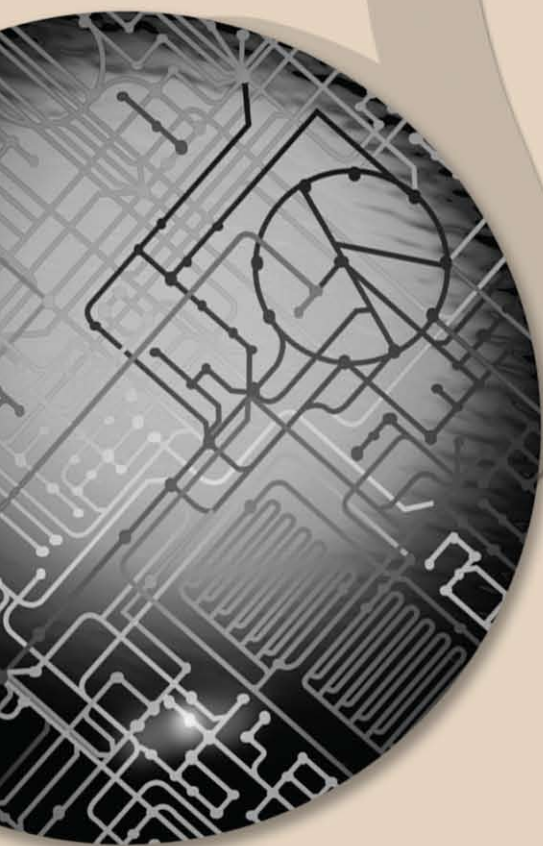


U.S. DEPARTMENT OF ENERGY OFFICE OF SCIENCE

# GENOMICS:GTL

## 2008 STRATEGIC PLAN



U.S. DEPARTMENT OF  
**ENERGY**

**DOE GENOMICS:GTL**  
SYSTEMS BIOLOGY  
FOR ENERGY AND  
ENVIRONMENT

# DOE Genomics:GTL Program

---

**Sharlene Weatherwax**

U.S. Department of Energy

Office of Science

Office of Biological and Environmental Research

genomics.gtl@science.doe.gov

**DOE GENOMICS:GTL**  
*Systems Biology for Energy and Environment*

**2008 STRATEGIC PLAN**

**U.S. Department of Energy**  
**Office of Science**  
**Office of Biological and Environmental Research**

**September 2008**  
[genomicsgtl.energy.gov/strategicplan/](http://genomicsgtl.energy.gov/strategicplan/)

## Table of Contents

<b>Mission-Inspired Science</b> .....	3
<b>GTL Concept and History</b> .....	4
National Academy of Sciences Evaluation of GTL Program.....	5
<b>GTL’s Fundamental Research Approach to Its Mission</b> .....	5
Hallmarks of the GTL Research Approach.....	6
<b>Multiscale Explorations</b> .....	6
Office of Biological and Environmental Research User Facilities .....	8
Metamethods: Metagenomics and Metaproteomics .....	10
<b>Robust Computational Framework for GTL Research</b> .....	11
GTL Information and Data Sharing Policy .....	13
<b>Elaboration of GTL Science for DOE Missions</b> .....	13
<b>Systems Biology for Energy Solutions</b> .....	13
<b>Systems Biology for Understanding the Carbon Cycle</b> .....	16
Planet-Transforming Microbes .....	16
<b>Systems Biology for Understanding Contaminant Transport in the Environment</b> .....	17
Microbe-Mineral Interface in Contaminated Environments .....	18
<b>DOE Office of Science</b> .....	19
<b>DOE Office of Biological and Environmental Research</b> .....	19
<b>For More Information</b> .....	19

## Mission-Inspired Science

The Department of Energy's (DOE) missions in energy security, climate protection, and environmental remediation have an enduring need for breakthrough technologies that require and facilitate expanding the frontiers of science. DOE's mission challenges for which biology can provide solutions include:

- Developing biofuels as a major secure national energy resource.
- Understanding relationships between climate change and Earth's ocean and terrestrial ecosystems and assessing options for biosequestration of carbon in these systems.
- Developing biological solutions for intractable environmental remediation problems.

Genomics:GTL\* (GTL) is a fundamental research program that aims to develop a predictive understanding of biological systems relevant to mission challenges (see figure at right). By revealing the genetic blueprint and fundamental principles that control biological systems, GTL will enable scientists to understand and use diverse biological capabilities to meet DOE and national needs.

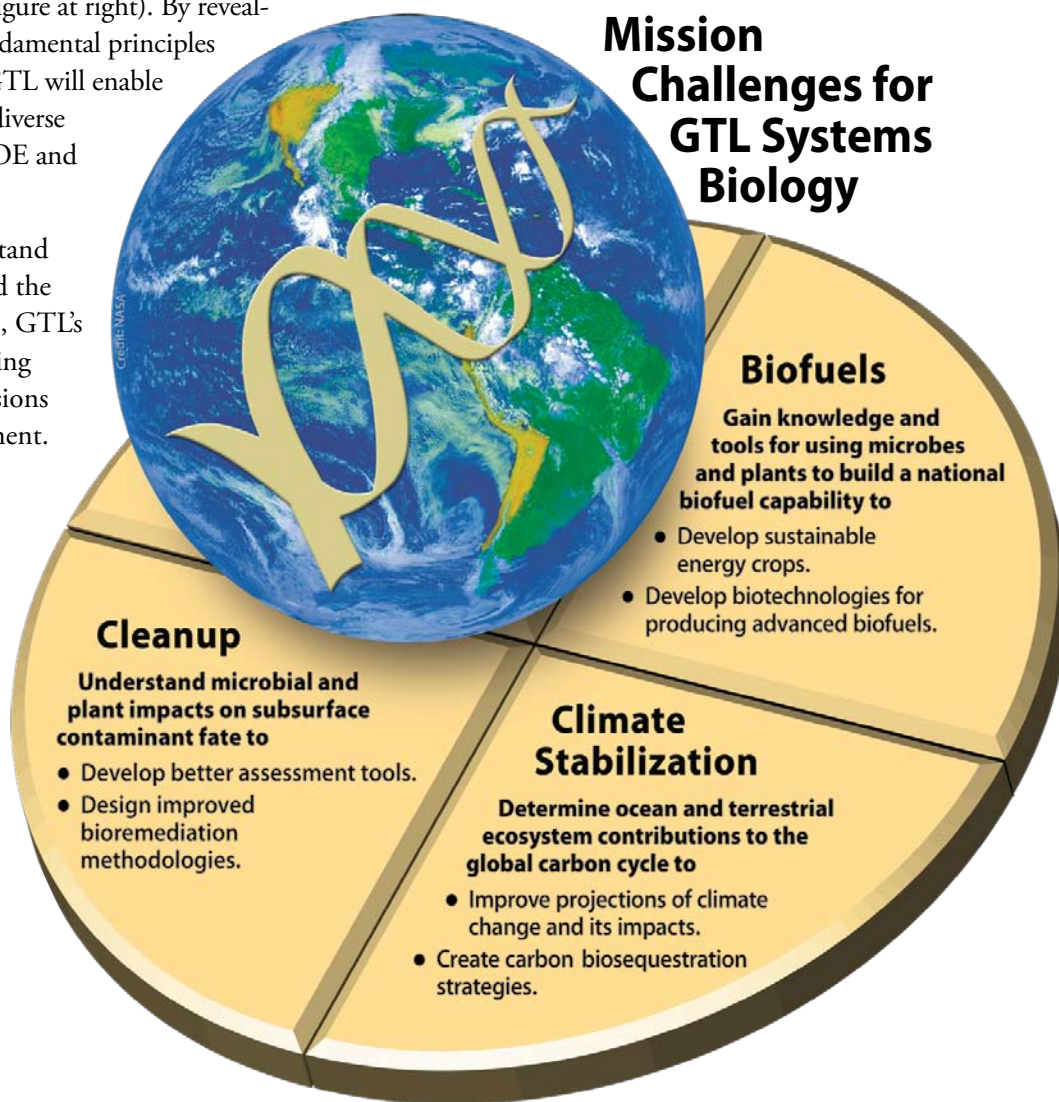
Just as DOE's mission to understand health impacts of energy inspired the Human Genome Project (HGP), GTL's systems approach to understanding biology is inspired by DOE missions in energy, climate, and environment. The HGP opened the door for improvements in human health, and GTL has become a gateway for biotechnological solutions to DOE mission challenges, stimulating upcoming generations of industrial biotechnology (see box above).

\*Formerly Genomes to Life.

## GTL Science to Catalyze Industrial Biotechnology

"Much as the Human Genome Project stimulated the growth of a biomedical biotechnology industry, the research laid out in (GTL) will catalyze growth in a new industrial biotechnology sector," Secretary of Energy Samuel W. Bodman said. "Microbes can be used for processes and products that can serve as an engine for economic competitiveness in the 21<sup>st</sup> Century."

