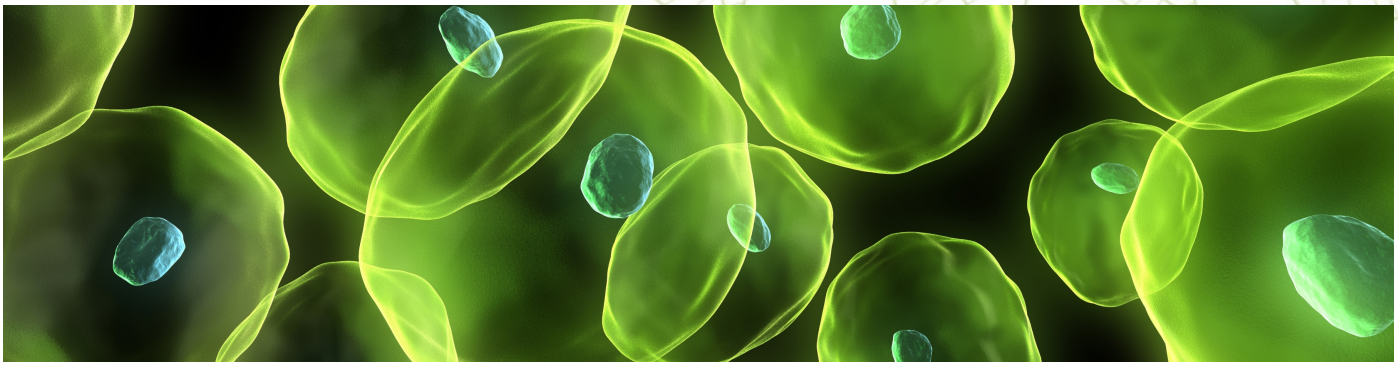
KBASE
predictive biology

DOE Systems Biology Knowledgebase



FACT SHEET

The new DOE Systems Biology Knowledgebase (KBase), is a collaborative, community-driven effort designed to accelerate our understanding of microbes, microbial communities, and plants.

Vision and Purpose

KBase supports DOE's aim to enable predictive understanding of microbial and plant biological systems relevant to DOE missions, including producing biofuels, sequestering carbon in terrestrial ecosystems, and cleaning up contaminated environments. KBase is a new community resource for predictive biology. It integrates a wide spectrum of data types across the microbial, microbial community, and plant domains, and ties this data into a varied set of powerful computational tools that can analyze and simulate data to predict biological behavior, generate and test hypotheses, design new biological functions, and propose new experiments. The KBase project has been conceived of and executed as a single coherent venture to create a resource for the research community that is integrated in all aspects of design, functionality and support. While we are leveraging the considerable work of many scientists and developers, the overarching objective is to provide a solid platform that supports predictive biology in a framework that does not require users to learn separate and varied systems in order to compose and execute the complex tasks that constitute many of the future analysis workflows to advance systems biology research.

History

The Office of Biological and Environmental Research (BER) in DOE's Office of Science has been on the path toward developing the Systems Biology Knowledgebase since at least 2005, when the concept appeared in the Genomics:GTL Roadmap report as the Systems Microbiology Knowledgebase. In May 2008 BER hosted the first Systems Biology Knowledgebase workshop that provided the first substantial vision and high-level requirements and roadmap for KBase. Then in 2009, DOE engaged the computing and experimental biological communities in an intensive series of workshops to articulate significant scientific objectives and develop a detailed work breakdown to guide actual development of KBase. In September 2010 the Implementation plan was published. This was the basis for a DOE Laboratory call that was ultimately awarded to the current KBase team (LBNL, ANL, BNL, and ORNL) in late summer 2011. We chose topics that can share a common infrastructure and build across the three domains while respecting the different breadth and depth of data and problems that exist in the microbial, community and plant domains.

DOE BER has also supported University Pilot projects that were funded specifically to develop functionality or tools useful for the future KBase system.

U.S. DEPARTMENT OF ENERGY | Office of Science

CONTACT US
URLs: kbase.science.energy.gov

EMAIL
outreach@kbase.us

Visit us at **outreach.kbase.us** to learn how you can provide input, become involved, and find out about upcoming events. Tell us what KBase can do for you by completing one of our online surveys.

Description of Services

Although our primary goals cut across the science domains, KBase addresses several science aims relevant to each area:

MICROBES:
KBase will maximize understanding of microbial system function, promote sharing of data and findings, and vastly improve the planning of effective experiments. Early efforts will target enabling the reconciliation of metabolic models with experimental data. The ultimate aim is manipulating microbial function for applications in energy production and remediation. In order to accomplish this, we will enable users to expand on a strong foundation of quality genome annotations, to reconstruct metabolism and regulation, to integrate and standardize omics data, and to construct models of genomes.

PLANTS:
A high priority of the plants research community is to link genetic variation, phenotypes, molecular profiles, and molecular networks, enabling model-driven phenotype predictions. A second goal will be to map plant variability onto metabolic models to create model-driven predictions of phenotypic traits. Initial work will focus on creating a workflow for rapidly converting sequencing reads into genotypes. We will also build tools for data exploration, and the linking of gene targets from phenotype studies such as genome-wide association studies, with co-expression, protein-protein interaction, and regulatory network models. Such data exploration will allow users to narrow candidate gene lists by refining targets, or be able to visualize a subnetwork of regulatory and physical interactions among genes responsible for a phenotype in question. Users can also highlight networks or pathways impacted by genetic variation.

COMMUNITIES:
Through comparative analysis of metagenomes acquired over different spatial, temporal, or experimental scales, it is now possible to define how communities respond to and change their environment. KBase will provide the computational infrastructure to research community behavior and to build predictive models of community roles in the carbon cycle, other biogeochemical cycles, bioremediation, energy production, and the discovery of useful enzymes. We are building the next-generation metagenomic platform that provides scalable, flexible analyses, data vectors for models, tools for model creation, data quality control, application programing interfaces, and GSC-compliant data and standards for data collection. Initial efforts will target the development of bioprospecting and experimental design tools.

Where we are now

(current status - Dec 2012)

KBase recently completed a successful November build, and we are preparing for our Version 1.0 release early next year. Version 1.0 of KBase will offer a unified environment for users to view, query, and download data for microbes, plants, and microbial communities. Users will also be able to apply a series of powerful demonstration workflows, including genome annotation, metabolic modeling, and phenotype analysis.

Version 2.0 of KBase, available in late 2013, will include a narrative interface that provides an online environment for users to collaboratively piece together a wide variety of KBase tools and data into a scientific narrative. Additionally, the software will provide extensive support for integrated analysis of thousands of closely related genomes, including comparative analysis of metabolism, regulation, and model-predicted phenotypes.

*The roadmaps and other documents are available at:

http://genomicscience.energy.gov/compbio/index.shtml

This project is supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research's Genomic Science program.

U.S. DEPARTMENT OF ENERGY | Office of Science

CONTACT US
URLs: kbase.science.energy.gov

EMAIL
outreach@kbase.us

Visit us at **outreach.kbase.us** to learn how you can provide input, become involved, and find out about upcoming events. Tell us what KBase can do for you by completing one of our online surveys.