

DOE Joint Genome Institute Strategic Planning for the Genomic Sciences

Workshop Report

DOE/SC-0152

EXECUTIVE SUMMARY



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research

DOE Joint Genome Institute Strategic Planning for the Genomic Sciences Report from the May 30–31, 2012, Workshop

Convened by

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

Co-Chairs

Jim Fredrickson Pacific Northwest National Laboratory
Michael Laub Massachusetts Institute of Technology
Jan Leach Colorado State University

Breakout Group Chairs

Richard Michelmore University of California, Davis
Kimmen Sjölander University of California, Berkeley
Tom Schmidt Michigan State University

Organizer

Daniel Drell

daniel.drell@science.doe.gov, 301.903.4742

Web address for entire document:

genomicscience.energy.gov/userfacilities/jgi/futuredirections/



About the Cover

Images on the cover represent a broad range of the complex scales encompassed by the science supported by the Office of Biological and Environmental Research (BER) within the U.S. Department of Energy's (DOE) Office of Science. These scales range from genes at the subcellular level to genomes of microbes and their communities, to the genomics of plant-microbe interactions and plants that could be feedstocks for bioenergy, to the scale of ecosystem and landscape function. The underlying DNA strand represents the DNA sequence data that provide the foundation for further systems-level experimentation. The images culminate in a wired cell that represents the predictive understanding of biological systems sought by BER programs. DNA sequence data generated by the DOE Joint Genome Institute user facility are having a major impact toward the achievement of this goal.

Image credits: Microscopic images of organic matter decomposers (copyright Corbis), fungal hyphae on a root surface (copyright Corbis), and cross-section of a switchgrass stem (DOE BioEnergy Science Center and National Renewable Energy Laboratory). Grassland habitat (U.S. Department of Agriculture Natural Resources Conservation Service). Spruce-peatland ecosystem (Oak Ridge National Laboratory). Aerial view of Arctic landscape (Oak Ridge National Laboratory). Cover developed at Oak Ridge National Laboratory.

Suggested Citation for Entire Report

U.S. DOE. 2012. *DOE Joint Genome Institute Strategic Planning for the Genomic Sciences: Report from the May 30–31, 2012, Workshop*, DOE/SC-0152, U.S. Department of Energy Office of Science.

EXECUTIVE SUMMARY

DOE Joint Genome Institute Strategic Planning for the Genomic Sciences

Report from the May 30–31, 2012, Workshop

Published August 2012

Convened by

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



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Executive Summary

The U.S. Department of Energy (DOE) Joint Genome Institute (JGI) Strategic Planning for the Genomic Sciences workshop was convened by the DOE Office of Science’s Biological and Environmental Research Program (BER) on May 30–31, 2012. The goal was to explore DOE JGI’s role in addressing DOE mission-critical scientific questions and in contributing data and knowledge to enable a new generation of systems biology research (see Sidebar 1, below).

DOE JGI has played a leadership role in genome sequencing, providing a foundation for complex biological studies. As DOE JGI moves forward into the next decade(s), it will have a continued and expanded role in genome sequencing. A major opportunity lies before DOE JGI as it seeks to build on its sequencing strength by providing high-throughput, “value-added” science that can be integrated with massive sequence datasets to accelerate the science underpinning DOE missions in bioenergy and the environment.

In particular, as a DOE scientific user facility that accelerates users’ research with capabilities not available in their own laboratories, DOE JGI could lower access barriers to cutting-edge capabilities and provide expertise to advance mission-relevant science by reducing the gap between genotype data (i.e., the gene sequence) and

phenotype data (what genes and their products do). Addressing this gap is of fundamental importance to the DOE mission (see report Appendix 1: Grand Challenges).

