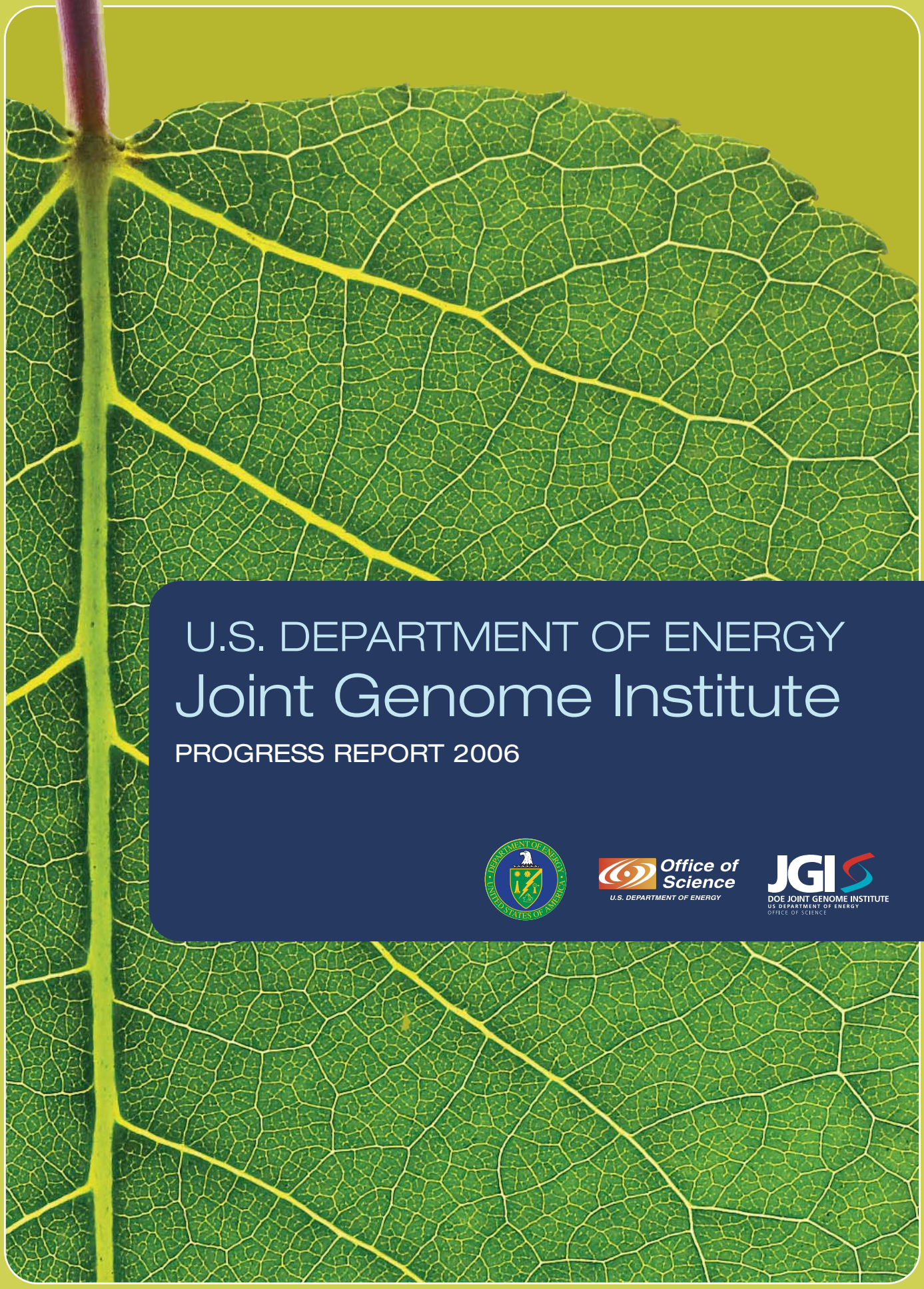


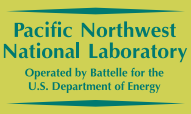
U.S. DEPARTMENT OF ENERGY
Joint Genome Institute
PROGRESS REPORT 2006

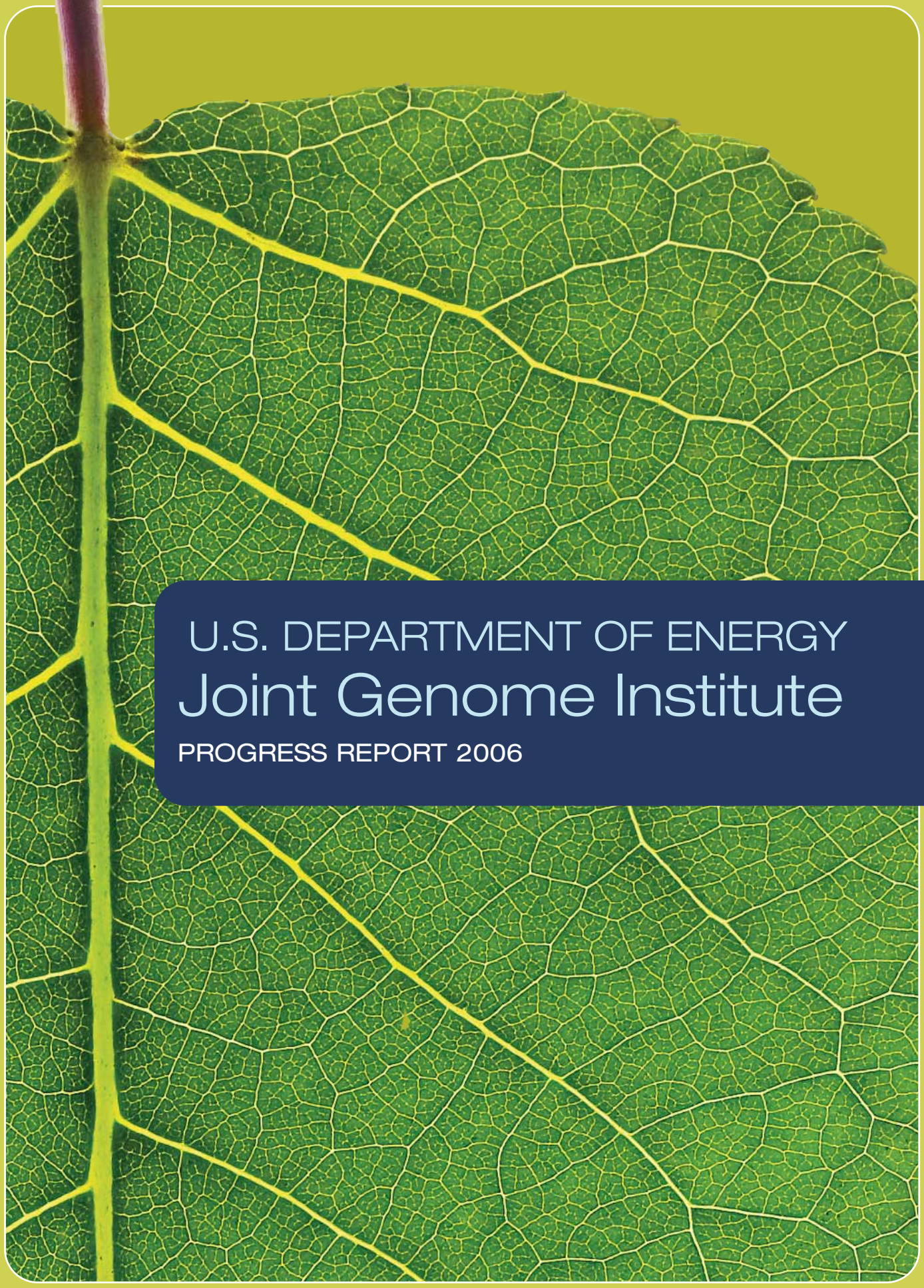
Office of Science
U.S. DEPARTMENT OF ENERGY

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

JGI's Mission

The U.S. Department of Energy Joint Genome Institute (JGI), supported by the DOE Office of Science, is focused on the application of Genomic Sciences to support the DOE mission areas of clean energy generation, global carbon management, and environmental characterization and clean-up. JGI's Production Genomics Facility in Walnut Creek, California, provides integrated high-throughput sequencing and computational analysis that enable systems-based scientific approaches to these challenges. In addition, the Institute engages both technical and scientific partners at five national laboratories, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, and Pacific Northwest, along with the Stanford Human Genome Center.





U.S. DEPARTMENT OF ENERGY

Joint Genome Institute

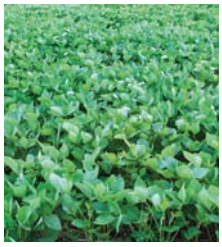
PROGRESS REPORT 2006

table of contents

Director's Perspective

By nearly all metrics, the last twelve months have been an extremely productive time for the DOE Joint Genome Institute (JGI). During this period, we completed the sequencing of the first tree genome, *Populus trichocarpa* (poplar). The sequencing and analysis of this bioenergy feedstock was highlighted by an article in the September 15, 2006, issue of *Science* (as well as the journal's cover). The publication was the result of a four-year scientific effort led by JGI and Oak Ridge National Laboratory, and represents a major step toward establishing the JGI as a leader in the sequencing and characterization of bioenergy feedstocks.

As reflected by the poplar study, the JGI is increasingly focused in the area of alternative energy production through the sequencing of relevant crops and microorganisms. Why do we believe that this emphasis on bioenergy genomics will contribute to the development of alternative fuels? It is, in part, based on observing the impact that the availability of the human genome has had on biomedical research. Nearly every



biomedical investigator's research, and most drugs presently in development, have been enabled or accelerated by scientists having access to the human genome sequence. The availability of key bioenergy-relevant genomes is already accelerating research in this field, supporting our confidence that genomics will contribute to the technologies needed to reduce the time until alternative fuels begin to play a significant role in meeting our energy needs.

In addition to poplar, the JGI has many other potential bioenergy feedstock plants as sequencing targets, including soybean, switchgrass, sorghum, maize, and cassava. Coupled with these feedstocks are a number of relevant microorganisms also targeted for sequencing by the JGI. Many of these microorganisms have been selected for sequencing because of their ability to break down cellulose and facilitate fermentation, two crucial biochemical processes that need to be optimized if cellulosic ethanol production is to contribute significantly to our fuel supply. The majority of these "bioenergy" projects have been proposed by outside scientists and approved for sequencing through the various JGI user programs.

Over the last year, in addition to increased efforts in bioenergy research, the JGI has continued genomic investigations of organisms involved in carbon cycling and bioremediation. Many of these studies have involved metagenomic approaches, a new genomic strategy which involves the sequencing of microbial communities as they are found in their natural environments without having to grow them in the laboratory. Metagenomics represents an emerging field of genomics in which the JGI is playing a leadership role. Our metagenomic projects have analyzed a variety of environments ranging from sewage slurry to termite hindguts. In the latter study, JGI microbial ecologists and their collaborators at Caltech are beginning to shed light on dozens of genes encoding novel cellulases and other enzymes involved in cellulose breakdown present in the microorganisms populating the insect's hindgut. Some of the gene products identified in the termite study are being explored for their potential to improve industrial cellulose processing.

Crucial to the exploitation of the JGI's high-throughput sequence is the development of user-friendly informatic tools enabling the exploration of data generated by the JGI. Integrated Microbial Genomes (IMG) and IMG Microbiome (IMG/M) are powerful genomic analysis tools developed by the JGI to encourage the comparative analysis of microbial and metagenomics data by both new and experienced users of the JGI. Just as the availability of the human genome's set of genes facilitated the diagnosis of clinical disorders and the development of new drugs, our ability to recognize the compendium of genes evolved in nature to capture carbon from the atmosphere, degrade biomass to fermentable sugars, and detoxify hazardous substances will facilitate harnessing these important proccesses in the national interest.

The developments of the past year, the majority of them led by participants in JGI user programs, point to the maturation of the JGI in its role as a genomic user facility, focusing on areas of science relevant to DOE's mission. By providing state-of-the-art sequencing and DNA analysis resources to a broad community of scientists, creative approaches to important societal issues are being generated. The continued involvement of a large and engaged user community will maximize the impact of the Institute as it moves forward in contributing to alternative fuels, reduced greenhouse gases, and a cleaner environment.

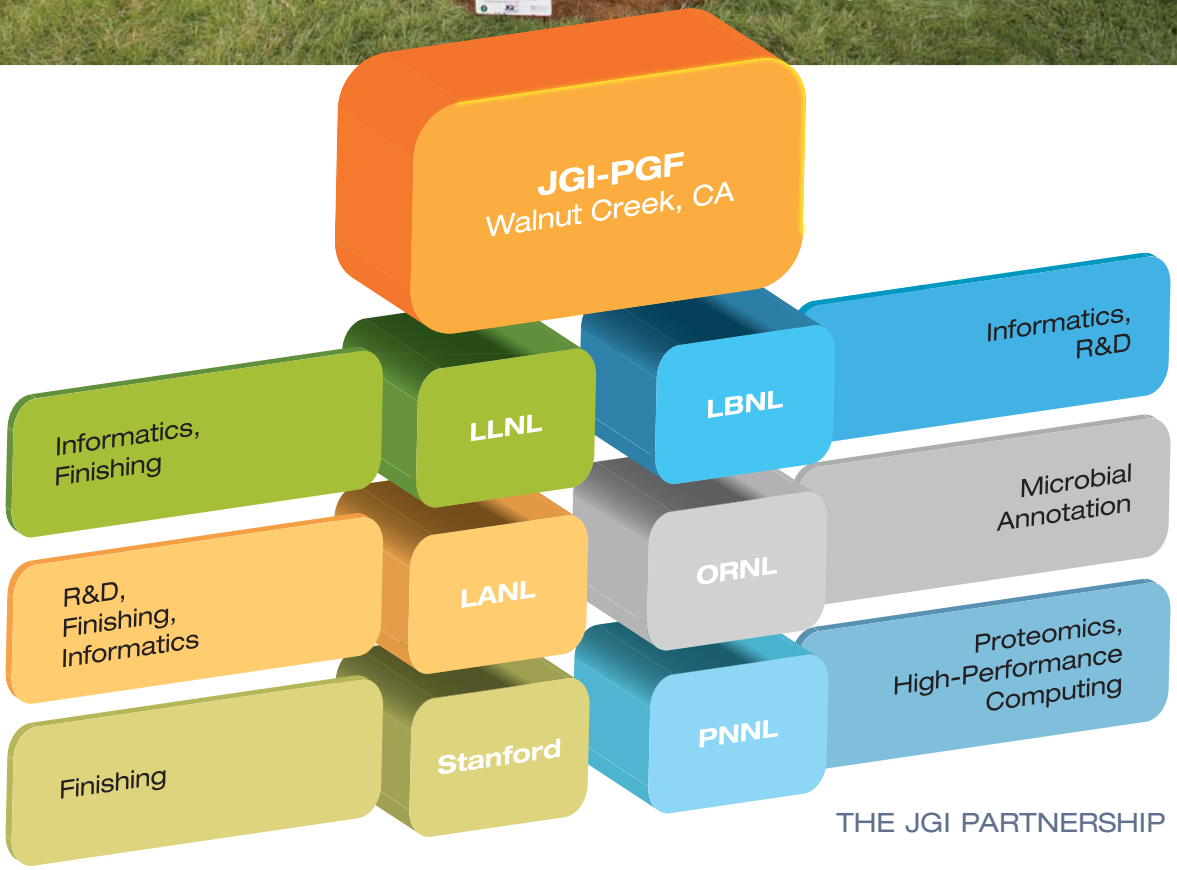
Edward M. Rubin, MD, PhD
Director
DOE Joint Genome Institute

JGI History

The U.S. Department of Energy Joint Genome Institute (JGI) was created in 1997 to unite the expertise and resources in genome mapping, DNA sequencing, technology development, and information sciences pioneered at the DOE genome centers at Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), and Los Alamos National Laboratory (LANL). In 1999, the University of California, which manages the three national labs for the DOE, leased 60,000 square feet of laboratory and office space in a light industrial park in Walnut Creek, California, to consolidate activities and accommodate JGI's 240 employees in what is now known as the Production Genomics Facility (PGF). A partnership with the Stanford University Human Genome Center serves JGI's goal of providing high-quality finished sequence to the greater scientific community. In 2003, with the successful completion of the human genome, JGI reoriented its focus to becoming a user facility driven by DOE mission needs in the areas of energy production, carbon management, and bioremediation. This transition has benefited from broader involvement of the DOE national laboratory system, including the formal participation of Oak Ridge National Laboratory (ORNL) and Pacific Northwest National Laboratory (PPNL) in the activities of JGI. The importance of the JGI as essential infrastructure for DOE science was codified through signing of a new MOU between the five partner labs in 2006.

For more information about JGI's history, facility, partners, and budget, visit our Web site at http://www.jgi.doe.gov/whoweare/index.html.

The poplar genome planting ceremony was held September 14, 2006, in celebration of the publication of the poplar genome in the journal *Science*.



THE JGI PARTNERSHIP

JGI Departments and Programs

Sequencing Department, Lead: Susan Lucas

The JGI Sequencing Department resides at the heart of the JGI Production Genomics Facility. The department generates high-quality sequence in a cost-efficient manner, expediting DNA through the process from library creation and sequencing preparation, to capillary sequencing and analysis. As genomics is a rapidly changing field, the department constantly adapts to take advantage of new technological developments that have substantially increased throughput while decreasing cost. The JGI Sequencing Department comprises several subgroups, including Mapping, Cloning Technologies, Library Support, Sequencing Preparation, Quality Control, Sequence Assessment and Analysis, Process Optimization, and Instrumentation. In FY06, the department produced 49 million sequencing reads and 33 billion base pairs of sequence. In FY07, the department will produce 36 billion basepairs.

Informatics Department, Lead: Darren Platt

The Informatics Department manages the tracking, assembly, analysis, and distribution of an ever-increasing stream of DNA sequence coming through the facility. There has been a steady increase over the past year in both volume of sequence and number of projects. To meet these challenges, there has been an ongoing investment in laboratory tracking systems, project management tools, better DNA assembly algorithms, annotation pipelines, and portals for distributing results. Recognizing the need to help the community deal with this tremendous volume of data, the Integrated Microbial Genomics (IMG) project, now in its seventh major release, holds the genomes of over 750 organisms, allowing researchers to make quick comparisons between whole genomes. In 2006, JGI finished 82 prokaryotic genomes representing 311 megabases of finished sequence, bringing the total to over 350 draft and finished microbial genomes. The Informatics Department released more than 25 new eukaryotic genomes to collaborators and the broader genomic community through JGI portals. Supporting this ever-growing volume of data, a new computing facility was brought online this year. Combining over 400 CPUs and 150 Tb of storage, computation involving tens of thousands of CPU days is routinely performed. The Informatics Department comprises several subgroups, including Production Informatics, Assembly, Comparative Genomics, Software Support, IT, Genome Data Systems, and Genome Annotation, and brings together a diverse set of computational skills to serve the JGI mission.

Computational Genomics Program, Lead: Dan Rokhsar

The Computational Genomics Program develops new analytical tools and data management systems that transform the raw data produced by JGI into biologically valuable information and insights. These tools are designed to facilitate the use of JGI data by the biological community. This work is essential for managing and visualizing the expanding body of genome-scale data and linking it to functional and phenotypic information generated at JGI and elsewhere. A major focus of the Computational Genomics group is to work with communities interested in the genomes of the different organisms sequenced by the JGI to bring these completed genomes to publication.

Genomics Technology Program, Lead: Paul Richardson

The Genomic Technologies Program works to make the sequencing and assembly process at JGI more efficient and to improve the quality of genomes produced. It accomplishes this by developing new protocols and evaluating new methods and instruments for use in production. Its efforts improve the overall quality of genomes produced at JGI and continually increase the Institute's capabilities. Notable initiatives include developing methods to improve read lengths, decrease reagent costs, and quickly close gaps in sequence data. This year, the group transferred a number of protocols to production, improving efficiency, expanding capabilities, and decreasing costs. One of the major accomplishments was the installation, testing, and

adoption of a new type of sequencing technology that differs from all previous Sanger-based methods. Sequencing by synthesis technology uses a massively parallel sequencing strategy that results in roughly 300,000 reads per run (~4.5 hours), resulting in 10–20 fold increases in capacity over current methods. This new technology is currently used for all microbial sequencing, decreasing the time required to generate a finished genome. The development of creative ways to use these new sequencing technologies will be an increasingly important focus of the Genomics Technology Group.

Genetic Analysis Program, Lead: Len Pennacchio

The Genetic Analysis Program is a scientific research group focused on making effective use of JGI generated sequence. One research area aims to identify sequence variants that may contribute to observed trait differences within a given species through the use of resequencing and genotyping. For instance, studies are ongoing to identify sequence variants that explain trait differences such as disease resistance within poplar trees *(Populus trichocarpa)*. In addition, the group explores the biological applications of new sequencing technologies (such as 454 and Solexa) that are capable of generating large datasets of short DNA sequence reads. Our goal is to guide JGI user programs by providing state-of-the-art examples of the utility of these brand new cutting-edge technologies.