– U.S. Department of Energy press release, October 3, 2005



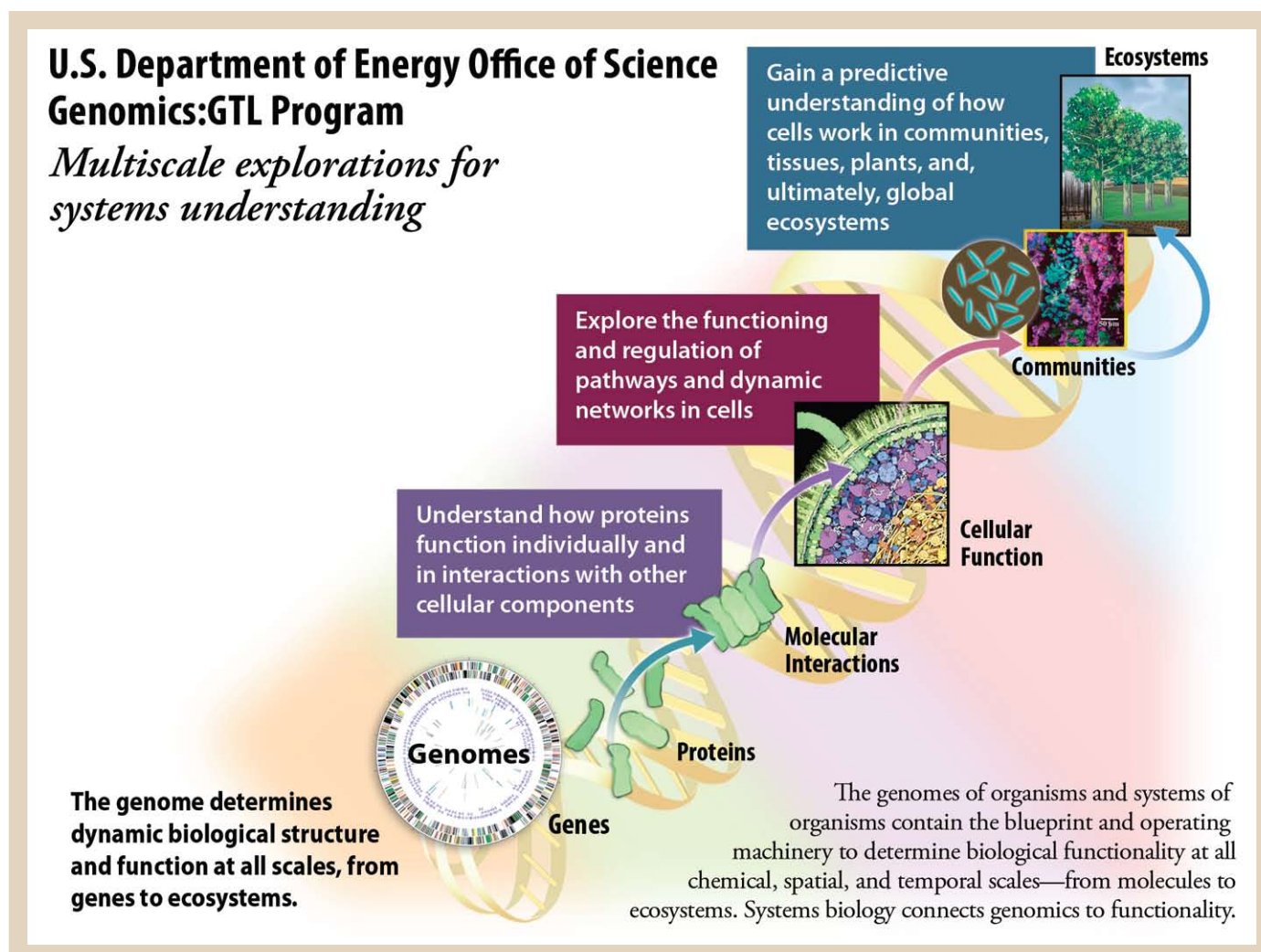


## GTL Concept and History

The DNA code—the genome—for each living organism provides the blueprint for the biochemical processes of life. It is the platform for understanding the functionality of whole biological systems as they respond to and modify their environments. The genomes of individual organisms and systems of organisms (e.g., communities and ecosystems) contain the information and operating capabilities that determine structure and function across multiple scales of biological organization. At these scales, including spatial and temporal dimensions (see figure below, Multiscale Explorations for Systems Understanding), scientists investigate the functioning of individual proteins and other molecules. They also study molecular interactions, pathways, and vast interconnected networks within whole cells, communities, and ecosystems.

Building on sequenced genomes, the GTL program aims to use systems biology research to understand how genomes are translated into functional responses. Central to this approach is having the ability to measure and then model, in a coordinated way, the biological functions of a system and establish a mechanistic connection to its genome sequence. A complete computationally predictive, whole-system understanding of microbes and plants eventually will allow the design and engineering of biobased solutions to DOE missions.

The scientific understanding of a system attained in pursuit of one DOE mission will aid breakthroughs in other mission areas because of common underlying biological principles. This is the ultimate efficiency of the GTL program and of biology in general. GTL's DOE mission focus and biological research approach have been affirmed by the National Academy of Sciences (see sidebar, National Academy of Sciences Evaluation of GTL Program, p. 5).



GTL systems biology research departs from traditional scientific methodology by addressing complex problems that require coordinated research among interdisciplinary teams from the life, physical, and computational sciences. To accomplish the desired in-depth understanding of biological systems in this new genomics-enabled era, an unprecedented integration of technologies, data and computing, theory, modeling, simulation, and experimentation must occur. Communication and research integration are even more important for GTL than for the HGP, which relied on only one dominant technology—DNA sequencing—and produced only one major dataset—DNA sequence. Myriad analytical technologies must be created, refined, and deployed to investigate and understand biological systems, linking different levels of biological discovery to gain a predictive understanding of whole systems—from cells to ecosystems.

### GTL's Fundamental Research Approach to Its Mission

DOE has a successful history of conducting multidisciplinary science and developing the tools and concepts needed to address high-risk, complex problems underly-

ing achievement of DOE mission goals. In this tradition, the GTL research enterprise, begun in FY 2002, is pursuing solutions to biological grand challenges with a scientific approach that has exhibited flexible scientific and managerial hallmarks (see Hallmarks of the GTL Research Approach, p. 6). GTL's mission is to,

“By 2015, provide sufficient scientific understanding of plants and microbes to develop robust new strategies to produce biofuels, clean up waste, or sequester carbon. This includes research that supports the development of computational models to direct the use and design of improved organisms carrying out these processes.”

— Report of the Biological and Environmental Research Advisory Committee: Review of the Life Sciences PART Measure, October 16, 2006

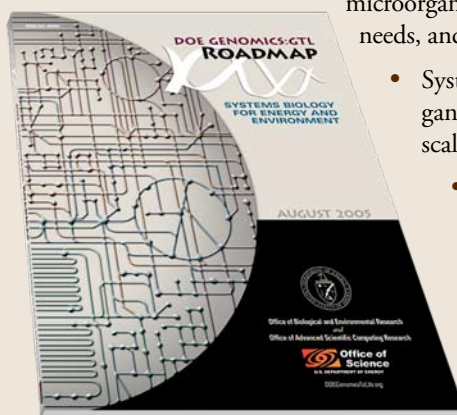
Derived from this mission are the GTL goal and objectives:

**Goal:** Achieve a predictive, system-level understanding of plants, microbes, and biological communities, via integration of fundamental science and technology development, to enable biological solutions to DOE mission challenges in energy, environment, and climate.

### National Academy of Sciences Evaluation of GTL Program\*

The 2005 GTL Roadmap described a new scientific approach—systems biology—to reveal and ultimately control and engineer the capabilities inherent in biological systems critical to DOE mission challenges. Less than a year after the roadmap was published, a National Academy of Sciences review affirmed the missions' needs (see figures, Alignment of GTL Goal and Objectives with DOE and Office of Science Goals, p. 7, and Genomics:GTL, A Mission-Inspired Fundamental Research Approach, p. 9) and this biological research approach. Biological capabilities thus were firmly established alongside more traditional physical sciences as the bases for technologies that could be developed by the department. Excerpts follow.

- “The use of systems and synthetic biology approaches in the Genomics:GTL program to address some of the most pressing issues in microbial genomics relevant to DOE's mission in energy security, environmental remediation, and carbon cycling and sequestration is not only appropriate but necessary.
- The study of individual components only does not provide knowledge on systems integration at the level of pathways, organisms, and microbial consortia—for example, on the effects of introducing new metabolites or new or engineered organisms to a community or on organism or community responses.
- Systems biology research is needed to develop models for predicting the behavior of complex biological systems, to engineer microorganisms for bioremediation and energy-related needs, and to understand carbon cycling.
- Systems biology research on plants and microorganisms is not likely to be conducted on a large scale without DOE's visionary thinking.
- Because the productive applications of new technologies to advance science will be hampered by the lack of appropriate tools, the committee finds that the concept of infrastructure for research and technology development offers a logical and even necessary pathway for achieving DOE's research goals.”



