellulose in plant cell walls; and fermenting sugars into biofuels.

1. Desert locust

Locust swarms can spread over as much as 20 percent of the world's land mass and consume the equivalent of their body mass daily, impacting the livelihoods of up to 10 percent of the world's population.

The locust's destructive properties could be harnessed to break down the tough cellulose in plant feedstocks for biofuel production.

For this CSP 2010 project, JGI will sequence both the locust's gut wall and the microbial community or metagenome inside the desert locust's gut to determine whether or not the ability to digest lignocellulose is conferred by microbes inside the desert locust or the insect itself. By using locusts from commercial breeding facilities and captured in the wild for the study, the sequencing information is expected to identify the contributions of both the microbes and the locust to the degradation process. Additionally, the sequence information might contain novel enzymes that could be useful for the commercial biofuel production. In addition, we plan to perform feeding experiments with different feedstocks, possibly *Miscanthus* and switchgrass.

The information could shed light on a pheromone known as guaiacol, which has been linked to the swarming behavior of the locust. A key component of the pheromone is produced by the microbes in the locust's gut, and the information could result in more effective biocontrol measures to reduce the incidence of crop destruction and famines caused by locust swarms.

The principal investigator is Falk Warnecke (University of Jena).

2. Microbial symbionts of New Zealand's endemic wood-degrading insects

When the Gondwana supercontinent broke up some 8 million years ago, New Zealand was separated from the other land masses. At the time, the only mammals there were bats and marine mammals. When humans arrived approximately 1,000 years ago, they brought more mammals, changing the island's ecosystems.

An adult *Prionoplus reticularis* beetle. Photo courtesy of Birgit Rhode, Landcare Research

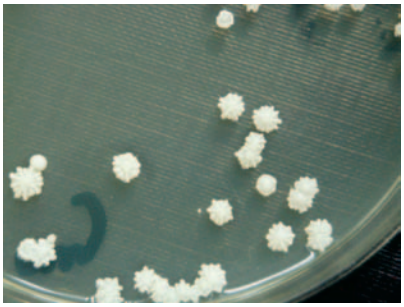
Millennia of separation led to the evolution of native species such as the huhu grub, a traditional food source for the Maori tribe, and the cricket-like weta that may not be found anywhere else in the world. Today New Zealand is considered one of the world's 25 biodiversity hotspots, and some researchers have said that as much as a quarter of the bird species, roughly 80 percent of the vascular plants, and more than 90 percent of the insects and shellfish in New Zealand cannot be found anywhere else in the world.

As a CSP 2010 project, JGI will perform 16S rRNA gene tag pyrosequencing on bacterial and archaeal samples from 11 endemic New Zealand insect species (in six orders) and three Australian termite species. The main focus of attention will be on microbial enzymes from the termite gut samples, as these may have novel cellulases useful for biofuel production.

The principal investigator is Mike Taylor (University of Auckland).

3. *Dekkera bruxellensis*

To wine and beer makers, *Brettanomyces custersii* is the invasive yeast that contaminates the fermentation process, ruining entire batches of their brews and leaving a distinctive odor of contamination.



Dekkera bruxellensis in culture. Photo courtesy of Chad Yakobson

To scientists, the species is known as *Dekkera bruxellensis*, and is the most common wild yeast found in fuel ethanol fermentations and blamed for the loss in ethanol yields. The yeast has a high tolerance for alcohol and simply waits for brewer's yeast to eliminate other microbes in the process before bumping the latter yeast off, spoiling the brew and reducing the production yield.

Sequencing *Dekkera's* genome is useful in bioenergy research and development because the biofuel production process, like making wine or beer, involves fermenting sugars from organic matter. The genomic information would give researchers insight into controlling the yeast to reduce production yield losses due to contamination. Additionally, *Dekkera* can work with both starch-based feedstocks and woody plant or lignocellulosic feedstocks.

There are also plans to compare the genomes of *Dekkera* with brewer's yeast to better understand their metabolic profiles. The sequence information could give biofuel producers insight into how to develop yeasts with enhanced abilities to handle lignocellulose fermentations, which could in turn lower the costs associated with the biofuel production process.

The principal investigator is Trevor Phister (North Carolina State University).

Carbon cycling

The global carbon cycle is heavily dependent on microorganisms fixing atmospheric carbon, promoting plant growth and degrading organic material. By understanding how microbes metabolize carbon, researchers can develop better predictive models that could lead to more effective methods of reducing the effect of increasing carbon dioxide emissions on the global climate. The DOE JGI is sequencing several microbes and microbial communities that influence carbon cycling.

Evolution Canyon. Photo courtesy of Michael Margulis

1. *Nostoc linckia*

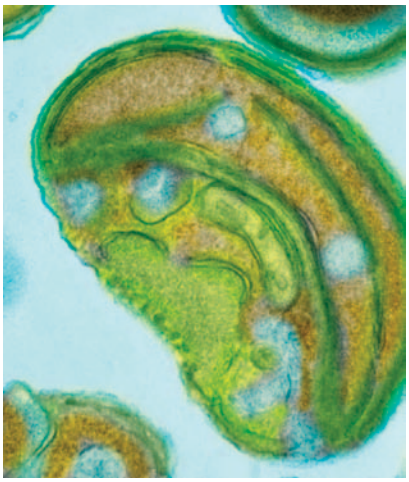
In the summer, the slopes of Lower Nahal Oren, in Mount Carmel, Israel, are dry. The warmer south side faces the African continent while the cooler north side faces Europe and has a more temperate climate. The slopes of what's been nicknamed "Evolution Canyon" receive significantly different amounts of sunlight, and the south side of Evolution Canyon receives 800 percent more solar radiation than the north side.

Nostoc linckia is a nitrogen-fixing cyanobacterium that can form unbranched filaments. The *Nostoc* species are key contributors to carbon sequestration in agricultural environments and can produce hydrogen. As a CSP 2010 project, draft genomes of *Nostoc* bacteria from both the southern and northern slopes will be sequenced for comparison, providing researchers with an opportunity to study the impact of long-term stress on the genome's evolution.

Due to the climate differences, variations in genome sizes and information are expected. For example, the *Nostoc* genome from the southern slope is expected to be slightly larger due to the additional solar radiation.

The principal investigator is Volodymyr Dvornyk (University of Hong Kong).

Ostreococcus micrograph, courtesy Wenche Eikrem and Jahn Throndsen, University of Oslo.



2. *Ostreococcus tauri*

Phytoplankton account for roughly half of the photosynthetic activity on Earth; land plants constitute the other half. In the oceans, picoeukaryotes make up a fraction of the marine biomass but are significant contributors to primary production in oceans worldwide. One of these microorganisms is *Ostreococcus*, a class of unicellular green algae with the smallest known nuclear genome for a photosynthetic free-living eukaryote. So far four groups of *Ostreococcus*, each with a single ancestor and adapted to ecologically different niches, have been identified.

Researchers want to learn more about the genetic diversity within each of these groups by studying several strains from a particular clade in the western Mediterranean. This CSP 2010 project focuses on sequencing several strains of a single species, *O. tauri*, collected from a variety of environments, to learn about the genetic diversity within a single group.

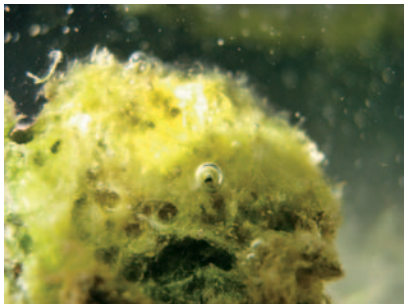
The genome sequence data provided by the DOE JGI allows researchers to take a population genomics approach with the data, analyzing the genetic variation within individuals from the same species at the genome level to determine which events impact the whole genome and which might target a particular base or loci.

The principal investigator is Hervé Moreau (CNRS, France).

3. Modern freshwater microbialites

Microbialites resemble coral reefs but can also be found in freshwater systems. Composed of sediments built up over many hundreds of years and by interactions between multiple micoorganisms such as diatoms, cyanobacteria, and minerals, they can sequester carbon through a process called biologically mediated carbonate precipitation, though just how this is done is still poorly understood. Some researchers think microbialites resemble structures that may have existed on Earth 450 million years ago, which could help them recognize signs of life on other planets.

By sequencing a handful of samples taken from several surface layers of an active, river-based microbialite in the Cuatro Cienegas Basin of northern Mexico, researchers hope to learn more about the microbial communities that are associated with these structures and find out what roles they play in sequestering carbon. Learning more about the genomes of the microbial communities in the microbialite would not only provide more information about the role microbialites play in the global carbon cycle, which could prove crucial as carbon dioxide levels continue to rise, but could also be useful to researchers who work with stromatolites and other, similar structures which can also sequester carbon.



Oxygen bubbling from a freshwater microbialite at Cuatro Cienegas, Mexico. Photo courtesy of Mya Breitbart, University of South Florida

The principal investigator is Mya Breitbart (University of South Florida).

3. *Curvularia protuberata*

Morning Glory Pool at Yellowstone National Park.

In 2008, drought conditions and high temperatures resulted in agricultural losses valued at $2 billion to states such as California, Texas, the Carolinas, and and Tennessee. Rising global temperatures mean species need to adapt to the changing ecosystem in order to thrive, and finding ways to help plants adapt to these conditions could help reduce food security concerns.

In Yellowstone National Park, where

the temperature of the soil can reach 55 degrees Celsius (131 degrees Fahrenheit), researchers found a fungus called *Curvularia protuberata* that can live in the tropical panic grass *Dichanthelium lanuginosum* without harming it. When the grass and fungus are separated, however, neither the fungus nor the grass can survive at temperatures exceeding 38 degrees Celsius.

Studies have shown the symbiosis is actually a three-way relationship, involving a double-stranded RNA virus in the fungus that is crucial to the heat-tolerance trait. Researchers do not know, however, just how the symbiotic relationship allows the plants to thrive in heated environments.

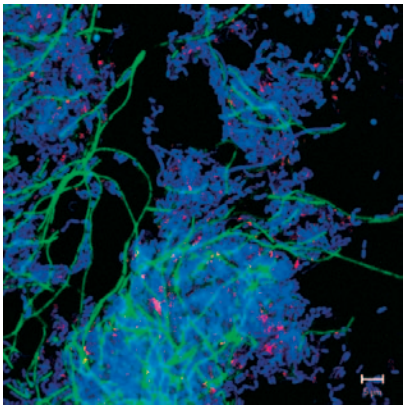
As a CSP 2010 project, sequencing the *C. protuberata* fungus will provide insight into how the genetic information is altered by the presence of the virus. The analysis will also be useful in understanding the molecular pathways involved in conferring heat tolerance, and this information in turn could be applied to other plants still adapting to a changing climate.

The principal investigator is Marilyn Roossinck (Samuel Roberts Noble Foundation).

Biogeochemistry

The field of biogeochemistry covers the biological, physical, geological, and chemical processes that regulate the natural environment. Breaking down or sequestering contaminants often depends on microbial communities or metagenomes working together. Bioreactors, systems that support biologically active environments, are typically studied as metagenomes because the microbes cannot be grown in isolation. The genetic data gleaned from such sequencing projects provide key insights into using these microbial communities to restore and maintain the environment.

1. *Dehalobacter*



Fluorescence in situ hybridization image of *Dehalococcoides* cells (pink). Image courtesy of Natuschka M. Lee (Techn. University of Munich, Germany) and Frank E. Löffler (Georgia Institute of Technology, USA)

One of the most common types of environmental contaminants, especially in groundwater, is chlorinated solvents, and studies have identified several microbial species that can break down these compounds and harness the energy for their own uses. The most studied dechlorinating organisms are from the *Dehalococcoides* genus.

Some chlorinating organisms have been shown to inhibit the growth of a second group of dehalorespiring organisms — these break down chlorinated ethanes and ethenes — known as *Dehalobacter*. One *Dehalobacter* organism has been found to be able to break down common chlorinated groundwater contaminants.

In sequencing *Dehalobacter* organisms as a CSP 2010 project, DOE JGI provides a reference genome from this anaerobic group that lends insight into its physiology. Additionally, the *Dehalobacter* sequence is being provided within the context of a microbial community, which allows researchers to study the group dynamics involved.

The primary investigator is Elizabeth Edwards (University of Toronto).

2. Organism: Halorespiring Firmicutes

Photo courtesy of Linda Schonknecht 2007/ Marine Photobank

Marine sponges are the oldest multicellular animals and are found in many tropical reef ecosystems. They can filter 24,000 liters of seawater per kilogram of sponge daily and as much as 60 percent of their biomass can be composed of

microorganisms, many of which are being studied for a number of medical applications.

Among the microbes found in sponges are halorespiring bacteria that play a key role in breaking down halogenated pollutants in anaerobic ecosystems ranging from subsurface soils to freshwater and marine sediments. Halorespiring bacteria also play a role in refueling the carbon cycle, though this process is not well understood.

Researchers are interested in learning more about the relationships between the microorganisms and their host sponges, and their roles in global cycles such as the carbon cycle and carbon sequestration. The genomic information gained from sequencing several bacterial strains could lead to the identification of bacterial enzymes that could treat sites contaminated with dioxins, pesticides like DDT, and other compounds found in coolants and lubricants. The pesticide DDT is of concern, for example, because it accumulates in animal fat tissue, harming the body, and can damage ecosystems.

The principal investigator is Hauke Smidt (Wageningen University).

3. *Ceratodon purpureus* (fire moss)

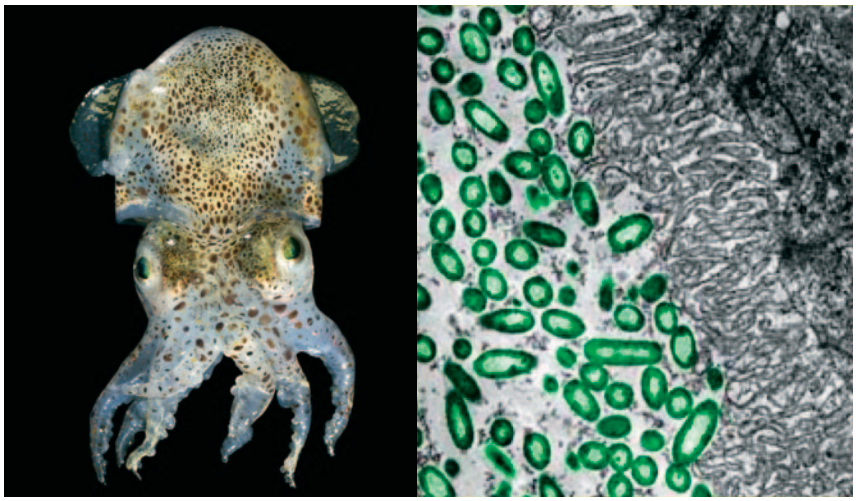
The moss *Ceratodon purpureus* has long been used as a developmental model system to discover novel genes because it can tolerate induced mutations. This trait allows *C. purpureus* to thrive in a wide range of habitats, from urban environments to soils contaminated with heavy metals such as areas near nuclear power or weapons manufacturing sites. Tests have also shown that the moss is resistant to conditions such as UV radiation, cold, salt, and pest infestations.

Fire moss. Image copyright Donegal Wildlife, www.donegal-wildlife.blogspot.com

By sequencing the moss genome, the DOE JGI will provide researchers with a genetic map that provides additional information on this tolerance trait. A second benefit of the genome involves learning more about the evolution of land plants.

In terrestrial biomes, more than 250,000 species of land plants serve as primary producers. Most genomic studies of land plants have focused on a few flowering plants. *C. purpureus* has male and female haploid genomes. In sequencing both genomes, the DOE JGI will have a second moss genome to phylogenomically compare with *Physcomitrella*, which produces both male and female gametes.

The principal investigator is Ralph Quatrano (Washington University in St. Louis).



Vibrio fischeri (right) is a bioluminescent symbiont with the Hawaiian bobtailed squid and is one of the selected CSP 2010 projects related to bioenergy. *V. fischeri* by Eric Stabb, University of Georgia; Hawaiian bobtailed squid by Hans Hillewaert

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

DOE JGI Programs

Photo by Roy Kaltschmidt, LBNL

Plant Genomics Program

The goal of the DOE JGI Plant Genomics Program is to shed light on the fundamental biology of photosynthesis and the transduction of solar to chemical energy. Other areas of interest include characterizing:

- Ecosystems and the role of terrestrial plants and oceanic phytoplankton in carbon sequestration
- The role of plants in coping with toxic pollutants in soils by hyperaccumulation and detoxification
- Feedstocks for biofuels, e.g., biodiesel from soybean; cellulosic ethanol from perennial grasses
- The ability to respond to environmental change (e.g., loss of diversity from monoculture produces vulnerabilities; nitrogen fixing nodules in legumes reduce fertilizer need)
- The generation of useful secondary metabolites (produced largely for disease resistance) for positive/negative control in agriculture, with attendant influence on global carbon cycle

The Plant Genomics Program accomplishes the above through the following activities:

1. DNA Sequence Generation. Produce genome sequences of key plant (and algal) species to accelerate biofuel development and understand their responses to climate change.
2. Function. Develop datasets (and synthetic biology tools) to elucidate functional elements in plant

genomes, with special focus on handful of "flagship" genomes.
3. Variation. Characterize natural genomic variation in plants (and their associated microbiomes), and relate to biofuel sustainability and adaptation to climate change.
4. Integration. Provide a centralized hub for the retrieval and deep integrated analysis of plant genome datasets.