The group also aids JGI users and internal programs in seamlessly accessing the JGI Production Sequencing Line. This is accomplished through integrated efforts with JGI's Laboratory Sequencing Program (LSP) and the Community Sequencing Program (CSP). This has resulted in facilitating programmatic growth through the hosting of JGI Workshops centered around DOE missions such as developing genomic insights into low-dose radiation effects as well as the advancement of national bioenergy feedstocks.

Finally, the Genetic Analysis Program contributes, through the use of comparative sequence analysis, to identifying functional sequences in the genomes of chordates sequenced by the JGI, including human (*Homo sapiens*), frog (*Xenopus tropicalis*), amphioxus (*Branchiostoma floridae*), sea squirt (*Ciona intestinalis*), and pufferfish (*Fugu rubripes*). Through these efforts, we have catalogued a large number of evolutionarily conserved modules of which a significant fraction has been shown to act as transcriptional enhancers of gene expression. These findings have led to a systematic effort to catalog gene enhancers through a transgenic mouse assay. The data can be found at http://enhancer.lbl.gov.

Microbial Ecology Program, Lead: Phil Hugenholtz

To date, molecular microbial ecology has relied heavily on small subunit ribosomal RNA (rRNA) sequence for culture-independent characterization of microbial communities. The Microbial Ecology Program (MEP) uses sequence-based technologies to obtain a deeper understanding of microbial communities through a combination of computational and experimental methods.

MEP has established a high-throughput pipeline for analysis of rRNA signature sequences and fluorescence in-situ hybridization (FISH) to visualize phylogenetic groups under the microscope. Building on this work, MEP is pioneering the emerging field of metagenomics—cloning, sequencing, and characterizing DNA extracted directly from environmental samples—to obtain an overview of community function and population dynamics. Since environmental shotgun sequencing is in its infancy, MEP is exploring ways to analyze and visualize metagenomic data together with the JGI's Microbial Genome Biology Program.

Microbial Genome Biology Program, Lead: Nikos Kyrpides

The identification of the complete set of functions of any organism provides the foundation upon which our understanding of the biology of that organism rests. In essence, it forms the basic framework that any genome project targets, and from which any biological interpretation originates. However, while the quality and quantity of sequencing data has dramatically increased during the last few years, their interpretation remains a major bottleneck. In fact, as more and more microbes are sequenced, the scientific community's efforts to assign functions to genes are lagging. In addition, the importance of comparative analysis and extensive sequence integration for a comprehensive genome analysis and reconstruction of the functional cellular subsystems (e.g., metabolic pathways, information processes, etc.) has been largely overlooked by most contemporary genome databases. To speed up annotation, group members are developing software tools for determining microbial gene function.

The Genome Biology Program accomplishments include the development of the Integrated Microbial Genome (IMG) system for interpreting newly sequenced genomes,

and for analysis of existing genomic sequence data on a comparative level. In addition, the group has developed the IMG/M system, which integrates microbial community (microbiome) genome data with IMG's isolate microbial genomes, and provides support for the comparative analysis of the aggregate microbiome genomes (metagenomes).

User Support Department, Lead: Jim Bristow

The User Support Department is charged with facilitating user interactions with the JGI. This group manages the application processes and peer review for the Community Sequencing, Laboratory Sequencing, and Microbial Genome programs. Once proposals are approved for sequencing, the department has responsibility for all aspects of project management, including project planning and initiation, communication of progress or problems with collaborators, coordination of data analysis, and project closeout. The Project Management Office is comprised of managers from LANL and the PGF.

The User Support Department also coordinates the annual user's meeting, which this year will bring more than 400 of JGI's users and potential users together with JGI staff to hear state-of-the-art talks on all aspects of sequence-based science as well as tutorials on JGI processes, user interfaces, and genomic tools such as the Integrated Microbial Genome tools.

Finally, the User Support Department serves as a conduit for feedback from users to the JGI through periodic user surveys, end-of-project questionnaires, and interaction with the Users Executive Committee.

Education Department, Lead: Cheryl Kerfeld

The JGI's recently created Education Department was conceived to develop educational programs and tools, centered on large-scale DNA sequencing and its bioinformatic analysis. It is expected that these programs and tools will be useful to a range of institution types, from research universities to community colleges. The products will also be adapted to high school outreach programs. Preliminary targets for the program are: (1) development of a genome annotation toolkit that will be used to train students to use search tools and databases sequenced specifically for this purpose by the JGI and (2) the development of online tools for undergraduate life sciences courses that leverage the existing Integrated Microbial Genome (IMG) and IMG/Microbiome (IMG/M) tools and databases. The JGI's Education Department will also help plan genome-based research projects at high schools and undergraduate institutions. The concept is to build a variety of genome-scale research projects specifically for students, and tailored to the existing curriculum and interests of educators.

Laboratory Science Program (LSP), Lead: Jerry Tuskan

The Laboratory Science Program (LSP), launched in 2006, is a new initiative from DOE to leverage JGI sequencing capacity, providing DOE national laboratory researchers with broader access to high-throughput DNA sequencing in support of mission-relevant projects.

The LSP will serve the national laboratories in two major ways. First, it will foster large-scale strategic sequencing projects, across the national laboratory system, that are aligned with future funding opportunities in DOE's biology programs. Second, it will provide small-scale sequencing that meets the needs of individual investigators at the national laboratories. The LSP is expected to use 15 to 20 percent of JGI's sequencing capacity, which is currently over 35 billion bases per year.

Sequencing to be carried out under the LSP will include genomes of entire microbial communities, and individual microbes and plants, useful in decreasing reliance on petroleum and petrochemicals by converting plant materials, such as soybeans, to "green" energy and chemical feedstocks. Sequencing will also focus on characterizing the variation in human susceptibility to nucleic acid damage by ionizing radiation.

LSP's lead, Gerald Tuskan, is a senior scientist in the Environmental Sciences Division at Oak Ridge National Laboratory. Tuskan will be responsible for developing, coordinating, and managing the LSP. For the list of LSP projects, see Appendix B.

JGI Users

First Annual JGI User Meeting

The first annual JGI User Meeting, March 29–April 1, attracted 274 participants to Walnut Creek to hear state-of-the-art, sequence-based science, including some of the most exciting JGI projects. The meeting also offered the opportunity to learn about new sequencing hardware, attend poster sessions, tour the Production Genomics Facility, and mingle with the JGI community.

After JGI Director Eddy Rubin kicked off the inaugural session, Penny Chisholm, Professor of Environmental Studies from the Massachusetts Institute of Technology (MIT), gave one of the keynote addresses, talking about the photosynthetic ocean microbe *Prochlorococcus* as a model for systems biology. She was followed by Drew Endy, Assistant Professor of Biological Engineering Division at MIT, discussing the new field of synthetic biology, a field that has been enabled by the availability of genome sequences.

Large eukaryotic genomics was the focus of Thursday morning's session, which included talks about the *Xenopus tropicalis* and *Chlamydomonas* genomes. Jerry Tuskan, from ORNL, described insights from the poplar genome and discussed how poplar research can be applied to both bioenergy and carbon cycling, aiding the understanding of global climate.

The focus of the afternoon of the first full-day session was microbial genomics, with talks ranging from what's being learned from the comparison of multiple genomes of *Shewanella* strains, to the role of mismatch repair in the evolution of the lactic acid bacterium *Oenococcus oeni*. Dan Cullen, from the USDA Forest Service, talked about filamentous white rot fungi that have the ability to degrade lignin present in plant cell walls. This is important because lignin is a major barrier to producing cellulosic ethanol on an industrial scale. The evening's poster session included a tour of the JGI Production Genomics Facility.

The second morning was devoted to microbial communities and metagenomics. JGI Microbial Ecology Program Head, Phil Hugenholtz, discussed his effort to use metagenomic datasets to provide genome-wide insights into the microbial population structure of waste-water treatment samples. Jared Leadbetter, from Caltech, talked about his journey to Costa Rica to collect termites for hindgut bacterial community sequencing, and this work's potential to reveal how termites degrade biomass into biofuels. Other talks covered the bacterial consortium in a gutless worm, green bacteria genomics, and surprises from superheated thermal aquifers.

The afternoon session focused on technology development, new sequencing platforms, and the integration of new technologies into the JGI microbial program. On Saturday morning, JGI Genome Biology Program Head Nikos Kyrpides hosted a popular Integrated Microbial Genomes (IMG) Workshop to discuss current and future developments in the JGI genome analysis pipeline.

Left to right: Luis Corrochano, Universidad de Sevilla; Ratnakar Deole, Oklahoma State University; Scott Baker, Pacific Northwest National Laboratory; Wouter Hoff, Oklahoma State University.

Thomas Schmidt, Michigan State University, and Dan Cullen, University of Wisconsin, Madison/USDA Forest Products Lab in Madison, Wisconsin.

Right, Kennan Kellaris, University of California, Berkeley

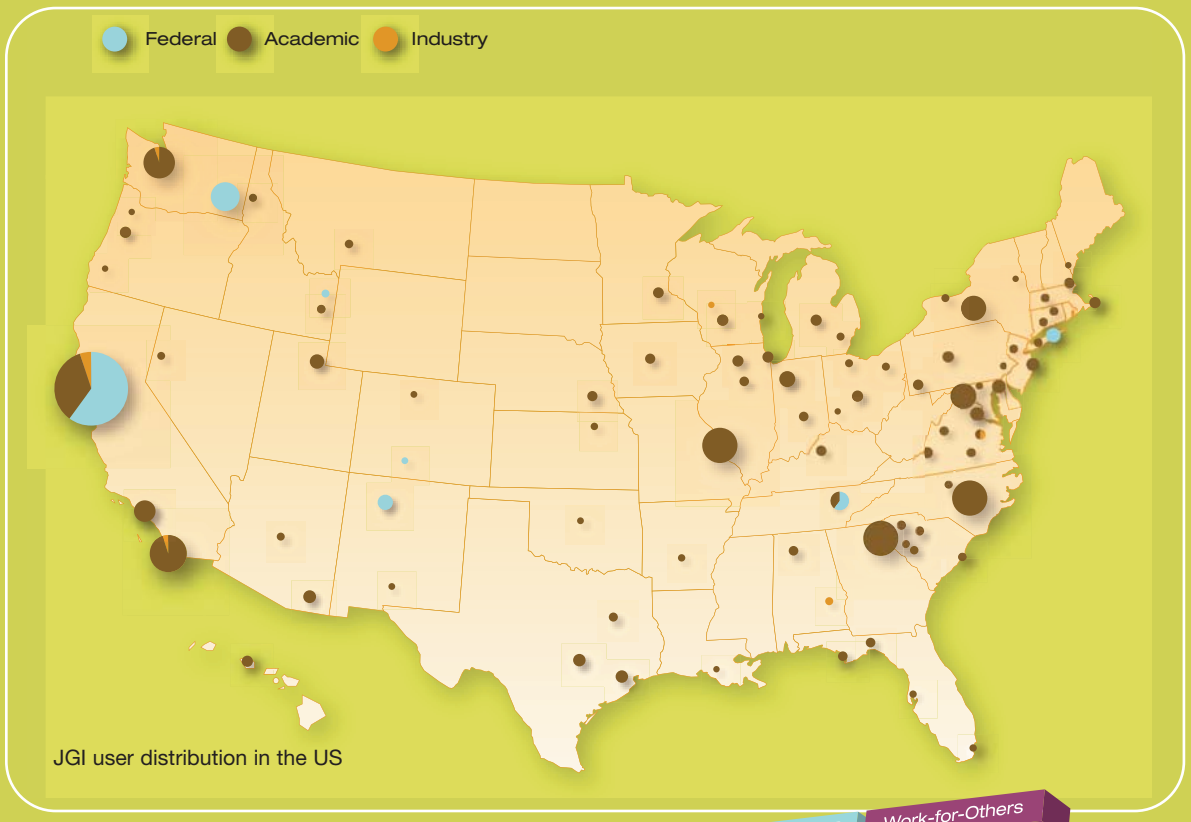
JGI USER STATISTICS

FY06 JGI Sequencing Capacity Allocation

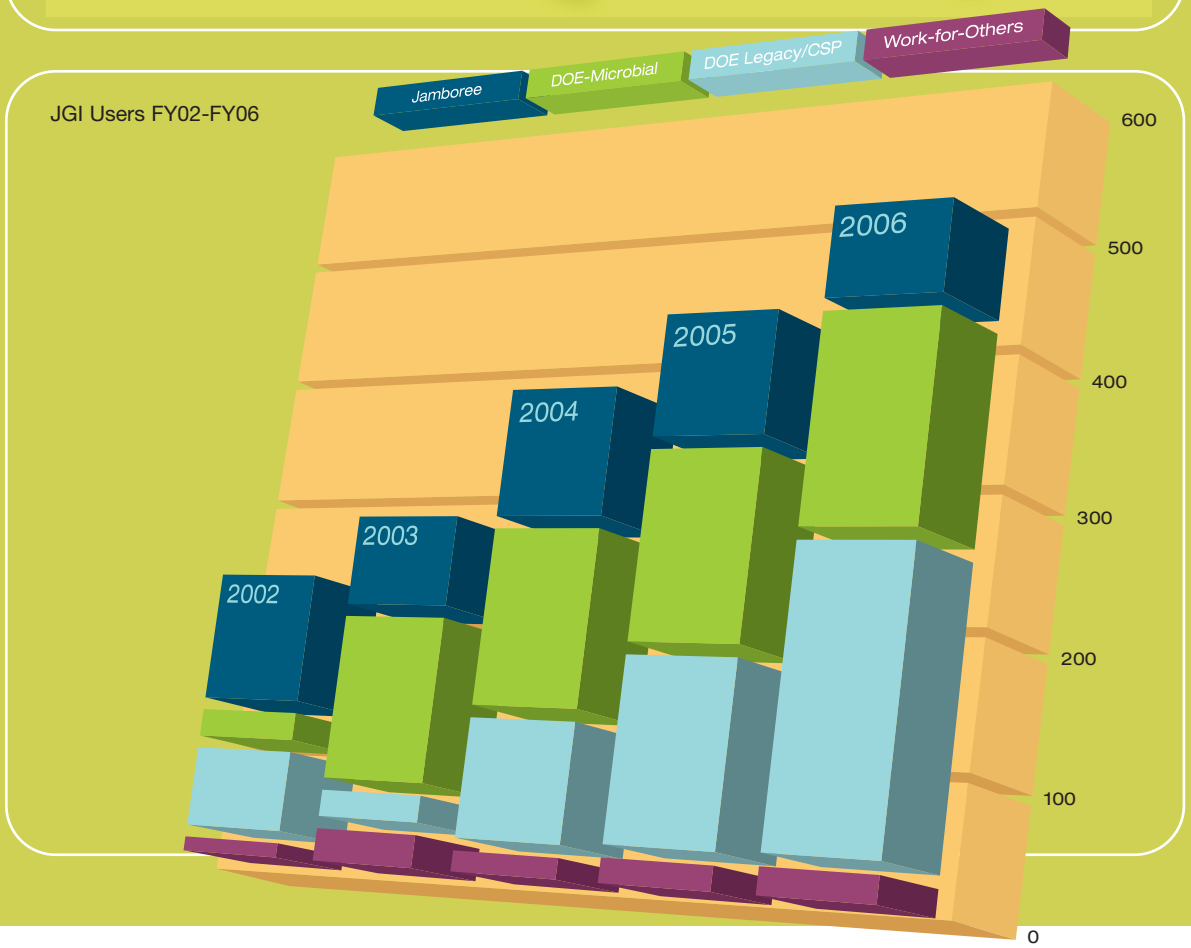
Major Sequencing Center Project Distribution

JGI, together with the other genome sequencing centers, is rising to the worldwide demand for sequence data.

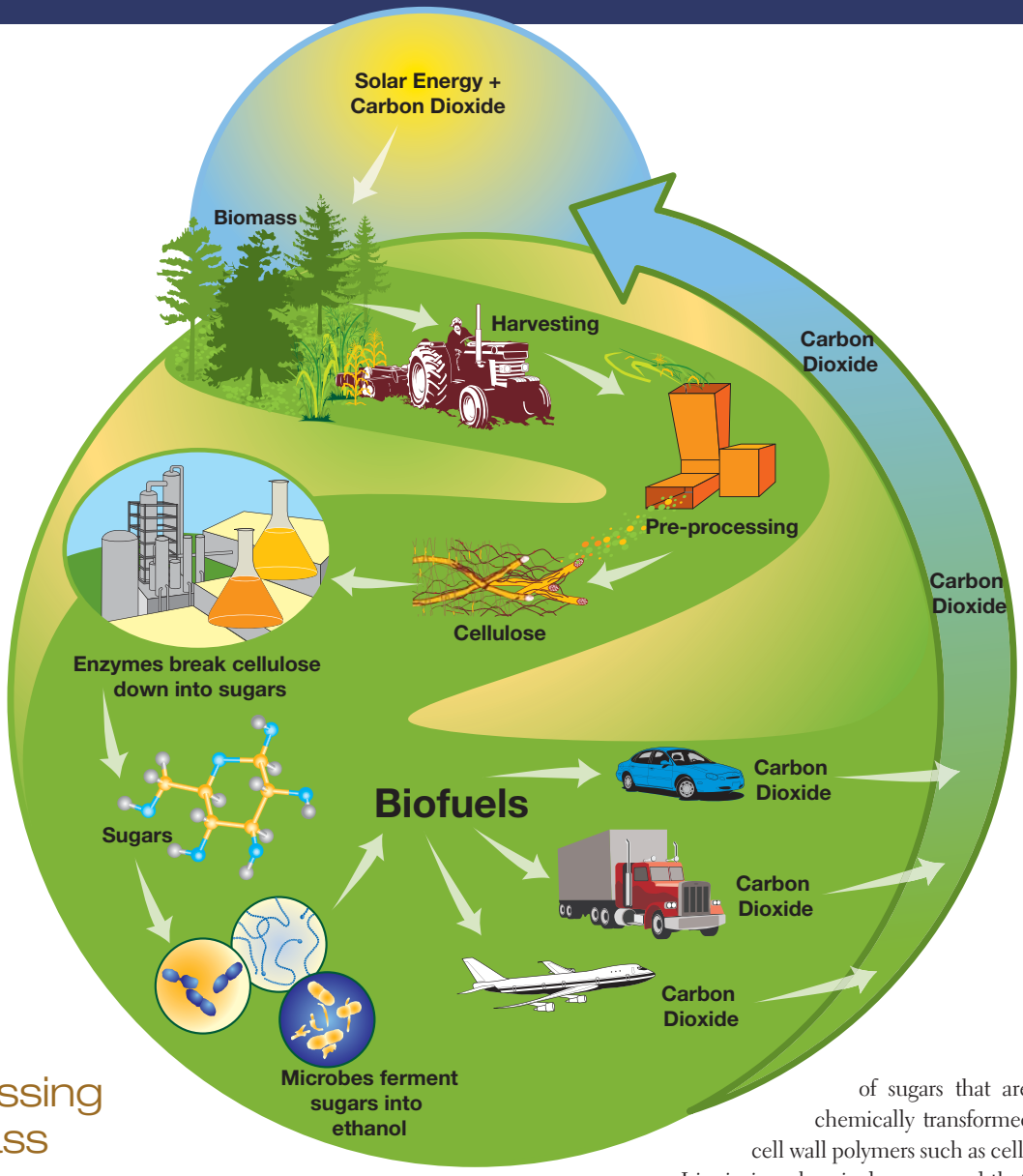
Data from GOLD: Genomes Online Database www.genomesonline.org



JGI user distribution in the US



The Benefits of Biofuels



Harnessing Biomass

Solid biomass, in the form of trees and other plants, can be converted into liquid fuels such as ethanol, methanol, and biodiesel. Selective breeding of crops has greatly enhanced corn and soybean yields over the last 50 years, but today's challenge lies in increasing dedicated bioenergy crop yields as well as improving the efficiency of biomass conversion to liquid fuels. Dramatic improvement in the efficiency of conversion needs to be made to ensure that the cost of biofuel production is competitive with or lower than that of petroleum fuels.

Plants store solar energy through photosynthetic production of sugars that are biochemically transformed into cell wall polymers such as cellulose. Lignin is a chemical compound that is an integral part of the cell walls of plants and trees, providing strength to cellulose fibers while conferring flexibility to the plant structure. Lignin makes up about 25-30% of the mass of dry wood, and most of the heat value associated with wood burning is derived from the combustion of lignin rather than cellulose. One challenge to a more efficient production of ethanol from woody plant matter is in learning how to break down the recalcitrant lignin polymer that complicates the efficient extraction of cellulose. DNA sequence information provided by JGI is poised to identify a novel set of microbial genes capable of degrading lignin and cellulose with the goal of increasing the efficiency of the conversion of biomass to fuels.

BIOSPEAK

In a conversation with Henry Ford and tire tycoon Harvey Firestone in 1931, shortly before Thomas Edison died, he said: "I'd put my money on the sun and solar energy. What a source of power! I hope we don't have to wait until oil and coal run out before we tackle that."