\*Review of the Department of Energy's Genomics: GTL Program, Committee on Review of the Department of Energy's Genomics:GTL Program, National Research Council (2006), [www.nap.edu/catalog/11581.html](http://www.nap.edu/catalog/11581.html).

## Hallmarks of the GTL Research Approach

### Science

- Uses global, genome-derived principles of microbial, plant, and community functions.
- Develops enabling experimental technologies and capabilities to provide comprehensive data.
- Uses modeling and simulation tools for predictive understanding across multiple scales of biological organization.
- Employs GTL Knowledgebase for facilitating data and information integration for modeling and comparative analyses (see figure, GTL Integrated Computational Environment for Biology, p. 12).

### Management

- Focuses on plant and microbial systems with capabilities relevant to DOE missions.
- Relies on collaborative and other integrative approaches.
- Maintains strategically managed research portfolio to respond to emerging national priorities and mission needs.
- Selects research based on scientific merit and peer review.
- Supports research conducted by individual investigators, collaborative teams, and research centers at DOE national laboratories, academic institutions, and industry.
- Establishes and utilizes user facilities, integrated capabilities, and centers.
- Leverages capabilities and resources across Office of Biological and Environmental Research programs and scientific user facilities.
- Encourages communication across the scientific community through the annual GTL program meeting, workshops, symposia, and exhibits at national meetings.
- Fosters an atmosphere of open access to data and information.
- Coordinates with other DOE programs and other federal agencies.

- **Objective 1:** Determine the genomic properties, molecular and regulatory mechanisms, and resulting functional potential of microbes, plants, and biological communities central to DOE missions.
- **Objective 2:** Develop the experimental capabilities and enabling technologies needed to achieve a genome-based, dynamic system-level understanding of organism and community functions.
- **Objective 3:** Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance the understanding, prediction, and manipulation of complex biological systems.

The figure, Alignment of GTL Goal and Objectives with DOE and Office of Science Goals, p. 7, illustrates how GTL objectives align with the specific overarching DOE and Office of Science goals that drive GTL science. As described in the Office of Science Strategic Plan (February 2004), DOE's Goal 5 is to provide world-class scientific research capability to support the higher DOE strategy of advancing energy-related biological and environmental research. This frames GTL research under the Office of Science Goal 2, "Harness the Power of Our Living World."

Focusing on DOE mission goals, GTL's approach in energy, carbon cycle, and environmental remediation involves (1) identifying research grand challenges through community input and workshops; (2) initiating proof-of-principle, high-risk pilot activities; (3) issuing solicitations for research strategies to build upon successful pilot projects; (4) establishing productive research partnerships; and (5) building a GTL Knowledgebase for integrating and sharing the biological information needed to solve these challenges.

GTL's systems biology approach requires a wide range of capabilities and research styles. The GTL program includes participants from national laboratory science focus areas, academia (both individual researchers and members of teams), and industry. GTL employs DOE's national user facilities [e.g., Joint Genome Institute (JGI) and Environmental Molecular Sciences Laboratory] and newly established Bioenergy Research Centers. The DOE Office of Biological and Environmental Research (OBER) partners with other organizations to help achieve GTL's mission. GTL also uses resources and findings supported through programs at other federal agencies, sometimes collaborating via interagency solicitations (see sidebar, OBER User Facilities, p. 8, and figure, GTL Research Enterprise, p. 9).

### Multiscale Explorations

Biology is a science of dynamic interacting systems, and no component of these systems operates in isolation. For biological systems central to DOE missions in energy, the carbon cycle, and environmental remediation, GTL analyzes properties and processes on three fundamental levels.

- **Molecular:** Focusing on genes, proteins, multicomponent protein complexes, and other biomolecules that provide structure and perform a cell's functions. Such an approach helps us understand how the genome determines dynamic biological structure and function at all scales, from genes



## Systems Biology for Energy and Environment

to ecosystems, and how proteins function individually or in interactions with other cellular components.

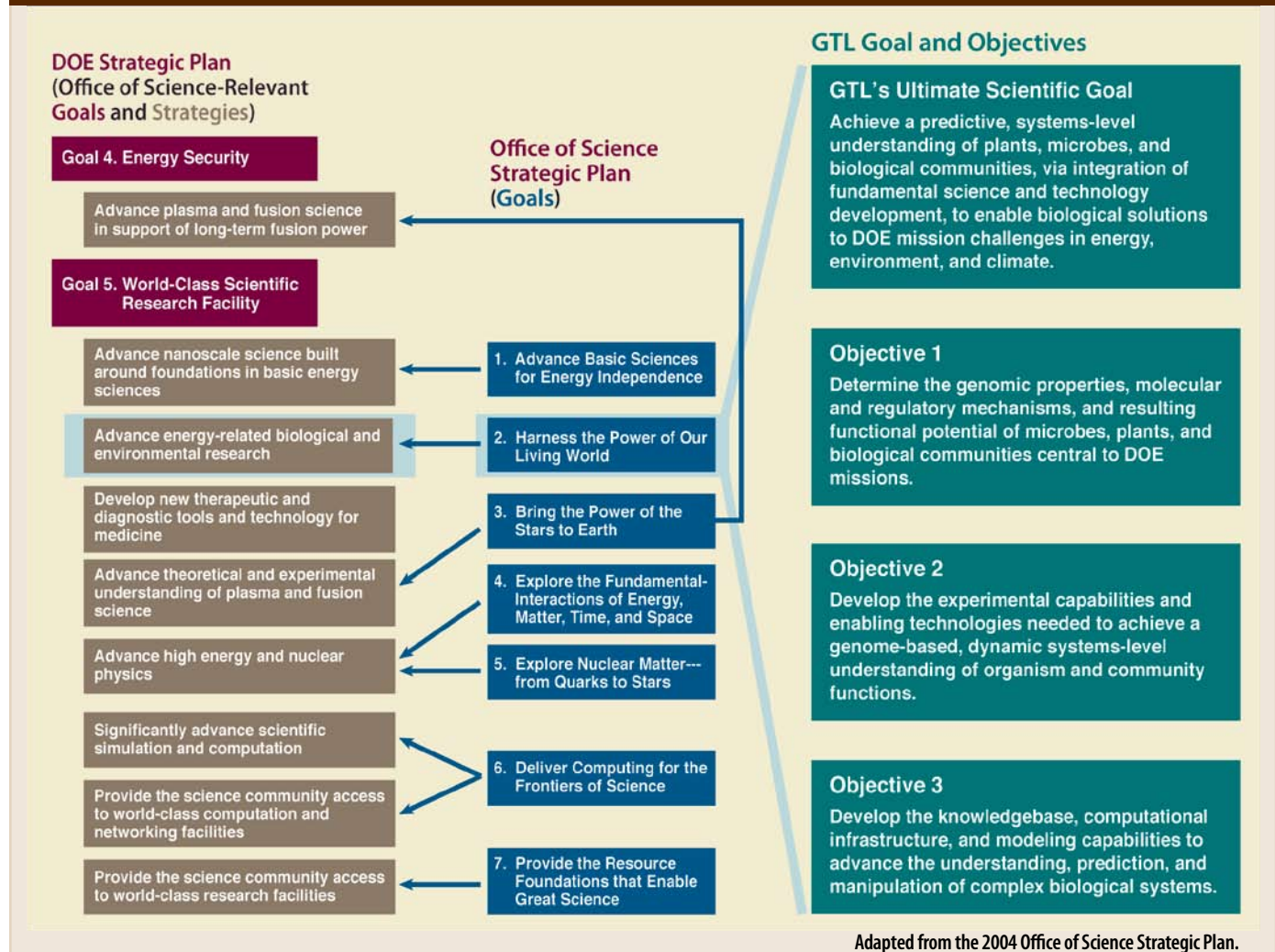
- **Whole cell:** Investigating dynamic molecular processes, networks, and subsystems controlled and coordinated to enable complex cellular processes such as growth and metabolism.
- **Microbial community and higher organisms:** Exploring diverse cellular systems that interact to carry out coordinated complex processes that both respond to and alter their environments to determine how cells work in communities, tissues, plants, and, ultimately, global ecosystems.

The myriad biological structures and processes that exist within these three system levels are interconnected and coordinated by an intricate set of regulatory controls and

continuous interactions with the environment. Exploring biology across all scales in a comprehensive and integrated way is essential to understanding how these systems operate in nature or in more application-oriented contexts related to new technology endpoints for DOE missions (see diagram, Genomics:GTL, A Mission-Inspired Fundamental Research Approach, p. 9).

To investigate biology at multiple scales, GTL is building new multidisciplinary research communities and advancing development of next-generation, automated technologies that increase sample throughput and reduce analysis time. Key research technology and methodology development areas for GTL include genomics, analytical “omics,” molecular imaging and structural analysis, modeling and simulation, and synthetic biology.