Image courtesy of Carla Wick

Schizophyllum commune (split-gill fungus)

Fungal Genomics Program

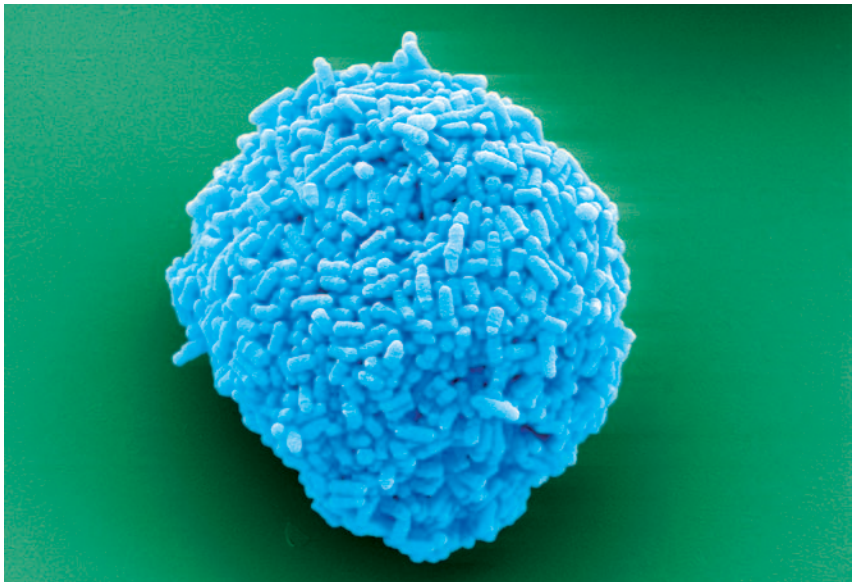
Encoded in the genomes of the organisms of the kingdom Fungi are biological processes with major relevance to the DOE missions in bioenergy, carbon cycling, and biogeochemistry. The economic footprint of the kingdom Fungi is enormous. Control of pathogens and symbionts is critical for sustainable growth of biofuel plant feedstocks. Fungi are amazingly efficient organisms for degrading biopolymers such as lignocellulose, which composes most of the plant cell wall. Future biorefineries will rely on fungi that efficiently produce and secrete cellulolytic enzymes and ferment sugars. Fungal genomics holds the key to our understanding of these important fungi and exploration of their phylogenetic diversity.

Built on established expertise in genomics, promises of new sequencing technologies, and strong connections with user communities, DOE JGI started the Fungal Genomics Program to scale up sequence and analysis of fungal genomes for DOE mission areas and develop the Genomic Encyclopedia of Fungi. The key areas of the program include:

- Plant feedstock health: mycorrhizal symbiosis, plant pathogenicity, biocontrol
- Biorefinery: lignocellulose degradation, sugar fermentation, industrial hosts
- Fungal diversity

Understanding molecular mechanisms of interactions between plants and fungi, symbionts, or pathogens is essential for control of plant health. Reference genomes of mycorrhiza and other soil-inhabiting fungi will also facilitate comprehensive metagenomics studies of rhizosphere currently mostly limited to analysis of bacterial communities. On one hand, the goal is to sample phylogenetically and ecologically diverse fungi to catalog processes and enzymes involved in lignocellulose degradation and sugar fermentation. On the other hand, understanding the biology of industrial strains of fungi is also essential for developing biomass-to-biofuel production on an industrial scale.

Employing new sequencing technologies powered by comparative genomics analysis enables the DOE JGI to approach large and complex sequencing projects — such as sampling broad phylogenetic and ecological diversity of fungi or exploring genomic variation in natural populations and engineered strains — to address biological questions in each of the programmatic areas and to build a foundation for translating the genomic potential of fungi to practical applications.



Tsukamurella paurometabola DSM 20162 by Manfred Rohde, Helmholtz Centre for Infection Research

Microbial Genomics Program

Since the completion of the *Haemophilus influenzae* genome in the summer of 1995, discoveries in the field of microbiology have been thriving, accelerated in recent years by the dramatically increasing amounts of bacterial and archaeal genomes generated.

The Microbial Genomics Program (MGP) is a key provider of high-quality bacterial and archaeal genomes and their analyses, supporting the DOE user communities while aligning with the DOE missions of clean bioenergy, carbon cycling, and biogeochemistry. The MGP performs genome sequencing and analysis on behalf of the DOE Bioenergy Research Centers, as well as through the Community Sequencing Program (CSP). The key elements of the program are genome sequencing, finishing, annotation, data submission, and analysis. Beyond these elements, the program continuously aims to improve various aspects of the current methodologies and adapt to new technologies.

GEBA (Genomic Encyclopedia of Bacteria and Archaea) is an MGP initiative that aspires to sequence thousands of bacterial and archaeal genomes from diverse branches of the Tree of Life. The overall aims of this initiative are to create phylogenetic anchors for metagenomic datasets; to improve annotation; to discover novel genes, protein families, and pathways; to improve our understanding on evolutionary diversification; and to drive the development of novel analysis tools at the DOE JGI. In a pilot GEBA study, the DOE JGI sequenced 53 bacterial and three archaeal novel and highly diverse genomes, representing a first step toward a phylogenetically balanced sequence space in the microbial Tree of Life. Community involvement to aid expanding the phylogenetic Tree of Life is our goal and we thus request that the research communities submit CSP proposals for the sequencing of highly DOE-relevant microbes that also fit into the GEBA mission, including single-cell genomes.

Metagenomics Program

Metagenome sequencing is no newcomer to the DOE JGI science portfolio. This diverse array of projects targets multiple scientific goals, but what the studies share is the sequencing of nucleic acids from a community of organisms rather than from a single isolate. This type of project offers a unique set of opportunities and challenges for the DOE JGI scientific effort. A primary motivation for metagenomics is that most microbes found in nature exist in complex, interdependent communities and cannot readily be grown in isolation in the laboratory. One can, however, isolate DNA or RNA from the community as a whole, and studies of such communities have revealed a diversity of microbes far beyond those found in culture collections. It is suspected that these uncultivated organisms must harbor considerable as-yet undiscovered genomic, functional, and metabolic features and capabilities. Thus to fully explore microbial genomics, it is imperative that we access the genomes of these elusive players.

Several early successes for the DOE

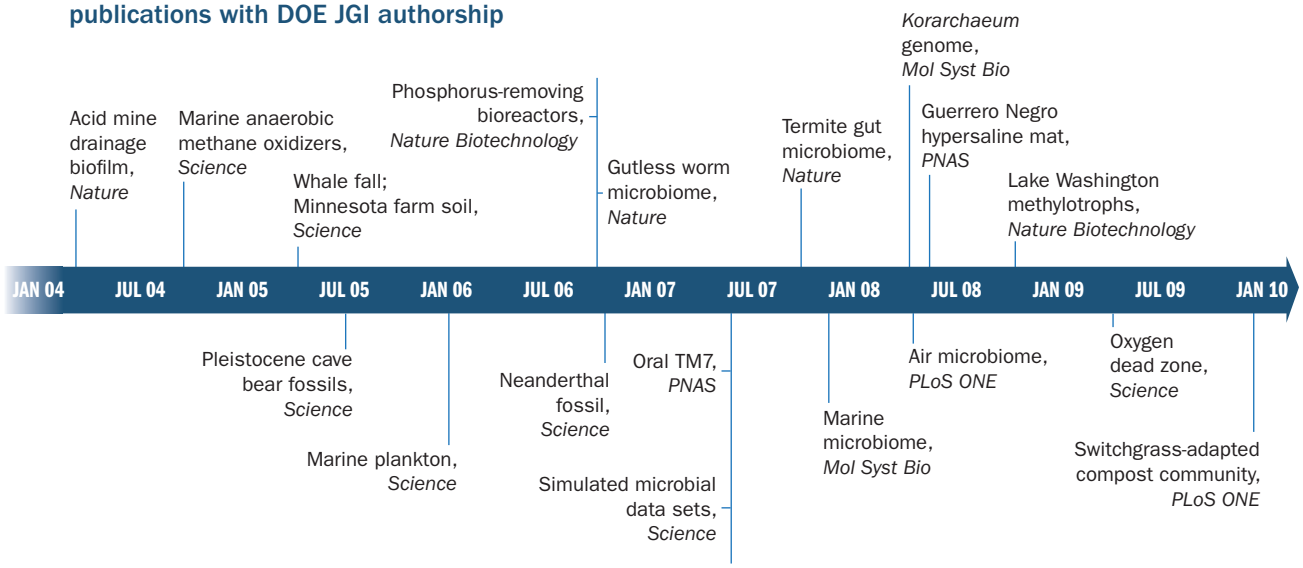
JGI metagenome research program sparked a surge in interest in metagenomics research both at the DOE JGI and in the research community as a whole. In 2007, a National Research Council report titled "The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet" proposed a global metagenomics initiative on the scale of the Human Genome Project. The first metagenomics proposals to the DOE JGI user programs came in 2005, and have continuously increased since then; in the CSP 2008 review, metagenomics projects were formally separated from the microbial genome projects and evaluated independently.

Several of the accepted proposals have culminated in high-profile publications, most notably, "Comparative Metagenomics of Microbial Communities," a gene-centric analysis of highly fragmented metagenomes, published in *Science* in April 2005; "Metagenomic Analysis of Two Enhanced Biological Phosphorus Removal (EBPR) Sludge Communities," published in *Nature Biotechnology* in October 2006; "Metagenomic and Functional Analysis of Hindgut Microbiota of a Wood-Feeding Higher Termite," published in *Nature* in November 2007; "High-resolution Metagenomics Targets Specific Functional Types in Complex Microbial Communities," a Lake Washington methylotroph community, published in *Nature Biotechnology* in August 2008; and a South African deep gold mine metagenomic analysis, "Environmental Genomics Reveals a Single-Species Ecosystem Deep Within Earth," published in *Science* in October 2008. These published projects showcase the application of metagenomics to the DOE mission areas: bioenergy, carbon cycling, and biogeochemistry. With subsequent calls for proposals, we have greatly expanded the metagenome portfolio in these mission areas.

Sequencing "products" also have expanded in the Metagenomics Program, from DNA shotgun sequencing ("standard" metagenomics) to RNA shotgun (metatranscriptomics), pooled fosmid sequencing, community profiling with 16S rRNA pyrotags, and most recently single-cell reference genomes to aid binning metagenomic data.

In common with the other programs at the DOE JGI, the Metagenomics Program is investing considerable time and energy in adapting methods and analytical pipelines to use next-generation sequencing (in particular Illumina sequencing) rather than traditional dye-terminator sequence. An important part of this process is production and analysis of benchmarking datasets for quality assurance.

Time line of DOE JGI metagenomic publications with DOE JGI authorship



Sequence Analysis Tools

IMG ER: Provides expert-driven QC for microbial genome information

After a genome is sequenced and automatically annotated, researchers often manually review the predicted genes and their functions in order to improve accuracy and coverage across the vast genetic code of the particular target organism or community of organisms. These annotations drive the publication of high-profile science relevant to advancing bioenergy research and our understanding of biogeochemistry—the biological, chemical, physical, and geological processes that regulate our environment.

Scientists at the DOE JGI and the Biological Data Management and Technology Center (BDMTC) at Lawrence Berkeley National Laboratory have launched the Expert Review (ER) version of the Integrated Microbial Genomes (IMG) system. IMG ER supports and enhances the review and revision of annotations for both publicly available genome datasets and those newly released from private institutions.

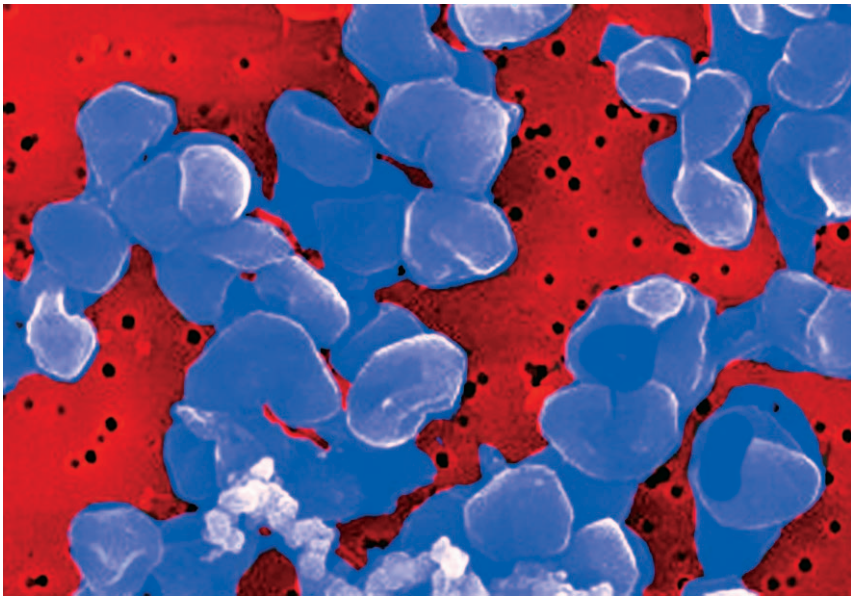
"IMG ER provides scientists with curation tools that improve the annotations of microbial genomes in the context of IMG's comprehensive collection of genomes," said Nikos Kyrpides, head of DOE JGI's Genome Biology Program. "As one of the leading microbial genome sequencing centers in the world, a core mission of the DOE JGI is to ensure the genome sequence data it makes publicly available is high quality."

DOE's National Energy Research Scientific Computing Center at LBNL's Magellan Cloud Computing System. Image by Roy Kaltschmidt, LBNL

The IMG system contains a rich collection of genomes from all three domains of life: as of February 2010, IMG included 1,749 bacterial, 77 archaeal, and 76 eukaryotic genomes, as well as 2,605 viruses and 1,051 plasmids. IMG and its companion metagenome system, IMG/M, have been cited in more than 250 publications and have been used in the analysis of dozens of genomes and metagenomes.

IMG ER curation tools allow detecting and then correcting of annotation problems, such as genes missed by gene prediction pipelines or genes without an associated function. Over the past year, IMG ER was used for reviewing the annotations of more than 150 microbial genomes.

Last year, IMG ER was used in research published online April 30 in the



Isolated from an Antarctic lake and used to study cold adaptation in archaea, *Methanococcoides burtonii* cells are imaged with a scanning electronic microscope. Photo by Dominic Burg, Cavicchioli Lab

International Society for Microbial Ecology (ISME) Journal, by a team led by Rick Cavicchioli at the University of New South Wales, in Sydney, Australia. They announced the completion of a comprehensive manual curation of the *Methanococcoides burtonii* genome. Anomalies found while combing through nearly 3,000 genes were corrected and recorded with IMG ER, leading to an overall greatly improved quality of the functional annotations.

M. burtonii was isolated from Ace Lake in the Vestfold Hills region of Antarctica and serves as a model organism to study the molecular mechanisms of cold adaptation.

Roughly three-quarters of the planet consists of extremely cold environments where temperatures hover around 5 degrees Celsius. Some of the microorganisms that have adapted to these conditions include methane-producing organisms (methanogens) that are capable of significant contributions to global carbon emissions.

"Understanding how methanogens in cold environments respond to changes in temperature is important," said Cavicchioli. "Doing so will not only help to forecast the levels of associated carbon emissions, but provides opportunities for determining effective ways of harnessing methane as an energy source. Understanding the enzymes and processes associated with growth and survival in the cold provides real opportunities for bioenergy and biotechnological innovation."

Polar environments are very sensitive to changes in global temperature and play critical roles in maintaining microbial processes that are essential for the health of the world's ecosystems, yet very little is known about polar microorganisms. Lakes in the Vestfold Hills are unique ecosystems that were once connected to the ocean. Because the lakes were cut off several thousand years ago, they represent a time capsule for studying the evolution of marine microorganisms.

M. burtonii was the first formally characterized organism of its class capable of growth and reproduction in such cold temperatures. As a model, its characterization has broad implications for other cold-adapted organisms that play critical roles in biogeochemical processes such as soil and ocean nitrification.

Victor Markowitz, head of BDMTC, said that Cavicchioli's research team was one of the first groups that started using IMG ER for reviewing genome annotations. "Cavicchioli and his colleagues used our system while it was still in development and their early experience with its tools and valuable feedback helped expand IMG ER's capabilities."

The IMG ER system can be accessed at **http://img.jgi.doe.gov/er**. First-time users need to request an account first at **http://img.jgi.doe.gov/request**. Upon submission of their genomes, users have password-protected access to their genomes and annotations. Subsequently, genomes become publicly available and are submitted to GenBank.