Biofuel refers to all fuels derived from organic matter (or *biomass*). The energy from sunlight is stored in the cells of plants, and this cellular energy can be converted into chemical fuels. *Ethanol* (or *bioethanol*) and *biodiesel* are the most common forms of biofuels at present, but converting the biomass into fuels is an expensive process. Both of these fuels can be used by many existing vehicles, and they reduce greenhouse gas emissions. Fossil fuels add carbon to the atmosphere when burned, but biofuels only release carbon that was captured from the atmosphere by the plant during photosynthesis. Today, most ethanol is blended with gasoline to reduce emissions and stretch gasoline supplies. According to the National Corn Grower's Association, 13.6% of the US corn crop in 2005 was used to produce 3.9 billion gallons of ethanol—using industrial corn varieties as opposed to edible sweet corn. Most ethanol in the US is currently made from corn, but other feedstocks with high starch content can be used with current technologies. Near-term research is focused on more efficient means to convert starchy plants into liquid fuels, but longer-term research is focused on cellulosic ethanol, which is made from abundant fibrous, woody, and generally inedible plant matter such as wood pulp, agricultural wastes (such as corn stover), and grasses (such as switchgrass).

For more information, see *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda:* http://genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml

JGI DNA SEQUENCING PROCESS

DNA: LIFE'S CODE

DNA (deoxyribonucleic acid), the information embedded in all living organisms, is a molecule made up of four chemical components—the nucleotides Adenine, Thymine, Cytosine, and Guanine—abbreviated A, T, C, G. These letters constitute the "rungs" of the double-helical ladder/backbone of the DNA molecule, with the As always binding with Ts, and Cs with Gs.

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 either a computer or a living organism, we must know the order, or sequence, of these informative bits.

Whole-genome shotgun sequencing is a technique for determining the precise order of the letters of DNA code of a genome. First, DNA received from JGI collaborators (1), or users, is sheared into small fragments that are easier for sequencing machines to handle (2).

These fragments are biochemically inserted into a plasmid vector (3)—a loop of nonessential bacterial DNA—and mixed into a solution of E. coli bacteria. An electric shock allows the plasmids to enter the bacterial cells (4). The bacterial cells are moved to an agar plate (5) and incubated with a nutrient and antibiotic to suppress the growth of unwanted cells. Over a night of incubation, colonies form, and each contains about a million bacteria. The clones in the colonies that contain the inserted DNA fragments required for sequencing are distinguished by color, picked (6), and incubated.with nutrients again to make many more copies, which can then be sequenced.

These DNA of interest are then duplicated, or amplified (7), through a process initiated with another enzyme called polymerase and an abundant supply of the chemical building-blocks of DNA, or nucleotides, from which the polymerase can assemble new copies. After many heating and cooling cycles, the ends of the DNA fragments are labeled with fluorescent markers (8). The DNA fragments are cleaned—isolated from the bacteria—using magnetic beads in an ethanol solution (9).

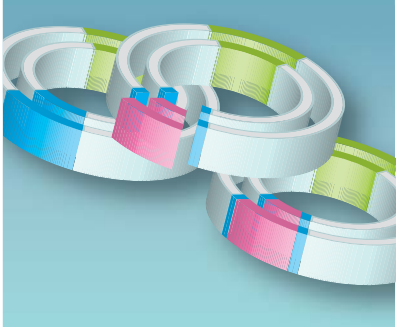
In a sequencing machine, an electrical charge is applied to the samples, pulling the DNA fragments through an assembly of fine glass tubes filled with a gel-like matrix, smaller fragments traveling faster than the larger, toward a laser detector system, which excites the fluorescent tags on each fragment by length and counts them to determine the sequence of As, Ts, Cs, and Gs (10).

During the assembly process, the DNA fragments are realigned based on overlaps in their sequences (11). Computer software uses the overlapping ends of different reads to assemble them into a reconstruction of the original contiguous sequence, then the annotated genome is made available to the scientific community (12).

Receive DNA from collaborators 1

Shear DNA 2

Insert DNA fragments into vectors 3



Introduce vectors in bacteria 4

Plate bacteria 5

Pick bacteria that contain vectors with inserts 6

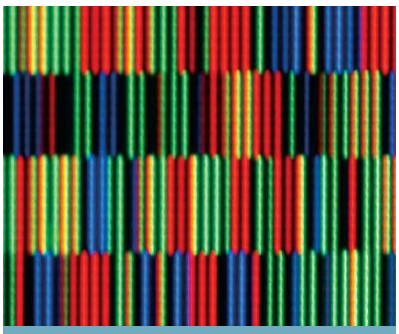
Amplify vectors with inserts 7

Produce dye-labeled DNA fragments 8

Clean up dye-labeled DNA fragments 9

Sequence by capillary electrophoresis 10

Assemble genome from sequence 11



Release genomic data to the world 12

SCIENCE BEHIND THE SEQUENCE
Highlights: Biomass to Biofuels

Energy Feedstocks

Raw plant materials, or dedicated bioenergy crops, provide the starting point for making biofuels. The US needs to grow a sufficient quantity of biomass in a profitable and sustainable fashion to develop a large biofuels industry, thereby reducing the need for imported oil for transportation fuels, reducing effects on global climate, and improving the economy. Cellulosic biomass crops, such as switchgrass and poplar, have been studied by JGI and its collaborators for their attractive environmental features and potential to improve yields. DNA sequence provided by JGI can be applied to the problem of improving biomass yield and the efficiency of processes used to convert plant materials into liquid fuels and valuable by-products. A 2005 joint study by DOE and USDA found that the US has enough agricultural and forestry land to support production of over one billion tons of biomass (www.eere.energy.gov/biomass), which could provide enough liquid biofuels to replace over a third of current transportation fuel consumption while continuing to meet food, feed, and export demands.

POPLAR: THE FIRST PUBLISHED TREE GENOME

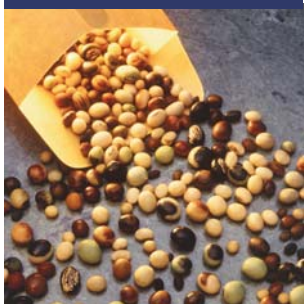
A major step in realizing the potential of cellulosic feedstocks for ethanol production was achieved with the completion and publication of the genome sequence of the poplar, *Populus trichocarpa*, the first tree to have its DNA decoded. The culmination of this four-year project, uniting the efforts of 40 institutions from around the world, led by JGI and Oak Ridge National Laboratory (ORNL), was published in the September 15, 2006 edition of the journal *Science*.

Poplar grows very quickly, and varieties can be grown throughout the US. As a result, it is a prime biomass feedstock. Research by JGI and its partners may increase this tree's ability to capture carbon dioxide from the atmosphere while improving the fast-growth qualities that make it a better renewable bioenergy resource.

Forest trees contain more than 90% of the Earth's terrestrial biomass, providing such environmental benefits as carbon capture, renewable energy supplies, improved air quality, and biodiversity. However, little is known about the biology of forest trees in comparison to the detailed information available for food crop plants.

The annotated poplar genome will facilitate rapid and effective analysis of the gene network underlying traits related to tree growth, drought tolerance, pest resistance, cell wall composition, and other traits relevant to the DOE mission.

SOYBEAN



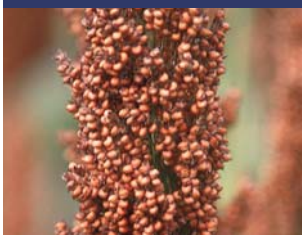
The soybean, *Glycine max*, is the principal source of the renewable alternative fuel, biodiesel, in the US. Biodiesel has the highest energy content of any alternative fuel and is much more environmentally friendly than comparable petroleum-based fuels. Biodiesel degrades rapidly in the environment and burns more cleanly than conventional fuels, releasing only half the pollutants and reducing the production of carcinogenic compounds by more than 80%.

Over 3.1 billion bushels of soybeans were grown in the US on nearly 75 million acres in 2004, with an estimated annual value exceeding $17 billion, second only to corn and approximately twice that of wheat. The soybean genome is about 1.1 billion base pairs in size—less than half the size of the maize or human genomes.

The soybean represents an excellent example of how JGI is playing a key role in "translational genomics," applying the tools of DNA sequencing and molecular biology to derive detailed knowledge about the soybean genetic code. This will allow for crop improvements and the effective application of this plant to clean energy generation. Knowing which

genes control specific traits, researchers could change the type and quantity of oil produced by the crop, and lead to soybean plants that are more resistant to drought and disease.

SORGHUM



In January of 2007, JGI completed 9x assembly of the sorghum genome. One of the world's leading grain crops, sorghum is also an important model for tropical grasses, and is a logical complement to Oryza (rice), the first monocot plant to be sequenced. Sorghum is representative of the tropical grasses in that it uses "C4" photosynthesis, with a complex combination of biochemical and morphological specializations resulting in more efficient carbon assimilation at high temperatures. By contrast, rice is more representative of temperate grasses, using "C3" photosynthesis.

In addition to its intrinsic value, the sorghum sequence will be a valuable reference for assembling and analyzing the four-fold larger genome of maize (corn), a tropical grass that is the leading US fuel ethanol crop (sorghum is second). Sorghum is an even closer relative of sugarcane, arguably the most important biomass/biofuels crop worldwide with annual production of about 140 million metric tons and a net value of about $30 billion. Sorghum and sugarcane are thought to have shared a common ancestor about 5 million years ago, but sorghum's genome is about 25% the size of human, maize, or sugarcane.

The sorghum project is a collaboration between Andrew H. Paterson (lead), John E. Bowers, and Alan R. Gingle (all three from the University of Georgia); C. Thomas Hash (International Crops Research Institute for the Semi-Arid Tropics); Stephen E. Kresovich (Cornell University); Joachim Messing (Rutgers); Daniel G. Peterson (Mississippi State University); and Daniel S. Rokhsar (JGI and University of California, Berkeley).

BRACHYPODIUM

While herbaceous energy crops (especially grasses) are poised to become a major source of renewable energy in the United States, we know very little about the genetic traits that affect their utility for energy production. The temperate wild grass species, *Brachypodium distachyon*, is a new model plant being studied by JGI for developing grasses into superior energy crops. *Brachypodium* is small in size, can be grown rapidly, is self-fertilizing, and has simple growth requirements. It can be used as a functional model to gain the knowledge about basic grass biology necessary to develop superior energy crops, like switchgrass and *Miscanthus*.

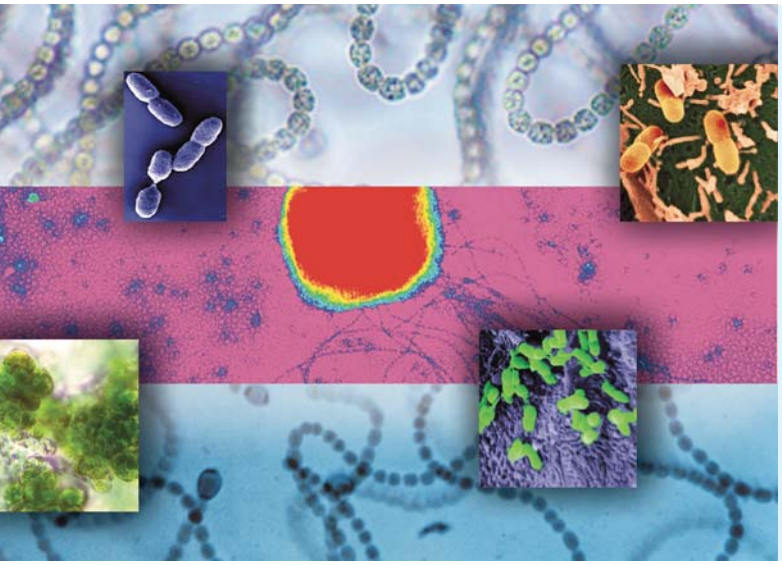
The sequencing of *Brachypodium* will be undertaken via a two-pronged strategy: the first, a whole-genome shotgun sequencing approach, is a collaboration between John Vogel and David Garvin, both of the USDA, and Michael Bevan at the John Innes Centre in England; and the second, an expressed gene sequencing effort, led by Todd Mockler and Jeff Chang at Oregon State University, with Todd Michael of The Salk Institute for Biological Studies, and Samuel Hazen from The Scripps Research Institute.

CASSAVA

JGI is also sequencing cassava *(Manihot esculenta)*, an excellent energy source and food for approximately one billion people around the planet. Its roots contain 20-40% starch, from which ethanol can be derived, making it an attractive and strategic source of renewable energy. Cassava grows in diverse environments, from extremely dry to humid climates, acidic to alkaline soils, sea level to high altitudes, and in nutrient-poor soil.

Sequencing the cassava genome will help bring this important crop to the forefront of modern science and generate new possibilities for agronomic and nutritional improvement. The cassava project will extend broad benefits to its vast research community, including a better understanding of starch and protein biosynthesis, root storage, and stress controls, enabling crop improvements while shedding light on similar mechanisms in related plants, including the rubber tree and castor bean.

The cassava project is led by Claude M. Fauquet, Director of the International Laboratory for Tropical Agricultural Biotechnology and colleagues at the Danforth Plant Science Center in St. Louis, and includes contributions from the USDA laboratory in Fargo, ND; Washington University, St Louis; University of Chicago; The Institute for Genomic Research (TIGR); Missouri Botanical Garden; the Broad Institute; Ohio State University; the International Center for Tropical Agriculture (CIAT) in Cali, Colombia; and the Smithsonian Institution.



100 FINISHED MICROBES

In 2006, JGI reached a new milestone in finishing the sequencing of its 100th microbial genome to an accuracy greater than 99.9999%, and that included the free dissemination of this information on the World Wide Web for the benefit of the global research community. These microbes cross all main branches of the tree of life: Eubacteria, Archaea, and even the Eukaryota, which include fungi, plants, and animals.

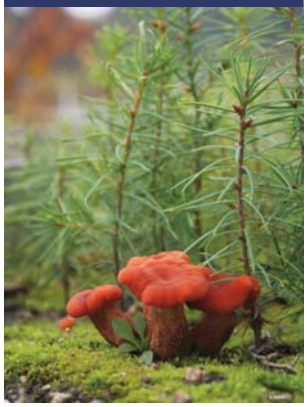
As microbial genomes range in size from typically five to tens of millions of bases, several microbes could be sequenced in one day. However, the sequencing process, in order to meet rigorous quality standards and to satisfy the demands of the scientific community, is an iterative one, requiring six- to eight-times coverage. The term "finished," associated with the 100 microbial genomes finished by JGI, is a technical designation referring to a standard of accuracy established for the Human Genome Project of tolerating no more than one mistake in 10,000 letters of genetic code without any gaps.

The power of JGI's sequencing of microbes, and other organisms, is that it provides the complete genomic "parts list" of those organisms. With this list in hand, researchers can explore how microbes use these parts to build and sustain key functions, including many of critical importance to DOE, such as the breakdown of plant materials to produce useful sources of energy, and for cleaning up toxic waste sites.

MICROBES THAT IMPACT FEEDSTOCKS

A large number of microbes have an impact on plants by providing essential nutrients as well as serving as agents of disease. Since feedstocks operate in a complex system that includes a large number of microbes, the JGI has a growing portfolio of plant-associated microbes.

LACCARIA BICOLOR: SYMBIOTIC TREE FUNGUS



The DNA sequence of *Laccaria bicolor*, a fungus that forms a beneficial symbiosis with poplar and other trees and inhabits one of the most ecologically and commercially important microbial niches in North American and Eurasian forests, has been determined by JGI. The complete *Laccaria* genome sequence was announced July 23, 2006, at the Fifth International Conference on Mycorrhiza in Granada, Spain, by an international consortium comprised of JGI, Oak Ridge National Laboratory (ORNL), France's National Institute for Agricultural Research (INRA), the University of Alabama in Huntsville (UAH), Ghent University in Belgium, and additional groups in Germany, Sweden, and France.

The *Laccaria* genome sequence will provide the global research community with a critical resource to develop faster-growing trees for producing more biomass that can be converted to fuels, and for trees capable of capturing more carbon from the atmosphere.

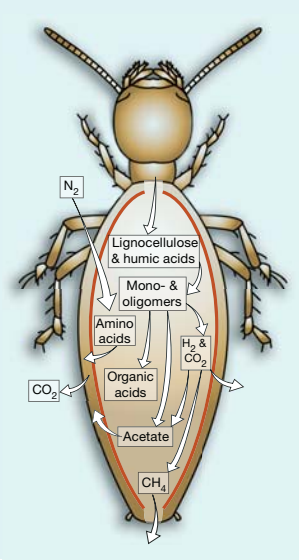
Key factors behind the ability of trees to generate large amounts of biomass, or store carbon, reside in the way that they interact with soil microbes known as mycorrhizal fungi, which excel at procuring necessary, but scarce, nutrients such as phosphate and nitrogen. (When *Laccaria bicolor* partners with plant roots, a mycorrhizal root is created, resulting in a mutual relationship and making these nutrients available to their host, and significantly benefiting both organisms.) The fungus within the root is protected from competition with other soil microbes, and gains preferential access to carbohydrates within the plant.

This research will advance the understanding of how functional genomics of this symbiosis enhances biomass production and carbon management, particularly through the interaction with the poplar tree, also sequenced by JGI. It will now be possible to harness the interaction between these species and identify the factors involved in biomass production by characterizing the changes that occur between the two genomes as the tree and fungus collaborate to generate biomass. It will also help scientists understand the interaction between these two symbionts within the context of the changing global climate.

THE MICROBIAL ETHANOL PRODUCERS

In 2006, JGI sequenced several microbes that should contribute to progress toward the goal of the DOE Office of Biomass Program to enable the development of economical processes for converting biomass to fuel ethanol to displace 30% of US 2004 transportation fuels by 2030. These projects included:

CELLULOSE DEGRADATION IN THE TERMITE'S HINDGUT



The termite is capable of producing two liters of hydrogen by fermenting just one sheet of paper, making it one of the planet's most efficient bioreactors. Termites accomplish this Herculean task by exploiting the metabolic capabilities of about 200 different species of microbes that inhabit their hindguts.

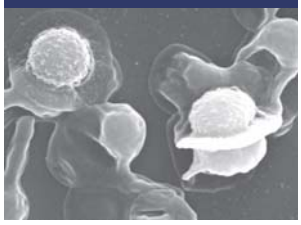
Hydrogen is normally created by using electricity to remove hydrogen molecules from water or natural gas, but the electricity is most often generated using fossil fuels that emit carbon pollutants. The microbial community in the termite gut efficiently manufactures large quantities of clean hydrogen. By sequencing the termite's microbial community, JGI can provide a better understanding of these biochemical pathways.

Termites eat wood, and bacteria in the termite's gut break down the complex lignocellulose polymers into simple sugars, using enzymes that produce hydrogen as a by-product. A second wave of bacteria uses the simple sugars and hydrogen to make the acetate the termite requires for energy. If JGI can help figure out which enzymes are used to create hydrogen, and which genes produce them, this process could be scaled up with bioreactors to generate hydrogen from woody biomass, such as poplar, in commercial quantities.

CLOSTRIDIUM THERMOCELLUM

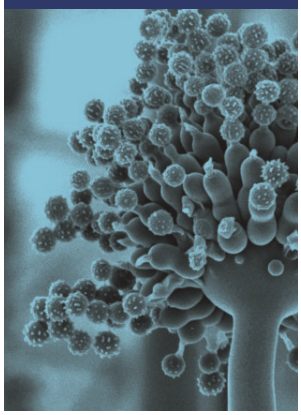
Clostridium thermocellum is an anaerobic bacterium capable of directly converting cellulose from biomass into ethanol. The degradation of cellulose is carried out by an extracellular cellulase system called the cellulosome, and continued research in this area should provide crucial information for better understanding the cellulolytic reaction—a key process in biomass conversion.

PICHIA STIPITIS



Pichia stipitis is a fungus that ferments xylose to ethanol, and degrades lignin and cellulose for the potential conversion of biomass to ethanol. The lack of industrial-grade microorganisms for converting biomass into fuel ethanol has traditionally been cited as a major technical roadblock to developing a bioethanol industry. The highest yields for the conversion of biomass to ethanol are expected to come from microorganisms such as *P. stipitis* that can ferment the sugar xylose.

