# Overview of the DOE Systems Biology Knowledgebase and Related Research Activities

KBASE predictive biology
DOE Systems Biology Knowledgebase

The Office of Biological and Environmental Research (BER) within the U.S. Department of Energy's (DOE) Office of Science advances world-class biological and environmental research and provides scientific facilities to support DOE missions in scientific discovery and innovation, energy security, and environmental responsibility. As a leader in systems biology, BER's Genomic Science program supports scientific research that seeks to achieve a predictive understanding of microbial and plant systems relevant to DOE missions (genomicscience.energy.gov). By revealing the genetic blueprints and fundamental principles that control the biological functions of these systems, the Genomic Science program advances the foundational knowledge underlying biological approaches to producing biofuels, sequestering carbon in terrestrial ecosystems, and cleaning up contaminated environments.

The program funds a portfolio of systems biology research that produces petabytes of data annually. Examples include genomic sequences on microbes, plants, and complex environmental samples; proteomic data; expression data; isotopic flux data for pathway analysis; protein binding data for functional annotation; data from imaging of proteins localized in subcellular compartments; and metadata associated with diverse experimental conditions and sampling techniques. Integrating and using these diverse data to develop predictive models of biological systems will require an integrated computational environment. To provide the research community with such a resource, the Genomic Science program is developing the DOE Systems Biology Knowledgebase (Kbase; see Fig. 1). A knowledgebase is a cyberinfrastructure consisting of a collection of data, organizational methods, standards,

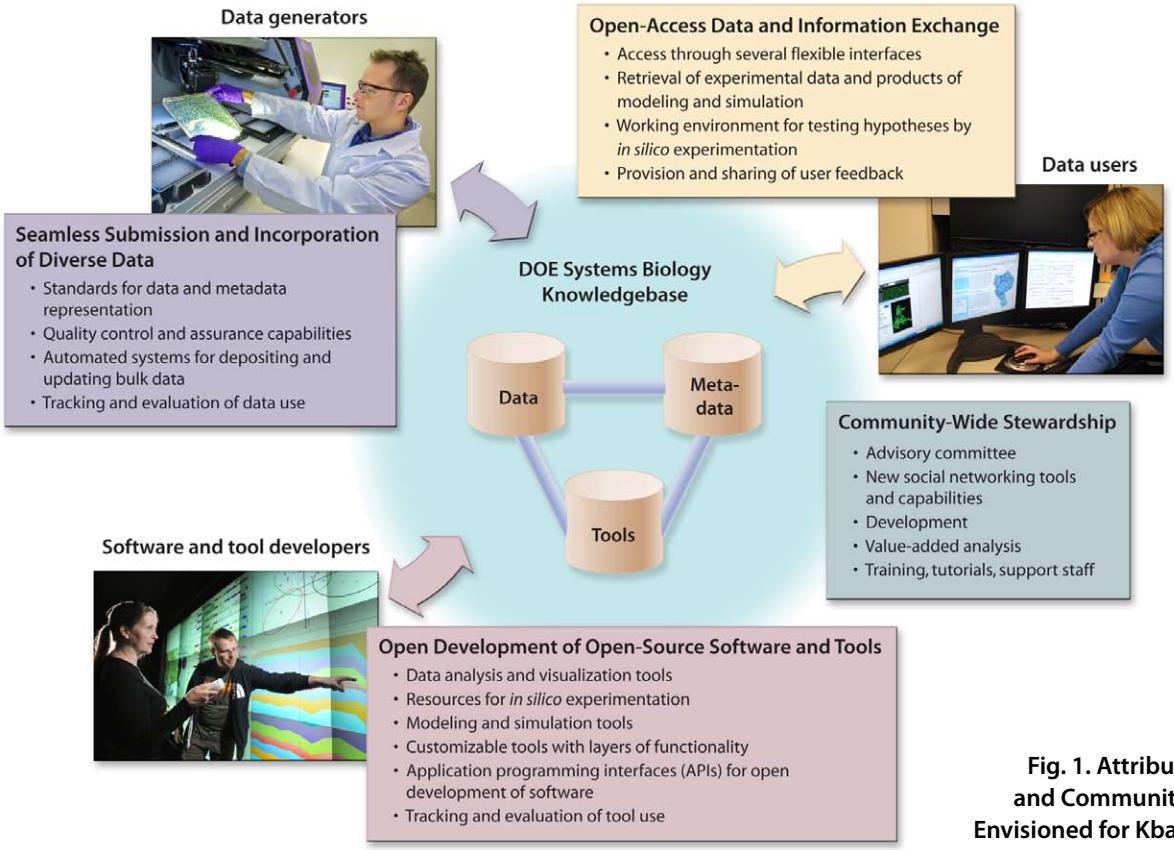



**Fig. 1. Attributes and Communities Envisioned for Kbase.**

analysis tools, and interfaces representing a dynamic body of knowledge.

The fully functional Kbase is a cyberinfrastructure for systems biology information and data that supports open community science. It not only will include data storage, retrieval, integration, and management, but also will enable new knowledge acquisition through free and open access to data, analysis tools, resources for modeling and simulation, and information for the research community. Kbase differs from current informatics efforts by bringing together research products from many different projects and laboratories to create a comprehensive computational environment focused on DOE scientific objectives in microbial, plant, and metacommunity (complex communities of organisms) research. By democratizing access to data and computational resources, Kbase will enable any laboratory or project, regardless of size, to participate in a transformative community-wide effort for advancing systems biology and accelerating the pace toward predictive biology.