Phytozome "Tree of Life"

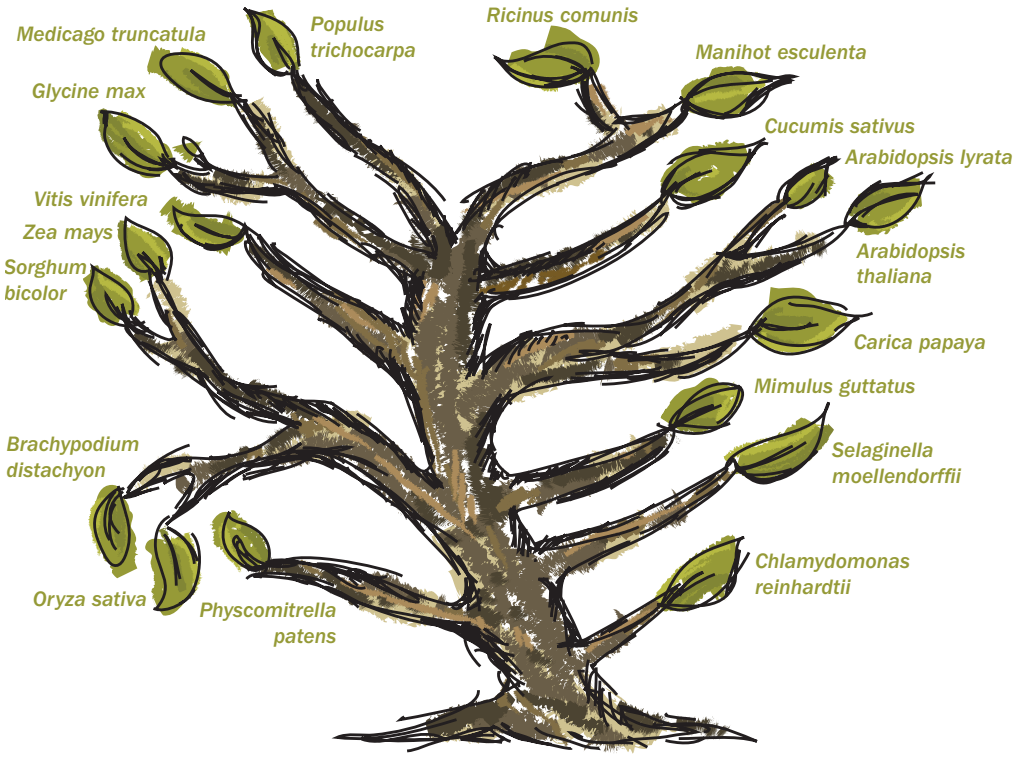


Image by Caitlin Youngquist, LBNL

Genome Portal

DNA sequence is just a starting point for genome exploration. Predicted genes and functions supported by genome-scale transcriptomics, proteomics, and genome variation studies offer large amounts of experimental and computational data that require advanced tools for their analysis. The Genome Portal (http://genome.jgi-psf.org/) offers interactive Web-based tools for genome analysis and comparative genomics and provides access to data from more than 70 eukaryotic organisms and numerous prokaryotes. Over 20,000 users use these tools in their research monthly. The Genome Portal combines the "genome-centric view," rich in experimental and community-generated data and focused on specifics of individual genomes, with the "comparative view," which places the analysis of each genome in context of comparative genomics.

The Genome Portal includes tools to explore genome structure, gene families, and pathways in a comparative fashion and to collect data, gene models, and annotations from users. It also serves as a community hub for genomics data equipped with a collection of tools for genome analysis and community-wide annotation.

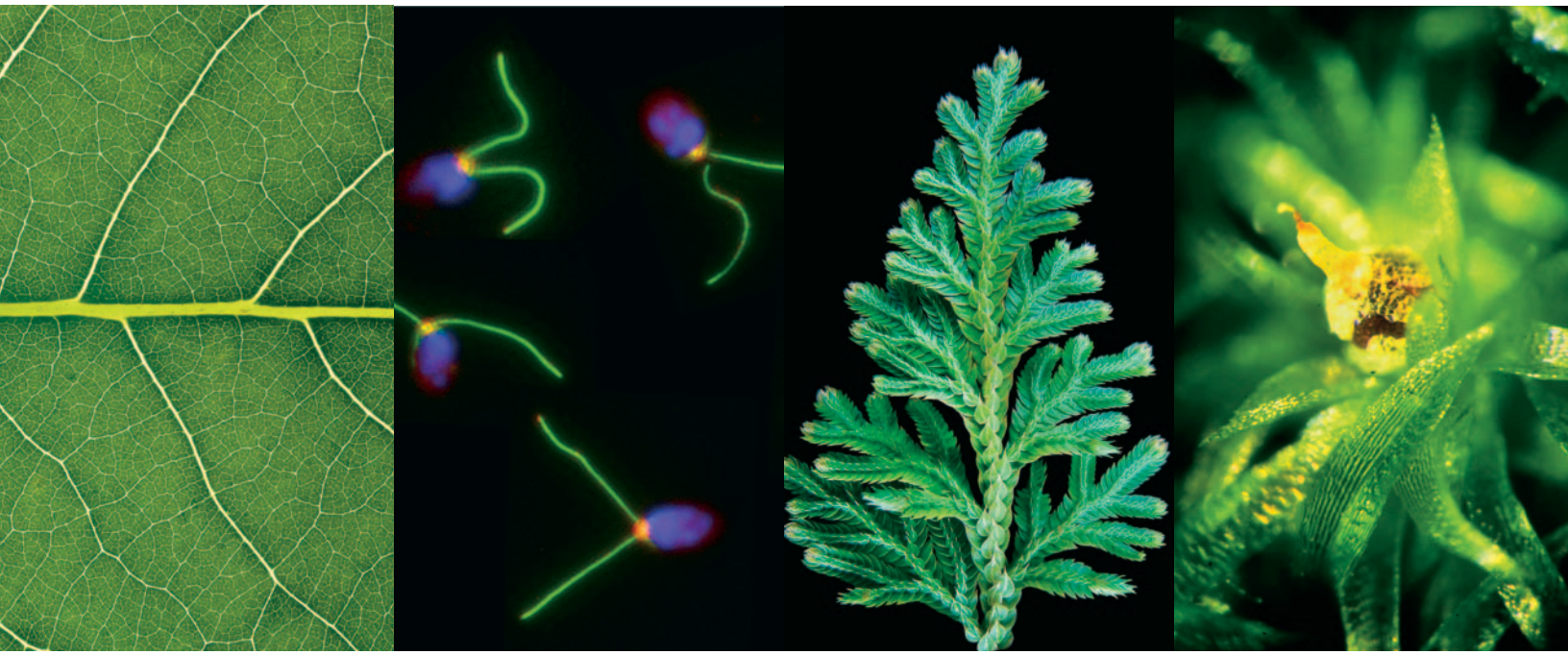
Last year, the Genome Portal was redesigned to boost its performance to better address the needs of its user communities. To support the growing community of users, DOE JGI offers online tutorials and hosts a series of interactive training sessions and jamborees.

Phytozome expanded—"Hub" for comparative plant genomics data

In 2009, the DOE JGI released an enhanced version of Phytozome.net, a Web portal for comparative plant genomics geared to advance biofuel, food, feed, and fiber research. Phytozome provides a central "hub" for Web access to a rapidly growing number of plant genomes, and includes tools for visualization of plant genomes and associated annotations, sequence analysis, and bulk — as well as targeted — plant data retrieval. The gene families available in Phytozome, defined at several evolutionarily significant epochs, provide a framework for the transfer of functional information to important biofuel and agricultural crops from model plant systems, and allow users to explore land plant evolution.

The 4.0 release of Phytozome spans 14 plant genomes, including eight that have been sequenced at the DOE JGI:

- *Populus trichocarpa*, the black cottonwood tree, the first tree sequenced and being explored as a feedstock for a new generation of cellulosic biofuels
- *Sorghum bicolor*, a drought-tolerant grass and the second-most prevalent biofuels crop in the United States.
- Soybean (*Glycine max*), the No. 2 U.S. crop in both harvested acreage and sales and the principal source of biodiesel, a renewable, alternative



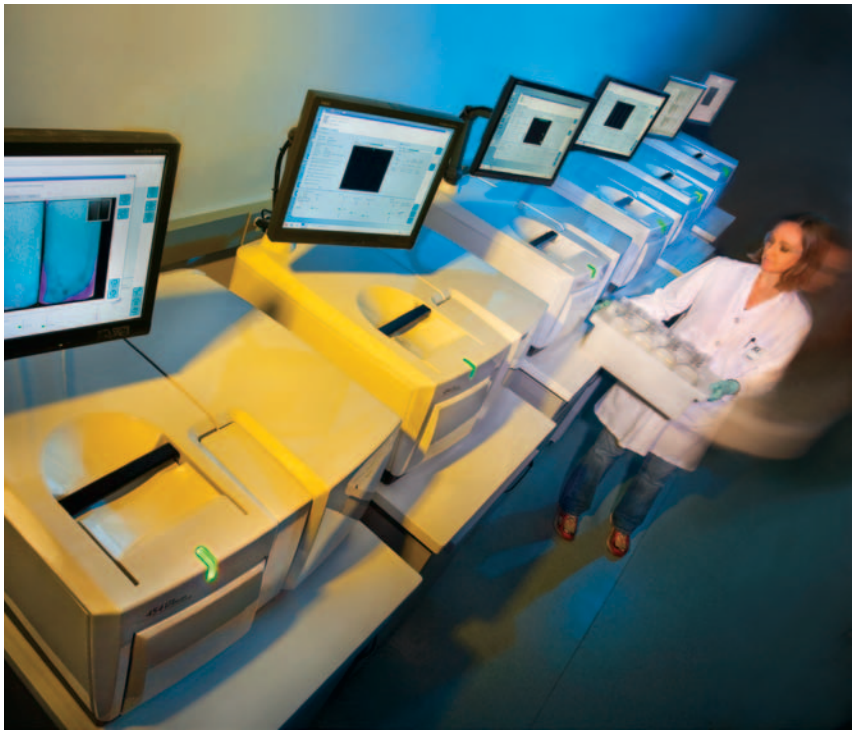
From left to right: *Populus trichocarpa* photo by Roy Kaltschmidt, LBNL; *Chlamydomonas reinhardtii* from EMBO Practical Course, University of Geneva, Switzerland; *Selaginella moellendorffii* by Jing-Ke Weng, Salk Institute; *Physcomitrella patens* by Harald Hedman, SLU, Sweden

fuel with the highest energy content of any current alternative fuel

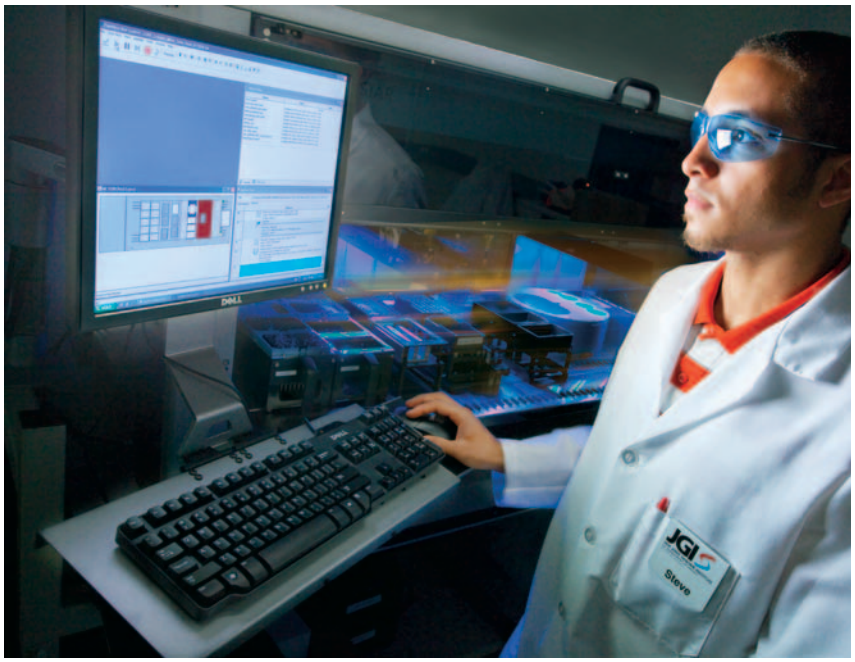
- *Chlamydomonas reinhardtii*, a single-celled green alga, a powerful model system for the study of photosynthesis and source of hundreds of genes associated with carbon dioxide capture and generation of biomass
- *Brachypodium distachyon*, a temperate wild grass and model plant for temperate grasses and herbaceous energy crops
- *Arabidopsis lyrata*, a close relative of the model plant *Arabidopsis thaliana* and a reference genome shedding light on the genetics, physiology, development, and structure of plants in general and how they respond to disease and environmental stress
- *Physcomitrella patens*, a moss widely recognized as an experimental organism of choice not only for basic molecular, cytological, and developmental questions in plant biology, but also as a key link in understanding plant genome evolution
- *Selaginella moellendorffii*, a spike-moss with a compact genome that is helping to define an ancient core of genes common to all vascular plants

Phytozome also includes the completed sequences of rice, papaya, grape, *Medicago* (the genus that includes alfalfa as a member), *Arabidopsis thaliana*, as well as maize bacterial artificial chromosome sequences from the Maize Genome Sequencing Project. The sequences are accessible to the public at **www.phytozome.net**.

Phytozome is a collaboration between scientists at the DOE JGI, Lawrence Berkeley National Laboratory, and the University of California, Berkeley, Center for Integrative Genomics. It was developed with funding from the DOE, the National Science Foundation, the National Institutes of Health, and the Gordon and Betty Moore Foundation.



Megan Kennedy prepares to load reagents into a Roche 454 sequencer. Photo by Roy Kaltschmidt, LBNL



Steven Wilson preparing the DNA sample picotiter plate for the Roche 454 platform. Photo by Roy Kaltschmidt, LBNL

Gap Resolution

Over the past year, the DOE JGI has modified its sequencing pipeline to take advantage of the benefits next-generation DNA sequencing technologies have to offer over traditional Sanger sequencing. Currently, standard 454 Titanium and paired-end 454 Titanium data are generated for all microbial projects and then assembled using the commercially available genome assembler, Newbler. This allows more efficient production of high-quality draft assemblies at a much greater throughput. However, it also presents new challenges—with increased throughput comes a larger number of gaps in the Newbler genome assemblies. Gaps in these assemblies are usually caused by repeats such as when Newbler collapses repeat copies into individual contigs, strong secondary structures, and artifacts of the PCR (Polymerase Chain Reaction) process (specific to 454 paired end libraries). To expedite gap closure and assembly improvement on the growing inventory of these assemblies, a team at the DOE JGI led by Alla Lapidus has developed a software called Gap Resolution, to address this issue. The code was written by Stephan Trong, Brian Foster, and Kurt LaButti with valuable contributions by Tom Brettin and Cliff Han. The software was made freely available to academic institutions in August 2009.

Education, Outreach, and Safety/ Ergonomics

The DOE JGI Undergraduate Research Program – Microbial Genome Annotation was held at the JGI on January 28-29, 2010. Photo by David Gilbert, DOE JGI

DOE JGI Education Program

To help fill a void in the life sciences training and help teaching institutions keep up with the rapid advances in sequencing technologies, the DOE JGI's Education Program develops programs and tools to train the next generation of genomic scientists on large-scale DNA sequencing and bioinformatic analysis by integrating the use of these materials in their research experiences.

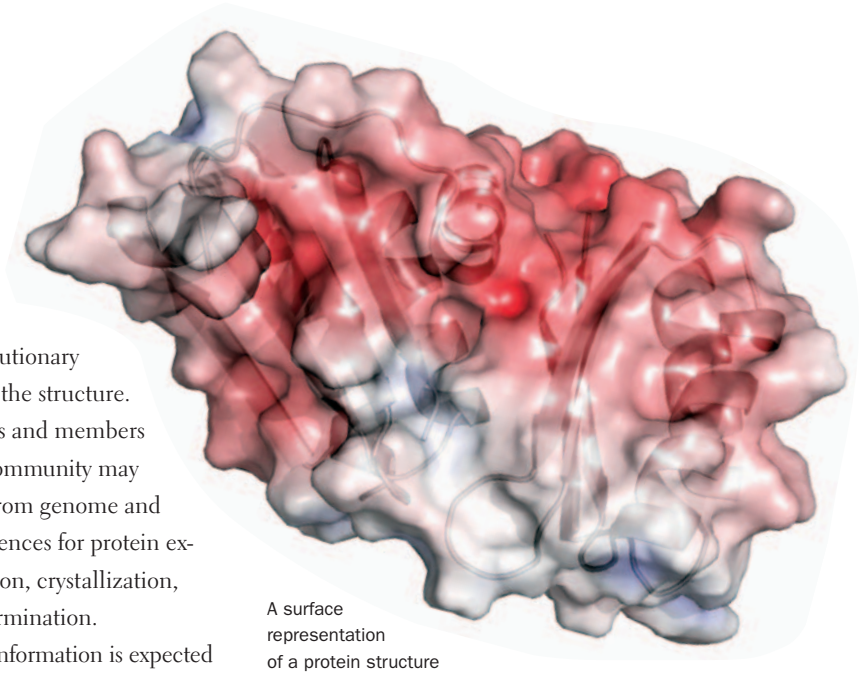
Most projects focus on undergraduates both at community colleges and research universities.