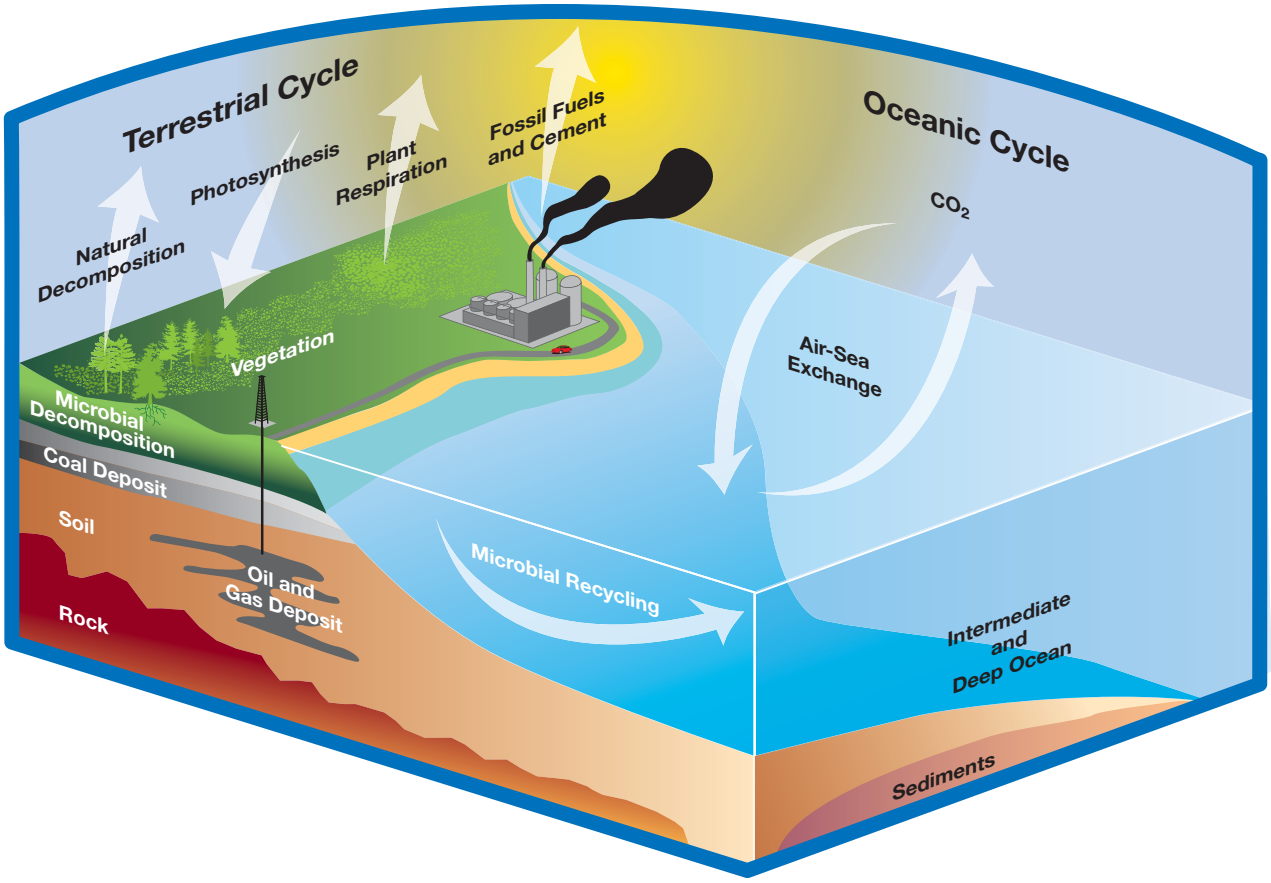
ASPERGILLUS NIGER



Aspergillus niger is a filamentous fungus that has been used by industry for nearly 100 years to make organic acids and several biomass-processing enzymes. JGI and the Fungal Biotechnology Team of the Chemical and Biological Process Development Group at Pacific Northwest National Laboratory (PNNL) released the first public sequence of *Aspergillus niger* in 2006.

At PNNL, *A. niger* is used as a model microbial system to make value-added products from lignocellulosic biomass to augment the product portfolio of the anticipated biorefinery industry. Just as the petroleum refinery produces a variety of non-fuel products from oil, filamentous fungi will be used to convert complex biomass via their specific hydrolyzing enzymes and/or co-products, such as organic acids and platform molecules for the biopolymer industry. The sequencing of *A. niger* is an example of the collaboration between the DOE Office of Science (OBER), the Energy Efficiency and Renewable Energy (EERE) Office of Biomass Program, and PNNL, all of which provided direct or indirect support of this project.

Carbon Cycling

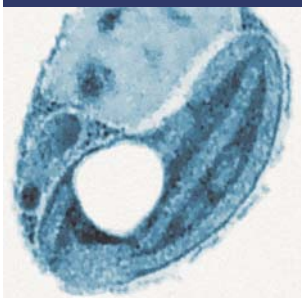


Investigations by JGI and its partners are shedding light on the cellular machinery of organisms and microbes that can be used to capture carbon dioxide from the atmosphere. While research continues into new sources of energy that emit little or no carbon, strategies must also be explored for capturing atmospheric carbon dioxide (CO_2)—a chief contributor to global climate change—generated by the use of fossil fuels. Research into the role that microorganisms play in the Earth's natural carbon cycle may lead to new strategies for reducing atmospheric carbon and other greenhouse gases.

TINY ORGANISMS WITH GLOBAL IMPLICATIONS

The ocean plays a key role in removing carbon dioxide from the atmosphere, with marine photosynthetic organisms consuming the carbon and releasing oxygen and phytoplankton in particular, accounting for almost half of total global photosynthesis. Species of the smallest known eukaryotes, the phytoplankton of the genus *Ostreococcus*, are being sequenced, characterized, and compared in a collaboration entailing researchers from 16 institutions led by JGI's Igor Grigoriev; Brian Palenik of the Scripps Institution of Oceanography, University of California, San Diego; and Hervé Moreau of CNRS and University of Paris.

OSTREOCOCCUS

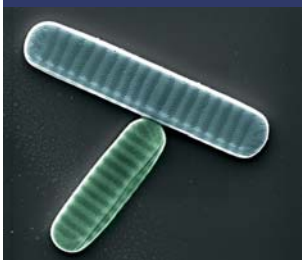


Ostreococcus belongs to the *Prasinophyceae*, an early-diverging class within the green plant lineage, and is globally abundant, single-celled alga thriving in the upper (illuminated) water column of the oceans. Three different ecotypes or potential species have been

defined, based on their adaptation to light intensity. One *(O. lucimarinus)* is adapted to highlight intensities and corresponds to surface-isolated strains. The second (RCC141) has been defined as low light and includes strains from deeper in the water column. The third *(O. tauri)* corresponds to strains isolated from a coastal lagoon and thus exists in various light conditions.

In related work by Alexandra Worden, at the University of Miami Rosenstiel School of Marine & Atmospheric Science, JGI is sequencing two strains of another organism, *Micromonas pusilla*, closely related to *Ostreococcus*, to conduct comparative analyses of these organisms to understand niche differentiation and their specific role in global carbon cycling.

DIATOMS

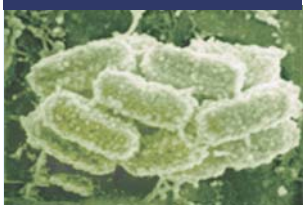


Diatoms are also eukaryotic, photosynthetic microorganisms found throughout marine and freshwater ecosystems that are responsible for around 20 percent of global primary productivity. The diatom *Phaeodactylum tricornutum* is the second diatom for which a whole genome sequence has been generated. The first was *Thalassiosira pseudonana*, sequenced by JGI and published in the journal *Science* in 2004. *P. tricornutum* has been found in several locations around the world, typically in coastal areas with wide fluctuations in salinity. Unlike other diatoms, it can exist in different morphotypes, and changes in cell shape can be stimulated by environmental conditions.

Polar diatoms are adapted to cold temperatures and high salinities, and they live both within the seawater and the brine channels formed within sea ice. They serve as the base of the entire polar food web. Polar regions are displaying the greatest responses to climate change, with increased melting of sea ice and the potential disruption of this sensitive ecosystem. The sequencing of a polar diatom was proposed by Ginger Armbrust of the University of Washington, Klaus Valentin of the Alfred-Wegener Institute for Polar and Marine Research, and Chris Bowler of the Ecole Normale Superieure, to gain insights into how these organisms have adapted to the extreme environments in which they thrive.

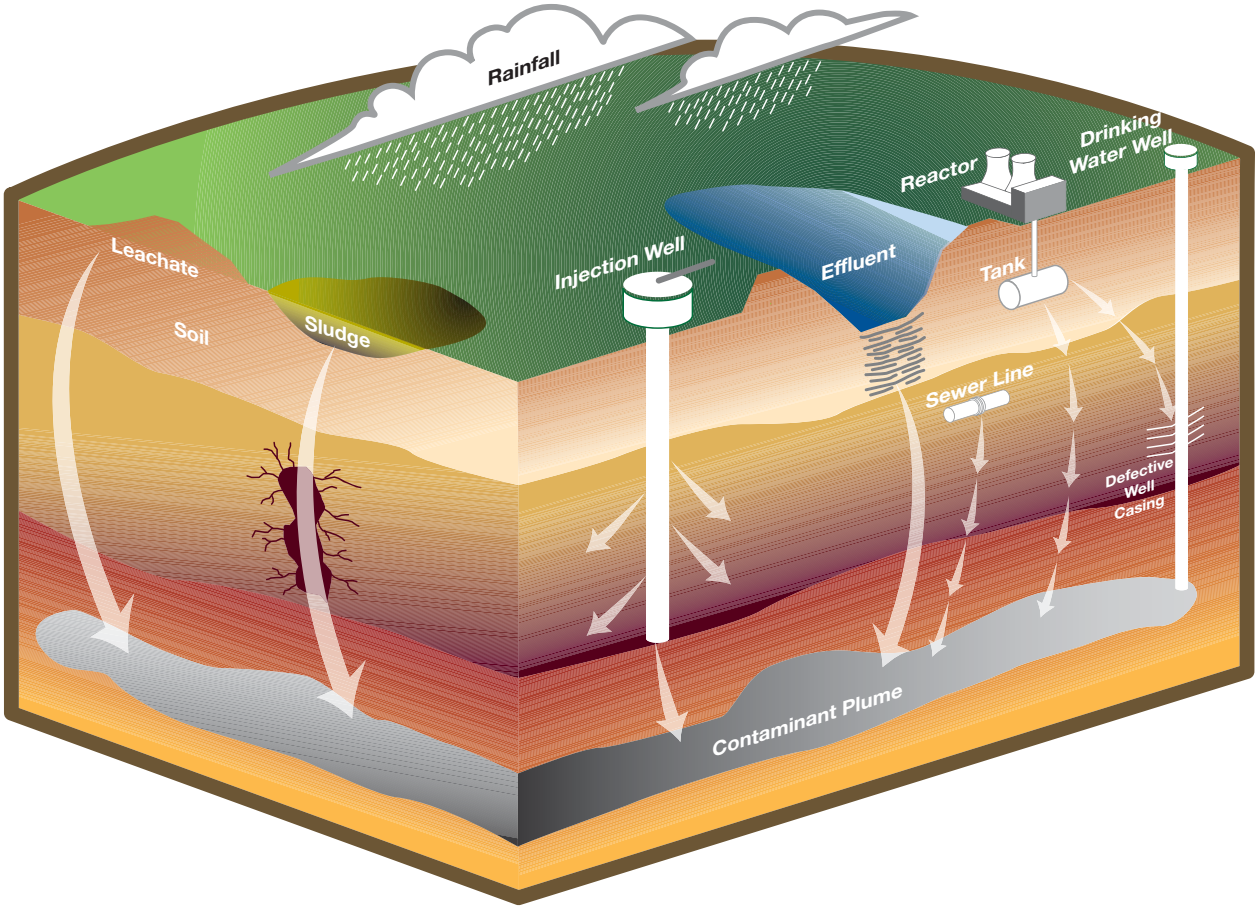
Fragilariopsis cylindrus is regarded as a typical cold-water species found in Arctic and Antarctic seawater and sea ice. It can form large ice-edge blooms and is regularly detected in sea ice investigations. The whole genome sequences of the temperate diatoms *Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, and *Pseudo-nitzschia multiseries* (another diatom genome slated to be sequenced by JGI) will provide a basis for comparative analysis with the cold-tolerant *F. cylindrus*. The sequence of *F. cylindrus* will lead to further insights into the unique physiology of polar-adapted eukaryotes, and allow for better predictions of the impact climate change may have on organisms that serve as the basis for all polar ecosystems.

CHLOROBI



Other important micromanagers involved in the carbon cycle being sequenced by JGI include the green sulfur bacteria (GSB) of the phylum *Chlorobi*, commonly found in illuminated, stratified, and aquatic environments deprived of oxygen such as sediments and sulfide-rich hot springs. Because of unique adaptations of their light-harvesting antennae, these bacteria are capable of growth at light intensities under which no other phototrophs can survive. In some aquatic environments, these organisms can account for up to 83% of the total annual productivity, and thus it is clear that these organisms can be the primary contributors of fixed carbon in certain ecological niches. Sequencing of *Chlorobi* is enabling JGI user Donald Bryant of Pennsylvania State University to conduct functional genomics studies to advance our understanding of the sulfur and ferrous iron oxidation, carotenoid and chlorophyll biosynthesis, photosynthetic light harvesting, oxygen tolerance, and many other aspects of the physiology and metabolism.

Bioremediation



Bioremediation is a technology that can be used to reduce, contain, or eliminate hazardous waste. Microorganisms can transform and degrade many contaminants in soil and water. Organic contaminants such as hydrocarbon fuels can be degraded into carbon dioxide, while toxic and radioactive metals can be immobilized or removed from the environment. These solutions are made possible by the microorganisms that have adapted to live in contaminated environments. DNA sequencing provides unique signatures of individual microbes and their communities, enabling JGI to help identify new solutions for cleaning up hazardous waste sites and restoring the environment.

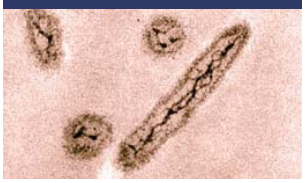
BURKHOLDERIA

Burkholderia species have tremendous versatility: they can efficiently degrade pollutants in water and soil, fix atmospheric nitrogen, or help plants fight against their pathogens; hence they contribute to a healthy environment. *Burkholderia* also occupy diverse habitats, from soil to rhizosphere (root zone) to water, as well as

intimate associations with plants and animals—even living intracellularly in amoebas, for example. Their genomes are among the largest of all known bacteria (7 to 9.7 Mb), consisting of three replicons, the smaller two exhibiting a faster evolutionary rate and considerable plasticity. This rather unique genome structure is thought to contribute to *Burkholderia's* characterization as a "versaphile."

Burkholderia xenovorans LB400 was isolated from PCB-contaminated landfill soil in upper New York State, and has been the subject of over 70 studies related to aerobic PCB degradation, making it the model organism for biodegradation of this pollutant. It is one of the largest bacterial genomes, with 9.73 million base pairs, and is the first non-pathogenic isolate of *Burkholderia* to be sequenced. Most of the *Burkholderia* in nature are non-pathogenic, but previous studies focused on plant pathogens and species causing animal and human disease. This work, led by Jim Tiedje, University Professor at Michigan State University and LLNL's Patrick Chain, was published in the *Proceedings of the National Academy of Sciences* (PNAS), October 17, 2006 (vol. 103, no. 42).

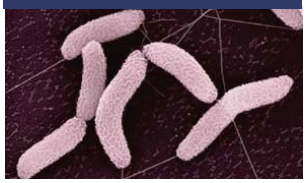
SHEWANELLA



Shewanella oneidensis MR-1 is a motile, facultative Proteobacterium with remarkable respiratory versatility; it can utilize a range of organic and inorganic compounds as terminal electron acceptors for anaerobic metabolism. The ability to grow aerobically, and to utilize metals and radionuclides as electron acceptors in the absence of oxygen, also makes *Shewanella* an excellent model organism for contaminant transformation investigations. To function and compete in environments that are subject to spatial and temporal environmental change in electron acceptors, *Shewanella* must be able to sense and respond to such changes and therefore require relatively robust sensing and regulation systems.

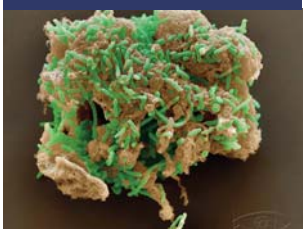
The overall goal of the multi-institutional *Shewanella* Federation is to apply the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella* being sequenced by JGI, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems, we propose to use genome-based approaches to investigate *Shewanella* as a system of integrated networks; first describing key cellular subsystems — those involved in signal transduction, regulation, and metabolism — then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, this project will employ complementary "top-down" — bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments. JGI is working with Pacific Northwest National Laboratory and the *Shewanella* Federation to characterize the genomes of multiple members of this genus under DOE's Genomics: GTL program. Developing integrated regulatory and metabolic network models of *Shewanella* at the whole-genome level will help researchers predict how this bacterium responds to different environmental conditions.

DESULFOVIBRIO DESULFURICANS G20



Desulfovibrio desulfuricans is a microbe with a robust appetite for such toxic metals as uranium and chromium, from a family known as sulfate-reducing bacteria, or SRB. Isolated from a corroded oil well in the late 1980s, the G20 strain was sequenced by JGI. Sulfate-reducing bacteria are conspicuous because the product of their respiration, hydrogen sulfide (H_2S), smells like rotten eggs. Producing this extremely reactive gas—toxic to plants and animals—these bacteria thrive in anaerobic conditions of deep marine sediments, where oil can be found. Another idiosyncrasy—one that researchers are seeking to tap—is their generosity in donating electrons, a process known as reduction. In association with organics in the soil, SRB convert sulfate to sulfide—hence the big stink. DOE's interest in these microbes hinges on their ability to also reduce uranium from its highly oxidized state. This reaction detoxifies the metal so that it is not as harmful to humans while making it insoluble in water, effectively immobilizing it in the soil.

GEOBACTER



G. metallireducens serves an important function in the carbon management and nutrient cycles of aquatic sediments, and in the bioremediation of organic and metal contaminants in groundwater. This bacterium

"Just as the availability of the human genome's set of genes facilitated the diagnosis of clinical disorders and the development of new drugs, our ability to recognize the compendium of genes evolved in nature to capture carbon from the atmosphere, degrade biomass to fermentable sugars, and detoxify hazardous substances will facilitate harnessing these important processes in the national interest."

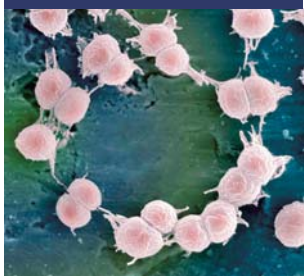
—Eddy Rubin

is able to gain energy through the reduction of iron, manganese, uranium, and other metals—essentially powering itself by using iron oxides in the same way that humans use oxygen. Its ability to harvest electricity from waste organic matter, combined with a demonstrated ability to transfer electrons onto the surface of electrodes, has made it possible to design microbial fuel cells.

For environmental restoration, *Geobacter* can quickly destroy petroleum contaminants in groundwater, degrading them into harmless carbon dioxide. It can also remove radioactive metal contaminants from groundwater.

The genomes of several *Geobacter* species have been sequenced by JGI, and a computer model of this information will predict *Geobacter* metabolism under a variety of environmental conditions. This systems biology approach is greatly accelerating the understanding of how this bacterium functions, and how it can be optimized for bioremediation and energy-harvesting applications.

DEINOCOCCUS GEOTHERMALIS



Deinococcus geothermalis is noteworthy for its extreme resistance to ionizing radiation, toxic chemicals, dehydration, and high temperatures. Related to *D. radiodurans*, which was named the "World's Toughest Bacterium" by the Guinness Book of Records, *D. geothermalis* not only endures high doses of radiation, but can also grow at temperatures as high as 55°C.

One of the biggest challenges in waste cleanup at a number of DOE sites is dealing with mixed waste—organic chemicals and radioactive wastes—stored together in underground tanks that are now leaking. Isolated from thermal springs in Naples, Italy, this microbe is now being considered for use in cleaning up radioactive mixed waste environments. Researchers are working on developing a strain of this bacterium that will not only resist destruction by radiation, toxic metals, and organic chemicals, but will also stay alive long enough to complete its waste cleanup task, perhaps using the waste as a food source to keep *D. geothermalis* alive and growing.

Exploratory Sequence-Based Science

While the vast majority of projects at the JGI relate directly to DOE mission areas, a small number of projects are either supported by non-DOE funding or are deemed to be pioneering next-generation applications of DNA sequencing by the JGI senior management.

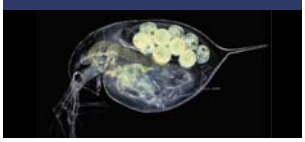
OLAVIUS ALGARVENSIS: GUTLESS WORM

Researchers have now characterized the unique lifestyle of a gutless worm. This worm commutes through marine sediments powered by a community of symbiotic microbial specialists harbored just under its skin, obviating the need for digestive and excretory systems that we normally think of as essential for animals.

Using DNA sequencing and other computational methods to reconstruct the genomes of the symbiont organisms, scientists from the US Department of Energy's Joint Genome Institute (JGI) have described this complex worm/microbe collaboration in a species of marine oligochaete worm isolated off of the coast of Elba, the Mediterranean island of Napoleon's exile. Their results were published in the September 17, 2006, edition of the journal *Nature*.

The work was conducted under the JGI Community Sequencing Program by postdoctoral fellow Tanja Woyke and colleagues from the Rubin lab at JGI, along with collaborators led by Nicole Dubilier at the Max Planck Institute for Marine Microbiology in Bremen, Germany. The team uncovered the unique method of waste management employed by *Olavius algarvensis* through metagenomics—a strategy pioneered by JGI and its collaborators. This is the first instance of such a symbiotic relationship being analyzed using a metagenomic shotgun sequencing approach, heralding a renaissance in symbiosis research.