This document provides an overview of the DOE Systems Biology Knowledgebase (p. 3), a collaborative effort between four national laboratories and many research universities. It also describes Kbase development efforts carried out during the past year and summarizes current research, including:

- **Knowledgebase R&D project.** Completed in September 2010, this effort included five pilot projects that laid the foundation for Kbase. Summaries of these early successes begin on p. 6.
- **University-led projects to develop computational biology and bioinformatic methods enabling Kbase.** Descriptions of the 11 funded projects awarded in 2010 begin on p. 9.

Together, these activities—along with existing user-community data and resources—underpin development of the Kbase program (see Fig. 2).

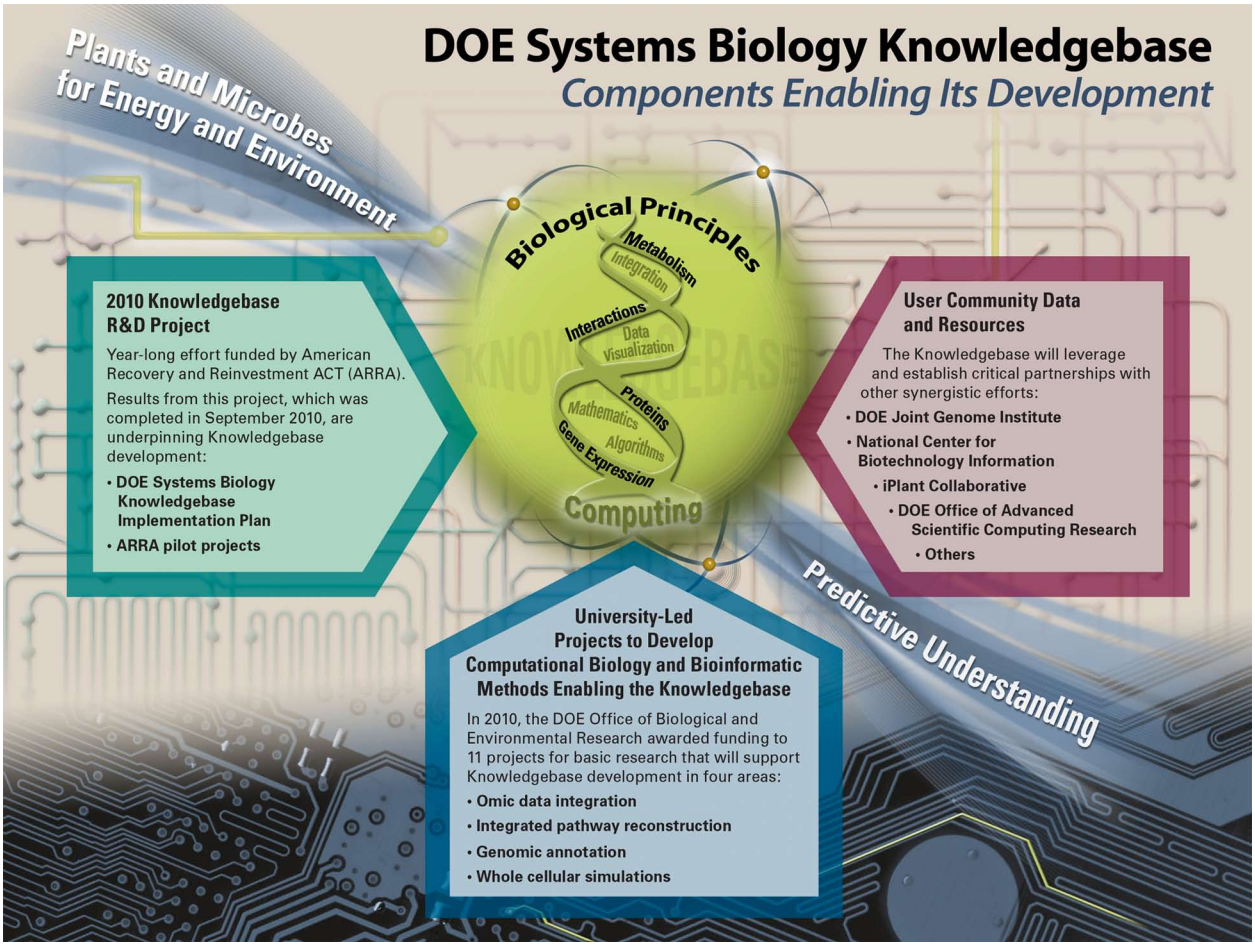



**Fig. 2. DOE Systems Biology Knowledgebase.**

# DOE Systems Biology Knowledgebase Project

Kbase will be a community-driven software framework enabling data-driven prediction of microbial, plant, and biological community function in an environmental context. Extensible and scalable, Kbase also will feature open architecture, source code, and development. This cyberinfrastructure will vastly improve algorithmic development and deployment efficiency, as well as access to and integration of data from heterogeneous sources. As a result, Kbase will provide the means for solving a wide range of biological problems that span understanding the reciprocal effects of community–environment interactions to creating advanced biofuels by leveraging laboratory microbes.

The Kbase framework is a collaborative partnership between DOE national laboratories and research universities. Leading the collaboration is principal investigator Adam Arkin of Lawrence Berkeley National Laboratory (LBNL), with co-principal investigators Rick Stevens of Argonne National Laboratory (ANL), Robert Cottingham of Oak Ridge National Laboratory (ORNL), and Sergei Maslov of Brookhaven National Laboratory (BNL).

