

KBase: The DOE Systems Biology Knowledgebase



An Overview

What is KBase?

The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of public data and analysis tools into an easy-to-use graphical user interface and leverages DOE computational infrastructure to perform sophisticated systems biology analyses. KBase is a freely available system that enables scientists to upload their own data, analyze it alongside collaborator and public data, build validated systems biology models, and share workflows and conclusions (Fig. 1).

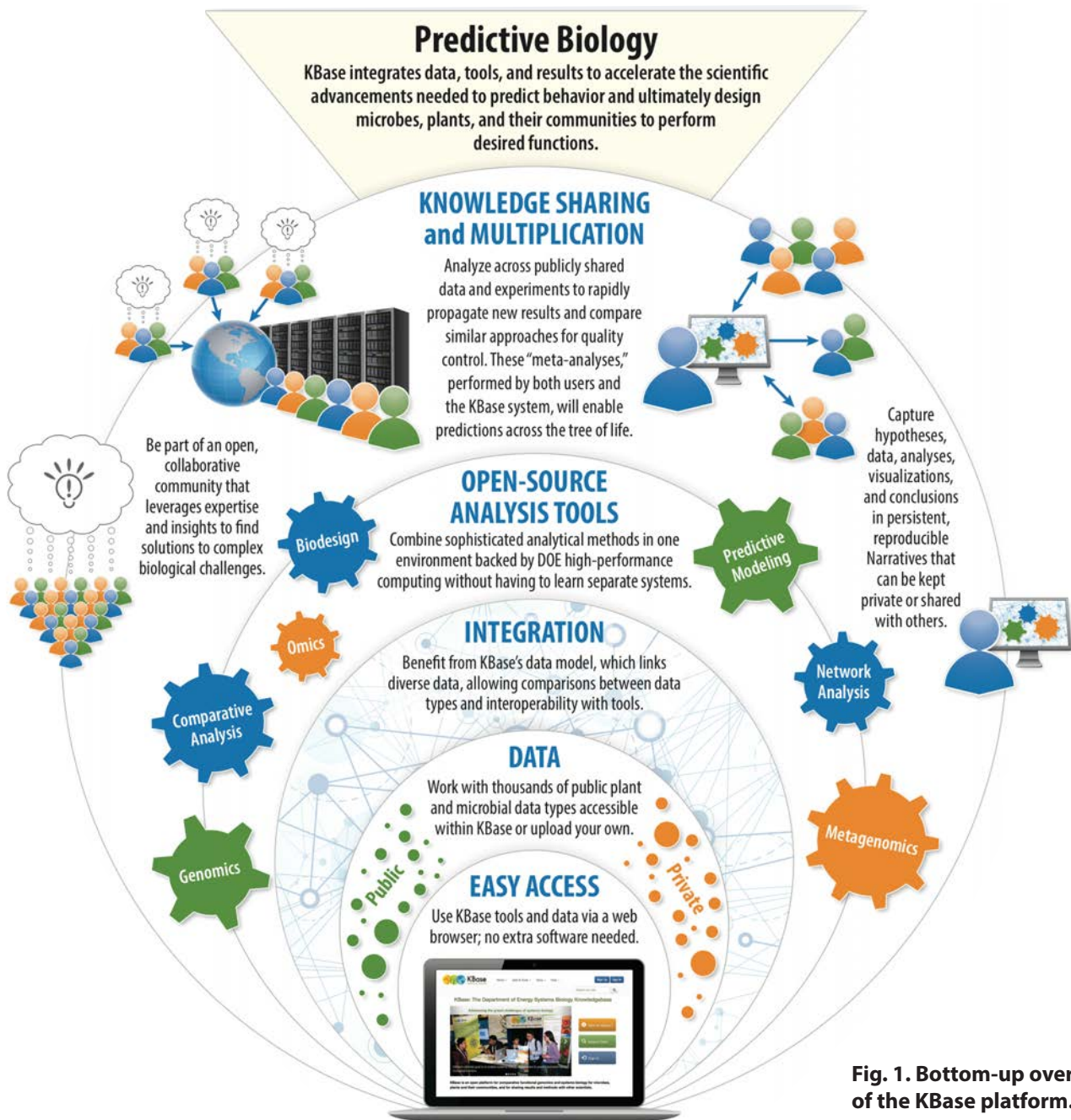


Fig. 1. Bottom-up overview of the KBase platform.