DAPHNIA PULEX: WATER FLEA



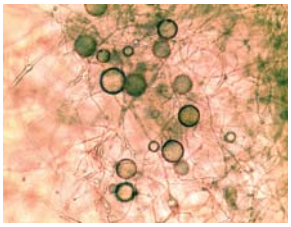
Tiny inhabitants of aquatic ecosystems can serve as sensitive early indicators of pollutants. Developmental biologists at JGI and its partner institutions study DNA sequence information from organisms such as *Daphnia pulex*, the water flea, to learn more about how they grow, develop, and cope with insults to their environment. Computer simulation of these types of model organisms helps researchers analyze the genetic control of cell growth, differentiation, and "morphogenesis," which is the process that creates tissues, organs, and anatomy. Ecologists and evolutionary biologists also want to learn more about how genetic variation is important for ecosystem adaptation, and how populations survive in a changing world.

Daphnia's short generation time and small genome—200 million base pairs—make it ideal for the study of how environments influence an organism's genetics. Despite their common name, water fleas are not insects, but crustaceans (like lobsters and crabs). *Daphnia* is the first crustacean genome to be sequenced.

PHYTOPHTHORA: PLANT DESTROYER

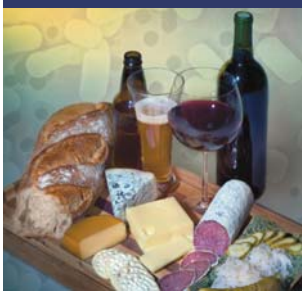
Thanks to DNA sequence provided by JGI, researchers are now closer to understanding two related plant pathogens that cause "Sudden Oak Death" (SOD) and a devastating soybean disease. JGI conducted this multi-agency effort in collaboration with the Virginia Bioinformatics Institute (VBI), the US Department of Agriculture (USDA), and the National Science Foundation (NSF). The availability of the genomes of the organisms should contribute to the development of new diagnostics and treatments for these diseases.

The aptly named *Phytophthora* is named after the Greek word for "plant destroyer." The genome sequence from JGI provides a framework for understanding how these plant pathogens cause disease and what can be done to control them. *Phytophthora* attacks a wide variety of plants, including agricultural crops, trees, and shrubs. One



of the *Phytophthora* species, *P. infestans*, was responsible for the mid-19th century Irish potato famine. *P. ramorum* causes Sudden Oak Death, concentrated in California and southern Oregon, but also killing trees across the country. Losses attributed to *P. sojae* infestation, known as *Phytophthora* root rot of soybean, exceeded $1 billion in 2003. This work was published in the September 1 issue of *Science*.

LACTIC ACID BACTERIA



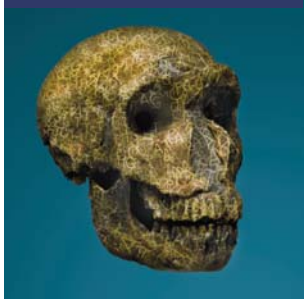
Researchers from JGI and the University of California, Davis, and their colleagues have characterized the genome sequences of nine different lactic acid-producing bacteria, or LAB, and published their findings in the October 17 edition of the *Proceedings of the National Academy of Sciences* (http://www.pnas.org/cgi/content/abstract/0607117103v1). The small LAB genomes encode a diverse repertoire of genes for efficient carbon and nitrogen acquisition from the nutritionally rich environments they inhabit and reflect a limited range of biosynthetic capabilities promising broad industrial applications.

Lactic acid-producing bacteria play a key role in the production of fermented foods and beverages, accounting for tens of billions of dollars in sales annually. Products constituting a fine repast, such as wine, salami, cheese, sourdough bread, pickles, yogurt, cocoa, and coffee are all enhanced by LAB, which ferment six-carbon sugars, or hexoses, to produce lactic acid.

Access to the genome sequences for these fermentative microorganisms will dramatically increase the understanding of their role in industrial food production, leading to more optimized production schemes. LAB are used for production of various commercial bioproducts such as dextran and antimicrobials. The availability of these genome sequences will foster development of additional production schemes for biofuels and other important chemicals.

The publication is the culmination of a multiyear effort by the Lactic Acid Bacteria Genome Consortium, a group of at least a dozen academic organizations formed in 2001. The functional classification embraced a variety of industrially important genera, including *Lactococcus*, *Enterococcus*, *Oenococcus*, *Pediococcus*, *Streptococcus*, *Leuconostoc*, and *Lactobacillus* species. These diverse species offered a window into the sugar metabolism and energy conversion systems of LAB, and the evolution of these systems, which helped identify key enzymes involved in the production of end products, including acetic acid, lactic acid, ethanol, and CO_2.

NEANDERTHAL



JGI scientists are constantly exploring new ways to use DNA sequence to gain insight into biological problems that cannot be solved in other ways. Such is the case of a project that generated much attention in the media last fall with the lifting of the veil of mystery surrounding our extinct hominid cousins, the Neanderthals. In a project funded by the National Institutes of Health, JGI Director Rubin's laboratory at LBNL, coupled with researchers at the JGI, generated genomic DNA sequence from a fossilized Neanderthal bone and published their findings in the November 17, 2006, issue of the journal *Science*. Their results show that the genomes of modern humans and Neanderthals are at least 99.5 percent identical, but despite this genetic similarity, and the fact that the two species inhabited the same geographic region for thousands of years, there is no evidence of any significant crossbreeding between the two. Based on these early results, *Homo sapiens* and *Homo neanderthalensis* last shared a common ancestor approximately 700,000 years ago.

Technology described in the paper marks an important advance in the field of metagenomics, which is increasingly being used to sequence the complex mixtures of DNA found in the environment.

Sequence Analysis

THE INTEGRATED MICROBIAL GENOMES (IMG) DATA MANAGEMENT SYSTEM

As the microbial world comes to light through DNA sequencing, the Integrated Microbial Genomes (IMG) data management system facilitates the delivery of valuable sequence information to the global research community. In 2006, Integrated Microbial Genomes for Microbiome Samples (IMG/M) became available for the management and analysis of metagenomics data.

A collaboration between JGI and the Lawrence Berkeley National Laboratory Biological Data Management and Technology Center (BDMTC), the IMG system was launched in 2005 as an easy-to-use tool for investigators wanting to extract information from microbial sequence information. IMG responds to the urgent and increasing need for a means to handle the vast and growing spectrum of datasets emerging from genome projects taken on by JGI and other public DNA sequencing centers. This important computational tool enables scientists to tap the rich diversity of microbial environments and harness the possibilities that they hold for addressing challenges in environmental cleanup, agriculture, industrial processes, and alternative energy production.

JGI is currently producing nearly one-quarter of the microbial genome projects worldwide, more than any other single institution. As the number of microbial genomes sequenced continues to rise, the genome analysis process becomes the rate-limiting step. By integrating publicly available microbial genome sequence with JGI sequence, the IMG system offers a powerful data management platform that supports timely analysis of genomes from a comparative, functional, and evolutionary perspective.

In December 2006, Version 2.0 of the Integrated Microbial Genomes (IMG) data management system was released to the public. The content of IMG 2.0 was entirely refreshed and extended with the latest versions of genomes available from the National Center for Biotechnology

Information's (NCBI) Reference Sequence collection (RefSeq).

IMG 2.0 featured the following enhancements:

- 1,541 new public microbial, viral, and eukaryotic genomes were added to IMG 2.0, bringing the total to 2,301 genomes (595 bacterial, 32 archaeal, 13 eukaryotic genomes, and 1,661 viruses) of which 2,058 are finished and 243 are draft.
- 79 finished and 98 draft genomes sequenced by JGI, bringing this total to 177 microbial genomes generated in-house.

IMG 2.0 extensions included gene-based links to NCBI's Entrez Gene, and other microbial genome systems, such as Lawrence Berkeley National Laboratory's MicrobesOnline and Argonne National Laboratory's PUMA. IMG is accessible to the public at http://img.jgi.doe.gov/.

IMG/M FOR METAGENOMICS

Integrated Microbial Genomes for Microbiome Samples (IMG/M) was released in 2006 as a metagenome data analysis and management system. IMG/M integrates data from diverse environmental microbial communities with isolated microbial genome data from JGI's IMG system, allowing the application of IMG's comparative analysis tools to metagenome data. New tools also enable the examination of functional annotation profiles across microbial communities and isolated organisms of interest, and the analysis of strain-level heterogeneity within a species population in metagenome data.

New Technology

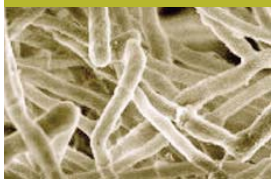
JGI's support comes primarily from the DOE Office of Science. This serves to support 106 capillary sequencing machines, 70 of which run 24 hours, seven days a week. Currently, on average, JGI is generating 3.1 billion bases of sequence per month, establishing JGI as one of the leaders in sequence generation, and with close to 400 genomes either in the works or sequenced, JGI leads all other centers worldwide. In addition, over the last year, two new Genome Sequencer 20 Systems, from 454 Life Sciences, make it possible for one person to prepare and sequence an entire genome without the time-consuming steps of cloning and colony picking. Running 100 times faster than the previous generation of sequencers used by JGI, the 454 instrument uses a parallel-processing approach to whole-genome shotgun sequencing to produce over 20 megabases (20,000,000 bases) of DNA sequence per 4.5-hour operating run. A prototype of another new platform from Solexa, based on sequencing-by-synthesis, has recently been installed at the JGI.

Jamborees 2005-2006

2005-2006 JAMBOREE ROUNDUP

Jamborees are working meetings at which members of a scientific community gather to discuss and annotate (assign function to) the genome of an organism (or family of organisms) of common interest. The goal is to identify genes and generate high-quality annotations, ultimately leading to publications in the scientific literature.

Trichoderma reesei
March 20-23, 2005
Asilomar, California



T. reesei is an industrially important cellulolytic filamentous fungus. In light of this organism's ability to secrete large amounts of cellulases and hemicellulases, DOE is supporting research into developing *T. reesei* as a host to produce low-cost enzymes for the conversion of plant biomass materials into industrially useful bioproducts such as sugars and bioethanol. This was followed by the second online jamboree using Access Grid Technology on November 7-8, 2005. Both Jamborees were organized by Diego Martinez, JGI-LANL.

Ostreococcus sp.
September 20-22, 2005 Walnut Creek, California

Ostreococcus belongs to the *Prasinophyceae*, an early-diverging class within the green plant lineage, and is reported as a globally abundant picoeukaryotic group (tiny algae) throughout the oceanic euphotic zone (the top of the water column, where green plants live). This project provided the opportunity to study the gene content of marine eukaryote species that are phylogenetically closely related but adapted to various ecological niches, a boon for the field of comparative and environmental genomics. This jamboree focused on comparative analysis of two species, *O.lucimarinus* and *O.tauri*, and was organized by Igor Grigoriev of JGI and Brian Palenik of Scripps Institute of Oceanography, San Diego.

Phaeodactylum tricornutum
March 22-24, 2006
Walnut Creek, California

Diatoms are eukaryotic photosynthetic microorganisms found throughout marine and freshwater ecosystems. Major players in the carbon cycle, diatoms are responsible for about 20% of global production of new biomass by photosynthesis.

Phaeodactylum tricornutum is the second diatom sequenced. This 30 million base genome, with the diatom *Thalassiosira pseudonana* (also sequenced by JGI), provides the basis for comparative genomics studies and will serve as a foundation for interpreting the ecological suc-

cess of these organisms. This Jamboree was organized by the principal investigator Chris Bowler of the Ecole Normale Superieure, and Igor Grigoriev of JGI.

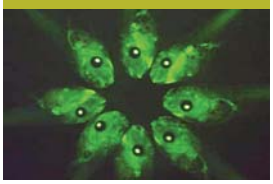
Laccaria bicolor
April 4-5, 2006
Nancy, France

Tree species dominating forest ecosystems in boreal, temperate, and montane regions (e.g., pines, oaks) develop symbiotic associations, so-called ectomycorrhizas, with soil fungi. The symbiotic fungus *Laccaria bicolor* is used in large-scale commercial inoculation programs in forest nurseries worldwide to enhance growth of tree seedlings. Characterization of the interactions between poplar and this fungus would allow in-depth exploration of the coordinated community response to these abiotic and biotic stresses, thus adding a needed dimension to climate change research, and providing another step in the quest for mechanistic modeling of ecosystem responses. This workshop was organized by the principal investigator Dr. Francis Martin of the INRA-Nancy.

Aspergillus niger
April 12-14, 2006
Vienna, Austria

Citric acid production by the filamentous ascomycete fungus *Aspergillus niger* represents the most efficient, highest yielding bioprocess in commercial practice. This process is a model for other filamentous fungal fermentation processes that will become a key part of DOE's vision of the biorefinery, where multiple products such as organic acids and ethanol are produced from renewable biomass. These products can be further refined for use as plastic monomers, solvents, or fuels, thereby decreasing dependence on petroleum, the traditional source of these products. As a common member of the microbial communities found in soils, *A. niger* also plays a significant role in the global carbon cycle. This jamboree was organized by Diego Martinez, JGI-LANL.

Xenopus tropicalis
April 24-28, 2006
Walnut Creek, California



Frogs and other amphibians have been referred to as the canary in the coal mine, and may provide an early warning to humans of environmental degradation or pending disaster. Surveys of local frog populations have been used to monitor water pollution around mining and industrial sites. The numbers of frogs (including *Xenopus tropicalis*) have been declining worldwide for the past several decades. Possible reasons for this decline include loss of habitat, pesticide pollution, parasitic infestations, and increased UV light exposure due to the decrease in the ozone layer. Having a better understanding of the frog genome through this project will give greater insight into and allow for a more detailed analysis of this troubling problem. The jamboree was organized by Paul Richardson of JGI; Richard Harland of University of California, Berkeley; and Robert Grainger of Washington University, St. Louis.

Mycosphaerella graminicola
June 7-9 2006
Walnut Creek, California

Mycosphaerella is one of the largest genera of plant pathogenic fungi, having more than 1,000 named species, many of which cause economically important diseases in temperate and tropical crops. Losses caused by *M. graminicola* cost United States wheat growers more than $275 million every year. The genus *Mycosphaerella* contains at least 10,000 species, making it the largest genus of plant pathogens. It causes extensive economic losses on a range of crops including tomatoes, strawberries, sugar beets, soybeans, and even tree crops. This jamboree was organized by Igor Grigoriev of JGI, Stephen Goodwin of USDA-ARS, and Gert Kema of the Plant Research International (Netherlands).

Physcomitrella patens
June 26-27, 2006
Walnut Creek, California

The moss *Physcomitrella patens* is becoming 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 evolution, especially genome evolution. *Physcomitrella* is well-placed phylogenetically to provide important comparisons with the flowering plants. The jamboree was organized by Jeffrey Boore of JGI; Brent Mishler of University of California, Berkeley; and Ralph Quatrano of Washington University, St. Louis.

Education and Outreach

The JGI Production Genomics Facility (PGF) is a magnet for educational and community partnerships. There is a strong interest on behalf of the PGF's neighbors, from high school students to retired citizens, in learning more about the power of genomics and the relevance it has for their health and that of the environment. Given the opportunity, visitors delight at witnessing industrial-scale DNA sequencing in action. For those unable to visit in person, the JGI Web site (http://www.jgi.doe.gov/index.html) offers detailed descriptions of the sequencing process and its applications. The PGF hosted more than 1,000 students from local high schools and colleges in 2005–2006.

Additional Education and Outreach activities included:

JGI Microbial Genomics Workshop for Educators

Every year, JGI reveals the wonders of DNA sequencing to thousands of avid students, teachers, researchers, and community members drawn through the portal of PGF. Visitors often ask the question: what becomes of those billions of letters of genetic information churning out of the DNA sequencers every month? On April 24-25, 2006 in the spirit of both Earth Day and National DNA Day, JGI hosted a Microbial Genomics Workshop geared toward revealing the rest of the story—the bioinformatics end—to educators who may then spread the word in their classrooms and inspire students to pursue careers in this burgeoning field. A team of JGI researchers demonstrated the computational methods that they and their collaborators around the world use to harness the potential of the largely untapped microbial world for the development of clean bioenergy alternatives, a better understanding of the global carbon cycle, and novel bioremediation applications.

The 2006 Hopkins Microbiology Course

The Hopkins Microbiology Course assumes the task of identifying unifying principles in microbial physiology, ecology, and evolution in order to understand the diversity of microbes observed in nature, specifically in environments characteristic to the Monterey Bay, including the nearby Elkhorn Slough. JGI staff participated in the development and teaching of the successful 2006 course.

The course lectures and labs in molecular microbial ecology techniques (construction of 16S rRNA gene clone libraries, construction and interpretation of phylogenetic trees, fluorescence *in situ* hybridization) and isolation of metabolically diverse microorganisms were applied to study the evolutionary forces shaping patterns of diversity within populations of marine *Vibrio* obtained from two Monterey Bay environments.

JGI in the Community

JGI volunteers also participated in the following:

- "Got Science," event hosted by Lawrence Livermore National Laboratory on October 21, 2006, at the Robert Livermore Community Center.
- Berkeley Lab Founders Day (pictured below), August 26, 2006, which marked the date 75 years ago that Ernest Orlando Lawrence received permission to open up a lab to pursue his studies of particle acceleration. To commemorate this occasion, the Lab hosted a daylong celebration for employees, retirees, and their families and friends, as well as special guests.
- The Contra Costa County Intel-affiliated Science & Engineering Fair, March 30-April 1, 2006.
- Orinda Intermediate School Family Science Night October 6, 2005.

LOOKING AHEAD

Genomics Education Partnership with UC Merced

The University of California, Merced, a UC campus with a significant student enrollment of minorities underrepresented in science, which opened in 2005, is the 10th UC campus and located about two hours' drive from the JGI Production Genomics facility. Former JGI researcher Mónica Medina is now on the faculty at UC Merced in the Natural Sciences Department and has been working with her colleagues at JGI to develop a Genome Biology course. Starting in January 2007, the course will run through the end of April and cover every aspect of the JGI sequencing and bioinformatics process. The lectures and labs will be conducted by JGI staff along with Mónica and her colleagues at UC Merced, with a view toward inspiring and preparing the next generation genomics workforce.

PROPOSED NEW IN-HOUSE EDUCATION CENTER

The JGI is designing an on-site genomics education center targeting hands-on education and training for both students and educators from high school through university graduate programs. The Center will entail an approximately 1,500 square-foot teaching space that can accommodate up to 24 students in a well-equipped, state-of-the art, molecular genetics laboratory. The undergraduate and graduate program will be developed and overseen by Cheryl A. Kerfeld, Ph.D., who recently joined JGI from the University of California, Los Angeles, where she was director of the UCLA Undergraduate Genomics Research Initiative. While at UCLA, she exploited the user facility resources of JGI to provide her students with an interdepartmental multi-course collaboration with the central theme of sequencing and analyzing the genome of a bacterium.

Safety/Ergonomics

Safety is of paramount importance throughout every activity of the JGI Production Genomics Facility. Due to the potential for repetitive strain injuries resulting from some of the tasks associated with the sequencing line, JGI has been vigilant about conducting proactive safety and ergonomic assessments in all workstations.

Ergonomics is about getting a good match between the work we do and what our bodies can handle. When there's a mismatch, it can lead to aches and pains, inefficiencies, and even to serious injuries if not promptly addressed. Unfortunately, ergonomics-related problems are now the most common type of injury at JGI, and for this reason, improving ergonomics has become a major focus here. JGI has designed and implemented an ergonomic "warm up par course" at the Production Genomics Facility, offering employees the opportunity for stretch breaks.

Genomics Glossary

ANNOTATION: The process of identifying the locations of genes in a genome and determining what those genes do.

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.

BAC: (Bacterial Artificial Chromosome) An artificially created chromosome in which large segments of foreign DNA (up to 150,000 bp) are cloned into bacteria. Once the foreign DNA has been cloned into the bacteria's chromosome, many copies of it can be made and sequenced.

BASE: A unit of DNA. There are 4 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.

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.

ELECTROPHORESIS: A process by which molecules (such as proteins, DNA, or RNA fragments) can be separated according to size and electrical charge by applying an electric current to them. Each kind of molecule travels through a matrix at a different rate, depending on its electrical charge and molecular size.

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.

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

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 relatively new field of genetic research allows the genomic study of organisms that are not easily cultured in a laboratory.

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.

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.

RCA: Acronym for Rolling Circle Amplification, a randomly primed method of making multiple copies of DNA fragments, which employs a proprietary polymerase enzyme and does not require the DNA to be purified before being added to the sequencing reaction.

READ LENGTH: The number of nucleotides tabulated by the DNA analyzer per DNA reaction well.

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.

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.

VECTOR: DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vector's capacity for self-replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, Bacterial Artificial Chromosomes (BACs), or Yeast Artificial Chromosomes (YACs).

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.