## Alignment of GTL Goal and Objectives with DOE and Office of Science Goals



## Office of Biological and Environmental Research User Facilities

The Department of Energy's (DOE) Office of Science creates, maintains, and operates state-of-the-art national user facilities key to continued U.S. leadership in physical and biological science research. Transformational scientific discoveries are resulting from biological research at the Office of Science's advanced scientific computing, synchrotron, advanced light source, and neutron facilities. These world-leading facilities are enabling scientists to follow biomolecular processes in real time and image biological materials at atomic resolution. User facilities focused on genome sequencing and environmental molecular research are supported by the Office of Biological and Environmental Research (OBER) within the Office of Science.

### DOE Joint Genome Institute

The DOE Joint Genome Institute (JGI) in Walnut Creek, California, is one of the world's largest and most productive public genome-sequencing centers. The JGI sequencing capacity exceeds 40 billion DNA base pairs per year and is growing with the incorporation of new technology. The majority of this capacity is focused on sequencing plants, microbes, and microbial community metagenomes relevant to DOE missions in energy, carbon cycling, and environmental cleanup. This high-quality genome-sequencing data, which JGI makes freely available to the scientific community, is the product of integrated genomic expertise at six partner institutions: Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, Stanford University, and Pacific Northwest National Laboratory.

JGI sequencing targets have included bacteria that detoxify hazardous substances, marine microbes that capture carbon from the atmosphere, plants that provide biomass for bioenergy production, and microbial communities with enzymes that can process plant biomass into ethanol and other biofuels. This growing collection of sequence data will enable discovery of the fundamental principles and details of systems biology needed to develop biobased solutions to critical national needs and DOE missions.



**High-Throughput DNA Sequencing.** Automated machines at the DOE JGI Production Genomics Facility sequence millions of DNA base pairs per day.

### DOE Environmental Molecular Sciences Laboratory

The William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), located in Richland, Washington, provides co-located experimental and computational resources to enable fundamental studies of physical, chemical, and biological processes and their interactions to address DOE's energy, environmental, and science missions. With more than 60 leading-edge experimental instruments and a high-performance computer, EMSL provides the scientific community with capabilities for investigating a wide range of issues and processes involved in systems biology studies, including microbial communities and their interfaces with minerals or contaminants.

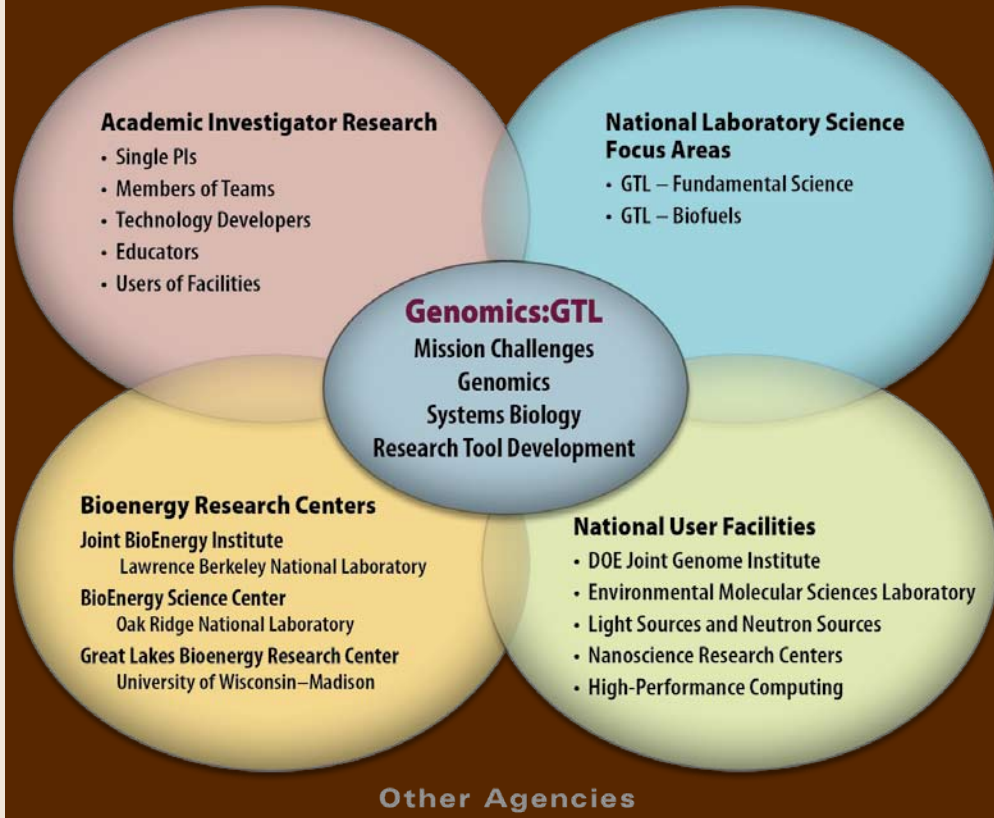
Relevant capabilities include high-resolution nuclear magnetic resonance (NMR) spectrometers for protein structure studies; high-performance mass spectrometers for proteomics analyses; Mossbauer spectrometers and surface-analysis tools (e.g., a cryotransmission electron microscope and other high-resolution microscopy systems and a secondary ion mass spectrometer) for characterizing or visualizing microbe-mineral or microbe-contaminant interfaces; a flow cell facility for studying particle and microbial transport through groundwater and sediments; and a supercomputer for conducting molecular-dynamics studies of protein-protein interactions and protein folding and other biological structure-function research and for simulating subsurface microbial and contaminant transport.



**EMSL's 900-MHz Nuclear Magnetic Resonance Spectrometer.** This instrument is used to provide atom-by-atom, three-dimensional views of protein complexes and other molecular structures.



# GTL Research Enterprise



## Genomics:GTL A Mission-Inspired Fundamental Research Approach

### GTL Systems Biology and Technology Development

### Core Science Goals

### Mission Grand Challenges for Biology

- Microbe genomics, plant genomics, metagenomics
- Analytical omics
- Molecular imaging
- Modeling and simulation
- Prediction and design
- Synthetic biology

Gain a predictive understanding of how cells work in communities, tissues, plants, and, ultimately, global ecosystems

Ecosystems

### Energy

Tools and concepts for designing and engineering bioenergy plant and microbial systems, including the mechanistic bases.

Explore the functioning and regulation of pathways and dynamic networks in cells

Communities

### Carbon Cycle

Tools and concepts to determine the carbon-cycling and sequestration processes of ocean and terrestrial ecosystems.

Understand how proteins function individually and in interactions with other cellular components

Cellular Function

Molecular Interactions

Genomes  
Genes

Proteins

The genome determines dynamic biological structure and function at all scales, from genes to ecosystems.

### Environmental Remediation

Microbial and plant modeling and experiments to predict and control contaminant fate and transport.

**Technology Endpoints**  
Payoffs for the Nation



Viable Biofuel Technologies

Earth System Modeling and Biosequestration Strategies

Improved Strategies for Bioremediation

## Metamethods: Metagenomics and Metaproteomics

Central to solving the biological grand challenges critical to DOE missions are understanding and using the myriad biochemical interactions associated with complex microbial and plant communities. The study of microbial ecology has been revolutionized by the pioneering development of metamethods that allow the genomic and molecular characterization of entire environmental communities. More than 99% of microbes have defied culturing in the laboratory, precisely because of the intimate interactions, critical to survival, that occur in complex communities and environments. Marine and terrestrial metagenomic and metaproteomic studies have discovered a stunning diversity (millions of unique genes),

revealing new mechanisms for survival and adaptation by massive genetic exchange and creation of new families of proteins within communities. In addition, new imaging techniques augmented by fluorescent probes with molecular resolving power now allow analysis of individual cells and processes in these complicated community and geochemical environments.

These metamethods can be applied to microbial communities in industrial settings as well (e.g., in biofuel production). Discoveries emanating from metamethods have immense fundamental and practical value, spurring both scientific discoveries and industrial applications that will provide mission solutions for DOE.

**Genomics and Metagenomics.** Sequencing and analyzing DNA from individual organisms (genomics) or microbial communities in environmental samples (metagenomics) form the foundation of systems biology. The sequences of DOE-relevant microbes have been provided largely by JGI, an important DOE user facility and a world leader in generating microbial, metagenome, plant, and other genome sequences (see sidebar on OBER user facilities, p. 8). In addition to sequencing and annotating genomes, GTL researchers are working with JGI to improve methods for rapidly identifying and characterizing functional and regulatory gene networks in microbes, microbial communities, and plants.

**Analytical Omics.** Transcriptomics, proteomics, and metabolomics—collectively described as “omics” analyses—identify and measure the abundance and fluxes of key molecular species indicative of organism or community activity at specific points in time. Global analysis of RNA transcripts, proteins, and metabolites informs scientists about organisms’ physiological status, provides insights into gene function, and indicates which genes are activated and translated into functional proteins as organisms and communities develop or respond to environmental cues. Metamethods that analyze DNA, RNA, and proteins extracted from complex microbial communities in marine and terrestrial environments are enabling discovery of new biological processes and providing novel insights into microbial diversity (see sidebar, Metamethods: Metagenomics and Metaproteomics, above).