Undergraduate research in microbial genome annotation

The DOE JGI's Microbial Genome Annotation program was developed to incorporate the use of annotation in undergraduate courses. Through the "Adopt a Genome" Program, students would study and annotate recently sequenced genomes – such as those of organisms from little-known branches of the Tree of Life selected as part of the DOE JGI's Genomic Encyclopedia of Bacteria and Archaea (GEBA) – in the context of their own classwork.

The research experience provides students with a "real-world" opportunity to study complete sequences of microorganisms and make novel discoveries that enrich the scientific community as a whole. Ultimately, the program's goal is to allow students nationwide to adopt and annotate GEBA genomes while learning about genomics and bioinformatics.

A total of 16 institutions, three of them from outside the United States, have been selected to participate in the 2010 Microbial Genome Annotation Program. As their annotation platform, students will use the Integrated Microbial Genomes Annotation Collaboration Tool (IMG-ACT), a wiki/Web portal fusion that lets them work with exist-



A surface representation of a protein structure determined by the DOE JGI partnership with the Midwest Center for Structural Genomics.

ing genome datasets and record their discoveries. IMG-ACT is in turn linked to other databases used in microbial genome annotation, including IMG/EDU, a six-frame visualization tool that introduces students to the fundamental principles of genome biology.

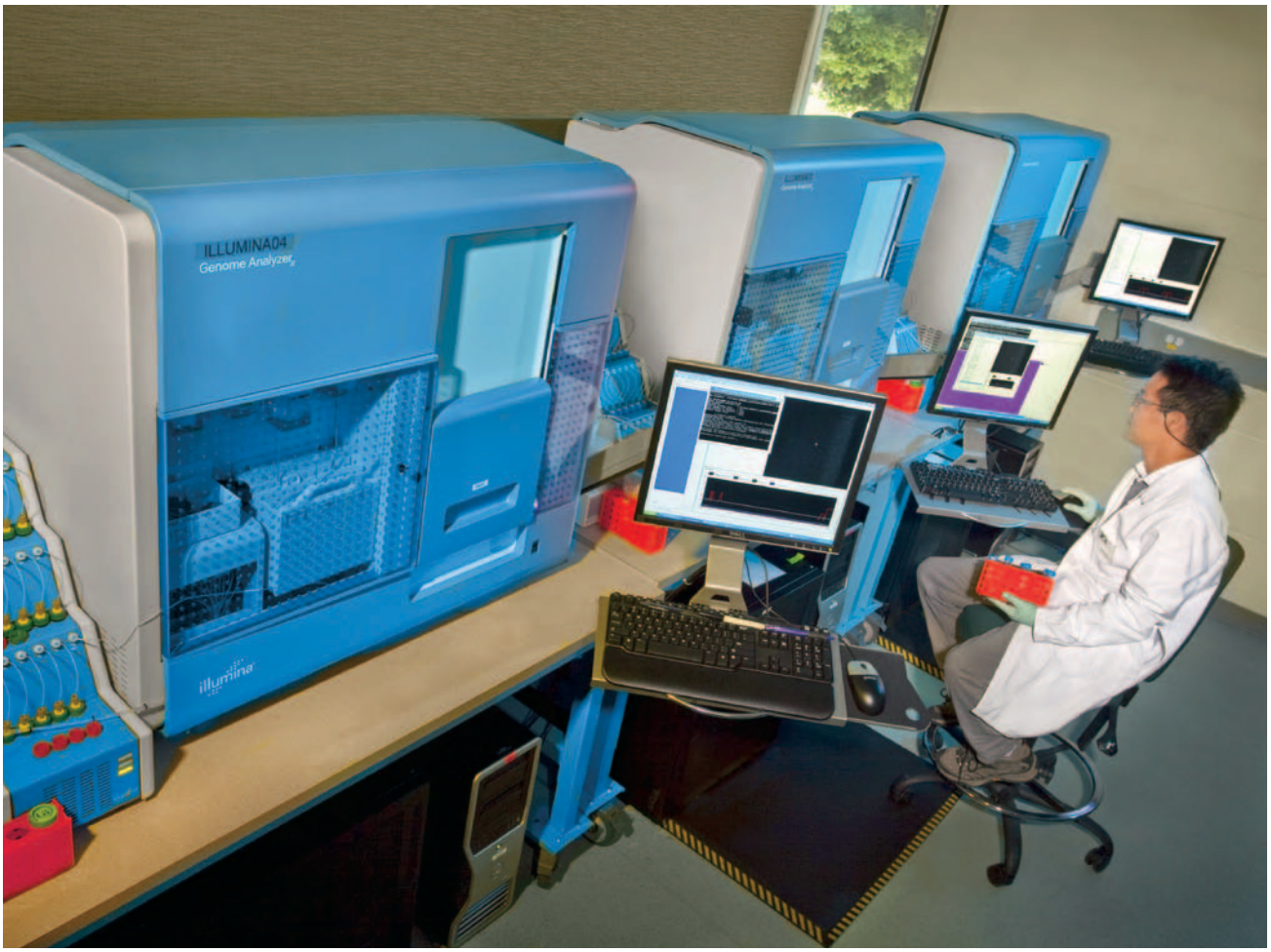
A complementary metagenomics annotation program is also being developed by the DOE JGI Education Program.

DOE JGI Pilot Structural Genomics Program

With the Midwest Center for Structural Genomics at Argonne National Laboratory, the DOE JGI is studying the potential impact of a widespread application of high-throughput structural biology to produce proteins and determine functions and evolutionary relationships from the structure. DOE JGI scientists and members of the JGI User Community may nominate targets from genome and metagenome sequences for protein expression, purification, crystallization, and structure determination.

The structural information is expected to allow researchers to learn more about the biochemical function of proteins, as well as to predict and model substrate interactions in enzymes. Among other things, the work could also help synthetic biologists design protein structures and functions and give scientists a better understanding of the molecular interactions and pathways in these structures.

Christopher Hack pipetting in the Roche 454 sequencing reaction preparation hood. Photo by Roy Kaltschmidt, LBNL



James Han monitors the progress of an Illumina sequencing platform run. Photo by Roy Kaltschmidt, LBNL

The ASM/JGI Bioinformatics Institute

To help undergraduate faculty in science, technology, engineering, and math (STEM) disciplines with little to no familiarity in the use of bioinformatics tools to understand, interpret, and use molecular sequence information, the American Society of Microbiology and the DOE JGI have developed a series of workshops that allow teachers to develop classroom activities and research projects for their students.

The workshops provide participants with hands-on experience in accessing the Internet for databases, tools, and resources to identify tools and resources for developing classroom activities and research projects.

To apply for the workshops, STEM faculty must be current, full-time teachers at community colleges, four-year colleges, or research universities. After participating in the Institute workshops, participants are expected to demonstrate the effectiveness of the training they received not just in developing curricula for use in their classes but also in sharing such modules with other STEM faculty in education publications, and in presenting a project at a national professional society meeting.

Undergraduate Research in Microbial Characterization

In collaboration with the University of Missouri, Columbia, and the University of South Florida, the DOE JGI is developing tools to help undergraduates isolate, characterize, and sequence the genomes of novel organisms that form the foundation of the food chain and the energy chain found in deep-sea vents.

The research project affords undergraduates the opportunity to understand how biological knowledge is created from isolating novel autotrophs and then sequencing their genomes, using bioinformatics to analyze and interpret their data and annotating the information.

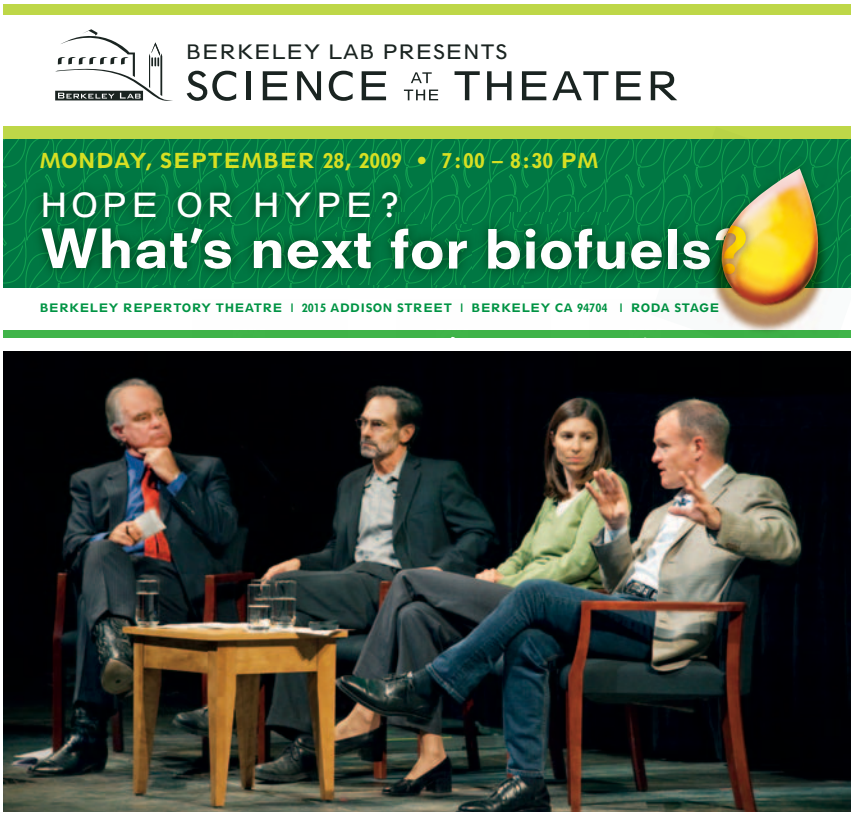
The modules focus on providing students with an understanding of processes such as gene evolution, protein structure and function, metabolic pathways, and ecological adaptation. These tools will be shared with faculty at diverse types of undergraduate institutions through a workshop in the summer of 2010 and are expected to be applicable to a variety of life science laboratory courses.

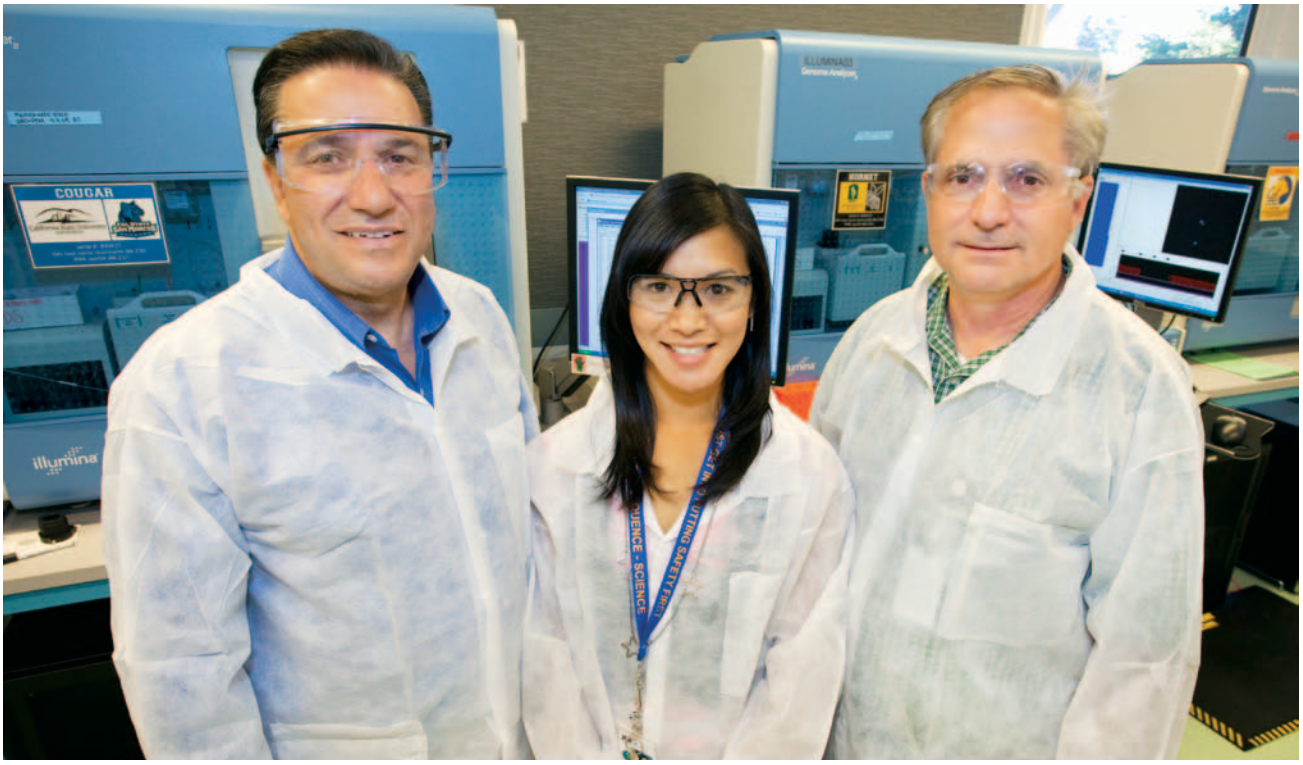
Undergraduate Research in Microbial Functional Genomics

Annotating the complete genome sequence of any organism is akin to conducting an experiment to test a hypothesis. Functional genomics applies techniques in molecular biology and microbiology to as many genes as possible. To bring functional genomics research into the undergraduate experience, the JGI is developing a pilot program with Hiram College in Ohio, and St. Cloud State University in Minnesota, to incorporate concepts such as reverse and forward genetics, protein overexpression, and protein crystallization in undergraduate lab experiences.

Community Outreach

DOE JGI's own Jim Bristow (center left) and Susannah Tringe (center right) appeared on a panel with Joint BioEnergy Institute CEO Jay Keasling (right) on September 28, 2009, at the Berkeley Repertory Theatre's Roda Stage. KTVU Channel 2 health and science editor John Fowler (left) moderated the talk titled: "From the Sun to Your Gas Tank: A New Breed of Biofuels May Help Solve the Global Energy Challenge and Reduce the Impact of Fossil Fuels on Global Warming," discussing ways to convert the solar energy stored in plants into liquid fuels. A video of the talk can be viewed at http://www.youtube.com/watch?v= mRTwuxVurIE





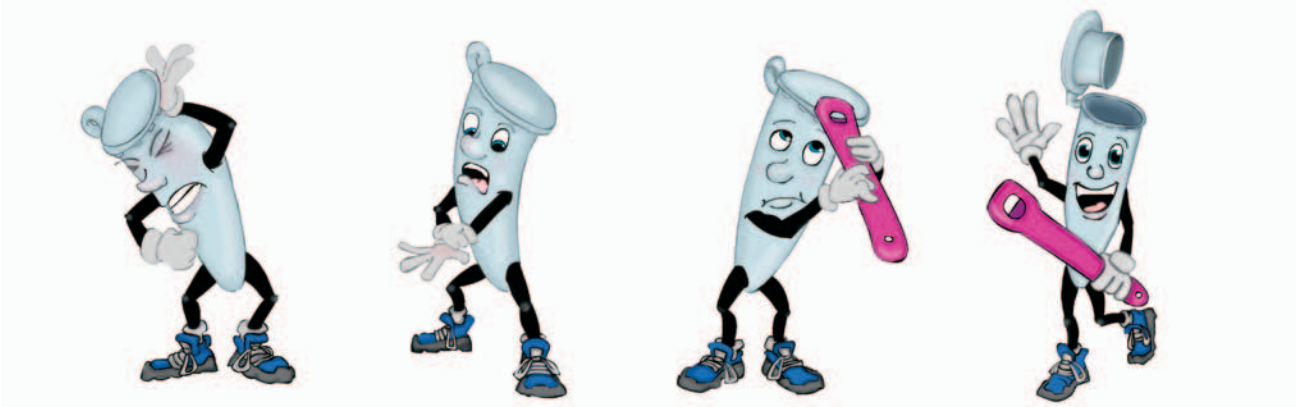
After the DOE JGI's ergonomic injuries hit double digits in 2008, management determined that the problems were caused in part by a reactive rather than a proactive safety approach, and a lack of communication between management and staff. As part of a proactive safety culture program developed to resolve these issues, 33 DOE JGI members were designated Area Safety Leaders (ASLs) to monitor the ergonomic habits of the DOE JGI employees with the help of an on-staff ergonomist. With the employee-driven Safety Culture Group, they also serve as liaisons to help the DOE JGI Safety Team of Vito Mangiardi, Deputy Director of Business Operations, Production, and Informatics (left); Safety Coordinator Stephen Franaszek (right); and Assistant Safety Coordinator Cheryl Chu spread the message of safety awareness between line management and staff. A year after the program was implemented, the list of ergonomic injuries at the DOE JGI has dropped by nearly 80 percent. Photo by Roy Kaltschmidt, LBNL

Safety/Ergonomics

Emphasizing employee safety in the workplace, the DOE JGI relies on the Integrated Safety Management Plan to ensure that all work processes are executed with the health and safety of employees, guests, the public, and the environment in mind. DOE JGI employees and their supervisors come to an agreement on how to work safely through a Job Hazards Analysis (JHA), a process developed by the Lawrence Berkeley National Laboratory's Environment, Health, and Safety team, to identify any work hazards and required training prior to starting work. A variation of the JHA is also used for any work done by contractors at the DOE JGI.