This four-lab partnership brings together core expertise and existing biological knowledge resources in scientific and computer science domains and also models the kind of community envisioned for Kbase—with developers, computational resources, and users distributed throughout the country but working together. The Kbase science team (LBNL lead) will collaborate with the infrastructure team (software and hardware; ANL lead) to encapsulate and develop powerful algorithms and workflows for predictive modeling of biological function. With a focus on plants, BNL also will work with these teams to develop tools for data integration, analysis, and modeling. This collaboration will ensure maximal accessibility and scalability of Kbase capabilities for the larger scientific community. Together with a tightly allied outreach program (ORNL lead), Kbase will train, collaborate with, and receive feedback from the larger community for constant improvement and rapid prototyping.

## *Core Missions and Science Applications*

Kbase will be an enabling technology for organizing, accessing, and interpreting the vast flow of information on microbes, plants, and microbial communities generated by the international community. This framework will provide the basis for using genomics information to predict how microbes, plants, and their environments transform each other and to control and design their functions. Genome sequences serve as foundational information for Kbase on which all other information rests, and the assessment of sequence function—whether in complete or partial genomes or from metagenomes—is paramount. Homology-based methods continue to improve and are being scaled up to meet the computational demands of exponentially increasing sequence data. Lagging behind, however, are methods for rapidly applying new experimental evidence to support or change gene and regulatory sequence annotation and to accurately and quickly transfer this information across the tree of life.

Therefore, the **first core scientific mission** of Kbase will be to create a flexible, extensible, and high-quality framework for the experimental, evidence-based functional annotation of genome sequences. A central goal for high-quality genome assessments, and one means of validating genome annotations, is to use sequence to predict overall organismal function and community structure-function relationships in diverse environments (i.e., phenotypes). The **second core mission** for Kbase is to enable the scalable and mostly automated creation of high-quality metabolic and regulatory models that may be validated against experimental data and used to generate scientific predictions and testable hypotheses. Forming the Kbase backbone will be software to facilitate the query, comparison, and visualization of gene function, genome and metagenome organization, and metabolic and regulatory models in phylogenetic context. The **third core Kbase mission** is to make all algorithms, software, and data openly accessible to the community. Community access will be realized (1) programmatically through software libraries and an open-development architecture for computational biologists; (2) functionally through well-designed user interfaces and workflows available to the biological applications community; and (3) socially through training, documentation, and software features that support social networking around scientific and algorithmic questions. These three missions underlie the overall goal of predictive biology, with a primary emphasis on improving scientific estimates of genome function.

## Science Applications

Kbase will provide access to a distributed, scalable computing resource for compute and data-intensive analyses and will support a large user community with tools and services consistent with the following scientific goals.

- **Microbial Systems:** Reconstruct and predict metabolic and gene expression regulatory networks for 100 to 1,000 microbes to manipulate microbial function.
- **Plant Systems:** Integrate phenotypic and experimental data and metadata for 10 key plants related to DOE missions to predict biomass properties from genotype and assemble regulatory data to enable analysis, cross-comparisons, and modeling.
- **Microbial Communities:** Model metabolic processes within 10 to 100 microbial communities with DOE relevance and mine metagenomic data to identify unknown genes.

## Guiding Principles and Vision

The Kbase vision is that users and developers working at the frontier of systems biology will have easy access to high-quality datasets, database integration, and analytical algorithms to aid their research. This access will enable navigation across data types—from genome sequences to proteins and complexes and to biochemical reactions, regulatory elements, and phenotypes. To accomplish this goal, Kbase will provide a thoughtful and well-designed interface to support a range of functions and services that push and facilitate the development of next-generation tools and workflows (see Fig. 3).

The guiding principles of Kbase are to develop an open-source, open-development software framework. The core architecture will be built to be extensible as the effort expands (e.g., "design for the future but build for today"). The project will begin by implementing a computational environment that supports integrated access to the basic data items needed to support construction of models (see Fig 4, p. 5). These items include genomic, expression, and phenotypic data as well as an encoding of relevant chemistry and reactions. To ensure