Apps (Fig. 3) are analysis tools that you can use in KBase. Apps interoperate seamlessly to enable a range of scientific workflows. Third-party developers can add their own open-source, open-license apps to the system using the **Software Development Kit** (kbase.us/developer).

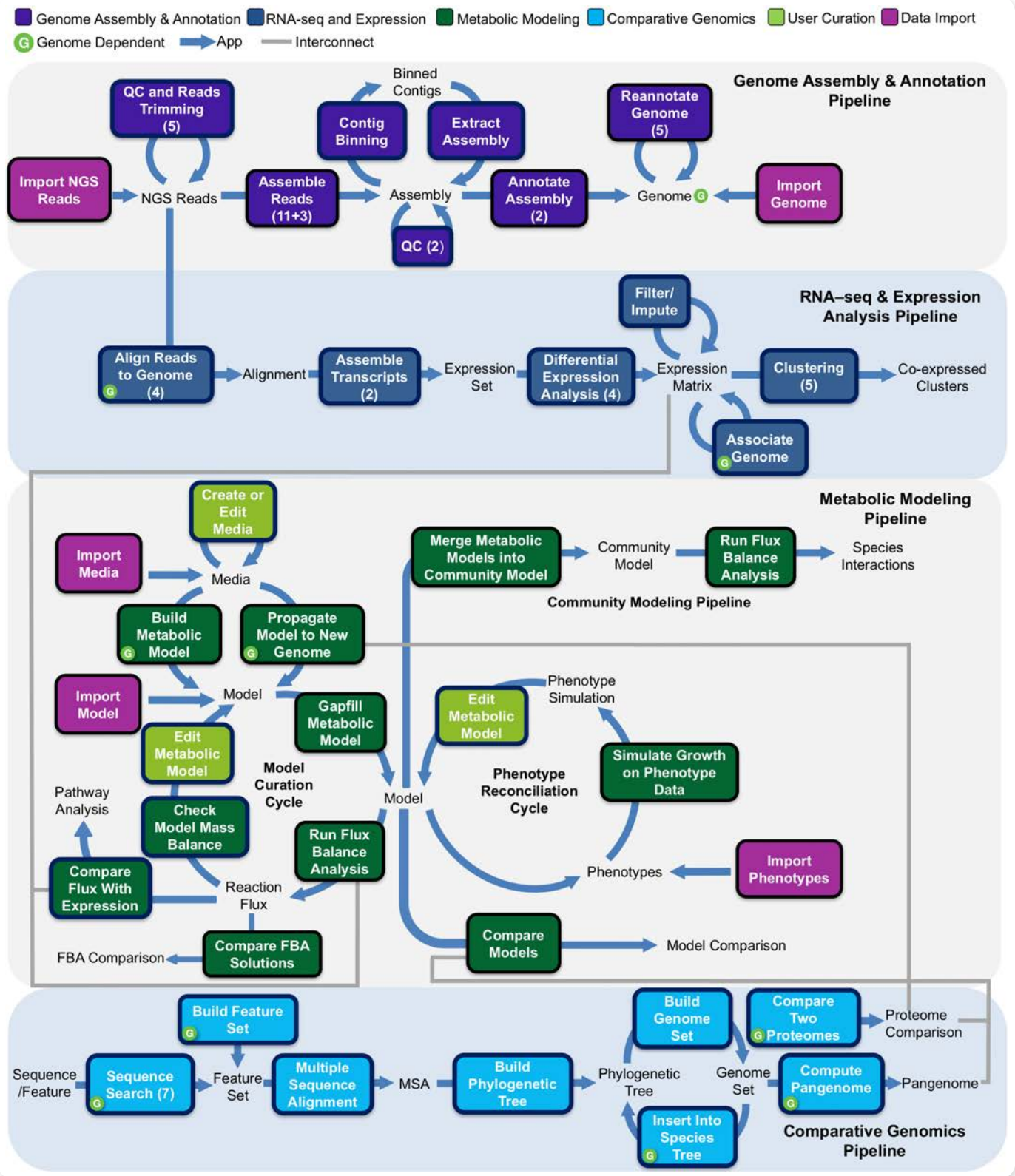


Fig. 3. Outline of the major workflows and datatypes in KBase. The unboxed labels represent datatypes, while each colored box represents a single App. The box colors signify the category of functionality (key at top), and the numbers in parentheses indicate the number of alternative apps that implement each function. Apps that require a genome as input are marked with a green "G" icon.

