

GET INVOLVED. HELP BUILD KBASE.

We invite you to partner with us.

We are building a strong group of collaborators to meet the needs of the community and to leverage existing efforts rather than duplicate them. Strong, open lines of communication with collaborators, future developers and users are very important to us. You can reach us directly via email and our other online communication tools, or join us at workshops and conferences.

The success of the KBase project depends not only on producing a large-scale open computational capability for systems biology research data management and analysis, but also on positioning these tools to be used by the community. Consequently, KBase is actively looking for collaborators who provide data, those who are tool builders, and those who are users of both data and tools.

Users will interact with KBase via the submission of new experimental data (e.g., microarrays, RNAseq, Biolog arrays, mutant phenotypes) or the integration of new analysis techniques and algorithms. All data and algorithms integrated into KBase will be fully accredited to the source.

TIMELINE

Feb 2012 • Development Release

prototype deployments, performance testing, architecture refinement

May 2012 • Alpha Release

draft tutorials, database loads, unified prototypes, workflow drafts, cloud and cluster services

Aug 2012 • Early Adopter Beta Testing

workflow functions, prototype plugin interfaces, prototype Galaxy support, performance testing

Nov 2012 • Public Beta Testing

draft website, draft documentation, functional API, draft UI, external beta testers

Feb 2013 • KBase Production Release

public website, unified API, initial production UI, database loads for science domains, demonstration workflows

The KBase collaboration is led by Lawrence Berkeley, Argonne, Brookhaven, and Oak Ridge national laboratories.

Also involved in the multi-institutional program are:

- Cold Spring Harbor Laboratory
- University of California, Davis
- Hope College
- University of Illinois at Urbana-Champaign
- Yale University

The Joint Genome Institute, the InterBRC Knowledgebase, and several university projects are also important contributors.

CONTACT US

Learn more about how you can be involved.

URLs: kbase.science.energy.gov
outreach.kbase.us

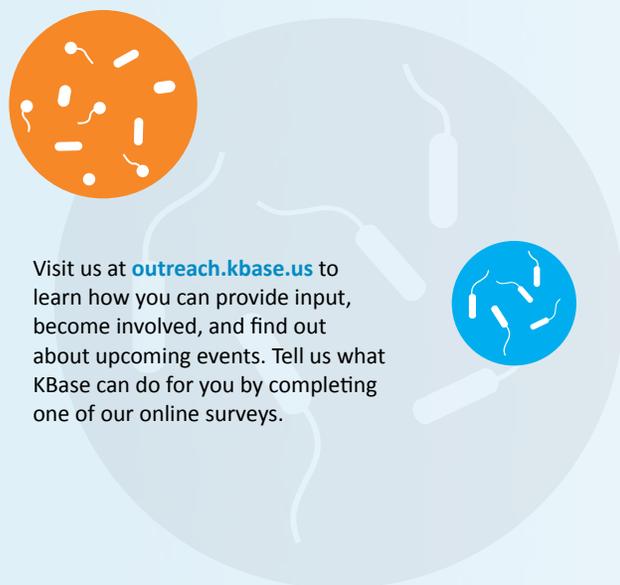
Email: outreach@kbase.us



DOE Systems Biology Knowledgebase

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

It will be an extensible and scalable open source software framework and application system. KBase will offer free and open access to data, models, and simulations, enabling scientists and researchers to build new knowledge, test hypotheses, design experiments, and share their findings.



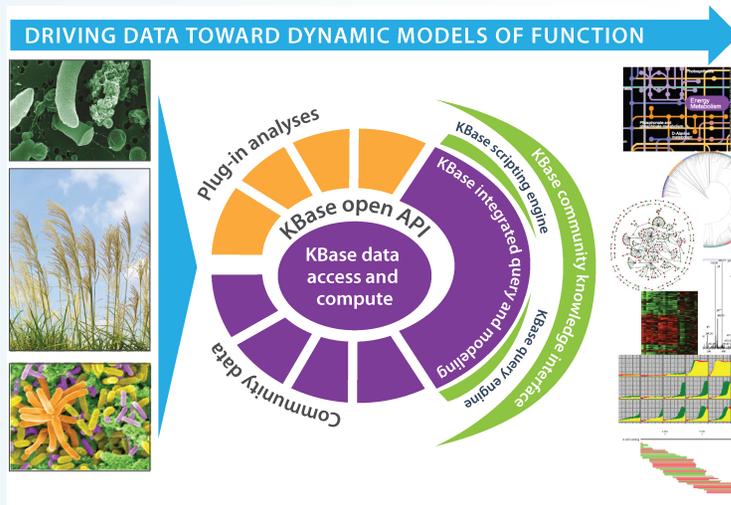
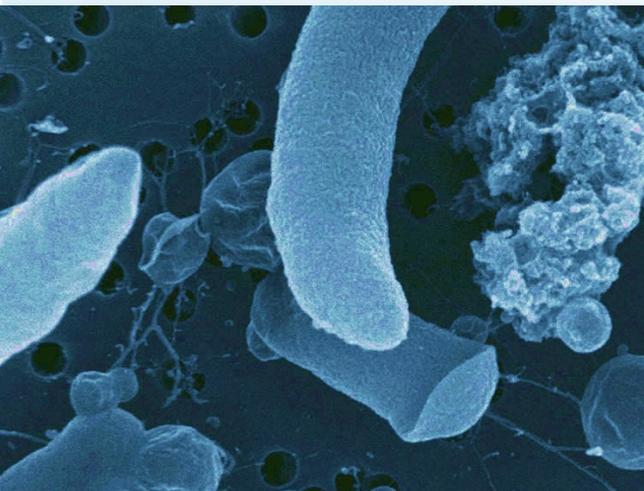
Visit us at outreach.kbase.us to learn how you can provide input, become involved, and find out about upcoming events. Tell us what KBase can do for you by completing one of our online surveys.



A RICH SET OF MODELING TOOLS FOR HYPOTHESIS TESTING AND EXPERIMENTAL DESIGN

KBase integrates commonly used tools and their associated data, and builds new capabilities on top of the combined data. New functionality allows users to visualize data, create powerful models, and design experiments based on KBase-generated suggestions. The project is about much more than data unification: KBase is focused on interpreting missing information necessary for predictive modeling; on aiding experimental design to test model-based hypotheses; and on delivering quality-controlled data. The project leverages cloud computing and high performance computing resources across the DOE laboratory complex to handle the anticipated rapid growth in data and computational needs.

KBase is composed of core biological analysis and modeling functions, including an application programming interface that can be used to connect different software programs within the community. These capabilities are constructed from the popular analysis systems at each of the KBase sites, such as the MG-RAST and SEED systems. Their integration into KBase will combine individual functions to create the next generation of biological models and analysis tools. The KBase application programming interface also enables our diverse community of users to design new functions. KBase is supported by a computing infrastructure based on the OpenStack cloud system software distributed across the core sites.



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 sub-

network 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 programming interfaces, and GSC-compliant data and standards for data collection. Initial efforts will target the development of bioprospecting and experimental design tools.

WHAT CAN KBASE DO FOR YOU?

KBase will enable users to do a number of tasks they could not do before, such as:

- Easily access and combine heterogeneous data types.
- Gain standardized access to bioinformatic and modeling analyses.
- Use evidence-supported annotations of genome structure and genetic function.
- Discover new associations and network structures in community and molecular networks.
- Effectively map genotype to complex organismal traits.
- Infer, refine, and design experiments for testing hypotheses on models of metabolism, regulation, and community function.
- Share data, hypotheses, and newly-generated knowledge across a community of scientists.