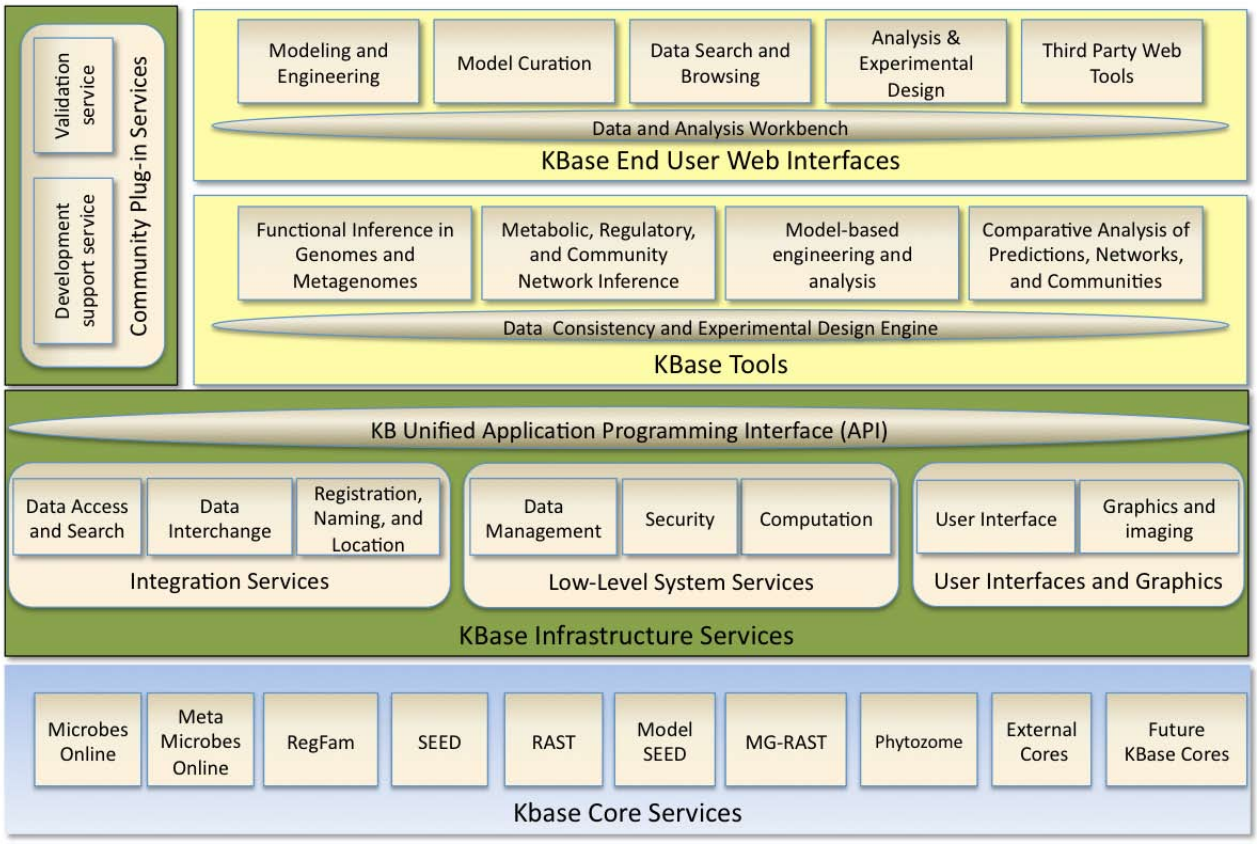



**Fig. 3. Kbase Architecture.**

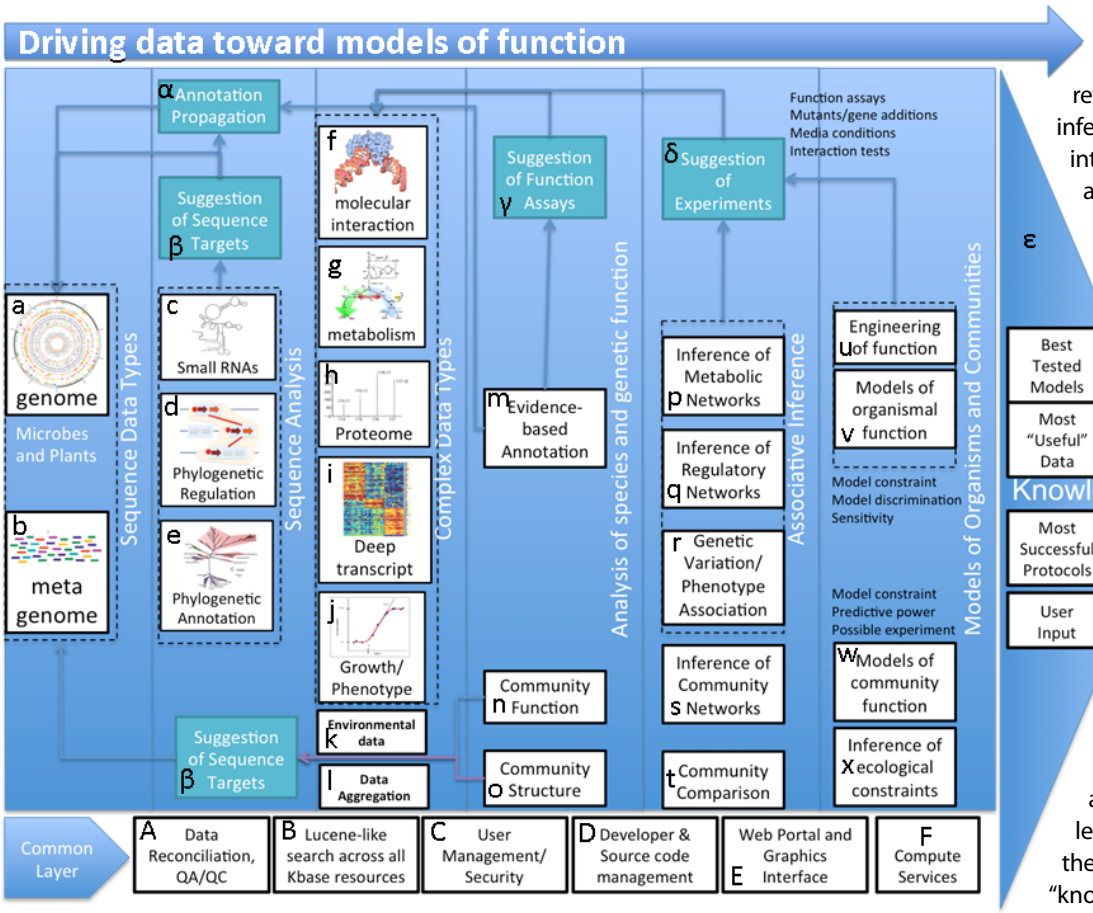



**Fig. 4. Toward Predictive Capability.** Kbase seeks to drive data toward dynamic, testable models of gene and network function and dynamic behavior in an environmental context. The progression is shown from left to right in the middle: sequence data is annotated first using comparative methods (a to e) and then increasingly with functional data (f to o). The annotations and data lead to inferences of relationships among sequences, functional data, and species, allowing refinement of genetic function inference and predictions of interactions and networks among both biological molecules and cellular species (p to t). These, in turn, lead to model-based predictions of function and phenotype (u to x). Kbase is distinguished from other biological data workbenches by its focus on elucidating missing information necessary for predictive modeling (e.g., boxes β), aiding experimental design to test model-based hypotheses (γ, δ), and organizing all models according to data confidence levels, thereby delivering the best-tested biological "knowledge" (ε).