2005–2007 Community Sequencing Program

	ORGANISM	COLLABORATOR	INSTITUTION
2005			
MICROBES	*Olavius algarvensis*	Dubilier	Max Planck Inst. of Marine Microbiology
	Crenarchaeota	Delong	Massachusetts Inst. of Technology
	Marinobacter aquaeolei	Edwards	Woods Hole Oceanographic Inst.
	Rhodocyclus-like polyphosphate	Hugenholtz	Joint Genome Inst.
	Rhodobacter	Kaplan	Univ. of Texas, Houston
	Contaminated groundwater	Zhou	Oak Ridge National Laboratory
	Lactobacillus reuteri (2 strains)	Tannock	Univ. of Otago, Dunedin, NZ
	Bacillus cereus	Sorokin	INRA, France
	Synechococcus (2 strains)	Palenik	Scripps Inst. of Oceanography
	Syntrophobacter fumaroxidans	McInerney	Univ. of Oklahoma
	Thermoanaerobacter ethanolicus	Fields	Miami Univ.
	Thiomicrospira (2 strains)	Scott	Univ. of South Florida
BASAL ORGANISMS	*Selaginella moellendorffii*	Banks	Purdue Univ.
	Trichoplax adhaerens	DellaPorta	Yale Univ.
	Sporobolomyces roseus	Wolfe	Trinity College, Dublin
	Reniera	Degnan	Univ. of Queensland, Australia
	Mycosphaerella 2 sp.	Goodwin	Purdue Univ.
	Spironucleus vortens	Cande	Univ. of California, Berkeley
	Naegleria gruberi	Cande	Univ. of California, Berkeley
HIGHER ANIMALS & PLANTS	*Physcomitrella patens*	Mishler	Univ. of California, Berkeley
	Lottia gigantea (limpet)	Edsinger-Gonzalez	Univ. of California, Berkeley
	Helobdella robusta (leech)	Weisblat	Univ. of California, Berkeley
	Capitella capitata	Cande	Univ. of California, Berkeley
ESTS & TARGETED SEQUENCING	*Alvinella pompejana*	Tainer	Scripps Research Inst.
	Mitochondria seed plant	Palmer	Indiana Univ., Bloomington
	Karenia brevis	Bhattacharya	Univ. of Iowa
	Dipteran fosmids	Eisen	Lawrence Berkeley National Laboratory
2006			
EUKARYOTES	*Sorghum sp.*	Paterson	Univ. of Georgia
	Arabidopsis lyrata	Weigel	Max Planck Inst. of Developmental Biology
	Mimulus guttatus	Willis	Duke Univ.
	Piromyces sp. E2	Baker	Pacific Northwest National Laboratory
	Hydractinia symbiolongicarpus	Buss	Yale Univ.
	Phycomyces blakesleeanus	Corrochano	Univ. of Seville
	Xanthoria parietina	Crittendon	Univ. of Nottingham
	Trichoderma virens	Ebbole	Texas A&M Univ.
	Mycosphaerella fijiensis	Goodwin	US Dept. of Agriculture—ARS, Purdue Univ.
	Phytophthora capsici	Kingsmore	National Center for Genome Resources
	Campanulales	Knox	Indiana Univ.
	Cichlid Lake Malawi	Kocher	Univ. of New Hampshire
	Ciona intestinales	Lemaire	CNRS, France
	Ictalurus punctatus	Liu-J	Auburn Univ.
	Ictalurus furcatus	Liu-J	Auburn Univ.
	Melampsora larici-populina	Martin	Inst. National de la Recherche Agronomique
	Ostreococcus low-light strain	Palenik	Univ. of California, San Diego
	Parhyale hawaiensis	Patel	Univ. of California, Berkeley
	Jassa slatteryi	Patel	Univ. of California, Berkeley
	Petrolisthes cinctipes	Stillman	Univ. of Hawaii
	Batrachochytrium dendrobatidis	Taylor	Univ. of California, Berkeley
	Triphysaria	Yoder	Univ. of California, Davis

2005–2007 Community Sequencing Program

	ORGANISM	COLLABORATOR	INSTITUTION
2006 *cont.*			
BACTERIA AND ARCHAEA	*Euryarchaeota* community	Baker	Univ. of California, Berkeley
	Polynucleobacter	Hahn	Inst. for Limnology, Austria
	Microbial soil (community Alaska)	Handelsman	Univ. of Wisconsin, Madison
	Salinospora tropicalis	Jensen	Scripps Inst. of Oceanography
	Salinospora arenicola	Jensen	Scripps Inst. of Oceanography
	Termite gut microbial (community)	Leadbetter	California Inst. of Technology
	Terephthalate (TA) (community)	Liu	National Univ. of Singapore
	Archaeal (sp. hyperthermo.)	Lowe	Univ. of California, Santa Cruz
	Bacterioplankton (Antarctic marine)	Murray	Desert Research Inst.
	Thermotogales (7 hyperthermo.)	Noll	Univ. of Connecticut
	Nitrosomonas	Norton	Utah State Univ.
	Microbial mats (hypersaline)	Pace	Univ. of Colorado
	Sinorhizobium medicae	Reeve	Murdoch Univ.
	Verrucomicrobium	Schmidt	Michigan State Univ.
	Bacillus coagulans	Shanmugam	Univ. of Florida
	Crenarchaeote (community)	Simon	Oregon Health & Science Univ.
	Verrucomicrobia (5 strains)	Smidt	Wageningen Univ.
	Acidovorax (sp.)	Stahl	Univ. of Washington
	Caulobacter (2 strains)	Stephens	Santa Clara Univ.
	Korarchaeota (community)	Stetter	Univ. Regensberg, Diversa Corp.
	Archaea (6 strains)	Woese	Univ. of Illinois, Urbana-Champaign
2007			
LARGE EUKARYOTES	*Brachypodium distachyon* (Poaceae)	Vogel	USDA-ARS
	Aquilegia formosa	Hodges	UC Santa Barbara
	Gossypium (cotton)	Paterson	Univ. of Georgia
	Manihot esculenta (cassava)	Fauquet	Danforth Plant Science Ctr.
SMALL EUKARYOTES	*Guillardia theta* and *Bigelowiella natans*	Archibald	Dalhousie Univ.
	Mating loci from *Volvox carteri* & *Chlamydomonas reinhardtii*	Umen	Salk Inst.
	Fragilariopsis cylindrus (a diatom)	Mock	Univ. of Washington
	Pleurotus ostreatus (oyster mushroom)	Pisabarro	Public Univ. of Navarre
	Thellungiella halophila	Schumaker	Univ. of Arizona
	Reef-building corals and dinoflagellate symbionts	Medina	Univ. of California, Merced
	Heterobasidion annosum	Stenlid	Swedish Univ. of Agricultural Sciences
	Peronosporomycete mtDNAs (26)	Hudspeth	Northern Illinois Univ.
	Switchgrass	Tobias	USDA-ARS
	Cryphonectria parasitica (chestnut blight fungus)	Nuss	Univ. of Maryland Biotech. Inst.
	Three species of *Neurospora*	Taylor	Univ. of California, Berkeley
BACTERIA AND ARCHAEA	*Methanomicrococcus blatticola*	Hackstein	Radboud Univ. Nijmegen
	Rhizobium leguminosarum bv trifolii (strains WSM1325 and WSM2304)	Reeve	Murdoch Univ.
	Methylocella silvestris BL2, Methylocapsa acidiphila B2, & *Beijerinckia* indica subsp. indica	Dunfield	Inst. of Geological & Nuclear Sciences, New Zealand
	Pedomicrobium manganicum	Mackenzie	Univ. of Texas, Houston
	Lithifying mat communities of marine stromatolites (6 bacterial strains)	Decho	Univ. of South Carolina
	Macropus eugenii	McSweeney	CSIRO
	Actinobacteria	Jansson	Swedish Univ. of Agricultural Sciences
	Anaerobic benzene-degrading methanogenic consortium	Edwards	Univ. of Toronto
	Haloalkaliphilic sulfur-oxidizing bacteria	Muyzer	Delft Univ. of Technology
	Dechlorinating community (KB-1)	Edwards	Univ. of Toronto
	Six freshwater iron-oxidizing bacteria	Emerson	Amer. Type Culture Collection
	Beggiatoa alba	Mueller	Morgan State Univ.
	Near-shore anoxic basin: Saanich Inlet	Hallam	Univ. of British Columbia

2005–2007 Community Sequencing Program

ORGANISM		COLLABORATOR	INSTITUTION
2007*cont.*			
BACTERIA AND ARCHAEA *cont.*	*Candidatus Amoebophilus asiaticus* and *C. Cardinium hertigii*	Horn	Univ. of Vienna
	Candidatus Endomicrobium trichonymphae, free-living strain Pei191	Brune	Max Planck Institute for Terrestrial Microbiology
	Burkholderia	Tiedje	Michigan State Univ.
	Thermolithobacter ferrireducens	Wiegel	Univ. of Georgia
	Crenothrix polyspora enrichment	Wagner	Univ. of Vienna
	Six *Cyanothece* strains	Pakrasi	Washington Univ.
	Thauera sp. MZ1T	Sayler	Univ. of Tennessee
	Microbial community in wastewater treatment plants	van der Meer	Univ. of Lausanne
	Symbiont from the basal clade of the Frankiaceae	Benson	Univ. of Connecticut

2006–2007 Laboratory Science Program

ORGANISM	PROPOSER	AFFILIATION	STATUS*	AMOUNT PROPOSED (MB)
ChIP-enriched binding sequences	Stubbs	LLNL	Pending	33.0
New Orleans Floodwaters	Andersen	LBNL	In Production	10.0
Columbia River Microbiota	Magnuson	PNNL	Pending	5.0
Microbial diversity in tropical and temperate forests	Pett-Ridge	LLNL	In production	10.8
Tallgrass Prairie soil	Bailey	PNNL	In Production	10.0
Low Dose DiTags	Anderson	BNL	Pending	180.0
Aspergillus terreus EST	Baker	PNNL	Pending	10.0
Tremella mesenterica	Heitman	Duke University	Pending	168.0
Thielavia terrestris	Berka	Novozymes	Pending	294.0
Cochliobolus heterostrophus	Turgeon	CornellUniversity	Pending	336.0
Soluble protein domains	Terwilliger	LANL	Pending	6.5
Viruses in nasopharyngeal swab samples from healthy subjects	Mariella	LLNL	Pending	10
Five Archaea *(Methanohalobium evestigatum, Methanobacterium formicicum, Halococcoides, Ferroglobus placidus,* and *Acidianus sp. JP7)*	Kyrpides	LBNL	Pending	50
Microbial community actively decaying poplar biomass	van der Lelie	BNL	In production	200

DOE Microbial Genome Projects

GENUS	SPECIES	COLLABORATOR	INSTITUTION
2002			
Dechloromonas	*aromatica*	Coates	Univ. of California, Berkeley
Desulfuromonas	*acetoxidans*	Lovely	Univ. of Massachusetts
Ehrlichia	2 strains	McBride	Univ. of Texas, Medical Branch
Geobacter	*metallireducens*	Lovely	Univ. of Massachusetts
Methanococcoides	*burtonii*	Sowers	Univ. of Maryland
Pseudomonas	*syringae*	Lindow	Univ. of California, Berkeley
Psychrobacter	sp.	Tiedje	Michigan State Univ.
Ralstonia	*eutropha*	Gonzalez	Pontificia Univ. Catolica de Chile
Streptococcus	*suis*	Gottschalk	Univ. of Montreal, Canada
2003			
Burkholderia	2 strains	Tiedje	Michigan State Univ.
Methylobium	*petroleophilum*	Kane	Lawrence Livermore National Laboratory
Prochlorococcus	sp.	Chisholm	Massachusetts Institute of Technology
Synechococcus	*elongatus*	Golden	Texas A&M Univ.
2004			
Burkholderia	2 strains	Tiedje	Michigan State Univ.
Clostridium	*phytofermentans*	Leschine	Univ. of Massachusetts, Amherst
Frankia	sp.	Tisa	Univ. of New Hampshire
Nitrobacter	*hamburgensis*	Arp	Oregon State Univ.
Nitrobacter	*winogradskyi*	Arp	Oregon State Univ.
Nitrosococcus	*oceani*	Arp	Oregon State Univ.
Nitrosomonas	*eutropha*	Arp	Oregon State Univ.
Nitrosospira	*multiformis*	Arp	Oregon State Univ.
Prochlorococcus	sp.	Chisholm	Massachusetts Institute of Technology
Shewanella	2 strains	Fredrickson	Pacific Northwest National Laboratory
Synechococcus	2 strains	Palenik	Scripps Institution of Oceanography
Syntrophobacter	*fumaroxidans*	McInerney	Univ. of Oklahoma
Thermoanaerobacter	*ethanolicus*	Fields	Miami Univ.
Thiomicrospira	2 strains	Scott	Univ. of South Florida
2005			
Acidiphilium	*cryptum*	Magnuson	Idaho State Univ.
Acidobacterium	Ellin345	Kuske	Los Alamos National Laboratory
Acidothermus	*cellulolyticus*	Berry	Univ. of California, Davis
Actinobacillus	*succinogenes*	Vieille	Michigan State Univ.
Aspergillus	*niger*	Baker	Pacific Northwest National Laboratory
Aureococcus	*anophageggerens*	Gobler	Southhampton College of Long Island Univ.
Bacillus	*selenitireducens*	Stolz	Duquesne Univ.
Bradyrhizobium	sp.	Sadowsky	Univ. of Minnesota
Burkholderia	*ambifaria*	Tiedje	Michigan State Univ.
Caldicellulosiruptor	*saccharolyticus*	Kelly	North Carolina State Univ.
Calyptogena	*magnifica*	Cavanaugh	Harvard Univ.
Chloroflexus	*aggregans*	Bryant	Pennsylvania State Univ.
Chloronema	sp.	Bryant	Pennsylvania State Univ.
Chlorothrix	*halophila*	Bryant	Pennsylvania State Univ.
Clostridium	sp.	Stolz	Duquesne Univ.
Dehalococcoides	2 strains	Spormann	Stanford University
Desulfotomaculum	*reducens*	Tebo	Univ. of California, San Diego
Flavobacterium	*johnsoniae*	McBride	Univ. of Texas, Medical Branch
Geobacter	sp.	Kostka	Florida State Univ.
Halorhodospira	*halophila*	Hoff	Univ. of Chicago
Heliothrix	*oregonensis*	Bryant	Pennsylvania State Univ.
Herpetosiphon	*aurantiacus*	Bryant	Pennsylvania State Univ.

DOE Microbial Genome Projects

GENUS	SPECIES	COLLABORATOR	INSTITUTION
2005 cont.			
Iron Mountain AMD	Site 1	Banfield	Univ. of California, Berkeley
Iron Mountain AMD	Site 2	Banfield	Univ. of California, Berkeley
Lake Washington	formaldehyde	Lidstrom	Univ. of Washington
Lake Washington	formate	Lidstrom	Univ. of Washington
Lake Washington	methane	Lidstrom	Univ. of Washington
Lake Washington	methanol	Lidstrom	Univ. of Washington
Lake Washington	methylamine	Lidstrom	Univ. of Washington
Methanosaeta	*thermophila*	Smith	Clemson Univ.
Micromonas	2 strains	Worden	Univ. of Miami
Mono Lake deltaproteobacter		Stolz	Duquesne Univ.
Mono Lake gammaproteobacter		Stolz	Duquesne Univ.
Mycobacterium	5 strains	Miller	Utah State Univ.
Nectria	*haematococca*	VanEtten	Univ. of Arizona
Obsidian Hot Spring		Mead	Lucigen
Phycovirus	11 strains	Wommack	Delaware Biotechnology Institute
Polaromonas	*napthalenivorans*	Madsen	Cornell Univ.
Postia	*placenta*	Cullen	U.S. Department of Agriculture
Pseudoalteromonas	*atlantica*	Karls	Univ. of Georgia
Pseudomonas	*putida*	Parales	Univ. of California, Davis
Psychromonas	*ingrahamii*	Staley	Univ. of Washington
Rhodopseudomonas	4 strains	Harwood	Univ. of Iowa
Roseiflexus	2 strains	Bryant	Pennsylvania State Univ.
Shewanella	7 strains	Fredrickson	Pacific Northwest National Laboratory
The Cedars Alkaline Springs		Nealson	Univ. of Southern California
Ultra back	C level 1	Banfield	Univ. of California, Berkeley
Xanthobacter	*autotrophicus*	Ensign	Los Alamos National Laboratory

BERAC Review of JGI

In July of 2006, the DOE Biological and Environmental Research Advisory Committee (BERAC) completed a performance review of JGI that focused on science, management, and operations. The BERAC Review Committee included members representing three of the major sequencing centers in the world—the Broad Institute, Boston; the Wellcome Trust Sanger Institute, Cambridge, England; and the Genome Sequencing Center at Washington University, St. Louis. In addition, the committee included experts in microbial ecology and genomic computation as well as a number of high-ranking managers from several national laboratories and DOE facilities, who reviewed management and operations aspects of the JGI.

In summary, the Committee's general assessment was that JGI is a well-run, highly productive and efficient facility, representing a major asset in DOE's portfolio, engaged in first-rate science, contributing to pioneering research in the areas of energy, carbon cycling, and bioremediation. The Committee was supportive of the management, and acknowledged the effectiveness of the team operating JGI.

The Committee's report can be downloaded from: http://www.sc.doe.gov/ober/berac/JGI_review.pdf

BERAC Committee members are listed below.

SUBCOMMITTEE MEMBERS AND OBSERVERS FOR THE BERAC COMMITTEE SITE VISIT

Mel Simon (Chair)
Biaggini Professor of Biology
California Institute of Technology

Bruce Birren
Co-Director, Sequence and Analysis Program
Broad Institute, Massachusetts Institute of Technology

Klaus Berkner
Ex-Deputy Director for Operations
Lawrence Berkeley National Laboratory (retired)

Bruce Chrisman
Associate Director, Administration
Fermilab

Linda Horton
Director, Center for Nanophase Materials Sciences, and
Materials and Engineering Physics Program
Oak Ridge National Laboratory

Richard Mural
Chief Scientific Officer
Windber Research Institute

Jane Rogers
Director of Sequencing
The Wellcome Trust Sanger Institute

Jim Tiedje
Center for Microbial Ecology
Michigan State University

Richard Wilson
Director, Genome Sequencing Center
Washington University St. Louis

BER STAFF

Dan Drell
Life Sciences Division, Office of Biological and
Environmental Research
Office of Science, US Department of Energy

Kent Lohman
Life Sciences Division, Office of Biological and
Environmental Research
Office of Science, US Department of Energy

NHGRI OBSERVER FOR BERAC SITE VISIT

Jane Peterson
Associate Director, Division of Extramural Research
National Human Genome Research Institute, National
Institutes of Health

Review Committees and Board Members

JGI Policy Board

The JGI Policy Board serves two primary functions:

1. To serve as a visiting committee to provide advice on policy aspects of JGI/PGF operations and long-range plans for the program, including the research and development necessary to ensure the future capabilities that will meet DOE mission needs.
2. To ensure that JGI/PGF resources are utilized in such a way as to maximize the technical productivity and scientific impact of the JGI now and in the future. The JGI Policy Board meets annually to review and evaluate the performance of the entire JGI, including its component tasks and leadership. It reports its findings and recommendations to the participating Laboratory Directors and to the DOE BER.

MEMBERS

Gerry Rubin
Howard Hughes Medical Institute, Chair

Professor Sallie W. (Penny) Chisholm
Massachusetts Institute of Technology

Stephen Quake
Stanford University

Dr. David Galas
Vice President and Chief Scientific Officer, Battelle

Ed DeLong
Massachusetts Institute of Technology

Professor Richard Gibbs
Baylor College of Medicine

Susan Wessler
University of Georgia

Chris Somerville
Stanford University

James Tiedje
Michigan State University

Professor Melvin Simon
California Institute of Technology

JGI Scientific Advisory Committee

The Scientific Advisory Committee (SAC) is a board that the JGI Director convenes to provide a scientific and technical overview of the JGI/PGF. Responsibilities of this board include providing technical input on large-scale production sequencing, new sequencing technologies, and related informatics; overview of the scientific programs at the JGI/PGF; and overview of the Community Sequencing Program (CSP). A crucial job of the committee is to take the input from the CSP Proposal Study Panel on prioritization of CSP projects and, with BER's concurrence, set the final sequence allocation for this program.

MEMBERS

Mark Adams
Case Western Reserve University

Ginger Armbrust
University of Washington

Joe Ecker
Salk Institute

Ed DeLong
Massachusetts Institute of Technology

Bruce Birren
Broad Institute

Eric J. Mathur
Synthetic Genomics

Jim Krupnick
Lawrence Berkeley National Laboratory

George Weinstock
Baylor College of Medicine

Barbara Wold
California Institute of Technology

Marco Marra
Simon Fraser University

Review Committees and Board Members

JGI User Committee

The JGI User Committee provides guidance for organizing the various JGI user programs and activities, including the JGI User Meeting.

MEMBERS

Ed DeLong
Chair, Massachusetts Institute of Technology

Jerry Tuskan
Oak Ridge National Laboratory

Andy Paterson
University of Georgia

David Mead
Lucigen Corporation

Cheryl Kerfeld
University of California, Los Angeles

David Mills
University of California, Davis

Scott Baker
Pacific Northwest National Laboratory

Jared Leadbetter
California Institute of Technology

The CSP Proposal Study Panel

The CSP Proposal Study panel reviews investigator-initiated JGI sequencing requests for scientific merit.

