KBase: An Integrated Systems Biology Knowledgebase for Predictive Biological and Environmental Research

Adam P. Arkin[1] (aparkin@lbl.gov), Rick Stevens[2], Bob Cottingham[3], Sergei Maslov[4,5], Fernando Perez[1], and the KBase Team at the following institutions

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4] University of Illinois at Urbana-Champaign; [5] Brookhaven National Laboratory, Upton, NY; [6] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [7] Hope College, Holland, MI; [8] University of Tennessee, Knoxville, TN.

kbase.us

Project Goals: KBase is a computational platform designed to meet the key challenges of systems biology: predicting and ultimately designing biological function. The project's long-term goal is to help researchers understand how plants and microbes interact with and affect each other and environmental processes, and how these interactions could be harnessed for sustainable bioenergy and environmental solutions. These objectives will be advanced by KBase capabilities for data analysis; model-inspired experimental design; and dissemination of data, results, conclusions, and tools. Within this environment, KBase seeks to encourage biologists to collaboratively integrate, analyze, and interpret heterogeneous data to develop predictive, multiscale models of biological function. This integration of new information and tools from both external reference sources and users will enable researchers to greatly amplify the results of their own work by more effectively leveraging that of others.

The DOE Systems Biology Knowledgebase (KBase) enables secure sharing of data, tools, methods, and conclusions in a unified, extensible system that allows researchers to collaboratively generate and test hypotheses about biological functions; perform large-scale analyses on scalable computing infrastructure; and combine multiple lines of evidence to accurately model plant and microbial physiology and community dynamics.

KBase enables users to upload their own data or access public data in KBase to execute and share customized analyses that target their specific systems biology hypotheses. These computational experiments or analyses are captured in dynamic, interactive documents called Narratives that promote collaboration and reproducibility of scientific results. In addition to data and analysis steps, Narratives can include images, notes, and links, and can be kept private, shared with colleagues and collaborators, or made public for the benefit of the wider research community.

Major classes of KBase analysis tools include microbial genome assembly and annotation, phylogenetics and comparative genomics, expression analysis, and metabolic modeling.

Capabilities for metabolic model reconstruction and flux balance analysis simulation can provide insight into the metabolic pathways and interactions between plants and microbes, enabling researchers to identify, for example, the biochemical reactions active in biomass production. KBase's recently developed RNA-Seq pipeline and downstream analysis tools enable users to quantify expression from RNA-Seq reads, thereby identifying differential expression between different tissues, developmental stages, environmental conditions, and genetic backgrounds.

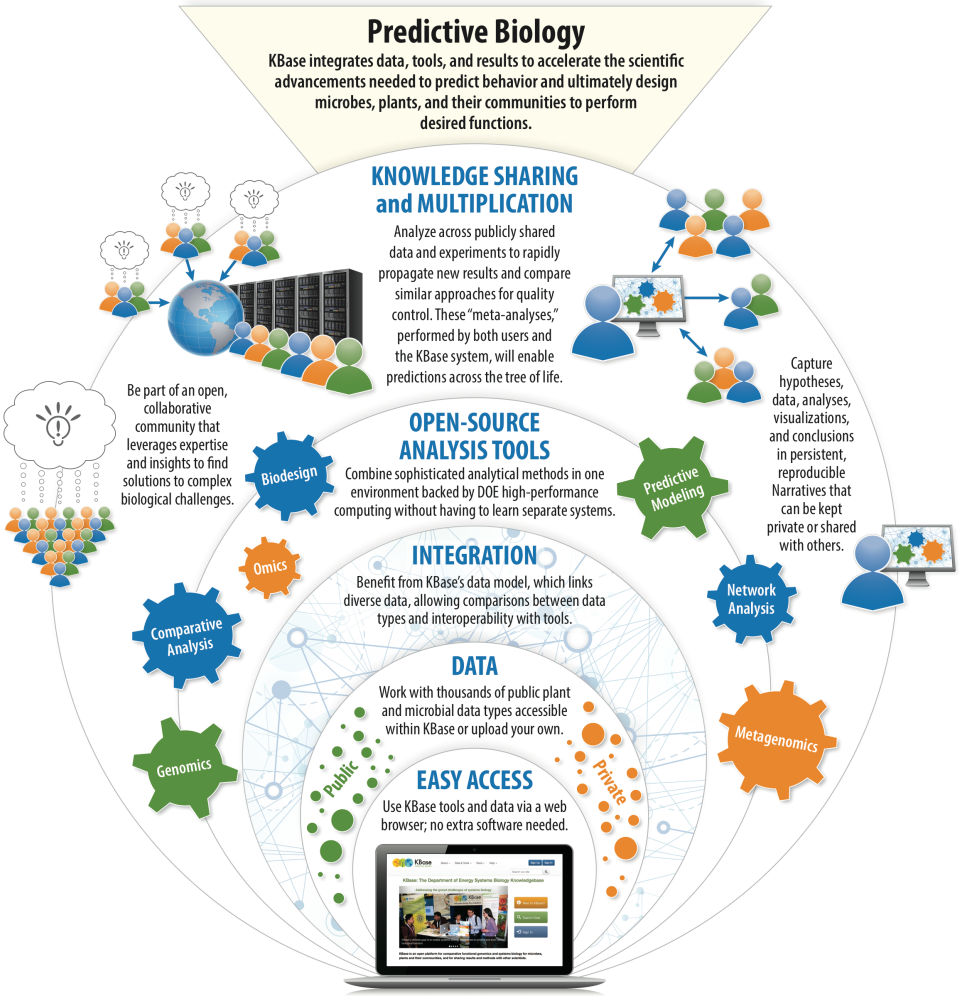

Fig. 1. Schematic overview of the KBase platform.

In addition to these resources, KBase has launched the Software Development Kit (SDK) that enables users to integrate other open source command-line tools and applications into the KBase platform. This new capability will accelerate the incorporation of new science functionality by allowing developers to more easily add—in a standardized and well-tested way—a wide variety of tools that then can be discovered and leveraged by other KBase users.

KBase is funded by the Genomic Science program within the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.