To elicit DOE JGI strategies for achieving these advances, BER invited participants from universities, DOE national laboratories, and other federal agencies with broad expertise in the biological sciences and bioinformatics. The workshop’s conclusions highlight the following capabilities: sequencing

Sidebar 1

Major Themes and Needs Emerging from the Workshop

1. Continued and increasing large-scale sequencing of biologically important organisms and communities. Extant sequencing targets range from viruses to complex, multicellular communities associated with a wide range of environments directly relevant to Department of Energy (DOE) and Biological and Environmental Research Program (BER) missions. Sequencing is becoming increasingly important as an experimental measurement, adding to or even replacing current measurement technologies—a quantitative “readout” rather than just an end in itself.
2. Large-scale functional genomics technologies for high-throughput functional annotation, informed by global measurements of actual cellular activities rooted in genome sequencing. Functional genomics technologies include those provided by whole-expression, proteomic, metabolomic, and genome-wide association studies, at several levels, including individual cells, multicellular organisms, and communities.
3. Extended and improved bioinformatics methods to enable integration and analysis of unprecedented quantities of data and to generate testable hypotheses critical to advancing DOE science.
4. Aggressively expanded capacity not only to sequence (“read”) DNA, but also to synthesize (“write”) DNA, enabling scientists to manipulate genomes. This capability will be essential for directed exploration of gene and genomic manipulations of biological functions relevant to DOE missions.
5. Improved automation of biological experiments to match the throughput now prevalent in sequencing. The costs of these technologies must be reduced so they can keep pace with improvements in sequencing technologies.
6. Communities of scientists led by the DOE Joint Genome Institute (JGI) and organized around key mission-relevant scientific questions. Addressing these questions and challenges will be feasible through the development of new biological approaches made possible by next-generation sequencing and follow-on technologies (that DOE JGI could establish and/or adopt). Novel biological approaches are critical to build an understanding of, and an ability to predict, biological behaviors required for complex applications such as biofuel production and understanding of biological feedbacks to the climate system.