**Molecular Imaging and Structural Analysis.** GTL investigators are developing and using new methods for characterizing the chemical reaction surfaces, organization, and

structural components in molecular complexes and tracking molecules to view cellular processes as they are occurring (see figure, Imaging Microbes, from Ecosystem to Molecular Scales, p. 11). Depending on the spatial scale, a variety of imaging technologies can be used to visualize the complex molecular choreography within biological systems. Some of these structural and imaging tools (e.g., synchrotrons, neutron sources, and electron microscopes) are available or are being developed at DOE Office of Science user facilities that provide photon, neutron, electron, magnetic, and mechanical instrumentation with state-of-the-art spatial, temporal, and chemical sensitivity.

**Modeling and Simulation.** Computational models are used to capture, integrate, and represent current knowledge of biology at various scales. For example, GTL researchers are using genome sequence and other comprehensive datasets (molecular, spatial, and temporal data) to build models of signaling networks, gene regulatory circuits, and metabolic pathways. Simulation couples individual processes and puts them into an environmental context to reveal emergent behaviors of biological systems, such as a cell or microbial community (see section, Enabling Predictive Understanding, p. 13).

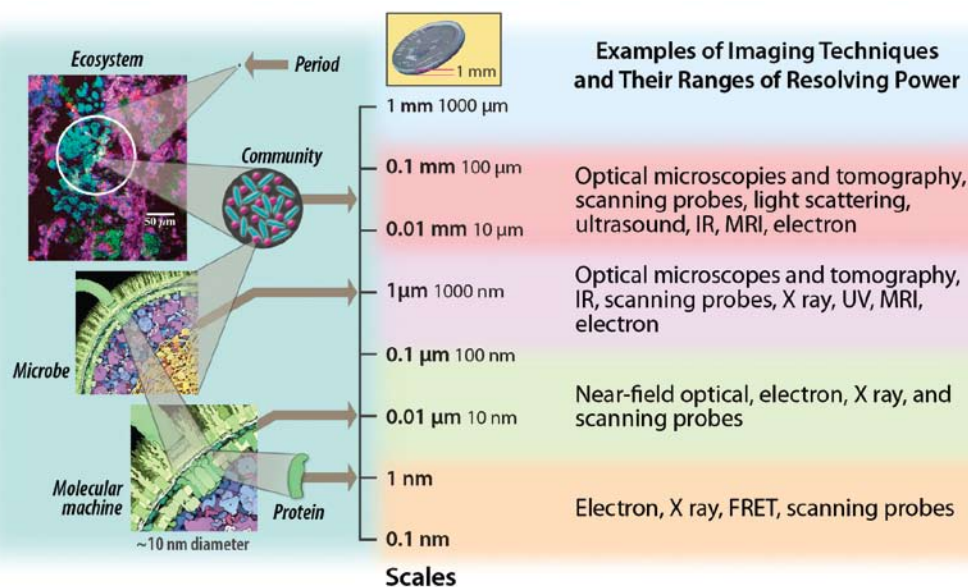
**Synthetic Biology.** Synthetic biology is the design and construction of novel organisms to generate products either not made by natural systems or for use as a research technique to test understanding. This research approach may involve building entirely new biological systems from a set of standard parts—genes, proteins, and metabolic pathways—or radically redesigning existing biological systems.



**Imaging Technologies.** A variety of imaging technologies can be used to investigate microbial community function, including microbe-plant interactions at different spatial scales. A microbial community isolated from a particular soil ecosystem can span microscale distances that are tiny fractions of a millimeter. At this microbial-community scale, technologies based on visible light and infrared radiation can be used to identify community members and their locations and link community structure to function. For bacterial cells, which are micrometers ( $\mu\text{m}$ , one-thousandth of a millimeter) in size, a mix of technologies ranging from optical microscopy to finer-resolution electron microscopies can be used to detect and track both extra- and intracellular states and to determine the dynamics of molecules involved in intercellular communication. Imaging technologies at the nanometer

(nm, one-thousandth of a micrometer) scale are used to detect, localize, and track individual molecules and subcellular structures within microbes. Preferably, measurements will be made in living systems over extended time scales and at the highest resolution. A number of techniques are emerging to address these demanding requirements; a brief listing is on the right side of the figure. [Figure source: GTL Roadmap, p. 176 (genomicsgtl.energy.gov/roadmap/).]

### Imaging Microbes, from Ecosystem to Molecular Scales



### Robust Computational Framework for GTL Research

GTL is addressing extremely complex science questions that span all scales of biology, requiring the collective expertise of scientists from many disciplines and the coordinated application of a wide range of technologies and experimental approaches—genome sequencing, gene expression profiling, proteomics, and metabolomics. The vast volume of heterogeneous data emanating from GTL research requires developing a robust computational framework for data integration and analysis, modeling, and the open sharing of information and data across the program and the scientific community.

At the heart of the computational framework will be the GTL Knowledgebase (see figure, GTL Integrated Computational Environment for Biology, p. 12). The knowledgebase is envisioned as a comprehensive collection of well-curated, integrated data repositories founded on genome sequences; powerful suites of software and other tools for data analysis and visualization, database querying, data mining, modeling, and simulation; and automated systems

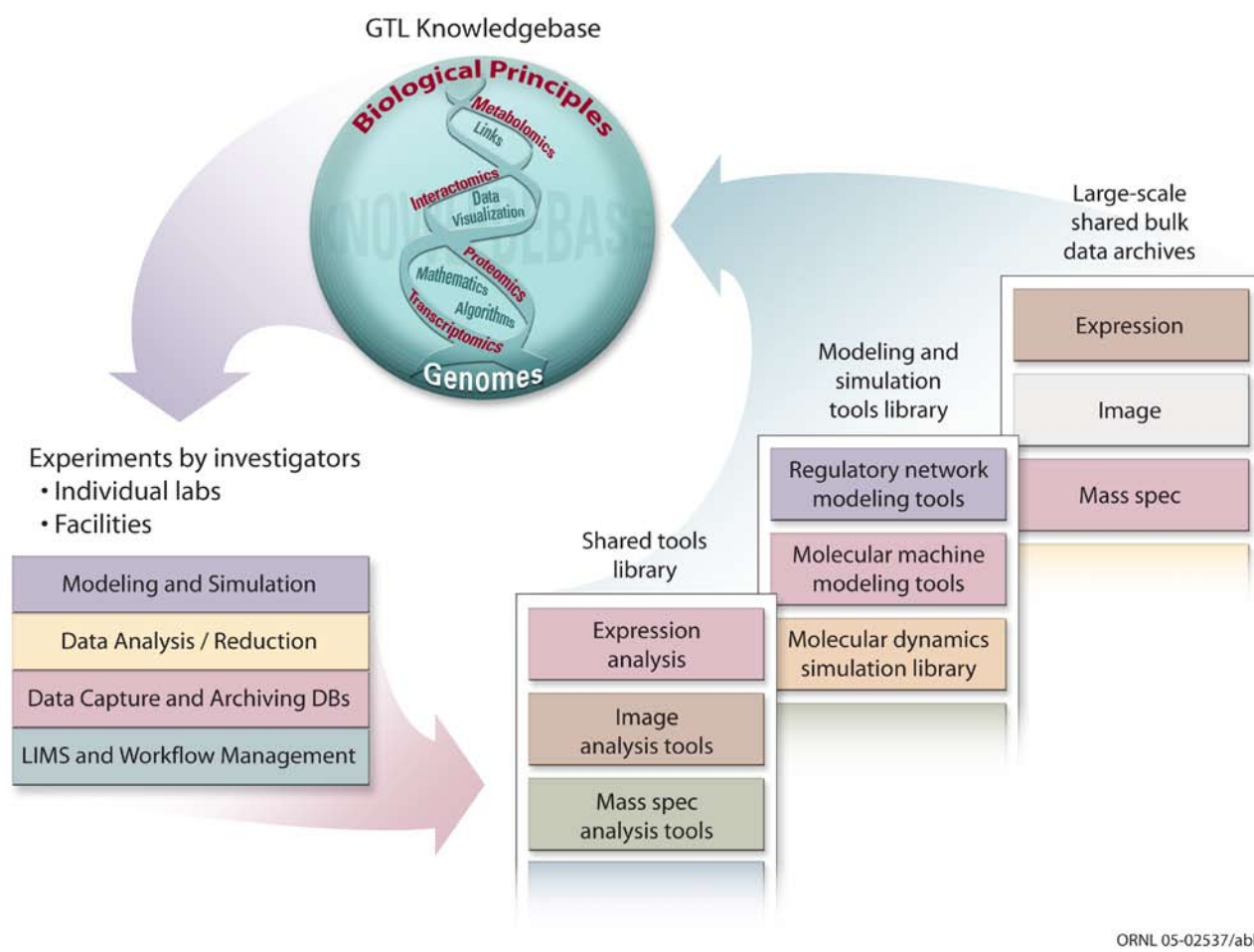
for generating value-added datasets and biologically meaningful information the scientific community can access and use freely. DOE OBER hosted a workshop in May 2008 to obtain input from leaders in the computational biology community who will help guide development of the GTL Knowledgebase. The workshop report will be available in spring 2009 at [genomicsgtl.energy.gov/compbio/](http://genomicsgtl.energy.gov/compbio/).