community data integration, Kbase will establish common modes for extracting and depositing data.

## *Leveraging and Integrating Existing Biological Data and Databases*

The first step in Kbase development is to integrate various services, databases, and application programming interfaces (APIs) from a set of existing systems in such a way that data can be reconciled among many different systems. This integration will require constructing a few services for registering and maintaining shared IDs, as well as a naming service that maps names from existing systems into a common name space as a way to register APIs and services. Once this initial integration is complete, Kbase will provide data, information, and analysis tools within a common framework. Early services that will be integrated together include MicrobesOnline, metaMicrobesOnline, RegPrecise, RegPredict, The SEED, RAST, ModelSEED, MG-RAST, Phytozome, and public datasets from the National Center for Biotechnology Information and the DOE Joint Genome Institute (JGI). An underlying principle of this effort is to avoid "reinventing the wheels" created by others and instead use existing tools to build an integrated resource that can more rapidly move science forward.

## *Building One Team to Build One System*

The Kbase strategy is to leverage existing integrated bioinformatics systems as core services and also enable research teams to focus on building new capabilities rather than duplicating existing ones. By providing an integrative platform to access complex data and a diverse set of advanced algorithms for comparative functional modeling, Kbase seeks to change the way biologists work with their data on a daily basis. Ultimately, Kbase will allow researchers to use modeling in a far more sophisticated experimental design mode than is currently possible.

# R&D Effort Results in Knowledgebase Implementation Plan and Pilot Projects

In 2009, funding provided by the American Recovery and Reinvestment Act (ARRA) was used to launch the year-long DOE Systems Biology Knowledgebase R&D project. This project consisted of research and development efforts to support the conceptual design and implementation planning necessary to develop Kbase. The implementation plan is a roadmap for creating Kbase and is available at two websites:

- genomicscience.energy.gov/compbio/kbase_plan/
- science.energy.gov/~/media/ber/pdf/kbase_plan.pdf

It articulates the scope and plans necessary to begin the Kbase effort and outlines a strategy for Kbase support of key research objectives in the microbial, plant, and metacommunity sciences. These objectives include metabolic reconstruction and modeling; inference of gene regulatory networks; linkage of phenotypic and experimental data and metadata; and assembly, integration, annotation, and mining of "omic" and other types of data. The implementation plan notes the need to leverage existing community resources and projects and describes the tasks, timelines, and plan for establishing Kbase's underlying infrastructure.


## *Early ARRA Successes*

In addition to informing and contributing to the implementation plan, several pilot projects developed software prototypes in conjunction with ARRA efforts in cloud computing funded by the DOE Office of Advanced Scientific Computing Research. Descriptions of these completed pilot projects follow.

### Developing Design Requirements and Prototypes of Workflows in the DOE Systems Biology Knowledgebase to Support Metabolic Pathway Engineering

- **Principal Investigator:** Adam Arkin (Lawrence Berkeley National Laboratory)

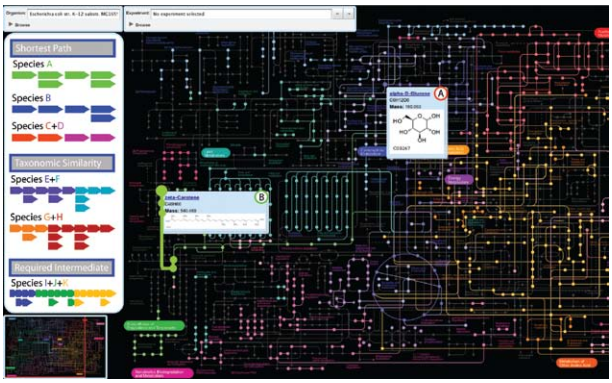
This project designed and implemented workflows for metabolic reconstruction within MicrobesOnline, a web portal for comparative and functional genomic analyses. Investigators began developing interfaces for navigating metabolic networks and experimental functional omic data using the Google-Like Application for Metabolic Maps or GLAMM (see Fig. 5, p. 7, and Bates, J. T., D. Chivian, and A. P. Arkin. 2011. "GLAMM: Genome-Linked Application for Metabolic Maps," *Nucleic Acids Research* **39**. DOI: 10.1093/nar/gkr433). GLAMM suggests pathways that may offer routes for retrosynthesis (e.g., how to build a pathway to convert feedstock X into chemical Y in organism Z).