Another important component of work safety is ergonomics, the science of designing equipment and practices to reduce musculoskeletal disorders in the workplace. As DOE JGI employees in both laboratory and office environments routinely perform highly repetitive tasks, developing ergonomic solutions suited to the worker and the environment are critical to ensuring worker and workplace safety. The DOE JGI has an on-site ergonomist who works with the members of the Ergonomics Working Group to promote awareness of ergonomic issues and determine how to resolve any concerns that might arise.

Several of the ergonomic solutions used by the DOE JGI have been featured at the annual Applied Ergonomics Conference in Dallas, Texas, and nominated for the prestigious Ergo Cup. At the 2010



Flippy the Lid Flipper and all the "ergo-characters" below by Micah Brown

Conference, two DOE JGI entries competed for Ergo Cups in separate categories: Ergonomic Program Improvement Initiatives and Workplace Solutions.

Titled "Empowering Employees in Ergonomics," the first entry focuses on employee-driven elements of the DOE JGI program. To help promote awareness of ergonomics and safety and encourage employee involvement, a grassroots movement established working groups in both safety and ergonomics composed of employees representing every department. These groups work on projects such as maintaining practice workstations, regularly updating safety and ergonomics flyers in high-traffic areas such as hallways and restrooms, on-the-job peer training, and events scheduled around Ergonomics Awareness Month.

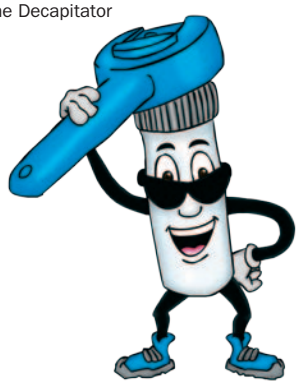
Other employee-driven initiatives encouraged by management and designed to reduce the total injuries resulting from performing repetitive and detail-oriented tasks also led to the central installation of computer usage tracking software that reminds and instructs employees to take regular breaks from their work, and the designation of an on-site relaxation and rejuvenation room.

Among the measurable results of increasing awareness of the DOE JGI's safety values and getting the employees to participate in the program are a 92 percent reduction rate in ergonomic-related reported incidents and a 72 percent reduction in total ergonomic injuries.

The DOE JGI also submitted an entry into the Workplace Solutions category of the Ergo Cup, offering an employee-driven solution to a repetitive and risky task for those working with new sequencing technologies: capping and decapping lids of hundreds of 2-milliliter tubes on a daily basis.

To easily open and close both flip-top and screw-cap tubes, production line employees designed and developed a lightweight tool. By eliminating the high levels of force needed to use a two-fingered grip and pry open manufacturer-tightened tubes, the solution presented in "Flip Your Lid" reduced the task's Strain Index from a hazardous 7.3 to a safe score of 0.2.

The Decapitator



Mr. Twister

Sgt. Swivel

Appendix A:

DOE JGI Sequencing Processes

DNA: Life's code

Deoxyribonucleic acid or DNA, the information embedded in all living organisms, is a molecule made up of four chemical components— the nucleotides Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). These letters constitute the "rungs" of the DNA molecule's double-helical ladder, with the A's always binding with T's, and C's with G's.

What is DNA sequencing?

Just as computer software is rendered in long strings of 0s and 1s, the "software" of life is represented by a string of the four chemicals, A, T, C, and G. To understand the software of a living organism, we must know the order, or sequence, of these informative bits.

DOE JGI Sanger sequencing process

Whole-genome shotgun sequencing is a technique for determining the precise order of the letters of DNA code of a genome. A decade ago the DOE JGI relied exclusively on this Sanger method to generate genome sequences by arranging and assembling DNA fragments based on their sizes. Sequencing technologies since adopted by the DOE JGI can sequence more samples simultaneously and at a much faster rate.

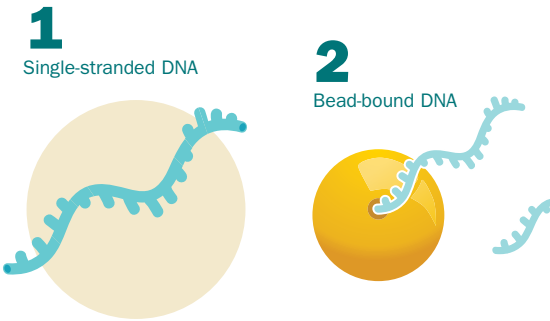
New sequencing technologies

Two sequencing technologies promise an even faster and more efficient future for sequencing at the DOE JGI. Both platforms take a sequencing-by-synthesis approach to build a picture of a newly synthesized DNA fragment one base at a time. The addition of each base is detected in real time, eliminating the need for separating molecules according to size using capillary electrophoresis. One method uses the Roche 454 machine and involves a process called emulsion PCR (Polymerase Chain Reaction) along with pyrosequencing. The other method uses the Illumina machine and involves bridge PCR along with reversible dye terminators.

Compared with the whole-genome shotgun approach, both new sequencing technologies eliminate the time-consuming steps of *in vivo* cloning, colony picking, and capillary electrophoresis. Additionally, both machines are able to produce approximately 1,000 times more bases per week than the previous generation of sequencers and both eliminate bias against *in vivo* genomic regions.

454 sequencing technology

Roche's 454 sequencing technology uses emulsion PCR DNA amplification combined with pyrosequencing, allowing just one person to prepare and sequence an entire genome. The 454 instrument uses a parallel-processing approach to produce an average of 300-400 megabases (300 million bases) of DNA sequence per nine-hour sequencing run. This means that a single 454 machine has the potential to sequence the equivalent of one human genome (3.1 billion bases) in one month.



1. A single-stranded DNA fragment is attached to one capture bead.
2. Each bead carries millions of copies of a unique single-stranded DNA.
3. PCR amplification is performed on each bead in separated water-in-oil droplet (emulsion) so that there are 10 million copies of a fragment on each bead, eliminating the need for cloning and robots.
4. Each of the DNA-coated beads are then individually loaded into one of the 3.2 million hexagonal wells of a fiber-optic slide.
5. Solutions containing a single nucleotide type (A, T, C, or G) are consecutively applied over the wells in cycles. As each base (A, T, C, or G) is incorporated into a new DNA strand, a CCD (Charged-Couple Device) camera records the light flashes generated by the reaction. The sequence of hundreds of thousands of individual reactions is determined simultaneously, producing millions of bases of sequence per hour from a single run.

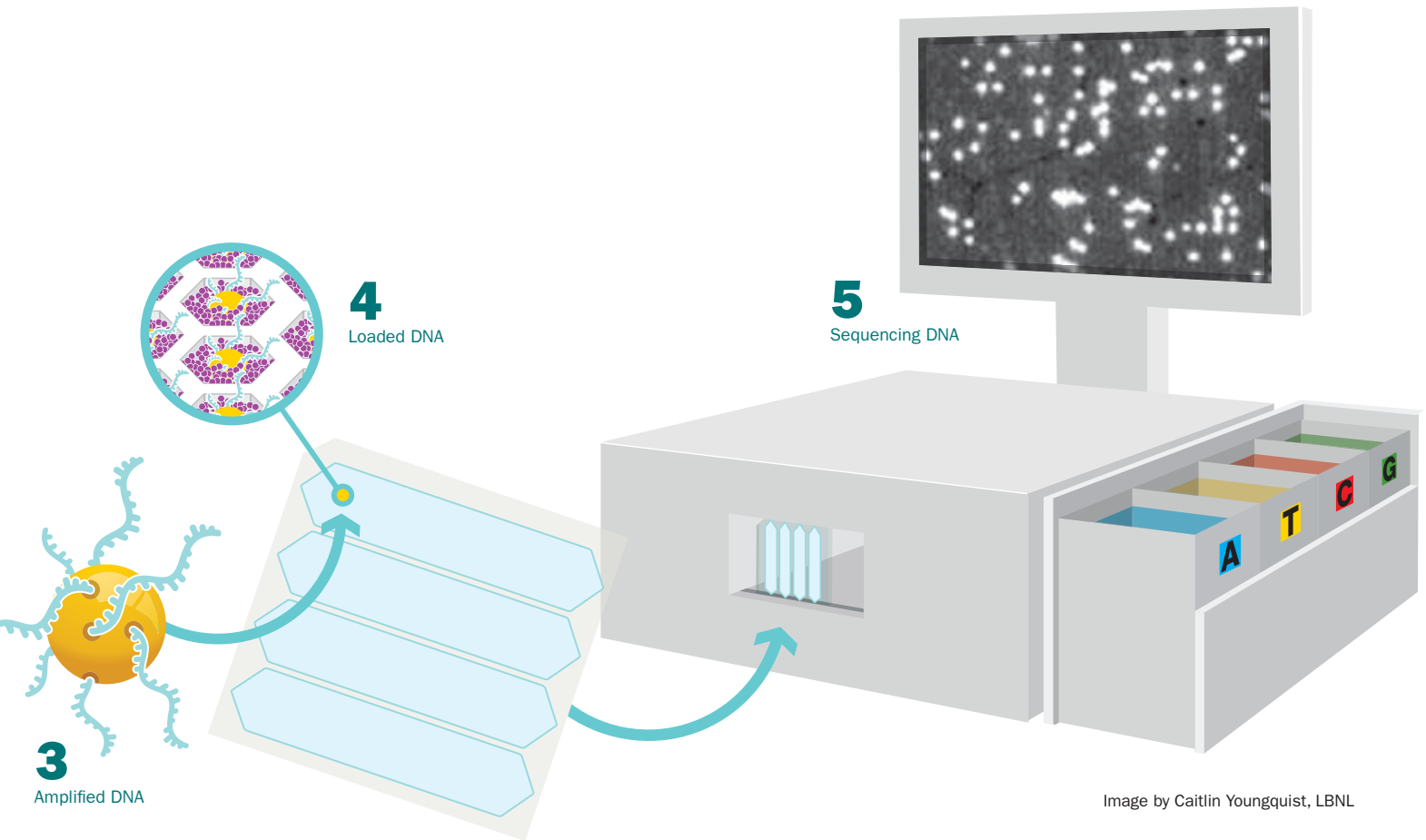
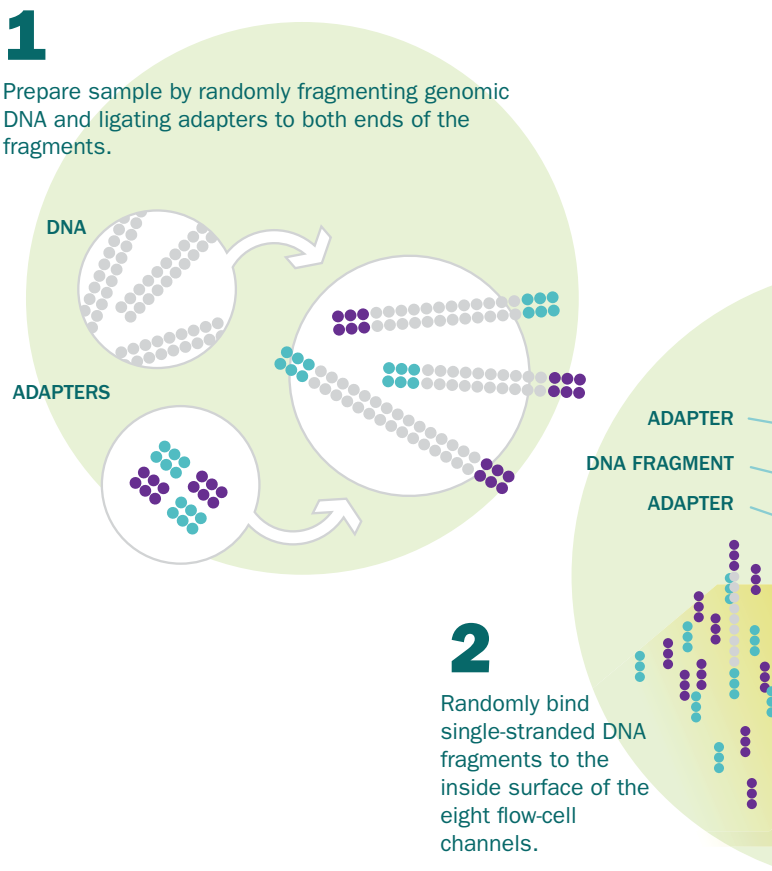


Image by Caitlin Youngquist, LBNL

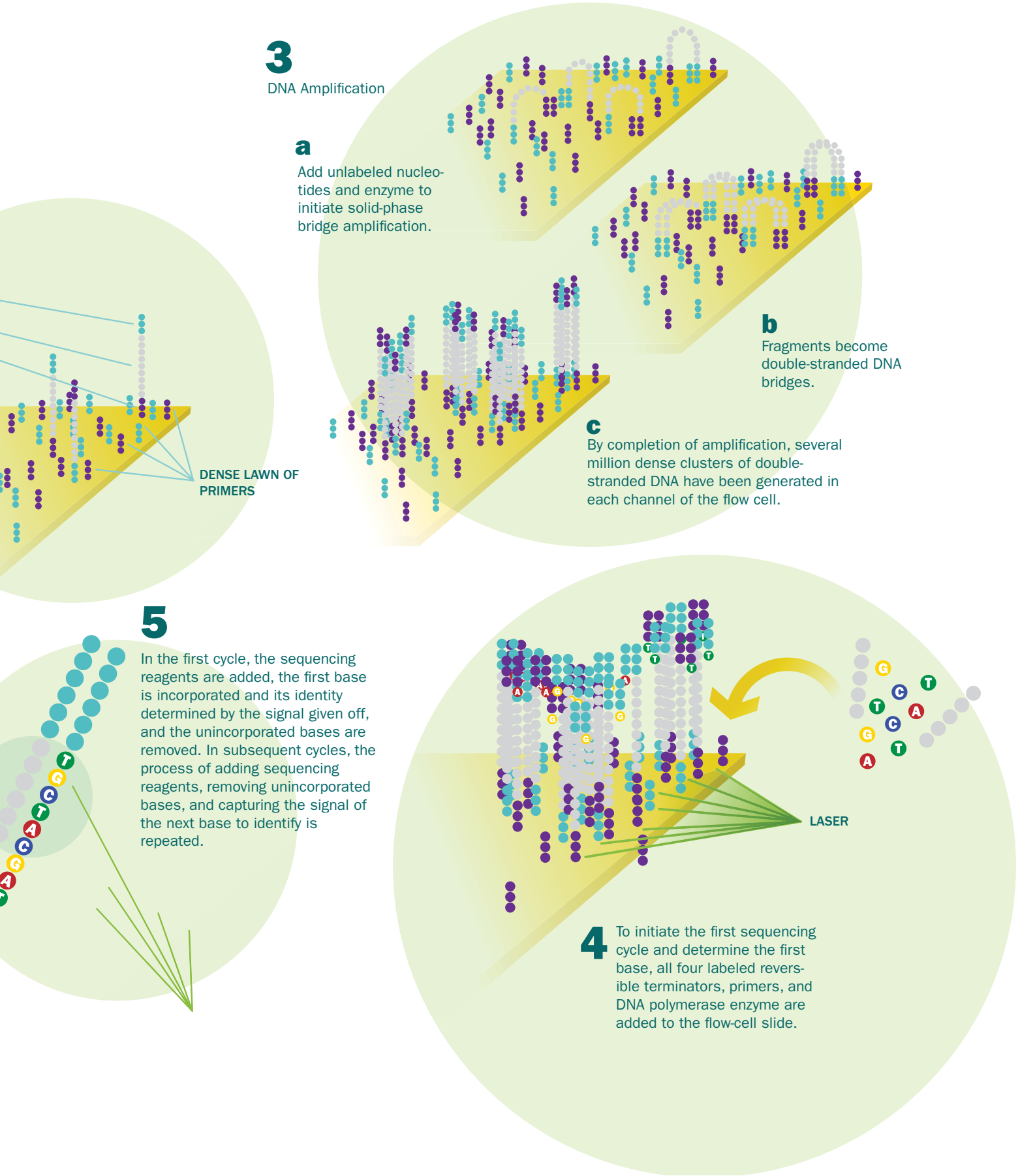
Illumina sequencing tehnology

This platform is based on parallel sequencing of millions of fragments using a proprietary Clonal Single Molecule Array technology—which amplifies the template by bridge PCR onto a glass flow-cell slide—and a novel reversible terminator-based sequencing chemistry that allows detection of the sequence in real time during the sequencing-by-synthesis process.