### Core GTL Computing Capabilities

Building the GTL Knowledgebase and other elements of a robust computational framework for systems biology requires continued development of several core capabilities. These core capabilities include tracking high-throughput experimental and contextual environmental data, capturing and archiving massive datasets, analyzing and reducing data, modeling and simulation, and facilitating community access to GTL data and resources.

**Tracking High-Throughput Experimental Data.** A Laboratory Information Management System (LIMS) is a computer program that tracks all samples throughout the complete experimental life cycle, monitors laboratory

## GTL Integrated Computational Environment for Biology



**GTL Knowledgebase.** The GTL Knowledgebase will assimilate a vast range of systems biology data and provide the community with data analysis, visualization, modeling, and simulation tools. Researchers will use the annotated database to develop models of biological systems that drive an iterative cycle of hypothesis formulation, experiment design, and large-scale data collection. This process will facilitate ongoing improvement of models, model-based predictions, and biological understanding. [From GTL Roadmap, 2005, p. 83 ([genomicsgtl.energy.gov/roadmap/pdf/GTL05\\_05Computing.pdf](http://genomicsgtl.energy.gov/roadmap/pdf/GTL05_05Computing.pdf))]

instrumentation, automates workflow, and coordinates experimental process management across distributed collaborations. LIMS is essential to managing the huge streams of data flowing from diverse high-throughput measurement of cellular activity. Several commercial LIMS products are available, but these systems must be customized to meet the needs of different GTL projects.

**Capturing and Archiving Massive Datasets.** The very large (petabyte) data volumes generated by systems biology research will require new representations of data from many different measurements (e.g., confocal imaging and mass spectrometry); identification of efficient formats for storing, retrieving, and

using different types of data; and development of new technologies for large-scale data transfer and storage.

**Analyzing and Reducing Data.** Basic research is needed to create new algorithms that handle diverse data types and improve data-reduction strategies for filtering out noise, compressing noncritical information, and identifying biologically meaningful subsets of data needed for subsequent analysis and modeling. Increasing analysis throughput is key. For example, time-intensive analysis of experimental images and video by individual researchers will need to be automated to increase throughput. Data analysis could be accelerated by research on grid-computing approaches and codes for clusters and massively parallel processing environments.

## GTL Information and Data Sharing Policy

The DOE Office of Biological and Environmental Research (OBER) has set the following guidelines and requirements for data and all publishable information resulting from GTL-funded research. The complete Information and Data Sharing Policy statement is available at [genomicsgtl.energy.gov/datasharing/](http://genomicsgtl.energy.gov/datasharing/).

- Notification to OBER when a publication reporting GTL-funded research appears in print.
- Submission of publication-related information (e.g., experimental protocols, raw or unprocessed data) to national or international public repositories within 3 months of publication. Accession numbers or other identifiers are to be reported to OBER.
- Use of community-recognized standards for ontologies and data formats when these standards exist.
- Involvement of the GTL research community in developing public repositories and standard ontologies for proteomics and other new and rapidly evolving research areas.
- Open-source availability (via license or submission to an open-source software community) of research software published in peer-reviewed publications.
- Storage of all experimental data in an archive maintained by a GTL project's principal investigator.
- Implementation of a Laboratory Information Management System (LIMS) or other electronic system to manage and archive experimental data for all projects with more than one senior investigator.

**Enabling Predictive Understanding.** Models and simulations represent an ultimate level of integrated understanding and enable scientists to test their knowledge via “virtual experiments” run on computers. To achieve GTL’s ultimate goal of a predictive understanding of biology, researchers must move beyond qualitative representations of biological systems. Creating models that predict bio-system response to untested conditions will require more quantitative details for biological processes (e.g., kinetic constants for enzyme activity). Also needed are new mathematical and machine-learning methods to address a biological system’s variable and evolving interactions with external signals from the environment. In addition to a lack of quantitative biological understanding, available computing power is a major limitation to the amount of biochemical detail, number of experimental conditions, and temporal and spatial ranges that can be included in biological models and simulations today.

**Facilitating Community Access to GTL Data and Resources.** When research information is not made publicly available to the global scientific community, a corresponding price is paid in lost opportunities, barriers to innovation and collaboration, and the obvious problem of unknowing repetition of similar work. The long-term success of GTL, and systems biology in general, depends on achieving high levels of data and information integration and sharing. OBER recently established an information and data sharing policy requiring public accessibility to all publishable information (see sidebar, GTL Information

and Data Sharing Policy, above). Ongoing development of this policy will help define standards and guidelines for establishing the GTL Knowledgebase—a one-stop shop for community access to GTL data, knowledge, and models.

## Elaboration of GTL Science for DOE Missions

GTL’s systems biology approach for each of the three missions follows.

### *Systems Biology for Energy Solutions*

By understanding the biological mechanisms underlying biofuel production, GTL will enable scientists to redesign, improve, and use biological systems to develop novel, efficient bioenergy strategies that can be replicated on an industrial scale (see figure, Cellulosic Biofuel Production Steps and Biological Research Challenges, p. 14). Systems biology tools and concepts for designing and engineering bioenergy plant and microbial systems will generate knowledge of the mechanistic bases for three key bioenergy challenges:

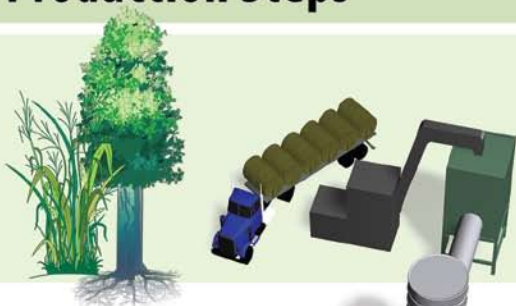
- High-yield, sustainable biomass crops with “designer” lignocellulosic composition.
- Consolidation of processes to use one microbial species or community for both degradation and fermentation.
- Production of biofuels beyond ethanol—improved energy density, processing, and handling.



## Cellulosic Biofuel Production Steps

### 1 Biomass Production and Delivery

Biomass is harvested, delivered to the biorefinery, and ground into particles.

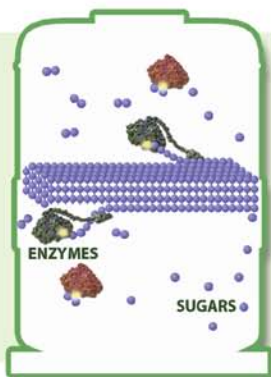


### 2 Pretreatment

Pulverized biomass is pretreated with heat and chemicals to make cellulose accessible to enzymes.

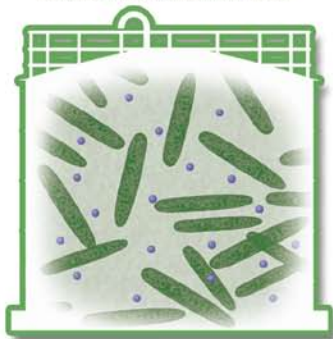
### 3 Cellulose Hydrolysis

Enzymes are added to break down cellulose chains into sugars.



### 4 Sugar Fermentation

Microbes ferment sugars into ethanol and other biofuels.



### 5 Biofuel Processing

Biofuels are extracted from the fermentation tank and prepared for distribution.



## Biological Research Challenges

### Biomass Production

- Sequence DNA from bioenergy crops.
- Identify genes and pathways that improve biomass productivity.
- Develop crops optimized for enzyme degradation.

### Pretreatment

- Identify enzymes that reduce the severity of thermochemical pretreatments.
- Minimize production of inhibitory by-products.

### Cellulose Hydrolysis by Enzymes

- Screen natural environments for the most efficient enzymes produced by fungi and bacteria.
- Understand how enzyme systems interact with cellulose.
- Increase the catalytic rate and thermal tolerance of enzymes.

### Consolidated Bioprocessing in Microbes

- Integrate biomass hydrolysis and fermentation into a single microbe or stable mixed culture.

### Sugar Fermentation by Microbes

- Engineer metabolic pathways to produce diverse biofuels.
- Increase product tolerance and yield.
- Develop microbes capable of efficiently fermenting a mix of all biomass sugars.

**Cellulosic Biofuel Production Steps and Biological Research Challenges.** This figure depicts some key processing steps in an artist's conception of a future large-scale facility for transforming cellulosic biomass (plant fibers) into biofuels. Three areas where focused biological research can lead to much lower costs and increased productivity include developing crops dedicated to biofuel production (see step 1), engineering enzymes that deconstruct cellulosic biomass (see steps 2 and 3), and engineering microbes and developing new microbial enzyme systems for industrial-scale conversion of biomass sugars into ethanol and other biofuels or bioproducts (see step 4). Biological research challenges associated with each production step are summarized in the right portion of the figure.