### Exploring Architecture Options for Workflows in a Federated, Cloud-Based Systems Biology Knowledgebase

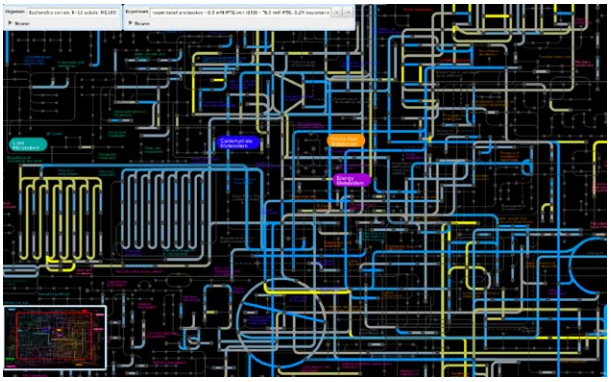
- **Principal Investigator:** Ian Gorton (Pacific Northwest National Laboratory)

This project involved investigating available mechanisms for storing and accessing biological data in a cloud computing environment and evaluating access to large archives of omic data using a cloud architecture to provide "Data As A Service." A use case scenario to identify and curate published genome annotations was established, and investigators implemented this workflow using a federated, cloud architecture, as proposed for Kbase (see Gorton, I., Y. Liu, and J. Yin. 2010. "Exploring Architecture Options for a Federated, Cloud-Based Systems Biology Knowledgebase," *2010 IEEE Second International Conference on Cloud Computing Technology and Science*. DOI: 10.1109/CloudCom. 2010.79).

**Fig. 5. Screenshots of GLAMM Retrosynthesis Interface (top) and GLAMM Functional Data Overlay (bottom).**




Routes from a starting metabolite to a "destination" metabolite may include retrosynthesis pathways with genes from other organisms. Genes, reactions, and metabolites are linked to MicrobesOnline.



Expression data from a single experiment is rendered using the metabolic reconstruction for *Escherichia coli* K12 MG1655. Genes are mapped to the pathways they are predicted to catalyze. Increase in expression relative to control is shown in yellow; decrease is shown in blue.

## Examining Technologies for Database Management Systems that Support Computational Biology and Bioinformatics Applications

- **Principal Investigator:** Victor Markowitz (Lawrence Berkeley National Laboratory and the DOE Joint Genome Institute)

This project focused on evaluating new database management system technologies that allow efficient analysis of very large datasets. Prototypes of a large database based on the DOE JGI's Integrated Microbial Genomes (IMG) data management system were implemented using several of these technologies. Performance tests of IMG "all versus all" data were conducted in Hbase on the DOE National Energy Research Scientific Computing Center's Magellan Hadoop cluster and on a smaller departmental Hadoop cluster. Results show that distributed tabular storage has significant long-term potential for Kbase but that it is not yet ready for large-scale production use. Investigators note that Hadoop and Hbase currently are undergoing rapid development, and they anticipate that stability issues will be addressed within the next 2 years. DOE JGI is now implementing the Magellan cloud infrastructure for microbial genome assembly and annotation based on the results of this pilot test.

# BER Funds University-Led Research Developing Computational Methods to Enable Kbase

In addition to the Kbase R&D project, the Genomic Science program in late 2010 began funding research leading to the development of new methods and analytics for creating Kbase. Eleven university-led projects were awarded in response to Funding Opportunity Announcement DE-FOA-0000143: *Computational Biology and Bioinformatic Methods to Enable a Systems Biology Knowledgebase*. Under this funding call, the Genomic Science program solicited applications for basic research in computational systems biology that both support Kbase development and address DOE missions in energy and the environment. New methods resulting from this research will be leveraged into the larger Kbase effort. The accepted projects exhibit strong collaboration among experimental data generators, bioinformaticists, computational biologists, and computer scientists in four areas:

- **Omic Data Integration.** New computational methods are desired for integrating multiple types of data such as genomic, metagenomic, proteomic, metabolomic, transcriptomic, expression, and phenotypic. These methods involve developing data standards, ontologies, and controlled vocabularies as well as assessing data quality. Also needed are methods that significantly improve data visualization and analysis, including new methods for complex web interfaces and third-party tool development. Methods for analyzing across different data types are priorities.
- **Genomic Annotation.** Also sought are new methods for computational gene annotation that include integrating data and information into gene functional assignments. New annotation methods are needed for capturing information such as cDNA, clustering and neighborhood gene analysis, expression and phenotypic data, protein folds and structures, and phylogenetic profiling data. Priorities include methods for estimating and embedding uncertainty and confidence levels in annotation assignments.
- **Integrated Pathway Reconstructions.** Significant improvements are needed in methodologies to couple metabolic and regulatory pathways and integrate associated data and information. These improvements include new methods in correlational and iterative analysis that would dynamically link data to model development. New methods in dynamical pathway reconstruction for on-the-fly pathway analysis also are being encouraged. Improvements supporting the integration of expression data (e.g., transcription and protein association and localization) with pathway simulations are priorities.
- **Whole Cellular Simulations.** New methods are needed for modeling complex cellular processes. These methods include integrating multiple data types such as two- and three-dimensional imaging and spectroscopic data with cellular models or simulations.

## *Summary of University-Led Projects*

### Enabling a Systems Biology Knowledgebase with Gaggle and Firegoose

- **Principal Investigator:** Nitin Baliga (Institute for Systems Biology)

This project will extend the existing Gaggle (open-source Java software system) and Firegoose (an extension to the Mozilla Firefox web browser) systems to develop an open-source technology that runs over the web and links desktop applications with many databases and software applications. Four specific aims are to (1) provide one-click mapping of genes, proteins, and complexes across databases and species; (2) enable multiple simultaneous workflows; (3) expand sophisticated data analysis for online resources; and (4) enhance open-source development of the Gaggle-Firegoose infrastructure.