Comparison of Sequencing Processes (as of February 2010)

PROCESS STEPS	454 – TITANIUM	ILLUMINA – GAII
Break up genome and isolate separate fragments	• Shear and ligate adapters • One fragment per capture bead	Shear and ligate adapters
Isolate DNA fragments	One bead per emulsion droplet	Hybridize to adaptor complementary oligos on flow-cell slide
Amplify fragments	Emulsion PCR	Fragments are now at one fragment per μm^2
Sequencing chemistry	Pyrosequencing One type of base per cycle	Bridge amplification
Detection of sequence	Real-time detection of luminescence each cycle	Real-time detection of fluorescence each cycle
Read length (max)	500-600 bases	35-100 bases (depending on run type)
Assembly	Single direction and paired reads available; assembled using Newbler program	Single-direction and paired reads available; various assembly programs
Bases per week per machine	2 billion bases (five runs)	>20 billion bases (two runs at 75 cycles)
Advantages	• Fast • Less expensive • Good for capturing uncloned regions	• Fast • Less expensive • Can distinguish bases in a homopolymer
Disadvantages	• Cannot accurately decipher repetitive regions • Cannot distinguish the number of nucleotides in a long homopolymer	• Cannot accurately decipher repetitive regions • Shorter reads


3
DNA Amplification
a
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.
b
Fragments become double-stranded DNA bridges.
c
By completion of amplification, several million dense clusters of double-stranded DNA have been generated in each channel of the flow cell.
DENSE LAWN OF PRIMERS
5
In the first cycle, the sequencing reagents are added, the first base is incorporated and its identity determined by the signal given off, and the unincorporated bases are removed. In subsequent cycles, the process of adding sequencing reagents, removing unincorporated bases, and capturing the signal of the next base to identify is repeated.
LASER
4
To initiate the first sequencing cycle and determine the first base, all four labeled reversible terminators, primers, and DNA polymerase enzyme are added to the flow-cell slide.

Appendix B:

Glossary

Annotation: The process of identifying the locations of genes in a genome and determining what those genes do to improve accuracy of genetic information collected.

Archaea: One of the three domains of life (eukaryotes and bacteria being the others) that subsume primitive microorganisms that can tolerate extreme (temperature, acid, etc.) environmental conditions.

Assembly: Compilation of overlapping DNA sequences obtained from an organism that have been clustered together based on their degree of sequence identity or similarity.

Barcoding: The practice of appending known, unique synthetic DNA sequences to sequencing libraries to allow pooling of libraries for next-generation sequencing, after which sequence data can be assigned to particular libraries or samples based on the barcode sequence.

Base: A unit of DNA. There are four bases: adenine (A), guanine (G), thymine (T), and cytosine (C). The sequence of bases is the genetic code.

Base pair: Two DNA bases complementary to one another (A and T or G and C) that join the complementary strands of DNA to form the double helix characteristic of DNA.

Bioremediation: Using microorganisms to breakdown contaminants and other unwanted substances in waste and other substances.

Biogeochemistry: A study of the biosphere's interactions with the Earth's chemical environment.

Bioinformatics: The use of computers to collect, store, and analyze biological information.

Bridge amplification: A proprietary technique used by Illumina sequencing platforms to generate single-stranded clusters of template DNA.

Cloning: Using specialized DNA technology to produce multiple, exact copies of a single gene or other segment of DNA to obtain enough material for further study.

Contig: Group of cloned (copied) pieces of DNA representing overlapping regions of a particular chromosome.

Coverage: The number of times a region of the genome has been sequenced during whole genome shotgun sequencing.

Curation: Analysis of genome annotations to improve and maintain data presentation.

Draft genome: The term for an incomplete genome sequence that can be applied to a wide range of sequences, from those that have the minimum amount of information needed for submission to a public database, to assembled genomes that have undergone manual and automatic review but still have sequence errors that need to be corrected.

Eukaryotes: The domain of life containing organisms that consist of one or more cells, each with a nucleus and other well-developed intracellular compartments. Eukaryotes include all animals, plants, and fungi.

Finished genome: In accordance with the 1996 Bermuda standard, this is a gapless sequence with a nucleotide error rate of one or less in 10,000 bases.

Flow cell: Resembles a microscopic slide with eight channels on which DNA samples are loaded for analysis on the

Illumina sequencing platforms.

Fosmid: A vector suitable for cloning genomic inserts approximately 40 kilobases in size.

GenBank: Open-access, publicly available collection of annotated sequences submitted by individual laboratories and large-scale sequencing centers that is overseen by the National Center for Biotechnology Information.

Library: An unordered collection of clones containing DNA fragments from a particular organism or environment that together represent all the DNA present in the organism or environment.

Mapping: Charting the location of genes on chromosomes.

Metagenomics (also Environmental Genomics or Community Genomics): The study of genomes recovered from environmental samples through direct methods that do not require isolating or growing the organisms present. This field of research allows the genomic study of organisms that are not easily cultured in a laboratory.

Methylotroph: A microorganism that uses one-carbon compounds such as methanol and methane as its main food source.

Microbiome: A defined environment within which a community of microbes exist and interact with each other.

Paired-end reads: DNA library preparation technique that lets researchers look at both the forward and reverse template strands of a large DNA fragment and provides positional information.

PCR: Acronym for Polymerase Chain Reaction, a method of DNA amplification.

Phylogeny: The evolutionary history of a molecule such as gene or protein, or a species.

Picotiter plate: Flat plate with multiple tiny wells that act as miniscule test tubes used to hold DNA fragments in sequencing platforms from 454 Life Sciences, a division of Roche.

Plasmid: Autonomously replicating, extrachromosomal, circular DNA molecules, distinct from the normal bacterial genome and nonessential for cell survival under nonselective conditions. Some plasmids are capable of integrating into the host genome. A number of artificially constructed plasmids are used as cloning vectors.

Polymerase: Enzyme that copies RNA or DNA. RNA polymerase uses preexisting nucleic acid templates and assembles the RNA from ribonucleotides. DNA polymerase uses preexisting nucleic acid templates and assembles the DNA from deoxyribonucleotides.

Prokaryotes: Unlike eukaryotes, these organisms, (e.g., bacteria) are characterized by the absence of a nuclear membrane and by DNA that is not organized into chromosomes.

Pyrosequencing: A method of DNA sequencing used by platforms from Roche's 454 Life Sciences to produce light that is captured by a camera to detect the activity of DNA polymerases.

Read length: The number of nucleotides tabulated by the DNA analyzer per DNA reaction well.

Sanger sequencing: The tried-and-true sequencing technique used by the JGI for several years in which a DNA strand is first treated with a variety of nucleotides, enzymes, and a specific primer to generate a collection of smaller DNA fragments. Each fragment is tagged by fluorescent markers that identify the last base of each segment. The fragments are then separated by size and pass through a detector connected to a camera that determines the original strand's sequence based on the order in which the fragments appear.

Sequence: Order of nucleotides (base sequence) in a nucleic acid molecule. In the case of DNA sequence, it is the precise ordering of the bases (A, T, G, C) from which the DNA is composed.

Sequencing by synthesis: Proprietary sequencing technique used by Illumina systems in which four fluorescently labeled nucleotides determine the sequence of a DNA fragment one base at a time.

Subcloning: The process of transferring a cloned DNA fragment from one vector to another.

Transformation: A process by which the genetic material carried by an individual cell is altered by the introduction of foreign DNA into the cell.

Transcriptome: A collection of all the RNA transcripts in a given cell that serves as a snapshot of global gene expression.

Vector: DNA molecule that can be inserted into another DNA fragment acting as a host cell without losing its ability to self-replicate.

Whole-genome shotgun: Semi-automated technique for sequencing long DNA strands in which DNA is randomly fragmented and sequenced in pieces that are later reconstructed by a computer.

Appendix C:

Community Sequencing Program 2010 Projects

Proposer	Affiliation	Project Description
EUKARYOTES		
Collier, Jackie	Stony Brook University	Four Labyrinthulomycete species
Cullen, Daniel	US Forest Service, Forest Products Laboratory	Homokaryotic derivative of *Postia placenta*
Cullen, Daniel	US Forest Service, Forest Products Laboratory	Lignin-degrading fungus *Phlebiopsis gigantea*
Goodwin, Stephen	USDA-ARS and Purdue University	Dothideomycetes plant pathogens
Grossniklaus, Ueli	University of Zurich	Apomictic plant *Boechera holboellii*
Koppisch, Andy	Los Alamos National Laboratory	Colony-forming microalga *Botryococcus braunii* var Showa
Kubisiak, Thomas	US Forest Service, Southern Research Station	Fusiform rust fungus *Cronartium quercuum f.sp. fusiforme*
Martin, Francis	Institut National de la Recherche Agronomique	Pan-global Basidiomycetes *Pisolithus tinctorius* and *Pisolithus microcarpus*
Moreau, Hervé	CNRS and UPMC	Photosynthetic marine eukaryotes phytopicoplankton
Paterson, Andrew	University of Georgia	Resequencing sorghum
Phister, Trevor	North Carolina State University	Completion of the *Dekkera (Brettanomyces) bruxellensis* genome sequence
Pringle, Anne	Harvard University	Cellulose-degrading fungus *Amanita thiersii*
Quatrano, Ralph	Washington University in St. Louis	*Ceratodon purpureus* (moss)
Reeve, Wayne	Murdoch University	Phytopathogenic oomycete *Phytophthora cinnamomi*
Roossinck, Marilyn	Samuel Roberts Noble Foundation	Alteration of *Curvularia protuberata* transcripts due to presence of *Curvularia* thermal tolerance virus
Vyverman, Wim	Ghent University	Diatom transcriptome and genome
Weeks, Donald	University of Nebraska-Lincoln	Transcriptome analyses of *Chlamydomonas* and *Chlorella*
Zhong, Shaobin	North Dakota State University	Fungal pathogen *Cochliobolus sativus*
BACTERIA AND ARCHAEA		
Auchtung, Jennifer	Michigan State University	Role of population microdiversity in adaptation to environmental redox gradients
Anderson, Iain	DOE Joint Genome Institute	Xylan degraders
Anderson, Iain	DOE Joint Genome Institute	Genomic survey of haloarchaeal genomes
Bayer, Travis	University of California, San Francisco	*Actinotalea fermentans*
Bollmann, Annette	Miami University	Five isolates from the contaminated subsurface sediment of Oak Ridge's FRC area
Brown, Igor	NASA Johnson Space Center	Two strains of Cyanobacteria for biological remediation
Bryant, Donald	Penn State University	Representative photosynthetic purple sulfur bacteria
Cavicchioli, Rick	University of New South Wales	Novel haloarchaea from Deep Lake
Coleman, Nicholas	University of Sydney	Ethene and vinyl chloride-oxidizing *Mycobacterium* strains
Cooper, Vaughn	University of New Hampshire	Adaptive mechanisms in *Burkholderia* biofilms

Copley, Shelley	University of Colorado at Boulder	*Sphingobium chlorophenolicum*
Daly, Michael	Uniformed Services University of the Health Sciences	Radiation-resistant bacterium *Deinococcus grandis*
Dopson, Mark	Umeå University	Psychrotolerant *Acidithiobacillus* species
Dvornyk, Volodymyr	University of Hong Kong	*Nostoc linckia* from "Evolution Canyon"
Edwards, Elizabeth	University of Toronto	Novel acetogenic bacterial isolates from dechlorinating microbial mixed cultures
Emerson, David	Bigelow Laboratory for Ocean Sciences	Two novel Zetaproteobacteria from the ocean
Green, Stefan	Florida State University	Denitrifying bacterial isolates
Grzymski, Joseph	Desert Research Institute	Microbes integral to the cycling of sulfate and iron
Haggblom, Max	Rutgers University	*Acidobacterium* species from Arctic tundra soils
Hedlund, Brian	University of Nevada, Las Vegas	Thermophiles in Great Basin hot springs
Kappler, Ulrike	The University of Queensland	Alkaliphilic sulfur oxidizing bacteria for sulfur pollution remediation
Lewis, Gillian	University of Auckland	Freshwater manganese depositing ß-proteobacterium (Siderocapsaceae)
Liao, James	University of California, Los Angeles	Reverse metabolic engineering of *Escherichia coli*
Liu, Wen-Tso	University of Illinois at Urbana-Champaign	Comparison of novel methanogens from peatlands and bioreactors
Liu, Wen-Tso	University of Illinois at Urbana-Champaign	Obligate syntrophic bacteria capable of phthalate isomer compound degradation in methanogenic conditions
Martinez, Robert	University of Alabama	ORFRC *Rahnella* sp. Y9602
Mavrommatis, Konstantinos	DOE Joint Genome Institute	Cyanobacteria (*Synechocystis*) transcriptome
Mayali, Xavier	Lawrence Livermore National Laboratory	Marine *Roseobacter* RCA cluster bacterial strain LE17
Mills, David	University of California, Davis	*Acetobacter aceti ATCC 23746*
Muyzer, Gerard	Delft University of Technology	Haloalkaliphilic chemolithoautotrophic *Thioalkalivibrio* bacteria
Nesbø, Camilla	University of Oslo	*Thermotogales* strain mesG1.Ag.4.1
Norton, Jeanette M.	Utah State University	*Nitrosomonas cryotolerans* and *Nitrosospira briensis* for comparative phylogenomics of ammonia-oxidizing bacteria
Pappas, Katherine	University of Athens	*Zymomonas mobilis* transcriptomes and resequencing *Z. mobilis* industrial strain ZM4
Reeve, Wayne	Murdoch University	Rhizobia of clover, pea/bean and lupin microsymbionts
Robb, Frank	Center of Marine Biotechnology	Carbon monoxide oxidizing thermophiles
Rodrigues, Jorge	University of Texas at Arlington	Genome closure of lignocellulosic degrader *Verrucomicrobium* sp. strain *TAV2*.
Sanchez Amat, Antonio	University of Murcia	Marine bacterial genus *Marinomonas*
Sello, Jason	Brown University	Biomass-degrading bacteria *Streptomyces viridosporus* ATCC 39115 and *Streptomyces setonii* ATCC 39116
Smidt, Hauke	Wageningen University	Halorespiring Firmicutes
Stabb, Eric	The University of Georgia Research Foundation	Mutations in *Vibrio fischeri*
Stein, Lisa	University of Alberta	Methanotrophic bacteria from diverse environments
Stepanauskas, Ramunas	Bigelow Laboratory for Ocean Sciences	Single-cell genome sequencing of mesopelagic bacterioplankton
Tisa, Louis	University of New Hampshire	An atypical *Frankia* isolate and non-*Frankia* Actinobacteria from actinorhizal Plants
Vieille, Claire	Michigan State University	Resequencing of *Actinobacillus succinogenes*
Ward, David	Montana State University	Synechococcus cyanobacterial isolates

METAGENOMES		
Breitbart, Mya	University of South Florida	Modern freshwater microbialites
Chistoserdova, Ludmila	University of Washington	Functional metagenomics of methane and nitrogen cycles in freshwater lakes
Davidson, Seana	University of Washington	Metagenome function of the earthworm egg capsule bacterial community
Deng, Li	University of Arizona	Viruses that infect freshwater Cyanobacteria
Edwards, Elizabeth	University of Toronto	Dehalobacter-containing dechlorinating community
Hedlund, Brian	University of Nevada, Las Vegas	Great Boiling Spring sediment and water microbial communities
Kirchman, David	University of Delaware	Metagenomic analysis of methane degradation in Arctic coastal waters and sediments
Madsen, Eugene	Cornell University	Naphthalene biodegrading microbial community
Mincer, Tracy	Woods Hole Oceanographic Institute	Natural microbial community associated with the cyanobacteria *Trichodesmium*
Moon, Christina	AgResearch Limited	Lignocellulolytic enzyme discovery from the rumen
Powell, Amy	Sandia National Laboratories	Eukaryotic microbial metatranscriptome of blue grama grass rhizosphere soils
Sullivan, Matthew	University of Arizona	Viral community in the Subarctic Pacific Ocean
Sullivan, Matthew	University of Arizona	Viral community in the Mediterranean Sea
Taylor, Mike	University of Auckland	Microbial symbionts of New Zealand's endemic wood-degrading insects
Waldrop, Mark	US Geological Services	Permafrost soil *Microbiota*
Ward, Naomi	University of Wyoming	Metatranscriptomic analysis of bacterial-algal interactions
Warnecke, Falk	Friedrich Schiller University of Jena	Desert locust (*Schistocerca gregaria*)
Worden, Alexandra	Monterey Bay Aquarium Research Institute	Metagenomics of uncultured marine eukaryotes

Appendix D:

Review Committees and Board Members

CSP 2010 Reviewers

Eukaryotic Proposal Reviewers

Kenneth Bruno
Pacific Northwest National Laboratory
Joan Bennett
Rutgers University
Daren Brown
USDA-ARS
Jeffrey Dean
University of Georgia
Charles Delwiche
University of Maryland
Maria Ghirardi
National Renewable Energy Laboratory
Stephen Goodwin
Purdue University
Timothy James
University of Michigan
Kevin McCluskey
University of Missouri, Kansas City
Monica Medina
University of California, Merced
Sabeeha Merchant
University of California, Los Angeles
Andrew Paterson
University of Georgia
Carolyn Silflow
University of Minnesota
Shauna Somerville
University of California, Berkeley
Joseph Spatafora
Oregon State University
Garret Suen
University of Wisconsin, Madison
Chris Town
J. Craig Venter Institute

Metagenome Proposal Reviewers

Eric Allen
University of California, San Diego
Donald Bryant
Pennsylvania State University
Ludmila Chistoserdova
University of Washington
Daniel Haft
J. Craig Venter Institute
Steven Hallam
University of British Columbia
Kirk Harris
University of Colorado
William Inskeep
Montana State University
Martin Klotz
University of Louisville
Wen-Tso Liu
University of Illinois
Todd Lowe
University of California, Santa Cruz
David Mead
Lucigen Corporation
William Mohn
University of British Columbia
Howard Ochman
University of Arizona
Mircea Podar
Oak Ridge National Laboratory
Frank Robb
University of Maryland
Matthew Sullivan
University of Arizona
Craig Stephens
Santa Clara University
Tamas Torok
Lawrence Berkeley National Laboratory
Rob Knight
University of Colorado at Boulder
Douglas Rusch
J. Craig Venter Institute
Patricia Sobecky
Georgia Tech
John Spear
Colorado School of Mines
Sergey Stolyar
Pacific Northwest National Laboratory
Daniel van der Lelie
Brookhaven National Laboratory
David Ward
Montana State University

Microbial Proposal Reviewers

Gary Andersen
Lawrence Berkeley National Laboratory
Donald Bryant
Pennsylvania State University
Jonathan Eisen
University of California, Davis
George Garrity
Michigan State University
Kirk Harris
University of Colorado
Phil Hugenholtz
DOE Joint Genome Institute
Martin Klotz
University of Louisville
Nikos Kyrpides
DOE Joint Genome Institute
Katherine McMahon
University of Wisconsin, Madison
Biswarup Mukhopadhyay
Virginia Tech
Frank Robb
University of Maryland
Naomi Ward
University of Wyoming

Joint Genome Institute Policy Board Membership

The DOE JGI Policy Board serves two primary functions. They are to:

1. Serve as a visiting committee that provides advice on policy aspects of JGI operations and long-range plans of the program, including the research and development necessary to ensure future capabilities will meet DOE mission needs;
2. Ensure that JGI resources are used to maximize the technical productivity and scientific impact of the DOE JGI now and in the future.