**Bioenergy Research Centers Established**

Alternative fuels from renewable cellulosic biomass—plant stalks, trunks, stems, and leaves—have the potential to significantly reduce U.S. dependence on imported oil while enhancing national energy security and decreasing the environmental impacts of energy use. To focus the most advanced biotechnology-based resources emanating from GTL on the biological challenges of biofuel production, DOE announced the establishment of three new Bioenergy Research Centers (BRCs) in June 2007. The BRCs were selected based on exceptional scientific merit determined by an independent peer-review panel. Pending continued availability of funding, each center will receive \$135 million over 5 years to pursue the basic research underlying a range of high-risk, high-return biological solutions for bioenergy applications (see table below).

The three centers are based in geographically diverse locations—the Southeast, Midwest, and West Coast—with partners across the nation. Oak Ridge National Laboratory was selected to lead the BioEnergy Science Center (BESC) in Tennessee; University of Wisconsin–Madison was selected to lead the Great Lakes Bioenergy Research Center (GLBRC); and Lawrence Berkeley National Laboratory was selected to lead the Joint BioEnergy Institute (JBEI). Each center represents a multidisciplinary partnership with expertise spanning the physical and biological sciences, including genomics, microbial and plant biology, analytical chemistry, computational biology and bioinformatics, and engineering. Institutional partners include DOE’s world-class national laboratories, universities, private companies, and nonprofit organizations (for more information, see <http://genomicsgtl.energy.gov/centers/brcbrochure.pdf>).

**DOE Bioenergy Research Center Strategies at a Glance**

The complexity of the three biological grand challenges that must be overcome to achieve industrial-scale bioenergy production requires the coordinated pursuit of numerous research approaches to ensure timely success. The DOE Bioenergy Research Centers\* represent a portfolio of diverse and complementary scientific strategies that will address these challenges on a scale far greater than any effort to date. Some strategies are listed briefly in the table below.

**Challenge: Development of Next-Generation Bioenergy Crops**

**Center Strategies**

- **BESC** – Decrease or eliminate harsh chemical pretreatments by engineering plant cell walls in poplar and switchgrass to be less recalcitrant; simultaneously increase total biomass produced per acre.
- **GLBRC** – Engineer “model” plants and potential energy crops to produce new forms of lignin and more starches and oils, which are more easily processed into fuels.
- **JBEI** – Enhance lignin degradation in “model” plants by changing cross-links between lignin and other cell-wall components; translate genetic developments to switchgrass.

**Challenge: Discovery and Design of Enzymes and Microbes with Novel Biomass-Degrading Capabilities**

**Center Strategies**

- **BESC** – Screen natural thermal springs to identify enzymes and microbes that effectively break down biomass at high temperatures; understand and engineer cellulosomes (multifunctional enzyme complexes for degrading cellulose).
- **GLBRC** – Identify combinations of enzymes and pretreatment needed to digest specific biomass types; express biomass-degrading enzymes in the stems and leaves of corn and other plants.
- **JBEI** – Improve performance and stability of enzymes harvested from the rainforest floor and other environments; engineer, through directed evolution, highly efficient cellulase enzymes.

**Challenge: Discovery and Design of Microbes that Transform Fuel Production from Biomass**

**Center Strategies**

- **BESC** – Reduce the number of cellulosic ethanol production steps by engineering a cellulose-degrading microbe to produce ethanol more efficiently.
- **GLBRC** – Reduce the number of cellulosic ethanol production steps by engineering an efficient ethanol-producing microbe to degrade cellulose.
- **JBEI** – Connect diverse biological parts and pathways to create new organisms that produce fuels other than ethanol; engineer organisms to produce and withstand high concentrations of biofuels; derive useful chemical products from lignin degradation.

\*BESC: BioEnergy Science Center; GLBRC: Great Lakes Bioenergy Research Center; JBEI: Joint BioEnergy Institute

## Systems Biology for Understanding the Carbon Cycle

Understanding biological influences on the global carbon cycle is critical to advancing climate change research and improving climate-model projections. GTL research can contribute by:

- Examining biological carbon sources and sinks in terrestrial and ocean systems that fix, transform, or reemit CO<sub>2</sub>.
- Facilitating connection of data across multiple scales of complexity: organism, community, ecosystem.
- Improving integration of experimental approaches and modeling efforts.
- Providing fundamental knowledge that will guide potential mitigation strategies.

Systems biology tools to determine the carbon cycling and biosequestration processes of ocean and terrestrial ecosystems include

- Comprehensive global ecosystem models incorporated into Earth System climate models to determine climate impacts and ecosystem feedbacks.
- Genome-based ecophysiology of plant and microbial communities for understanding their carbon cycling

abilities as the bases for CO<sub>2</sub> biosequestration strategies and technologies.

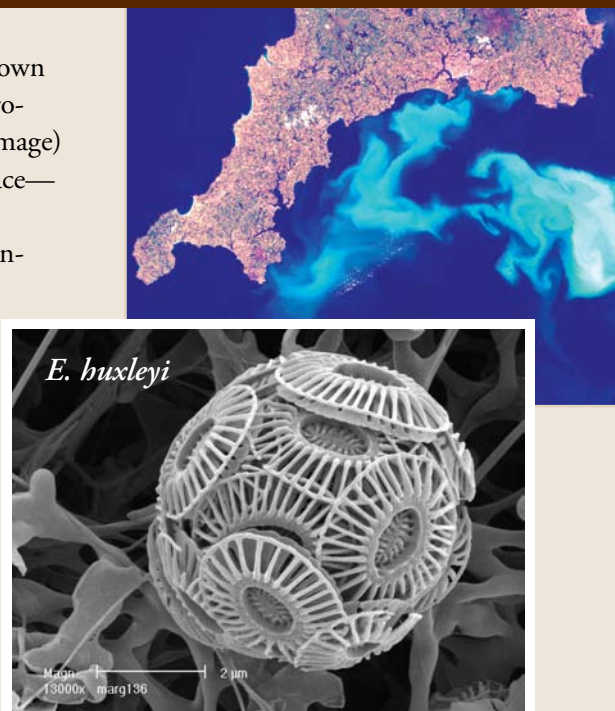
Determining the safe level of CO<sub>2</sub> in the atmosphere and the timing for achieving it is the primary driver for understanding how net energy emissions to the atmosphere affect climate and the global carbon cycle. A major mission of OBER research programs focuses on increasing understanding of carbon cycling in the Earth's marine and terrestrial ecosystems, examining potential means of biological sequestration of carbon, and determining how climate change affects biological processes that influence carbon cycling and biosequestration. Of particular interest is the linkage of global biogeochemical processes to genome-based biological functions of plant and microbial communities. These studies will be critical in developing increasingly sophisticated models of global biogeochemical cycling and climate change and will shape potential carbon biosequestration strategies (see sidebar, Planet-Transforming Microbes, below).

The following are key carbon cycle science and management questions:

- What are the sources and sinks in the global carbon cycle and how will they respond to changes in conditions

## Planet-Transforming Microbes

**B**looms of *Emiliania huxleyi*, captured by satellite, are shown just off the coast of the United Kingdom. Though microscopic, these carbon cycling coccolithophores (bottom image) are present in such large numbers that they are visible from space—an indicator of their pervasiveness and thus influence on ocean ecosystems (top image). Their shells are made of calcium carbonate, and over the ages their deposits have created the White Cliffs of Dover on the southern coast of England. Understanding the planet-transforming capabilities of these and other ocean microbes—that is, how they affect ocean ecosystems by cycling carbon and other important elements—is a focus of the GTL program. *E. huxleyi*'s genome was sequenced by DOE's Joint Genome Institute. For more information, see <http://genome.jgi-psf.org/Emihu1/Emihu1.home.html> and <http://www.noc.soton.ac.uk/soes/staff/tt/>.



Top – Landsat 7 photo of July 24, 1999, by S. Groom, Plymouth Marine Laboratory Remote Sensing Group; bottom – J. Young, Natural History Museum, London

brought about by climate change and expanded human activities? What is the mechanistic basis of these processes?

- What is the annual capacity of global ecosystems to absorb anthropogenic carbon emissions? How do human land-use choices influence this capacity?
- How will global carbon cycling systems and their properties and phenomena be altered by climate change? What is the mechanistic basis of these transformations?
- What measurement and analysis options are available to explore the behaviors of ecosystems in previously unseen conditions?
- Are there potential pathways to massive carbon release from global systems and how would they be recognized?
- What impact will substantial to profound climate change have on carbon stores and their dynamics?
- Will net carbon release from ecosystems under future scenarios of climate change lead to substantial positive feedbacks to the climate system?
- What are the key factors in soil ecology that control terrestrial carbon stores and their dynamics?
- How can modern tools of biology be used to elucidate the biological ecophysiology of such systems and predict their global behaviors?
- How can systems of global sources and sinks be managed and manipulated to optimize carbon capture and storage for mitigation purposes? How can large-scale use of energy crops be designed for vast carbon capture and storage?