### Tools and Models for Integrating Multiple Cellular Networks

- **Principal Investigator:** Mark Gerstein (Yale University)

This application will develop computational tools to link metabolic pathways with regulatory pathways and physical (protein-protein) interaction data. The three specific aims are to (1) develop computational tools for analyzing integrated networks; (2) conduct correlative and topological analysis using these tools, in combination with other genomic information (see Fig. 6); and (3) carry out dynamical and evolutionary modeling of the integrated network.

### Development of a Knowledgebase to Integrate, Analyze, Distribute, and Visualize Microbial Community Systems Biology Data

- **Principal Investigator:** Jill Banfield (University of California, Berkeley)

This project will develop a web-based knowledgebase that integrates metagenomic data with metaproteomic and metabolomic data from microbial communities, particularly from acid mine drainage. Three specific resources and capabilities will be developed: (1) a centralized database to integrate various omic datasets, (2) tools for mapping and representing proteomic and genomic datasets comprising orthologous genes in the presence of genomic variation, and (3) a metabolite atlas of the acid mine drainage microbial community.

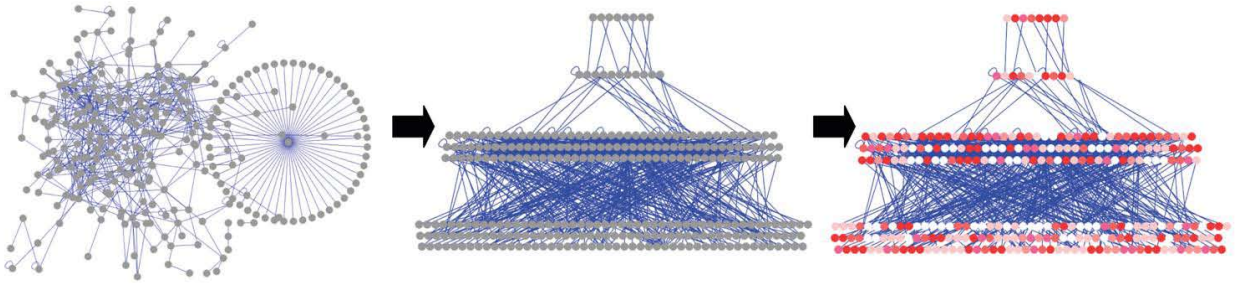


**Fig. 6. Schematic Representing Effects of Rewiring a Transcriptional Regulatory Network in a Hierarchical Context.** Related to the second aim, researchers performed a topological network analysis in correlation with genome-wide phenotypic data. (1) The transcriptional regulatory network is first arranged into a hierarchy, with the regulatory edges pointing only downward. To mimic the commonplace government and corporate hierarchy, no regulator controls any gene above it in the hierarchy. (2) Phenotypic effects of tampering with various nodes and edges (e.g., their removal or additions) on cell growth and survival are overlaid onto the hierarchy. Different genes have different effects on cell growth. (3) Node color (scaling from white to red) indicates the effect of tinkering on cell growth and survival. Tampering with white nodes has a minimal effect (cell grows normally), whereas red nodes are the genes that, upon deletion, affect cell growth adversely (cell grows at a slower rate or dies). Researchers find that tampering with upper-level nodes affects cell growth more adversely. [From Bhardwaj, N., et al. 2010. "Rewiring of Transcriptional Regulatory Networks: Hierarchy, Rather Than Connectivity, Better Reflects the Importance of Regulators," *Science Signaling* **3**(146), ra79. Reprinted with permission from AAAS.]

### Curation and Computational Design of Bioenergy-Related Metabolic Pathways

- **Principal Investigator:** Peter Karp (SRI International)

This project will develop an enhancement in the MetaCyc Pathway Tools (a set of metabolic pathway and enzyme tools) aimed specifically at bioenergy-related processes. Two specific goals are to (1) enhance MetaCyc data and generate a bioenergy-related pathway and genome database and (2) develop computational tools for engineering metabolic pathways that satisfy specified design goals.

### Computational Modeling of Fluctuations in Energy and Metabolic Pathways of Methanogenic Archaea

(Jointly funded with the DOE Office of Advanced Scientific Computing Research)

- **Principal Investigator:** Zaida Luthey-Schulten (University of Illinois, Urbana-Champaign)

This project will develop methodology and corresponding computational tools to simulate a population of microbes, particularly the methanogenic archaea *Methanosarcina* species. Specific aims are to (1) construct an integrated stochastic and systems model of *Methanosarcina,* (2) investigate how an *in silico* population of the microbe's cells respond to environmental fluctuations, and (3) validate the computational methodology and demonstrate its applicability to other biological systems.

### A Systems Biology Knowledgebase: Context for Content

- **Principal Investigator:** Bernhard Palsson (University of California, San Diego)

This project will develop a portal and computational tools for integrating multiple omic data to reconstruct transcriptional regulatory networks of microbes relevant to DOE (e.g., *Escherichia coli*, *Geobacter*, and *Thermotoga*). The data include protein binding (ChIP-chip), gene expression (microarrays and RNA-Seq), transcriptional start sites (sequencing), peptide (LC-FTICR-MS), and gene annotations. Three specific aims for the project are to (1) develop computational tools to integrate omic data for genome annotation and transcription, (2) develop a genome-scale knowledgebase to provide operational constraints on cellular function, and (3) formulate *in silico* models to enable genome-scale queries.