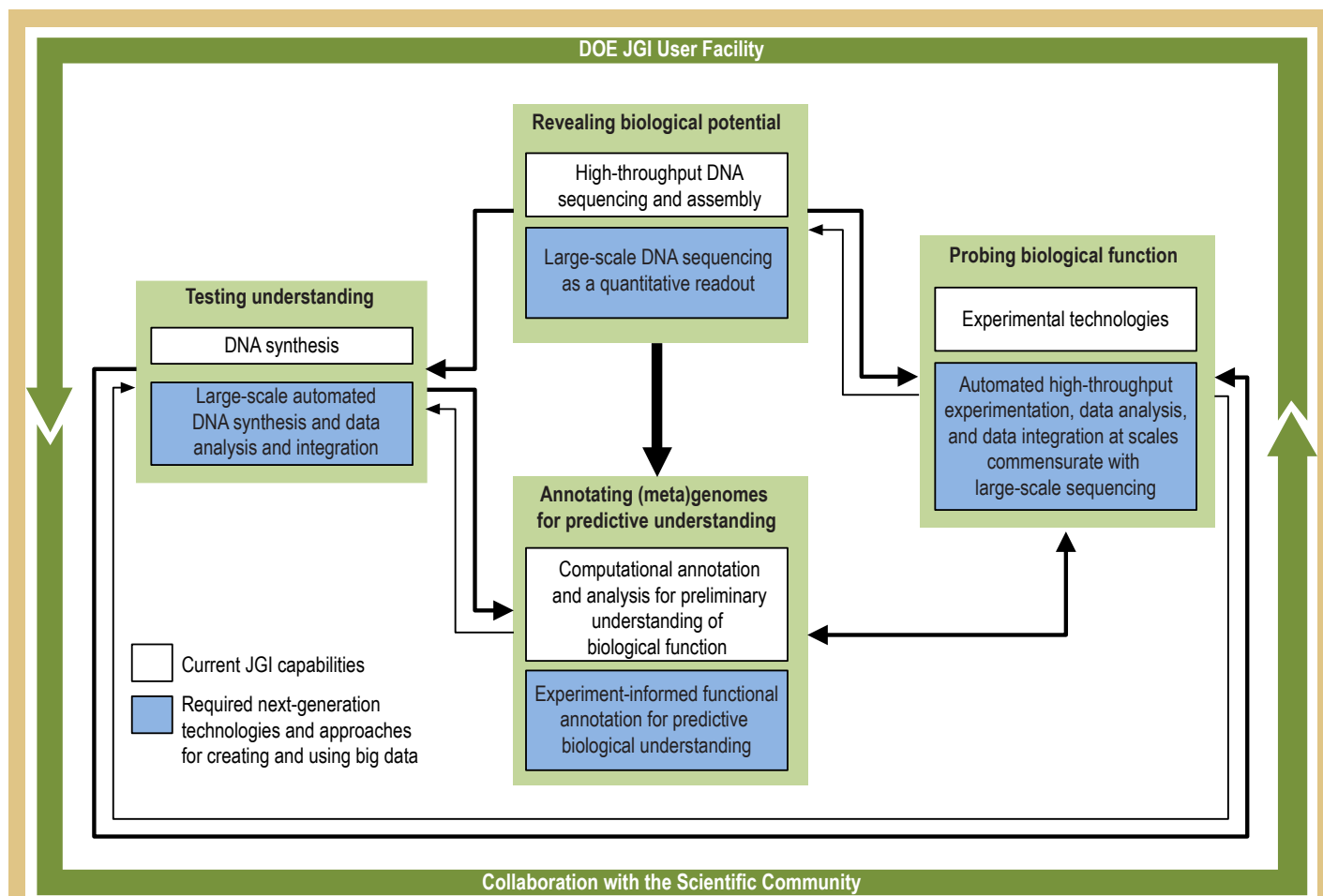


Fig. 1. Pathway to Predictive Biological Understanding. This figure depicts an integrated view of the current and future capabilities to be explored by the DOE JGI user facility in collaboration with the research community. These capabilities are required for improving the accuracy, efficiency, and effectiveness of annotations resulting from computational analyses. DNA sequence data provide the foundation for predictive understanding, revealing the biological potential in a genome or genomes. Experimentation probes the behavior of biological systems under different environmental conditions and informs hypotheses generated from sequencing and analyses. Because of the output of current and imminent sequencing technologies, however, high-throughput experimental data analysis and data integration capabilities are needed at lower cost and greater level of automation. New understanding gained through these integrated efforts can then be tested by additional experiments, including the use of synthetic techniques to build test systems. The thickness of the arrows represents the level of current challenges to information flow. Importantly, the setting of this conceptual pathway inside of the larger scientific community implies a strong supportive and collaborative relationship. The resulting experiment-informed functional annotations will enable research into more complex biological systems that could not otherwise be studied, leading to the predictive understanding required for DOE missions in bioenergy and the environment.

DNA, annotating DNA, addressing “Big Data” opportunities and challenges, writing DNA and developing associated technologies, implementing high-throughput experimentation, and building research communities (see Fig. 1. Pathway to Predictive Biological Understanding, above).

Continuing to Sequence DNA

The impressive accumulated sequencing accomplished to date by DOE JGI and other genome centers is insignificant compared with the diversity and sheer number of microbes, fungi, plants, other eukaryotes, and particularly complex

communities that remain unexplored. This diversity and variation is critical to biological functions relevant to DOE missions and is only beginning to be sampled by current sequencing efforts. Consequently, there was strong consensus among workshop participants that DOE JGI should continue high-throughput and high-quality sequencing. DOE JGI thus will play an important role in refining and developing sequencing-related technologies, including single-cell sequencing and metatranscriptome analysis of diverse species and communities.

Although sequencing capacity continues to grow rapidly, future efforts need to be guided by scientific questions relevant to DOE missions in bioenergy and the environment. To obtain the most value from continued sequencing, carefully selected model species and “model” environments should be identified to provide basic sequence knowledge and to nucleate better functional and structural genome annotations. Continued sequencing will supply the foundational “raw material” for DOE JGI contributions to:

- Building mechanistic models for biological processes.
- Generating a deep understanding of key (“flagship”) organisms.
- Reducing the fraction of genes whose functions remain unknown through improved function prediction protocols.
- Supporting investigations into the individual genetic variation within cells in a species, within species in a population, and between populations.
- Monitoring changes in expression profiles over time and other high-volume, sequence-critical biological measurements.