MEMBERS

Nina Agabian
University of California, San Francisco

Chris Amemiya
Benaroya Research Institute at Virginia Mason

Gary L. Andersen
Lawrence Berkeley National Laboratory

Jo Ann Banks
Purdue University

John Battista
Louisiana State University

Fred Brockman
Pacific Northwest National Laboratory

Zac Cande
University of California , Berkeley

Patrick Chain
Lawrence Livermore National Laboratory

Jonathan C. Cohen
UT Southwestern Medical Center

Nigel Dunn-Coleman
Genencor International

Joe Ecker
The Salk Institute for Biological Studies

Katrina Edwards
Woods Hole Oceanographic Institution

Kelly Frazer
Perlegen Sciences, Inc.

Richard Harland
University of California, Berkeley

Derek Lovley
University of Massachusetts

David Mills
University of California, Davis

Review Committees and Board Members

Alison Murray
Desert Research Institute

Arend Sidow
Stanford University

Nipam Patel
University of California, Berkeley

Karin Remington
Venter Institute

John Taylor
University of California, Berkeley

Naomi Ward
The Institute for Genomic Research

Bart Weimer
Utah State University

JGI EX OFFICIO MEMBERS

Jim Bristow
CSP Chairman, DOE Joint Genome Institute

Paul Richardson
DOE Joint Genome Institute

Dan Rokhsar
DOE Joint Genome Institute

Eddy Rubin
Director, DOE Joint Genome Institute

DOE REPRESENTATIVE

Daniel Drell
U.S. Department of Energy

Genomic Technologies for Improving Bioenergy Feedstocks Meeting

January 18-19, 2007
This meeting is intended to engage the JGI in providing genomic resources for ongoing national feedstock programs. In addition to DNA sequencing needs, this workshop seeks to identify what other high throughput technologies that the JGI should consider developing to aid in more effectively moving this field forward.

Ed Buckler
Cornell University

Brian Diers
University of Illinois

Steve Difazio
University of West Virginia

Joe Ecker
Salk Institute

Ken Feldmann
Ceres

Steve Goff
Syngenta

Sarah Hake
University of California, Berkeley

Ed Kaleikau
U.S. Department of Agriculture

Elliott Meyerowitcz
California Institute of Technology

Dana Nelson
University of California, San Francisco

Andy Paterson
University of Georgia

Pam Ronald
University of California, Davis

German Spangenberg
Victorian AgriBiosciences Centre

Gary Stacey
University of Missouri

Tim Tschaplinski
Oak Ridge National Laboratory

John Vogel
U.S. Department of Agriculture

Sharlene Weatherwax
US Department of Energy

Jerry Tuskan
Oak Ridge National Laboratory

Len Pennacchio
DOE Joint Genome Institute

Paul Richardson
DOE Joint Genome Institute

Dan Rokhsar
DOE Joint Genome Institute

Jim Bristow
DOE Joint Genome Institute

Review Committees and Board Members

Computational Review

February 14-15, 2007
This group's goal is to comment on and provide guidance for optimizing the various JGI informatic activities.

David Jaffe
Broad Institute

Bill Pearson
University of Virginia

Steve Lincoln
Affymetrix

Toby Bloom
Broad Institute

Peter Cartwright
SRI International

Steve Jones
BC Cancer Genome Center

Jonathan Eisen
University of California, Davis

Asif Chinwalla
Washington University

Suzanna Lewis
Lawrence Berkeley National Laboratory

JGI Publications 2005–2006

2005

Allen, E.E., et al.
Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology* 2005, 3: 489–498.

Banfield, J.F., et al.
Proteogenomic approaches for the molecular characterization of natural microbial communities. *OMICSI* 2005, 9: 301–333.

Baroukh N., et al.
Comparative genomic analysis reveals a distant liver enhancer upstream of the COUP-TFII gene. *Mammalian Genome* 16(2): 91–95.

Barns, S.M., et al.
Detection of diverse new Francisella-like species in environmental samples. *Applied and Environmental Microbiology* 2005, 71: 5494–5500.

Bergmann, D.J., et al.
Structure and sequence conservation of genes in the hao cluster of autotrophic ammonia-oxidizing bacteria: Evidence for their evolutionary history. *Applied and Environmental Microbiology* 2005, 71: 5371–5382.

Bertilsson, S., et al.
Release of dissolved organic matter by *Prochlorococcus*. *Vie et Milieu (Life + Environment)* 2005, 55(3–4): 225–231.

Boore, J.L., et al.
Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Genomics Division* 2005, 22: 2104–2112.

Boore, J.L., et al.
The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). *Gene* 2005, May 9, 350 (2): 117–128.

Boore, J.L., et al.
The first complete mitochondrial genome sequences for stomatopod crustaceans: Implications for phylogeny. *Molecular Phylogenetics and Evolution*, Elsevier.

Boore J.L., et al.
Sequencing and comparing whole mitochondrial genomes of animals. In *Molecular Evolution: Producing the Biochemical Data, Part B.* E.A. Zimmer and E. Roalson, Eds., Methods in Enzymology Series 395, Elsevier: Burlington, Massachusetts, pp 311–348.

Boore, J.L., et al.
Rolling circle amplification of metazoan mitochondrial genomes. *Molecular Phylogenetic Evolution* 2005, December 13.

Braun, T.F., et al.
Flavobacterium johnsoniae gliding motility genes identified by mariner mutagenesis. *Journal of Bacteriology* 2005, 187: 6943–6952.

Buchan, A., et al.
An overview of the marine Roseobacter lineage. *Applied and Environmental Microbiology* 2005, 71: 5665–5677.

Cameron, A., et al.
Unusual gene order and organization of the sea urchin hox cluster. *Journal of Experimental Zoology, Part B: Mol. Dev. Evol.* 2005, October 11, 306B(1).

Carapelli, A., et al.
Relationships between hexapods and crustaceans based on mitochondrial genomics. In *Crustacean and Arthropod Relationships.* S. Koenemann, R.A. Jenner, Eds., 2005, pp 295–306.

Chain, P.S.G., et al.
Whole-genome analyses of the speciation events in the pathogenic Brucellae. *Infect. Immun.* 2005, 73: 8353–8361.

Claude, W., et al.
Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 2005, 22(10): 1948–1963.

Clerico, E.M., et al.
Specialized techniques for site-directed mutagenesis in cyanobacteria. In *Methods in Molecular Biology.* E. Rosato, Ed., Humana Press: Totowa, NJ, 2005.

Conlan, S., et al.
Rhodopseudomonas palustris regulons detected by cross-species analysis of alphaproteobacterial genomes. *Applied and Environmental Microbiology* 2005, 71:7442–7452.

Cottrell, M.T., et al.
Sequence and expression analyses of Cytophaga-like hydrolases in a Western arctic metagenomic library and the Sargasso Sea. *Applied and Environmental Microbiology* 2005, 71: 8506–8513.

Couronne, O., et al.
Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. *Human Molecular Genetics* 2005, 14(20): 3057–3063.

Dehal, P., et al.
Two rounds of genome duplication in the ancestral vertebrate genome. *PLoS Biology* 3(10): e314.

De Tomaso, A.W., et al.
Isolation and characterization of a protochordate histocompatibility locus. *Nature* 2005, 438: 454–459.

Doyle, S.A., et al.
High-throughput cloning for proteomics research. *Methods Mol. Biol.* 310: 107–113.

Doyle, S.A., et al.
High-throughput purification of hexahistidine-tagged proteins expressed in *E. coli*. *Methods Mol. Biol.* 310: 123–130.

Doyle, S.A., et al.
Screening for soluble expression of recombinant proteins. In *Chemical Genomics: Reviews and Protocols,* J. Walker, Ed., The Humana Press, Inc.

Dubchak, I., et al.
SNP-VISTA: An interactive SNPs visualization tool. *BMC Bioinformatics* 6: 292.

Evans, K., et al.
A bacteriophytochrome regulates the synthesis of LH4 complexes in *Rhodopseudomonas palustris*. *Photosynth. Res.* 2005, 85:169–180.

Feil, H., et al.
Comparison of the complete genome sequences of *Pseudomonas syringae pv. syringae B728a* and pv. tomato DC3000. *Proc. Natl. Acad. Sci.* USA 102: 11064–11069.

Francino, M.P., et al.
An adaptive radiation model for the origin of new gene functions. *Nature Genetics* 2005 June, 37 (6): 573–577.

APPENDIX F

Frigaard, N.U, et al.
Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 439: 847–850.

Gatesy, J., et al.
Hidden likelihood support in genomic data: Can forty-five wrongs make a right? *Systematic Biology* 2005, 54 (3): 483–492.

Genoux, A., et al.
Transcriptional regulation of apolipoprotein A5 gene expression by the nuclear receptor RORα. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2005, 25(6): 1186–92.

Giraud, E., et al.
A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodopseudomonas palustris. J. Biol. Chem.* 2005, 280: 32389–32397.

Grosskopf, I., et al.
Apolipoprotein A-V deficiency results in marked hypertriglyceridemia attributable to decreased lipolysis of triglyceride-rich lipoproteins and removal of their remnants. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2005, 25(12): 2573–2579.

Harrison, F.H., et al.
The pimFABCDE operon from *Rhodopseudomonas palustris* mediates dicarboxylic acid degradation and participates in anaerobic benzoate degradation. *Microbiology* 2005, 151:727–736.

Hood, L., et al.
Unusual gene order and organization of the sea urchin hox cluster. *Journal of Experimental Zoology* 2005, 306B (1): 45–58.

Jansen R.K., et al.
Methods for obtaining and analyzing whole chloroplast genome sequences. *Molecular Evolution: Producing the Biochemical Data, Part B (Methods in Enzymology)* 2005, 395: 348–384.

Klaenhammer, T.R., et al.
Genomic features of lactic acid bacteria effecting bioprocessing and health. FEMS *Microbiology Reviews* 29: 393–409.

Keys, D.N., et al.
A saturation screen for cis-acting regulatory DNA in the Hox genes of *Ciona intestinalis. Proc. Natl. Acad. Sci.* 2005, January 18, 102 (3): 679–683.

Kropat, J., et al.
A regulator of nutritional copper signaling in *Chlamydomonas* is an SBP-domain protein that recognizes the GTAC core of the CuRE. *Proc. Natl. Acad. Sci.* USA 2005, 102: 18730–18735.

Kyrpides, N.C., et al.
The Integrated Microbial Genomes (IMG) System. *Nucleic Acids Research*, 34.

Laird, D.J., et al.
Stem cells are units of natural selection in a colonial ascidian. *Cell* 2005, 123: 1351–1360.

Laird, D.J., et al.
Predatory stem cells in the non-zebrafish chordate, *Botryllus schlosseri. Zebrafish* 2005, 1: 357–361.

Leebens-Mack, J., et al.
Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 2005, 22: 1948–1963.

Lindell, D., et al.
Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 2005, 438: 86–89.

Macey, J.R., et al.
The complete mitochondrial genome of a gecko and the phylogenetic position of the Middle Eastern *Teratoscincus keyserlingii. Molecular Phylogenetics and Evolution* 2005 July, 36(1): 188–193.

Macey, J.R., et al.
Plethodontid salamander mitochondrial genome: A parsimony evaluation of character conflict and implications for historical biogeography. *Cladistics* 2005, 21: 194–202.

Mackey, S.R., et al.
Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. In *Methods in Molecular Biology*, E. Rosato, Ed., Humana Press: Totowa, NJ, 2005 (in press).

Magie, C.R., et al.
Genomic inventory and expression of Sox and Fox genes in the cnidarian *Nematostella vectensis. Dev. Genes Evol.* 2005, 215: 618–630.

Marçais, C., et al.
ApoAV Q139X truncation predisposes to late-onset hyperchylomicronemia due to lipoprotein lipase impairment. *The Journal of Clinical Investigation* 2005, 115(10): 2862–2869.

Martin, B.L., et al.
A novel role for lbx1 in *Xenopus* hypaxial myogenesis. *Development* 2005, 133: 195–208.

Medina, M., et al.
Genomes, phylogeny and evolutionary systems biology. *Proc. Natl. Acad. Sci.* 2005, 102: 6630–6635.

Medina M, et al.
(2005) Phylogeny of the sea hares in the Aplysia clade based on mitochondrial DNA sequence data. *Bulletin of Marine Science*, 76(3): 691–698, 2005.

Merkel, M., et al.
Apolipoprotein A5 accelerates plasma hydrolysis of triglyceride rich lipoproteins by interaction with heparan sulfate proteoglycan bound lipoprotein lipase. *Journal of Biological Chemistry* 2005, 280(22): 21553–21560.

Meyer, T.E., et al.
Stabilization against hyperthermal denaturation through increased CG content can explain the discrepancy between whole genome and 16S rRNA analyses. *Biochem.* 2005, 44: 11458–11465.

Mills, D.A., et al.
Genomic analysis of *Oenococcus oeni* PSU-1 and its relevance to winemaking. *FEMS Microbiology Reviews* 2005, 29: 465–475.

Mueller, R.L., et al.
Molecular mechanisms and evolutionary causes of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Molecular Biology and Evolution* 2005, 22(10): 2104–2112.

Noonan, J.P., et al.
Genomic sequencing of Pleistocene cave bears. *Science* 2005, July 22, 309 (5734): 597–600.

Nowak, M., et al.
Insulin-mediated down-regulation of apolipoprotein A5 gene expression through the phosphatidylinositol 3-kinase pathway: Role of upstream stimulatory factor. *Molecular and Cellular Biology* 2005 25(4):1537–1548.

Oda, Y., et al.
Functional genomic analysis of three nitrogenase isozymes in *Rhodopseudomonas palustris. J. Bacteriol.* 2005, 187: 7784–7794.

Oren, A., et al.
How to be moderately halophilic with a broad salt tolerance: Clues from the genome of *Chromohalobacter salexigens.*

Extremophiles 2005 August, 9(4): 275–279.

Pazour, G.J., et al.
Proteomic characterization of a eukaryotic cilium. *Journal of Cell Biol.* 2005, 170: 103–113.

Parham, J.F., et al.
A reassessment of the referral of sea turtle skulls to Osteopygis (Late Cretaceous, New Jersey, USA). *Journal of Vertebrate Paleontology* 2005, 25 (1): 71–77.

Peñas, M., et al.
Identification and functional characterisation of ctr1, a *Pleurotus ostreatus* gene coding for a copper transporter. *Molecular Genetics and Genomics* 2005, 274: 402–409.

Pennacchio, L.A., et al.
Comparative genomic analysis reveals a distant liver enhancer upstream of the COUP-TFII gene. *Mammalian Genome* 2005 February, 16(2): 91–95.

Pennacchio, L.A., et al.
In vivo characterization of a vertebrate ultraconserved enhancer. *Genomics* 2005 June, 85(6): 774–781.

Place, A.R., et al.
Genetic markers in blue crabs *(Callinectes sapidus)*. II: Complete mitochondrial genome sequence and characterization of genetic variation. *Journal of Experimental Marine Biology and Ecology* 2005, 319: 15–27.

Poulin, F., et al.
In vivo characterization of a vertebrate ultra-conserved enhancer. *Genomics* 2005, 85(6): 774–781.

Ram, R.J., et al.
Community proteomics of a natural microbial biofilm. *Science* 2005, 308: 1915–1920.

Rasko, A.R., et al.
Genomics of the bacillus cereus group of organisms. *FEMS Microbiology Reviews* 2005, 29: 303–329..

Richards, S., et al.
Comparative genome sequencing of *Drosophila pseudoobscura:* Chromosomal, gene, and cis-element evolution. *Genome Res.* 2005, January 15, 1: 1–18.

Rivera, A.S., et al.
Characterization of Notch-class gene expression in segmentation stem cells *and segment founder cells in Helobdella robusta* (Lophotrochozoa; Annelida; Clitellata; Hirudinida; Glossiphoniidae). *Evol. and Dev.* 2005, 7(6): 588-599.

Rogers, Y.C., et al.
Closing bacterial genomic sequence gaps with adaptor-PCR. *BioTechniques* 2005, 39: 31–34.

Rubin, E.M., et al.
Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* 6: 805-814.

Rubin, E.M., et al.
Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. *Genome Research* 2005 July, 15(7): 928–935.

Samanta, S.K., et al.
Use of the *Rhodopseudomonas palustris* genome to identify a single amino acid that contributes to the activity of a coenzyme A ligase with chlorinated substrates. *Mol. Microbiol.* 2005, 55: 1151–1159.

Schwartz, J., et al.
Coral reef genomics: Developing tools for functional genomics of coral symbiosis. *Proceedings of International Coral Reef Symposium* 2005. Okinawa, Japan.

Sullivan, M.B., et al.
Three *Prochlorococcus cyanophage* genomes: Signature features and ecological interpretations. *PLoS Biology* 2005, 3(5): e144.

Swarbrick, M., et al.
Lack of support for the association between GAD2 polymorphisms and severe human obesity. *PLoS Biology* 2005, 3(9): e315.

Shah, N., et al.
SNP-VISTA: An interactive SNP visualization tool. *BMC Bioinformatics* 2005, 6(1): 292.

Tringe, S., et al.
Comparative metagenomics of microbial communities. *Science* 2005, April 22, 308 (5721): 554–557.

Tyson, G.W., et al.
Environmental shotgun sequencing. *Encyclopedia of Genomics, Proteomics and Bioinformatics.* John Wiley & Sons, Ltd., 2005.

Tyson, G.W., et al.
Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl. Environ. Microbiol.* 2005 October, 71(10): 6319–2634.

Tyson, G.W., et al.
Cultivating the uncultivated: a community genomics perspective. *Trends in Microbiology* 2005, 13, 411–415.

Urbich, C., et al.
Cathepsin L is required for endothelial progenitor cell-induced neovascularization. *Nature Medicine* 2005, 11(2): 206–213.

Vrdoljak, G., et al.
Characterization of a diesel sludge microbial consortia for bioremediation. *Scanning* 2005 January–February, 27(1): 8–14.

Waidner, L.A., et al.
Aerobic anoxygenic photosynthetic genes and operons in uncultured bacteria in the Delaware River. *Environ. Microbiol.* 2005, 7: 1896–1908.

Weng, L., et al.
Lack of MEF2A mutations in coronary artery disease. *Journal of Clinical Investigation* 2005 April, 115(4): 1016–1020.

Wolf, P.G., et al.
The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae): Implications for land plant phylogeny. *Gene* 2005, May 9, 350(2): 117–128.

APPENDIX F

2006

Ahituv, N., et al.
A PYY Q62P variant linked to human obesity. *Human Molecular Genetics* 2006, February 1, 15(3): 387–391.

Baker, B.J., et al.
Lineages of acidophilic archaea revealed by community genomic analysis. *Science* 2006, December 22, 314(5807): 1933–1935.

Bell, S.G., et al.
Cytochrome P450 enzymes from the metabolically diverse bacterium *Rhodopseudomonas palustris. Biochem. Biophys. Res. Commun.* 2006, 342: 191–196.

Beller, H.R., et al.
The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium *Thiobacillus denitrificans. Journal of Bacteriology* 2006, 188: 1473–1488.

Beller, H.R., et al.
Whole-genome transcriptional analysis of chemolithoautotrophic thiosulfate oxidation by *Thiobacillus denitrificans* under aerobic vs. denitrifying conditions. *Journal of Bacteriology* 2006, 188: 7005–7015.

Bely, A.E., et al.
Lessons from leeches: A call for DNA barcoding in the lab. *Evol. and Dev.* 2006 November/December, 8(6): 491.

Boore, J.L., et al.
Requirements and standards for organelle genome databases. *OMICS* 2006 June, 10(2): 119–126.

Boore, J.L., et al.
The Complete sequence of the mitochondrial genome of *Nautilus macromphalus* (Mollusca: Cephalopoda). *BMC Genomics* 2006, 7: 182.

Boore, J.L., et al.
The use of genome-level characters for phylogenetic reconstruction. *Trends in Ecology and Evolution*, 21(8): 439–446.

Brotcke, A., et al.
Identification of MglA-regulated genes reveals novel virulence factors in *F. tularensis. Infect. Immun.* 2006, 74: 6642–6655

Burghardt, T., et al.
The dominating outer membrane protein of the hyperthermophilic *Archaeum Ignicoccus hospitalis:* A novel pore-forming complex. *Mol. Microbiol.* 2006 (in press).

Cai, Z.C., et al.
Complete plastid genome sequences of Drimys, Liriodendron, and Piper: Implications for the phylogeny of magnoliids. *BMC Evolutionary Biology* 2006, 6: 77.

Carapelli, A., et al.
The mitochondrial genome of the entomophagous endoparasitoid *Xenos vesparum* (Insecta: Strepsiptera). *Gene* 2006, 376(2): 248–259.

Chain, P.S.G., et al.
Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc. Natl. Acad. Sci.* 2006, 103: 15280–15287.

Chain, P.S.G., et al.
Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: Evidence of gene reduction in an emerging pathogen. *J. Bacteriol.* 2006, 188: 4453–4463.

Chen, F.K., et al.
Induction of prophages from a marine bacterium: A genomic approach. *Appl. Environ. Microbiol.* 2006, 72: 4995–5001.

Choi, V.M., et al.
Developmental expression of FoxJ1.2, FoxJ2, and FoxQ1 in *Xenopus tropicalis. Gene Expr. Patterns* 2006 June, 6(5):443–447.