### Integrated Approach to Reconstruction of Regulatory Networks

- **Principal Investigator:** Dmitry Rodionov (Burnham Institute)

This project will extend research to identify regulons for regulatory network reconstruction and develop a method for comparing regulatory networks across microbial species. Specific aims are to (1) develop an integrative platform for genome-scale regulon reconstruction; (2) infer regulatory annotations for several groups of bacteria related to DOE missions; (3) develop a knowledgebase for microbial transcriptional regulation data and analysis; (4) develop a platform that integrates experimental and computational data on transcriptional regulation in microbes; and (5) allow any user to upload data (public or private), perform analyses with the data, and compare them to the analysis work conducted by the researcher who generated the data for a particular experiment.

### An Open-Source Platform for Multiscale Spatially Distributed Simulations of Microbial Ecosystems

- **Principal Investigator:** Daniel Segrè (Boston University)

This project will develop an open-source platform for simulating microbial ecosystems (see Fig. 7). Specific aims are to (1) modify a current dynamic flux balance analysis (dFBA) program to include spatially structured interacting metabolite dynamics of the microbial system and (2) study interactions in terms of dynamically changing colony morphology by modeling the simultaneous growth of mutualistic pairs of microbes.

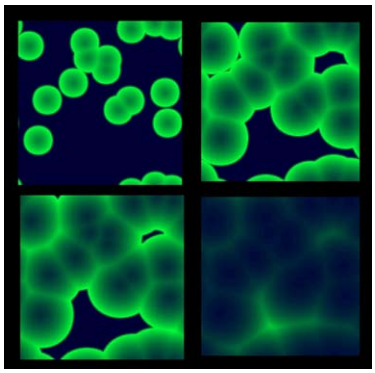


**Fig. 7. Snapshots Showing Simulation of Microbial Growth.** This simulation was generated using a spatially distributed dynamic flux balance analysis approach. The software platform, called COMETS (Computation Of Microbial Ecosystems in Time and Space), is being developed by the Segrè laboratory at Boston University. Here, in a first version of the software built by William Riehl, a central carbon metabolism model of *Escherichia coli* is used to simulate colony growth.

### Phylogenomic Tools and Web Resources for the Systems Biology Knowledgebase

- **Principal Investigator:** Kimmen Sjölander (University of California, Berkeley)

This project will develop new methods to functionally annotate microbial species based on phylogenomic relationships and using the hidden Markov model (HMM) methodology based on the structural information of families of homologous genomes. Three primary objectives for the project are to (1) extend the PhyloFact annotation method to include new microbial data and related database information such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), PFAM, Gene Ontology (GO), experimental evidence codes, and structural information; (2) develop a new HMM algorithm to create novel gene trees; and (3) apply the PhyloFact annotation pipeline to collaborative marine microbial systems.

### Development of an Extensible Computational Framework for Centralized Storage and Distributed Curation and Analysis of Genomic Data and Genome-Scale Metabolic Models

- **Principal Investigator:** Rick Stevens (University of Chicago)

This work will develop a computational framework that combines a centralized extensible database for integrating omic and sequence data with a distributed pipeline for using these data to annotate genomes and to reconstruct and analyze new genome-scale metabolic models. Specific project objectives are (1) an improved infrastructure to enhance the framework's extensibility, accessibility, and scalability; (2) an extended database to accommodate new predicted and experimental biological data types such as microbial transcriptional regulatory networks, genome-scale metabolic models, experimental evidence (e.g., microarray data, ChIP-chip data, and equilibrium constants), eukaryote genomes, and growth phenotype data (e.g., biology array data, culture conditions, growth rates, and gene essentiality); and (3) a new application programming interface to provide remote access to the database and tools, including RAST annotation of raw genome sequences, automated reconstruction of draft genome-scale metabolic models, flux balance analysis of such models, and querying of all data.

### Gene Ontology (GO) Terms and Automated Annotation for Energy-Related Microbial Genomes

- **Principal Investigators:** Biswarup Mukhopadhyay, Brett Tyler, and João Carlos Setubal (Virginia Polytechnic Institute and State University)

This effort will develop a set of GO terms for describing energy-related microbial processes. Two specific aims are to (1) develop MENGO terms (ontologies for microbial energy processes) and host a series of tutorials and workshops at key meetings to inform and train microbiologists on these terms and (2) develop a database and web interface for storing and displaying these terms and microbial annotations.

## DOE BER Contacts

**Dr. Susan Gregurick,** 301.903.7672, susan.gregurick@science.doe.gov

**Dr. Pablo Rabinowicz**, 301.903.0379, pablo.rabinowicz@science.doe.gov

## Reports

***DOE Systems Biology Knowledgebase Implementation Plan*, September 2010**

- genomicscience.energy.gov/compbio/kbase_plan/
- science.energy.gov/~/media/ber/pdf/kbase_plan.pdf

**Individual workshop reports contributing to Kbase Implementation Plan**

- genomicscience.energy.gov/compbio/ARRA/kbase_ARRA_workshops.shtml

**Summary reports of Kbase R&D pilot projects**

- genomicscience.energy.gov/compbio/ARRA/kbase_ARRA_pilots.shtml

**DOE Systems Biology Knowledgebase for a New Era in Biology, March 2009**

- genomicscience.energy.gov/compbio/workshop08/

## Websites

**DOE Office of Science**

- science.energy.gov

**DOE Office of Biological and Environmental Research**

- science.energy.gov/ber/

**Genomic Science program**

- genomicscience.energy.gov

**August 2011**