### Microbes and the Carbon Cycle

Extremely diverse ocean and terrestrial microbial communities serve fundamentally different roles in the carbon cycle, as primary photosynthetic producers of biomass in the ocean biological “pump” and as carbon and nutrient managers and decomposers in terrestrial systems. Microbes cycle immense volumes of carbon in the process of recycling most of Earth’s biomass: They can fix CO<sub>2</sub> by light-driven (photoautotrophy) and geochemically driven (lithoautotrophy) reactions, generate methane, produce CO<sub>2</sub> as they decompose organic matter, precipitate carbonate minerals, and catalyze the polymerization of plant polymers into recalcitrant pools of carbon in soil.

### Research Needs Identified at Workshop

To engage researchers in a dialogue on future directions of biology-based carbon cycle research, DOE OBER held the workshop, Carbon Cycling and Biosequestration, in March 2008. Its aims included identifying research needs and opportunities for understanding biological carbon cycling and biosequestration, providing an assessment of current science and technology, as well as barriers to progress, and describing new directions for fundamental research. Areas described for study include terrestrial plant productivity and carbon biosequestration, biological cycling of carbon in terrestrial and oceanic environments, effects of climate change on carbon cycling and biosequestration, and crosscutting science. The report will be available at <http://genomicsgtl.energy.gov/carboncycle/>.

### Systems Biology for Understanding Contaminant Transport in the Environment

Legacy waste from Cold War weapons production, processing, and storage activities has left many DOE sites with numerous and unique soil and groundwater contamination challenges for which no practical remediation solutions currently exist. It is therefore critical that DOE understand the behavior of environmental contaminants and provide science-based information for sound decision making for remediation strategies and long-term stewardship of DOE sites (see sidebar, Microbe-Mineral Interface in Contaminated Environments, p. 18).

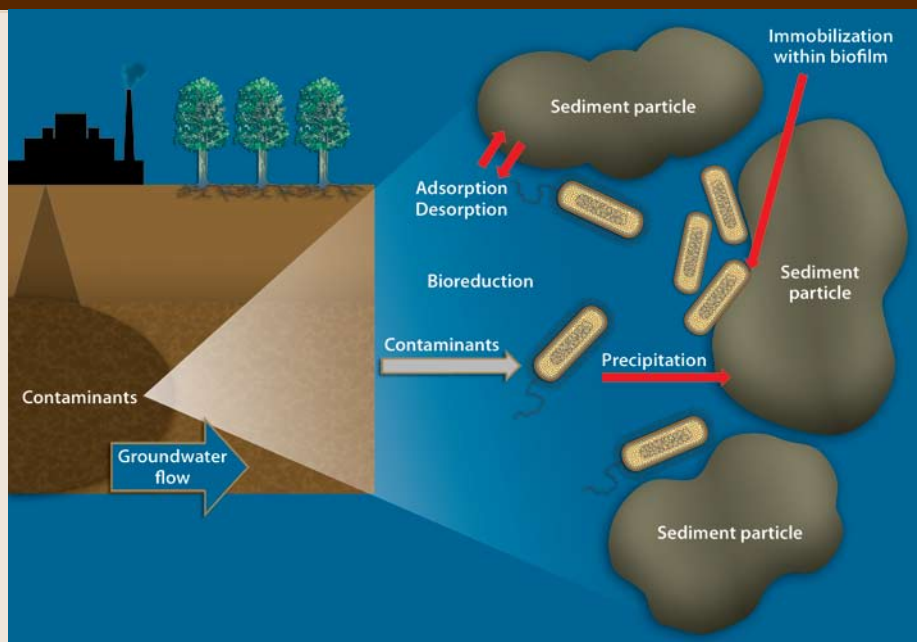
Numerous processes influence contaminant transport in the environment, but microbiological activity is arguably the most complex and least-understood component of current contaminant transport models or in situ environmental remediation technologies. Microbial activity can affect the mobility of environmental contaminants profoundly, yet the tools to investigate and understand such activity in a mechanistic manner are only now possible through a systems biology approach.

To facilitate incorporation of systems biology techniques into environmental descriptions of microbial processes, the GTL program maintains strong links with the multidisciplinary efforts within OBER’s Environmental Remediation Sciences Program (ERSP). GTL will develop methods to relate genome-based understanding of molecular processes

## Microbe-Mineral Interface in Contaminated Environments

**B**iogeochemical processes driven by interactions at the microbe-mineral interface in soils and sediments influence contaminant behavior. These reactions occur at the level of the individual cell or groups of cells, forming an architecture assembled against the mineral surface. The cells interact with each other and the mineral surface, creating a dynamic, microscale domain that controls the kinetics of biogeochemically mediated reactions.

Before large-scale restoration strategies can be implemented, such processes must be better understood to predict contaminant transport in situ under natural and induced conditions. (“Induced conditions” refers to the addition of nutrients, oxygen, or other electron donors and acceptors to increase microbial activity.) Achieving this level of knowledge requires new tools to characterize and resolve fundamental phenomena at the molecular, cellular, and community levels. Microbial information must be integrated with the subsurface strata’s geochemical characterization and the system’s hydrological properties at the sediment mineral and pore scales.



to long-term conceptual and predictive models for simulating contaminant fate and transport and development of remediation strategies. Challenges for GTL and its collaborators include:

1. Understanding the complex interactions of microbes with contaminants and the subsurface environment will allow predictions of microbial behavior to be based on fundamental knowledge of integrated biological, geochemical, and hydrological processes controlling contaminant biogeochemistry in the subsurface.
2. Developing new suites of biosensors and performance-assessment tools and analyzing natural microbial communities’ functions using genome-based, multidisciplinary, field-oriented approaches will advance development of robust, in silico models. Such models will allow a mechanistic and predictive understanding of microbial responses to contaminants or nutrients. These integrated models describing the fate and transport of environmental contaminants could save billions of dollars by supporting decisions to exploit natural attenuation alternatives, use environmental remediation for previously intractable problems, or improve the efficiency of conventional technologies.
3. Characterizing biogeochemical processes from the fundamental molecular to community levels to describe contaminant-transformation processes coinciding with simulated changes in microbial community composition and structure.



## DOE Office of Science

The Department of Energy's (DOE) Office of Science is the nation's single-largest supporter of basic research in the physical sciences and helps ensure U.S. world leadership across a broad range of scientific disciplines. The Office of Science supports a diverse portfolio of research at more than 300 colleges and universities nationwide and manages 10 world-class national laboratories with unmatched capabilities for solving complex interdisciplinary scientific problems. It also builds and operates the world's finest suite of scientific facilities and instruments used by more than 21,000 researchers in FY 2007 to extend all areas of science.

## DOE Office of Biological and Environmental Research

The Office of Science's Office of Biological and Environmental Research (OBER) is the home of the Genomics:GTL program. OBER supports DOE's mission of world-class fundamental scientific research through peer-reviewed multidisciplinary projects in the following areas.

- **Life Sciences:** Provide the fundamental scientific understanding of plants and microbes necessary to develop new robust and transformational basic research strategies for producing biofuels, cleaning up waste, and sequestering carbon.
- **Climate Change Research:** Deliver improved scientific data and models on the potential response of the Earth's climate and terrestrial biosphere to increased greenhouse gas levels so policymakers can determine safe levels of such gases in the atmosphere.
- **Environmental Remediation:** Provide sufficient scientific understanding to allow DOE sites to incorporate coupled physical, chemical, and biological processes into decision making for environmental remediation and long-term stewardship.
- **Medical Applications and Measurement Science:** Develop intelligent biomimetic electronics that can both sense and correctly stimulate the human nervous system.
- **Facilities:** Manage facility operations to the highest standards of overall performance using merit evaluations with independent peer review.

### For More Information

**Genomics:GTL**  
[genomicsgtl.energy.gov](http://genomicsgtl.energy.gov)

**GTL Roadmap**  
[genomicsgtl.energy.gov/roadmap/](http://genomicsgtl.energy.gov/roadmap/)

**DOE Mission Focus: Biofuels**  
[genomicsgtl.energy.gov/biofuels/](http://genomicsgtl.energy.gov/biofuels/)

*Breaking the Biological Barriers to Cellulosic Ethanol:  
A Joint Research Agenda*  
[genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml](http://genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml)

**DOE Bioenergy Research Centers**  
[genomicsgtl.energy.gov/centers/](http://genomicsgtl.energy.gov/centers/)

**DOE-USDA Plant Feedstock Genomics for Bioenergy**  
[genomicsgtl.energy.gov/research/DOEUSDA/](http://genomicsgtl.energy.gov/research/DOEUSDA/)

**DOE Joint Genome Institute**  
[jgi.doe.gov](http://jgi.doe.gov)

**Environmental Molecular Sciences Laboratory**  
[www.emsl.pnl.gov](http://www.emsl.pnl.gov)

**GTL Image Gallery**  
[genomics.energy.gov/gallery](http://genomics.energy.gov/gallery)

**DOE Office of Biological and Environmental Research**  
[science.doe.gov/ober/](http://science.doe.gov/ober/)

**DOE Office of Science**  
[science.doe.gov](http://science.doe.gov)

**U.S. Department of Energy**  
[energy.gov](http://energy.gov)