Chourrout, D., et al.
Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* 2006, August 10, 442(7103): 684–687.

Chumley, T.W., et al.
The completed chloroplast genome sequence of *Pelargonium x hortorum:* Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution* 2006, 23: 1–16.

Cohen, S.M., et al.
Genome-wide sequence and functional analysis of early replicating dna in normal human fibroblasts. *BMC Genomics* 2006, November 29, 7: 301.

Coleman, M.L., et al.
Genomic islands and the ecology and evolution of *Prochlorococcus. Science* 2006, 311: 1768–1770.

Collins, A.G., et al.
Cnidarian phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. *Systematic Biology* (in press).

Connelly, H.M., et al.
Characterization of pII family (GlnK1, GlnK2, and GlnB) protein uridylylation in response to nitrogen availability for *Rhodopseudomonas palustris. Anal. Biochem.* 2006, 357: 93–104.

Cristescu, M.E.A., et al.
A microsatellite-based genetic linkage map of the waterflea, *Daphnia pulex:* On the prospect of crustacean genomics. *Genomics* 2006 October, 88(4): 415–430.

Cui, L., et al.
ChloroplastDB: The chloroplast genome database. *Nucleic Acids Research* 2006, 34: D692–D696.

Dehal, P., et al
The PhIGs (phylogenetically inferred groups) database: A resource for phylogenomics. *BMC Bioinformatics* 7: 201.

DeLong, L., et al.
Mutations of GUCY2F, EPHA3 and NTRK3 in human cancers. *Human Mutation* 2006, 27(10): 1060–1061.

DeLong, E.F.
Community genomics among stratified microbial assemblages in the oceans interior. *Science* 311: 496–503.

Denslow, N.D. et al.
Selection of surrogate animal species for comparative toxicogenomics. In *Emerging Molecular and Computational Approaches for Cross-Species Extrapolations.* W. Benson and R. Di Giulio, Eds., SETAC Press: Florida, 2006.

Doggett, N.A., et al.
A 360-kb interchromosomal duplication of the human HYDIN locus. *Genomics* 2006 December, 88(6): 762–771.

Doyle, C.K. et al.,
Differentially expressed and secreted major immunoreactive protein orthologs of *Ehrlichia canis* and *E. chaffeensis* elicit early antibody responses to epitopes on glycosylated tandem repeats. *Infect Immun.* 2006 January, 74(1): 711–720.

Dyer, P.S., et al.
Discussion group: Lichen-fungal genome sequencing project. In *8th*

International Mycological Congress: Congress Handbook + Abstracts Book. Cairns, Australia, 2006, p. 161.

Eads, B.D., et al.
Ecological genomics in *Daphnia:* Stress responses and environmental sex determination. *Heredity* 2006 (in press).

Fletcher, R.B., et al.
FGF8 spliceforms mediate early mesoderm and posterior neural tissue formation in *Xenopus. Development* 2006 May, 133(9): 1703-1714.

Floyd, S.K., et al.
The ancestral developmental genetic toolkit of land plants. *Int. J. Plant Sci.* 2006 (in press).

Floyd, S.K., et al.
Evolution of Class III Homeodomain–Leucine Zipper Genes in Streptophytes. *Genetics* 2006, 173: 373–388.

Frank, P., et al.
A new structural motif for biological iron: Iron K-edge XAS reveals a $[Fe_4$-μ-$(Or)_5(Or)_{9–10}]$ cluster in the ascidian *Perophora annectens. Inorganic Chemistry* 2006, 45: 3920–3931.

Ganesan, B., et al.
Carbohydrate starvation causes a metabolically active but nonculturable state in *Lactococcus lactis. Appl. Environ. Microbiol.* 2006 (in press).

Ganesan, B., et al.
Identification of the leucine-to-2-methylbutyric acid catabolic pathway of *Lactococcus lactis. App. Environ. Microbiol.* 2006, 72: 4264–4273.

Gonsalves, F.C., et al.
MAPK regulation of maternal and zygotic Notch transcript stability in early development. *Proc. Natl. Acad. Sci.* 2006 (in press).

Goodwin, S.B., et al.
Identification and genetic mapping of highly polymorphic microsatellite loci from an EST database of the septoria tritici blotch pathogen *Mycosphaerella graminicola. Fungal Genetics and Biology*, published online October 30, 2006 (in press).

Groover, A.T., et al.
The Populus homeobox gene ARBORKNOX1 reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Molecular Biology* 2006, 61: 917–932.

Grzymski, J.J., et al.
Comparative genomics of DNA fragments from six Antarctic marine, planktonic bacteria. *Appl. Environ. Microbiol.* 2006, 72(2):1532–1541.

Hallam, S.J., et al.
Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum. Proc. Natl. Acad. Sci.* 2006, November 28, 103(48): 18296–18301.

Hamell, L.-P., et al.
Ancient signals: Comparative genomics of plant MAPK and MAPKK gene families. *TRENDS in Plant Science* 2006 April, 11(4): 192–198.

Han, C.S., et al.
The pathogenomic sequence analysis of *B. cereus* and *B. thuringiensis* isolates closely related to *Bacillus anthracis. J. Bacteriology* 2006, 188(9): 3382–3390.

Han, C.S.
Finishing repeat regions automatically with Dupfinisher. In *Proceeding of the 2006 International Conference on Bioinformatics and Computational Biology.* R. Hamid, H.R. Arabnia, and H. Valafar, Eds., CSREA Press: 2006, pp 141–146.

Howard, E.C., et al.
Bacterial taxa that limit sulfur flux from the ocean. *Science* 2006, 314: 649–652.

Ivleva, N.B., et al.
Protein extraction, fractionation and purification from cyanobacteria. *In Methods in Molecular Biology*, E. Rosato, Ed., Humana Press: Totowa, NJ, 2006(in press).

Johnson, Z., et al.
Niche partitioning among *Prochlorococcus ecotypes* along ocean-scale environmental gradients. *Science* 2006, 311: 1737–1740.

Kane, S.R., et al.
Whole-genome analysis of methyl tert-butyl ether (MTBE)-degrading beta-proteobacterium *Methylibium petroleiphilum* PM1. *J. Bacteriol.*, published online December, 8, 2006, doi:10.1128/JB.01259-06.

Klotz, M.G., et al.
The complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium *Nitrosococcus oceani* ATCC19707. *Appl. Environ. Microbiol.* 2006, 72: 6299–6315.

Krampis, K.B, et al.
Extensive variation in nuclear mitochondrial DNA content between the genomes of *Phytophthora sojae* and *Phytophthora ramorum. Mol. Plant Microbe Interact.* 2006 December, 19(12): 1329–1336.

Lee, P.N., et al.
A WNT of things to come: Evolution of Wnt signaling and polarity in cnidarians. *Seminars in Cell and Develop. Biology* 2006, 17, 157–167.

Liang, Y., et al.
Biochemical pathway and enzymatic study of pyrene degradation by Mycobacterium sp. KMS. *Appl. Environ. Microbiol.*, published online October 13, 2006.

Loots, G.G., et al.
Array2BIO: From microarray expression data to functional annotation of co-regulated genes. *BMC Bioinformatics* 2006, 7:307.

Maeder, D.L., et al.
The *Methanosarcina barkeri* genome: Comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. *J. Bacteriol.* 2006, 188(22): 7922–7931.

Makarova, K., et al.
Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci.* 2006, 103: 15611–15616.

Martiny, A.C., et al.
Phosphate acquisition genes in *Prochlorococcus ecotypes:* Evidence for genome-wide adaptation. *Proc. Natl. Acad. Sci.* 2006, 103(33): 12552–12557.

Matus, D. Q., et al.
"Dorso-ventral" genes are asymmetrically expressed and involved in germ layer demarcation during cnidarian gastrulation. *Current Biology* 2006, 16: 499–505.

Matus, D.Q, et al.
Molecular evidence for deep evolutionary roots of bilaterality in animal development. *Proc. Natl. Acad. Sci.* 2006, 103: 11195–11200.

Matus, D., et al.
Expression of Pax gene family members in the anthozoan cnidarian *Nematostella vectensis. Evol. Dev.* (in press)

APPENDIX F

Matus, D.Q., et al.
FGF signaling in gastrulation and neural development of *Nematostella vectensis*, an anthozoan cnidarian. *Dev. Genes Evol.* (in press).

Mavromatis, K., et al.
The genome of the obligately intracellular bacterium Ehrlichia canis reveals themes of complex membrane structure and immune evasion strategies. *J. Bacteriol.* 2006, 188: 4015–4023.

Martin, H.G., et al.
Metagenomic analysis of phosphorus removing sludge communities. *Nature Biotechnology* 2006, 24: 1263–1269.

Mazza, M.E., et al.
Genomic organization, gene structure, and developmental expression of three clustered otx genes in the sea anemone *Nematostella vectensis*. *J. Exp. Zool. (Mol. and Dev. Evol.)* (in press).

McHardy, A.C., et al.
Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods* 2006, 4: 63–72.

McNeal, J.R., et al.
Using partial genomic fosmid libraries for sequencing complete organellar genomes. *Biotechniques* 2006, 41(1): 69–73.

Medina, M., et al.
Naked stony corals: Skeleton loss in Scleractinia. *Proc. Natl. Acad. Sci.* 2006, 103(24): 9096–9100.

Merchant, S.S., et al.
Trace metal utilization in chloroplasts. In *The Structure and Function of Plastids*, R.R. Wise and J.K. Hoober, Eds., Springer: Netherlands, 2006, pp 199–218.

Merchant, S.S., et al.
Between a rock and a hard place: Trace element nutrition in *Chlamydomonas*. *Biochim. Biophys. Acta* 2006, 1763: 578–594.

Kuske, C.R., et al.
Environmental survey for four pathogenic bacteria and closely related species using phylogenetic and functional genes. *J. Forensic Sciences* 2006, 51(3): 548–558.

Miller, T.R, et al.
Bacterial motility is important in the interaction of *Silicibacter sp.* TM1040 with *Pfiesteria dinoflagellates*. *Environ. Microbiol.* 2006, 8: 1648–1659. (Cover photograph.)

Nelson, S.S., et al.
Mutations in *Flavobacterium johnsoniae* secDF result in defects in gliding motility and chitin utilization. *J. Bacteriol.* 2006, 188: 348–351.

Nicole, M.C., et al.
MAP-ping genomic organization and organ-specific expression profiles of poplar MAP kinases and MAP kinase kinases. *BMC Genomics* 2006, August 31, 7: 223.

Noonan J., et al.
Sequencing and analysis of neanderthal genomic DNA. *Science* 314(5802): 1113–1118.

Normand, P., et al.
Genome characteristics of three facultatively symbiotic *Frankia sp.* strains reflect host range and host plant biogeography. *Genome Research* 2006 (in press).

Nyholm, S.V., et al.
Fester, a candidate allorecognition receptor from a primitive chordate. *Immunity* (2006) 25: 163–173.

Parham, J.F., et al.
The complete mitochondrial genome of the enigmatic bigheaded turtle (Platysternon): Description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. *BMC Evolutionary Biology* 2006, 6: 11.

Parham, J.F., et al.
The phylogeny of Mediterranean tortoises and their close relatives based on complete mitochondrial genome sequences from museum specimens. *Molecular Phylogenetics and Evolution* 2006, 38(1): 50–64.

Park, S.-K., et al.
Genetic linkage map and expression analysis of genes expressed in the lamellae of the edible basidiomycete *Pleurotus ostreatus*. *Fungal Genetics and Biology* 2006, 43: 376–387.

Pennacchio, L.A., et al.
Human cathepsin L rescues the neurodegeneration and lethality in cathepsin B/L double deficient mice. *Biological Chemistry* 2006 July, 387(7): 885–891.

Pennacchio, L.A., et al.
In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 2006, November 23, 444: 499–502.

Pennacchio, L.A.,
Predicting tissue-specific enhancers in the human genome. *Genome Research* (in press).

Podsiadlowski, L., et al.
The mitochondrial genomes of *Campodea fragilis* and *Campodea lubbocki* (Hexapoda: Diplura): High genetic divergence in a morphologically uniform taxon. *Gene* 2006, October 15, 381: 49–61.

Prabhakar, S., et al.
Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Research* 2006, 16(7): 855–863.

Prabhakar, S., et al.
Accelerated evolution of conserved noncoding sequences in the human genome. *Science* 314(5800): 786.

Prigge, M.J., et al.
Evolution of the class III HD-Zip gene family in land plants. *Evol. Dev.* 2006, 8: 350–361.

Qin, J., et al.
Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci.* 2006, 103: 2075–2080.

Reeve, W.G., et al.
The Sinorhizobium medicae WSM419 lpiA gene is transcriptionally activated by FsrR and required to enhance survival in lethal acid conditions. *Microbiology* 2006, 152: 3049–3059.

Rey, F.E., et al.
Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. *J. Bacteriol.* 2006, 188: 6143–6152.

Romagnoli, S., et al.
A novel three-protein two-component system provides a regulatory twist on an established circuit to modulate expression of the cbbI region of *Rhodopseudomonas palustris* CGA010. *J. Bacteriol.* 2006, 188: 2780–2791.

Romagnoli, S., et al.
Phosphotransfer reactions of the CbbRRS three-protein two-component system from Rhodopseudomonas palustris CGA010 appear to be controlled by an internal molecular switch on the sensor kinase. *J. Bacteriol.*, published ahead of print on October 27, 2006.

APPENDIX F

Shi, H.
A report on the wild hybridization between two species of threatened Asian box turtles (Testudines: Cuora) on Hainan Island, China. *Amphibia-Reptilia* (in press).

Simison, W.B., et al.
Rolling circle amplification of metazoan mitochondrial genomes. *Molecular Phylogenetics and Evolution* 2006, 39(2): 562–567.

Scheuring, S., et al.
The photosynthetic apparatus of *Rhodopseudomonas palustris:* Structures and organization. *J. Mol. Biol.* 2006, 358: 83–96.

Scott, K.M., et al.
The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. *PLoS Biology* 2006 December, 4(12): e383.

Sevenich, L., et al.
Human cathepsin L rescues the neurodegeneration and lethality in cathepsin B/L double deficient mice. *Biological Chemistry* 2006, 387(7): 885–891.

Slavotinek, A.M., et al.
Array comparative genomic hybridization in patients with congenital diaphragmatic hernia: Mapping of four CDH-critical regions and sequencing of candidate genes at 15q26.1-15q26.2. *European Journal of Human Genetics* 2006, 14(9): 999–1008.

Starkenburg, S.R., et al.
The genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Appl. Environ. Microbiol.* 2006, 72: 2050–2063.

Steglich, C.
Genome-wide analysis of light sensing in *Prochlorococcus. J. Bacteriol.* 2006(in press).

Stillman, J.H., et al.
Construction and characterization of EST libraries from the porcelain crab, *Petrolisthes cinctipes. Integrative and Comparative Biology* 2006 (in press, available online).

Sullivan, M.B., et al.
Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biology* 2006, 4(8): 1345–1357.

Tartaglia, M.,
Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome. *Nature Genetics* 2006, 39: 75–79.

Timme, R.E., et al.
A comparison of the first two sequenced chloroplast genomes in Asteraceae: Lettuce and Sunflower. *American Journal of Botany* 2006 (in press).

Tolonen, A.C., et al.
Gene expression of *Prochlorococcus* ecotypes on different nitrogen sources and during nitrogen starvation. *Molecular Systems Biology* 2006 (in press).

Tuskan, G.A., et al.
The genome of black cottonwood, *Populus trichocarpa* (Torr. + Gray). *Science* 2006, 313: 1596–1604.

Tyler, B.M., et al.
Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 2006, 313: 1261–1266.

Vallès, Y.
Lophotrochozoan mitochondrial genomes. *Integrative and Comparative Biology*, Allen Press (in press).

Vanden Wymelenberg, A., et al.
Computational analysis of the *Phanerochaete chrysosporium* v2.0 genome database and mass spectrometry identification of peptides in lignin-olytic cultures reveal complex mixtures of secreted proteins. *Fungal Genetics and Biology* 2006 May, 43(5): 343–356.

Van Driessche, G., et al.
GHP: A new c-type heme protein from *Halochromatium salexigens* and other proteobacteria. *FEBS J.* 2006, 273: 2801–2811.

Van Passel, M.W., et al.
The reach of the genome signature in prokaryotes. *BMC Evol. Biol.* 2006, 6: 84.

Ventura, M., et al.
Comparative genomics and transcriptional analysis of prophages identified in the genomes of *Lactobacillus gasseri, Lactobacillus salivarius,* and *Lactobacillus casei. Appl. Environ. Microbiol.* 2006 May, 72(5): 3130–3146.

VerBerkmoes, N.C., et al.
Determination and comparison of the baseline proteomes of the versatile microbe *Rhodopseudomonas palustris* under its major metabolic states. *J. Proteome Research* 2006, 5: 287–298.

Visel, A., et al.
VISTA enhancer browser: A database of tissue-specific human enhancers. *Nucleic Acids Research*, published online November 27, 2006, doi: 10.1093/nar/gkl822.Nov 27.

Wang, Q.F., et al.
Primate-specific evolution of an LDLR enhancer. *Genome Biol.* 2006, August 2, 7(8): R68.

Whitaker, RJ., et al.
Population genomics in natural microbial communities. *Trends in Ecology and Evolution* 2006 September, 21(9): 508–516.

Whitham, T.G., et al.
Community and ecosystem genetics: a framework for integrating from genes to ecosystems. *Nature Reviews Genetics* 2006, 7: 510–523.

Wills, A., et al.
Twisted gastrulationis required for forebrain specification and cooperates with Chordin to inhibit BMP signaling during X. *tropicalis* gastrulation. *Dev. Biol.* 2006, 289: 166–178.

Worden, A.Z., et al.
In-depth marine microbial community genomics. *Trends in Microbiology* 14: 331–336.

Woyke, T., et al.,
Symbiosis Insights through Metagenomic Analysis of a Microbial Consortium. *Nature*, published online September 17, 2006.

Xueming, Wei, et al.
Transcript profiles of *Nitrosomonas europaea* during growth and upon deprivation of ammonia and carbonate. *FEMS Microbiol. Lett.* 2006, 257: 76–83.

Yang, P., et al.
Radial spoke proteins of *Chlamydomonas flagella. J. Cell Sci.* 2006, 119: 1165–1174.

Yeager, C.M., et al.
Three genera of cultured heterocystous cyanobacteria comprise the dominant N2-fixing members of biological soil crusts of the Colorado Plateau, USA. *FEMS Microbiol. Ecol.* 2006.

Zhan, J., et al.
Significant difference in virulence

between MAT1-1 and MAT1-2 isolates in the wheat pathogen *Mycosphaerella graminicola*. *Fungal Genetics and Biology* (in press).

Zhang K., et al.
Sequencing genomes from single cells via polymerase clones. *Nature Biotechnology* 2006, 24: 680–686.

Zhengqiu, C., et al.
Complete chloroplast genome sequences of Drimys, Liriodendron, and Piper: Implications for the phylogeny of magnoliids and the evolution of GC content. *BMC Evolutionary Biology* (in press).

The best friend on earth of man is the tree: when we use the tree respectfully and economically we have one of the greatest resources of the earth.

— Frank Lloyd Wright

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 Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231, and Los Alamos National Laboratory under Contract No. W-7405-ENG-36.

For more information about JGI, contact:
David Gilbert, Public Affairs Manager
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
email: gilbert21@llnl.gov
phone: (925) 296-5643

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

Published by the Berkeley Lab Creative Services Office in collaboration with JGI researchers and staff for the U.S. Department of Energy.

CSOJO12572

IMAGE CREDITS

Science cover September 2006, Poplar (Page 18): Reprinted with permission from *Science* Vol. 313, 15 September 2006. Copyright 2006 AAAS

Brachypodium distachyon (Page 19): Courtesy of Todd Mockler, Oregon State University

Laccaria bicolor (Pages 20 and 31): ©INRA

Clostridium thermocellum (Page 21): Courtesy of Edward Bayer, The Weizmann Institute of Science

Aspergillus niger (Pages 21 and 31): Courtesy of Nick Read, University of Edinburgh

Ostreococcus (Pages 22 and 30): Wenche Eikrem and Jahn Throndsen, University of Oslo

Fragilariopsis cylindrus (Page 23): Courtesy of Henrik Lange and Gerhard Dieckmann, Alfred-Wegener Institute for Polar and Marine Research, Germany

Deinococcus geothermalis (Page 26): Courtesy of Raija Peura, University of Oulu

Olavius algarvensis (Page 27): Courtesy of Nicole Dubilier, Max Planck Institute of Marine Microbiology

Daphnia pulex (Page 27): Courtesy of Paul Hebert, University of Guelph

Phytophthora ramorum (Page 27): Courtesy of Matteo Garbelotto, UC Berkeley

Trichoderma reesei (Page 30): Courtesy of Irma Salovuori, VTT Biotechnology

Phaeodactylum tricornutum (Page 30): Courtesy of Chris Bowler, Ecole Normale Supérieure

Mycosphaerella graminicola (Page 31): Courtesy of Keith Duncan, DuPont; Stephen B. Goodwin, USDA-ARS

*Physcomitrella patens (*Page 31): Courtesy of D.J. Cove, Leeds, UK