The DOE JGI Policy Board meets annually to review and evaluate the performance of the entire JGI. Findings and recommendations are reported to the participating Laboratory Directors and to the DOE Office of Biological and Environmental Research.

Members

Gerry Rubin
Howard Hughes Medical Institute (Chair)
Susan Wessler
University of Georgia
Mel Simon
California Institute of Technology
Richard A. Gibbs
Baylor College of Medicine
Sallie W. (Penny) Chisholm
Massachusetts Institute of Technology
James Tiedje
Michigan State University
Edward F. DeLong
Massachusetts Institute of Technology
Chris Somerville
Stanford University
David Galas
Battelle Memorial Institute, Columbus, Ohio, and Institute for Systems Biology, Seattle, Washington
Stephen Quake
Stanford University

Scientific Advisory Committee (SAC)

The Scientific Advisory Committee is a board convened by the DOE JGI Director to provide a scientific and technical overview of the DOE JGI. Among the board's responsibilities are: providing technical input on large-scale production sequencing, new sequencing technologies, and related informatics; overview of the DOE JGI's scientific programs; and an overview of the Community Sequencing Program (CSP). One of the most important tasks of the SAC is to set the final sequence allocation for the CSP based on input from the CSP Proposal Study Panel on CSP project prioritization, and the concurrence of the DOE Office of Biological and Environmental Research.

Members

Bruce Birren
Broad Institute
Edward F. DeLong
Massachusetts Institute of Technology
Joseph R. Ecker
Salk Institute for Biological Studies
Anantha Krishnan
Lawrence Livermore National Laboratory
Jim Krupnick
Lawrence Berkeley National Laboratory
Eric J. Mathur
Synthetic Genomics
Julian Parkhill
The Sanger Institute
Doug Ray
Pacific Northwest National Laboratory
Toby Bloom
Broad Institute

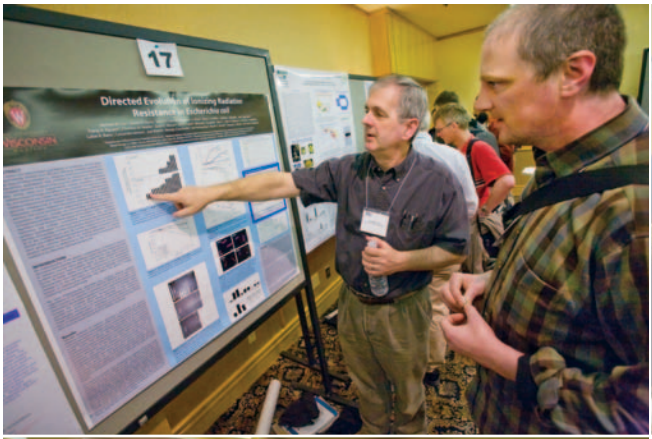
Appendix E:

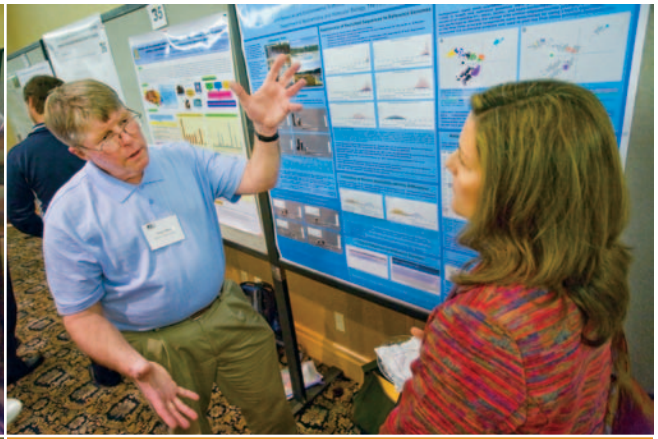
2009 DOE JGI User Meeting Speakers

The DOE JGI Fourth Annual Genomics of Energy and Environment User Meeting took place March 25-27, 2009, in Walnut Creek, California. The keynote speakers were Chris Somerville from the Energy Biosciences Institute, George Church from Harvard University, and J. Craig Venter from the J. Craig Venter Institute.

Other featured speakers were:

Edward M. Rubin
Director, DOE Joint Genome Institute
Jim Bristow
DOE Joint Genome Institute
Nikos Kyrpides
DOE Joint Genome Institute
Igor Grigoriev
DOE Joint Genome Institute
Jeff Dangl
University of North Carolina
Jamie Cate
Energy Biosciences Institute
Bryan O'Neill
Sapphire Energy
Cameron Currie
University of Wisconsin-Madison
Joe Ecker
Salk Institute
Dick Smith
Pacific Northwest National Laboratory
Dan Rokhsar
DOE Joint Genome Institute
Ashlee Earl
Harvard University
Jonathan Eisen
University of California, Davis
James Galagan
Broad Institute
John Willis
Duke University
Ginger Armbrust
University of Washington
Len Pennacchio
DOE Joint Genome Institute
Steve Turner
Pacific Biosciences
Pavel Pevzner
University of California, San Diego
Gary Andersen
Lawrence Berkeley National Laboratory
Sabeeha Merchant
University of California, Los Angeles
Peg Riley
University of Massachusetts
Lucy Shapiro
Stanford University
Jeffrey Miller
University of California, Los Angeles
Rod Wing
University of Arizona





Videos of the 2009 User Meeting talks are available on the DOE JGI's SciVee channel at: http://www.scivee.tv/node/10579/video

Scenes from the 2009 DOE JGI Genomics of Energy and Environment User Meeting. Photos by Roy Kaltschmidt, LBNL

Appendix F:

2009 Publications

Aklujkar M et al. The genome sequence of *Geobacter metallireducens:* features of metabolism, physiology and regulation common and dissimilar to *Geobacter sulfurreducens. BMC Microbiol.* 2009 May 27;9:109.

Allen MA et al. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii:* the role of genome evolution in cold adaptation. *ISME J.* 2009 Sep;3(9):1012-35.

Allgaier M et al. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One.* 2010 Jan 21;5(1):e8812.

Alvarez Retuerto AI et al. Association of common variants in the Joubert syndrome gene (AHI1) with autism. *Hum Mol Genet.* 2008 Dec 15;17(24):3887-96.

Anderson IJ et al. The complete genome sequence of *Staphylothermus marinus* reveals differences in sulfur metabolism among heterotrophic *Crenarchaeota*. *BMC Genomics.* 2009 Apr 2;10:145

Anderson IJ et al. Genomic characterization of methanomicrobiales reveals three classes of methanogens. *PLoS One.* 2009 Jun 4;4(6):e5797.

Anderson IJ et al. Complete genome sequence of *Methanoculleus marisnigri* (Romesser et al. 1981) Maestrojuan et al. 1990 type strain JR1. *Standards in Genomic Sciences.* 2009:1(2):183-188.

Anderson IJ et al. Complete genome sequence of *Methanocorpusculum labreanum* Zhao et al. 1989 type strain Z. *Standards in Genomic Sciences.* 2009:1(2):189-196.

Anderson IJ et al. Complete genome sequence of *Staphylothermus marinus* Stetter and Fiala 1986 type strain F1. *Standards in Genomic Sciences.* 2009:1(2):197-203.

Anderson IJ et al. Complete genome sequence of *Halorhabdus utahensis* type strain (AX-2T). *Standards in Genomic Sciences.* 2009:1(3):218-225.

Anderson SJ et al. Development of simple sequence repeat markers for the soybean rust fungus, *Phakopsora pachyrhizi. Mol Ecol Res.* 2008 Nov;8(6):1310-1312.

Barabote RD et al. Complete genome of the cellulolytic thermophile *Acidothermus cellulolyticus* 11B provides insights into its ecophysiological and evolutionary adaptations. *Genome Res.* 2009 Jun;19(6):1033-43.

Bickel RD et al. Contrasting patterns of sequence evolution at the functionally redundant bric a brac paralogs in *Drosophila melanogaster. J Mol Evol.* 2009 Aug;69(2):194-202.

Birney E et al. Prepublication data sharing. *Nature.* 2009 Sep 10;461(7261):168-70.

Blazewicz J et al. Whole genome assembly from 454 sequencing output via modified DNA graph concept. *Comput Biol Chem.* 2009 Jun;33(3):224-30.

Bluhm BH et al. Analyses of expressed sequence tags from the maize foliar pathogen *Cercospora zeae-maydis* identify novel genes expressed during vegetative, infectious, and reproductive growth. *BMC Genomics.* 2008 Nov 4;9:523.

Bowler C et al. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*. 2008 Nov 13;456(7219):239-44.

Bressler J et al. The INSIG2 rs7566605 genetic variant does not play a major role in obesity in a sample of 24,722 individuals from four cohorts. *BMC Med Genet.* 2009 Jun 12;10:56.

Breznak JA et al. *Spirochaeta cellobiosiphila* sp. nov., a facultatively anaerobic, marine spirochaete. *Int J Syst Evol Microbiol.* 2008 Dec;58(Pt 12):2762-8.

Cai ZQ et al. Extensive reorganization of the plastid genome of *Trifolium subterraneum (Fabaceae)* is associated with numerous repeated sequences and novel dna insertions. *J Mol Evol.* 2008 Dec;67(6):696-704.

Calton MA et al. Association of functionally significant Melanocortin-4 but not Melanocortin-3 receptor mutations with severe adult obesity in a large North American case-control study. *Hum Mol Genet*. 2009 Mar 15;18(6):1140-7.

Campbell BJ et al. Adaptations to submarine hydrothermal environments exemplified by the genome of *Nautilia profundicola*. *PLoS Genet.* 2009 Feb;5(2):e1000362.

Carbon S et al. AmiGO: online access to ontology and annotation data. *Bioinformatics.* 2009 Jan 15;25(2):288-9.

Chain PS et al. Genomics. Genome project standards in a new era of sequencing. *Science*. 2009 Oct 9;326(5950):236-7.

Chivian D et al. Environmental genomics reveals a single-species ecosystem deep within Earth. *Science*. 2008 Oct 10;322(5899):275-8.

Chovatia M et al. Complete genome sequence of *Thermanaerovibrio acidaminovorans* type strain (Su883T). *Standards in Genomic Sciences.* 2009:1(3):254-261.

Chun J et al. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci.* 2009 Sep 8;106(36)15442:15447.

Clum A et al. Complete genome sequence of *Acidimicrobium ferrooxidans* type strain (ICPT). *Standards in Genomic Sciences.* 2009:1(1):38-45.

Clum A et al. Complete genome sequence of *Pirellula staleyi* type strain (ATCC 27377T). *Standards in Genomic Sciences.* 2009:1(3):308-316.

Coleman JJ et al. The genome of *Nectria haematococca:* contribution of supernumerary chromosomes to gene expansion. *PLoS Genet.* 2009 Aug;5(8):e1000618.

Copeland A et al. Complete genome sequence of *Atopobium parvulum* type strain (IPP 1246T). *Standards in Genomic Sciences.* 2009:1(2):166-173.

Copeland A et al. Complete genome sequence of *Catenulispora acidiphila* type strain (ID 139908T). *Standards in Genomic Sciences.* 2009:1(2):119-125.

Copeland A et al. Complete genome sequence of *Desulfomicrobium baculatum* type strain (XT). *Standards in Genomic Sciences.* 2009:1(1):29-37.

Cordeddu V et al. Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nat Gen.* 2009 Sep; 41(9): 1022-U95.

Coyne RS et al. Refined annotation and assembly of the *Tetrahymena thermophila* genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. *BMC Genomics.* 2008 Nov 26;9:562.

De Val S et al. Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. *Cell*. 2008 Dec 12;135(6):1053-64.

Drainas C et al. Characterization and whole genome sequencing of *Arthrobacter phenanthrenivorans*, a new phenanthrene degrading bacterium. *FEBS J.* 2009 Jul;276(Suppl. 1):101-101.

Dubchak I et al. Multiple whole-genome alignments without a reference organism. *Genome Res.* 2009 Apr;19(4):682-9.

Duncan KE. Biocorrosive thermophilic microbial communities in Alaskan North Oil Slope facilities. *Environ Sci Technol.* 2009 Oct 15;43(20):7977-7984.

Engelbrektson A et al. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J.* 2010 Jan. 21.

Engelbrektson A et al. Probiotics to minimize the disruption of faecal microbiota in healthy subjects undergoing antibiotic therapy. *J Med Microbiol.* 2009 May;58(Pt 5):663-70.

Field D et al. Meeting reports from the Genomic Standards Consortium

(GSC) Workshops 6 and 7. *Standards in Genomic Sciences.* 2009:1(1):21-28.

Foster JT et al. Whole-genome-based phylogeny and divergence of the genus *Brucella*. *J Bacteriol.* 2009 Apr;191(8):2864-70.

Gray J et al. A recommendation for naming transcription factor proteins in the grasses. *Plant Physiol.* 2009 Jan;149(1):4-6.

Gomez LD et al. Analysis of saccharification in *Brachypodium distachyon stems* under mild conditions of hydrolysis. *Biotechnol Biofuels.* 2008 Oct 22;1(1):15.

Grimson A et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*. 2008 Oct 30;455(7217):1193-7.

Han C et al. Complete genome sequence of *Pedobacter heparinus* type strain (HIM 762-3T). *Standards in Genomic Sciences.* 2009:1(1):54-62.

Han C et al. Complete genome sequence of *Kangiella koreensis* type strain (SW-125T). *Standards in Genomic Sciences.* 2009:1(3):226-233.

Harris DR et al. Directed evolution of ionizing radiation resistance in *Escherichia coli*. *J Bacteriol.* 2009 Aug 19;1(16):5240-52.

Herlemann DP et al. Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly termite group 1). *Appl Environ Microbiol.* 2009 May;75(9):2841-9.

Hill KK et al. Recombination and insertion events involving the botulinum neurotoxin complex genes in *Clostridium botulinum* types A, B, E, and F and *Clostridium butyricum* type E strains. *BMC Biol.* 2009 Oct 5;7:66.

Hooper SD et al. Microbial co-habitation and lateral gene transfer: what transposases can tell us. *Genome Biol.* 2009;10(4):R45.

Hooper SD et al. Integration of phenotypic metadata and protein similarity in Archaea using a spectral bipartitioning approach. *Nucleic Acids Res.* 2009 Apr;37(7):2096-104.

Hooper S et al. Estimating DNA coverage and abundance in metagenomes using a gamma approximation. *Bioinformatics.* 2009:1(1):63-67.