Annotating DNA

A consensus of workshop participants agreed that functional annotation following genomic sequencing and structural annotation is one of the biggest challenges confronting the entire biology community (not just DOE JGI). Functional annotations for newly identified genes will require a combination of computational and experimental methodologies, enabling DOE JGI to generate testable hypotheses of function and to capture experimental data where available.

A number of avenues are available for DOE JGI to become a more involved participant in functional annotation and to narrow the gap between generating and understanding genome sequence. Ideas emerging from the workshop included DOE

JGI (1) having a leadership role in integrating different data types from sources worldwide; (2) engaging the scientific community to reveal genomic “dark matter” (e.g., conserved hypothetical genes), perhaps by evolving a competition process or exploiting crowd-sourcing and social networking technologies; (3) exploring the utility of new experimental and computational technologies to integrate data and functional inference protocols for genes, genomes, and metagenomes; (4) generating or acquiring validated high-accuracy genome and gene reference datasets; (5) seeding the development of active end-user communities involved in functional annotation; and (6) encouraging and assisting the development of improved experimental design approaches.

Functional annotations could benefit from novel bioinformatics technologies to accelerate experimental validation of putative functional assignments, as well as from better ways of linking plant or microbial genetic diversity with ecosystem function.

Addressing “Big Data” Opportunities and Challenges

As sequencing technologies have increased in speed and throughput, data generation has exceeded both storage and analysis capabilities. New computational tools to acquire, curate, analyze, and distribute information from these datasets are needed for all DOE JGI user communities. Addressing scientific questions relevant to DOE missions will require effective data integration. This task is dependent on the generation of complementary data types (e.g., transcriptomics and proteomics) in appropriate volumes; intelligent merging and association of disparate data types from a wide variety of national and international entities; and development of appropriate conventions for controlled vocabularies, genome descriptions, and metadata. To achieve these capabilities, DOE JGI must promote resource integration involving other facilities and utilization of other DOE Office of Science assets (see report Appendix 2: Department of Energy Assets, p. 29).

Writing DNA and Developing Associated Technologies

A clear message from the workshop was that DOE JGI should establish complementary technological capabilities, including “on demand” DNA synthesis and miniaturization of appropriate analysis technologies. “Writing” DNA is viewed

as a necessary complement to sequencing in ways that would accelerate DOE JGI science. Several applications are centered on generating desired sequences, introducing them into an appropriate cell, and exploring the effects on physiology, metabolism, cellular architecture, and responses to stimuli. By altering genetic information in defined ways, the process of linking the behaviors of gene products to sequence variants will be accelerated dramatically. Similarly, the relation of sequence to three-dimensional (3D) structures, of both proteins and chromosomes, can be explored much more effectively. DNA synthesis is a powerful tool for investigating biological processes, ranging from the individual gene or gene product (protein) to protein complexes, metabolic pathways, regulatory networks, and even an entire cell and complex communities. This scale remains difficult and challenging but now can be addressed with new high(er)-throughput approaches.

Implementing High-Throughput Experimentation

An important opportunity for DOE, BER, and DOE JGI is to combine the ideas and algorithms underlying the automation of hypothesis testing with the economies of scale possible with microfluidics devices. These combined

capabilities could dramatically reduce the cost of doing science, by sifting the credible hypotheses (the wheat) from the less credible (the chaff), improving the pace of discovery in functional genomics, directed evolution, and biological design. DOE JGI can accelerate this hypothesis-generation process with advanced automated experimental technologies that contribute to the understanding of genome sequences.

Building Research Communities

DOE JGI should continue to serve as a user facility, providing a valuable service to the biological community and participating in active collaborations to achieve science that biologists could not readily carry out in their individual laboratories. To this end, DOE JGI should stimulate and support community efforts for large science by providing high-throughput sequencing and other value-added services focused on problems of scale and complexity that exceed the ordinary. This endeavor will require a sustained commitment of time and resources and take advantage of DOE JGI's unique abilities and expertise. DOE JGI should actively seed and promote interdisciplinary teams and initiate stable collaborations among individual researchers, laboratories, and institutions that focus on mission-relevant science.