Horton A et al. Conservation of linkage and evolution of developmental function within the Tbx2/3/4/5 subfamily of T-box genes: implications for the origin of vertebrate limbs. *Dev Genes Evol.* 2008 Dec;218(11-12):613-28.

Hugenholtz P et al. Focus: *Synergistetes*. *Environ Microbiol.* 2009 Jun;11(6):1327-9.

Hugenholtz P et al. A changing of the guard. *Environ Microbiol.* 2009 Mar;11(3):551-3.

Ivanova N et al. Complete genome sequence of *Sanguibacter keddieii* type strain (ST-74T). *Standards in Genomic Sciences.* 2009:1(2):110-118.

Ivanova N et al. Complete genome sequence of *Leptotrichia buccalis* type strain (C-1013-bT). *Standards in Genomic Sciences.* 2009:1(2):126-132.

Ivanova N et al. Spreading of aqueous solutions of trisiloxanes and conventional surfactants over PTFE AF coated silicone wafers. *Langmuir.* 2009 Apr 9;25(6):3564-70.

Isabel S et al. Divergence among genes encoding the elongation factor Tu of *Yersinia* species. *J Bacteriol.* 2008 Nov;190(22):7548-58.

Jiang Z et al. Old can be new again: HAPPY whole genome sequencing, mapping and assembly. *Int J Biol Sci.* 2009;5(4):298-303. Review.

Chun J et al. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *PNAS.* 2009 Sep 8;106(36):15442-7.

Kataeva IA et al. Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. *J Bacteriol.* 2009 Jun;191(11):3760-1.

Kawashima T et al. Domain shuffling and the evolution of vertebrates. *Genome Res.* 2009 Aug;19(8):1393-403.

Kazakov AE et al. Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria. *J Bacteriol.* 2009 Jan;191(1):52-64.

Keim P et al. The genome and variation of *Bacillus anthracis*. *Mol Aspects Med.* 2009 Sep 1. [Epub ahead of print]

Kislyuk A et al. Frameshift detection in prokaryotic genomic sequences. *Int J Bioinform Res Appl.* 2009;5(4):458-77.

Klein MG et al. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J Mol Biol.* 2009 Sep 18;329(2):319-33.

Kouvelis VN et al. Complete genome sequence of ethanol producer *Zymomonas mobilis* NCIMB 11163. *J Bacteriol.* 2009 Sep 18.

Kristiansson E et al. ShotgunFunctionalizeR: An R-package for functional comparison of metagenomes. *Bioinformatics.* 2009 Aug 20.

Kunin V et al. Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. *Environ Microbiol.* 2009 Aug 27.

Kunin V et al. A bioinformatician's guide to metagenomics. *Microbiol Mol Biol Rev.* 2008 Dec;72(4):557-78.

Kyrpides NC. Fifteen years of microbial genomics: meeting the challenges and fulfilling the dream. *Nat Biotechnol.* 2009 Jul;27(7):627-32.

LaButti K et al. Complete genome sequence of *Anaerococcus prevotii* type strain (PC1T). *Standards in Genomic Sciences.* 2009:1(2):159-165.

Land M et al. Complete genome sequence of *Beutenbergia cavernae* type strain (HKI 0122T). *Standards in Genomic Sciences.* 2009:1(1):21-28.

Land M et al. Complete genome sequence of *Actinosynnema mirium* type strain (101T). *Standards in Genomic Sciences.* 2009:1(1):46-53.

Lane TW et al. Digital transcriptomic analysis of silicate starvation induced triacylglycerol formation in the marine diatom *Thalassiosira pseudonana. Phycologia.* 2009 Jul; 48(4):71-71. Suppl. S.

Lang E et al. Complete genome sequence of *Dyadobacter fermentans* type strain (NS114T). *Standards in Genomic Sciences.* 2009:1 (2):133-140.

Lapidus A et al. Complete genome sequence of *Brachybacterium faecium* type strain (Schefferle 6-10T). *Standards in Genomic Sciences.* 2009:1(1):3-11.

Larsson P et al. Molecular evolutionary consequences of niche restriction in *Francisella tularensis*, a facultative intracellular pathogen. *PLoS Pathog.* 2009 Jun;5(6):e1000472.

Lauro FM et al. A tale of two lifestyles: the genomic basis of trophic strategy in bacteria. *Proc Natl Acad Sci* U S A. 2009 Sep 15;106(37):15527-15533.

Le Crom S et al. Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. *Proc Natl Acad Sci U S A.* 2009 Sep 22; 106(38): 16151-16156.

Leary RJ et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc Natl Acad Sci U S A.* 2008 Oct 21;105(42):16224-9.

Lin-Jones J et al. Identification and localization of myosin superfamily members in fish retina and retinal pigmented epithelium. *J Comp Neurol.* 2009 Mar 10;513(2):209-23.

Liolios K et al. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 2009 Nov 13;19(4):682-9.

Macey JR et al. Socotra Island the forgotten fragment of Gondwana: Unmasking chameleon lizard history with complete mitochondrial genomic data. *Mol Phylogenet Evol.* 2008 Dec;49(3):1015-8.

Malfatti S et al. Complete genome sequence of *Halogeometricum borinquense* type strain (PR3T). *Standards in Genomic Sciences.* 2009:1(2):150-158.

Markowitz VM et al. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics.* 2009 Sep 1;25(17):2271-8.

Markowitz VM et al. The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res.* 2009 Oct 28;25(17):2271-8.

Martin F et al. The long hard road to a completed *Glomus intraradices* genome. *New Phytol.* 2008 Nov 6;180(4):747-50.

Martinez D et al. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci U S A.* 2009 Feb 10;106(6):1954-9.

Masta SE et al. Arachnid relationships based on mitochondrial genomes: Asymmetric nucleotide and amino acid bias affects phylogenetic analyses. *Mol Phylogenet Evol.* 2009 Jan;50(1):117-28.

Mavrommatis K et al. Complete genome sequence of *Cryptobacterium curtum* type strain (12-3T). *Standards in Genomic Sciences.* 2009:1(2):93-100.

Mavrommatis K et al. Complete

genome sequence of *Capnocytophaga ochracea* type strain (VPI 2845T). *Standards in Genomic Sciences.* 2009:1(2):101-109.

Mavrommatis K et al. Standard operating procedure for the annotations of microbial genomes by the Production Genomic Facility of the DOE JGI. *Standards in Genomic Sciences.* 2009:1(1):63-67.

Mavrommatis K et al. Genome analysis of the anaerobic thermohalophilic bacterium *Halothermothrix orenii. PLoS One.* 2009;4(1):e4192.

Mavrommatis K et al. Gene context data analysis in the Integrated Microbial Genomes (IMG) Data Management System. *PLoS ONE.* 2009 Nov 24:4(11):e7979.

McBride MJ et al. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* revealed by genome sequence analysis. *Appl Environ Microbiol.* 2009 Aug 28.

McMurdie PJ et al. Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides. PLoS Genet.* 2009 Nov;5(11):e1000714.

Merritt WM et al. Dicer, Drosha, and outcomes in patients with ovarian cancer. *N Engl J Med.* 2008 Dec 18;359(25):2641-50.

Munk C et al. Complete genome sequence of *Stackebrandtia nassauensis* type strain (LLR-40K-21T). *Standards in Genomic Sciences.* 2009:1(3):292-299.

Nolan M et al. Complete genome sequence of *Rhodothermus marinus* type strain (R-10T). *Standards in Genomic Sciences.* 2009:1(3):283-291.

Nolan M et al. Complete genome sequence of *Streptobacillus moniliformis* type strain (9901T). *Standards in Genomic Sciences.* 2009:1(3):300-307.

Novichkov PS et al. Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. *J Bacteriol.* 2009 Jan;191(1):65-73.

Novichkov PS et al. ATGC: a database of orthologous genes from closely related prokaryotic genomes and a research platform for microevolution of prokaryotes. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D448-54.

Novichkov PS et al. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.* 2009 Nov 1;191(1):65-73.

Oda Y et al. Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. *Proc Natl Acad Sci U S A.* 2008 Nov 25;105(47):18543-8.

O'Rourke JA et al. Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genomics.* 2009 Aug 13;10(376).

Panisko EA et al. "Genomes to proteomes (book chapter)." In *Automation in Proteomics and Genomics: An Engineering Case-Based Approach* by Wiley. 2009. pp. 21-29.

Paterson AH et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature.* 2009 Jan 29;457(7229):551-6.

Pati A et al. Complete genome sequence of *Saccharomonospora viridis* type strain (P101T). *Standards in Genomic Sciences.* 2009:1(2):141-149.

Pavlopoulos GA et al. jClust: a clustering and visualization toolbox. *Bioinformatics.* 2009 Aug 1;25(15):1994-6.

Pearson T et al. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol.* 2009 Nov 18;7:78.

Penn K et al. Genomic islands link secondary metabolism to functional adaptation in marine *Actinobacteria. ISME J.* 2009 May 28.

Perez G et al. Telomere organization in the ligninolytic *Basidiomycete Pleurotus ostreatus. Appl Environ Microbiol.* 2009 Mar;75(5):1427-36.

Peterson SB et al. Environmental distribution and population biology of *Candidatus Accumulibacter*, a primary agent of biological phosphorus removal. *Environ Microbiol.* 2008 Oct;10(10):2692-703.

Pierce E et al. The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ Microbiol.* 2008 Oct;10(10):2550-73.

Podar M et al. A genomic analysis of the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans. Genome Biol.* 2008;9(11):R158.

Pukall R et al. Complete genome sequence of *Slackia heliotrinireducens* type strain (RSH 1T). *Standards in Genomic Sciences.* 2009:1(3):234-241.

Pukall R et al. Complete genome sequence of *Jonesia denitrificans* type

strain (Prevot 55134T). *Standards in Genomic Sciences.* 2009:1(3):262-269.

Rahimov F et al. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet.* 2008 Nov;40(11):1341-7.

Ratnere I et al. Obtaining comparative genomic data with the VISTA family of computational tools. *Curr Protoc Bioinformatics.* 2009 Jun;Chapter 10:Unit 10.6.

Rodrigues DF et al. Architecture of thermal adaptation in an *Exiguobacterium sibiricum* strain isolated from 3 million year old permafrost: A genome and transcriptome approach. *BMC Genomics.* 2008 Nov 18;9:547.

Romeo S et al. Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *J Clin Invest.* 2009 Jan;119(1):70-9.

Romeo S et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet.* 2008 Dec;40(12):1461-5.

Salinero KK et al. Metabolic analysis of the soil microbe *Dechloromonas aromatica* str. RCB: indications of a surprisingly complex life-style and cryptic anaerobic pathways for aromatic degradation. *BMC Genomics.* 2009 Aug 3;10:351.

Satou Y, Mineta K, Ogasawara M, et al. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis:* new insight into intron and operon populations. *Genome Biol.* 2008 Oct 14;9(10):R152. [Epub ahead of print]

Saunders E et al. Complete genome sequence of *Eggerthella lenta* type strain (IPP VPI 0255T). *Standards in Genomic Sciences.* 2009:1(2):174-182.

Schubbe S et al. Complete genome sequence of the chemolithoautotrophic marine *Magnetotactic Coccus* strain MC-1. *Appl Environ Microbiol.* 2009 Jul;75(14):4835-52.

Schmutz J et al. Genome sequence of the palaeopolyploid soybean. *Nature.* 2010 Jan 14;463(7278):178-83.

Sela DA et al. The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A.* 2008 Dec 2;105(48):18964-9.

Seidl V et al. Transcriptomic response of the mycoparasitic fungus *Trichoderma atroviride* to the presence of a fungal prey. *BMC Genomics.* 2009 Nov 30;10:567.

Silby MW et al. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens.* *Genome Biol.* 2009;10(5):R51.

Sims D et al. Complete genome sequence of *Kytococcus sedentarius* type strain (strain 541T). *Standards in Genomic Sciences.* 2009:1(1):12-20.

Slavotinek AM et al. Sequence variants in the HLX gene at chromosome 1q41-1q42 in patients with diaphragmatic hernia. *Clin Genet.* 2009 May;75(5):429-39.

Smith DR et al. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res.* 2008 Oct;18(10):1638-42.

Spring S et al. Complete genome sequence of *Desulfotomaculum acetoxidans* type strain (5575T). *Standards in Genomic Sciences.* 2009:1(3):242-253.

Theofilopoulos S et al. Novel function of the human presqualene diphosphate phosphatase as a type II phosphatidate phosphatase in phosphatidylcholine and triacylglyceride biosynthesis pathways. *Biochim Biophys Acta.* 2008 Nov-Dec;1781(11-12):731-42.

Tindall BJ et al. Complete genome sequence of *Halomicrobium mukohataei* type strain (arg-2T). *Standards in Genomic Sciences.* 2009:1(3):270-277.

Tringe SG et al. A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol.* 2008 Oct;11(5):442-6. Review.

Tschöp MH et al. Getting to the core of the gut microbiome. *Nat Biotechnol.* 2009 Apr;27(4):344-6.

Tsiamis G et al. Prokaryotic community profiles at different operational stages of a Greek solar saltern. *Res Microbiol.* 2008 Nov-Dec;159(9-10):609-27.

Valdés J et al. *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics.* 2008 Dec 11;9:597.

Van der Auwera GA et al. Plasmids captured in *C. metallidurans CH34:* defining the PromA family of broad-host-range plasmids. *Antonie Van Leeuwenhoek.* 2009 Aug;96(2):193-204.

Visel A et al. Genomic views of distant-acting enhancers. *Nature.* 2009 Sep 10;461:199:205.

Visel A et al. Functional autonomy of

distant-acting human enhancers. *Genomics*. 2009 Jun;93(6):509-13.

Visel A et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009 Feb 12;457(7231):854-8.

Vogel CJ et al. *Brachypodium distachyon:* a new model for biomass crops *in vitro. Cell Dev Biol Anim. Spr* 2009; 45: S6-S6

Wagner-Dobler I et al. The complete genome sequence of the algal symbiont *Dinoroseobacter shibae:* a hitchhiker's guide to life in the sea. *ISME J.* 2009 Sep 10.

Walsh DA et al. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science*. 2009 Oct 23;326(5952):578-82.

Wang S et al. Assembly of 500,000 interspecific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biol.* 2010 Jan 22;11(1):R8.

Ward NL et al. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol.* 2009 Apr;75(7):2046-56.

Warnecke F et al. A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J Biotechnol.* 2009 Jun 1;142(1):91-5.

Wattam AR et al. Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. *J Bacteriol.* 2009 Jun;191(11):3569-79.

Worden AZ et al. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*. 2009 Apr 10;324(5924):268-72.

Woyke T et al. Assembling the marine metagenome, one cell at a time. *PLoS One.* 2009;4(4):e5299. Epub 2009 Apr 23.

Wright JC et al. Exploiting proteomic data for genome annotation and gene model validation in *Aspergillus niger. BMC Genomics.* 2009 Feb 4;10:61.

Wrighton KC et al. A novel ecological role of the Firmicutes identified in thermophilic microbial fuel cells. *ISME J.* 2008 Nov;2(11):1146-56.

Wu D et al. Complete genome sequence of the aerobic CO-oxidizing thermophile *Thermomicrobium roseum. PLoS One.* 2009;4(1):e4207.

Wu M et al. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 2008 Oct 13;9(10):R151.

Wu D. A phylogeny-driven genomic encyclopedia of bacteria and archaea. *Nature*. 2009 Dec 24;462(7276):1056-60.

Yagi JM et al. The genome of *Polaromonas naphthalenivorans* strain CJ2, isolated from coal tar-contaminated sediment, reveals physiological and metabolic versatility and evolution through extensive horizontal gene transfer. *Environ Microbiol.* 2009 Sep;11(9): 2253-2270.

Yan HH et al. Intergenic locations of rice centromeric chromatin. *PLoS Biol.* 2008 Nov 25;6(11):e286.

Yoder-Himes DR et al. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci U S A.* 2009 Mar 10;106(10):3976-81.

Young M et al. Genome sequence of the Fleming strain of *Micrococcus luteus,* a simple free-living actinobacterium. *J Bacteriol.* 2009 Nov 30:1(1):63-67.

Zhang Z et al. Massively parallel sequencing identifies the gene Megf8 with ENU-induced mutation causing heterotaxy. *Proc Natl Acad Sci U S A.* 2009 Mar 3;106(9):3219-24.

Zhou K et al. Proteomics for validation of automated gene model predictions. *Methods Mol Biol.* 2008 Oct 24;492:447-52.

For more information about DOE JGI, contact:

David Gilbert, Public Affairs Manager
DOE Joint Genome Institute
2800 Mitchell Drive / Walnut Creek, CA 94598

e-mail: degilbert@lbl.gov
phone: (925) 296-5643

DOE JGI Web site:
http://www.jgi.doe.gov/

Published by the Berkeley Lab Creative Services Office in collaboration with DOE JGI researchers and staff for the U.S. Department of Energy. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract number DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract number DE-AC52-07NA27344, Los Alamos National Laboratory under contract number DE-AC02-06NA25396 and HudsonAlpha Institute for Biotechnology under DOE grant number DE-FG02-09ER64726.

CSO JO18476