Narratives for Reproducible Science

Narratives in KBase (Fig. 2) capture computational experiments in dynamic, interactive documents that promote collaboration and reproducibility of scientific results. In addition to data and analysis steps, Narratives can include user images, notes, and commentary. They can be kept private, shared with colleagues and collaborators, or be made public for the benefit of the research community. Because Narratives are built using the Jupyter Notebook, users can also write custom scripts to interact with the data and tools contained within each Narrative.

The screenshot shows a KBase Narrative interface. On the left, there are panels for 'DATA' and 'APPS'. The 'DATA' panel lists several Shewanella-related datasets. The 'APPS' panel shows a list of tools including Annotation, Assembly, Communities, Comparative Genomics, Expression, Metabolic Modeling, Reads, Sequence, Uncategorized, Upload, and Util. The main workspace displays a narrative titled 'Shewanella Comparative Genomics' with a text description, a microscopic image of bacteria, and a 'Species Tree' application. The 'Species Tree' application shows a phylogenetic tree with various Shewanella species. Callouts with icons point to different parts of the interface: 'Data' (folder icon), 'Apps' (refresh icon), 'Analysis Steps' (cubes icon), 'Sharing' (share icon), 'Commentary' (speech bubble icon), 'Visuals' (chart icon), and 'Custom Scripts' (code icon).

Fig. 2. A Narrative in KBase.

What Can You Do in KBase?

- Perform quality control and de novo assembly on next-generation sequencing reads.
- Annotate prokaryotic genomes and plant transcripts.
- Compare and contrast growth phenotypes under hundreds of conditions.
- Explore the comparative genomic organization, phylogeny, and gene content of organisms.
- Identify the enriched species and functions among sets of metagenomes.

Reasons to Use KBase

- Enables collaboration and sharing of results in a transparent and reproducible format.
- Provides access to enterprise-class computing through a graphical user interface.
- Hosts a powerful and user-extensible suite of systems biology and modeling tools.
- Allows users to explore results in the context of available knowledge in the field.
- Lowers the barrier for integrating complex and diverse datasets from multiple sources.

Joint Genome Institute Data Integration (Coming Soon)

Quickly and easily transfer high-throughput sequencing data from the JGI Genome Portal to KBase for analysis.

Open Source

Nearly all KBase software is publicly available through GitHub (www.github.com/kbase), where it can be reviewed and extended by the community.

Data in KBase

KBase provides a single comprehensive resource that enables users to analyze their own data as well as a wide range of public bioinformatics data. KBase supports a variety of data types including:

- Sequencing reads
- Assemblies
- Genomes
- Metabolic models
- Media
- Pangenomes
- Phylogenetic trees
- Taxonomic and functional profiles

A distinguishing feature of KBase is its data model, which integrates diverse biological datasets and represents them as meaningfully rich types that describe relationships among data. This integration enables comparison across domains and interoperability with both standard and next-generation tools. With the data model, KBase also can add value to new information, treating it as additional evidence that enriches the understanding of previously acquired data.

Users can import their own sequencing reads, assemblies, genomes, plant transcripts, media, flux balance analysis models, phenotype sets, and expression matrices for analysis. Efforts are under way to support additional data types from both public sources and users, including providing search and import capabilities for a wide range of JGI data. Also coming soon is a new system-wide search interface that will provide unparalleled ability to access and explore data throughout KBase.

Public Data Sources in KBase

kbase.us/data-policy-and-sources/

Data Summary

kbase.us/data-summary/

Getting Started

To begin exploring and using KBase, go to kbase.us and click “Get started” (see Fig. 4). From here, you can register for a user account, access a Quick Start guide, and learn more about features such as Narratives, apps, and data in KBase.



Fig. 4. Getting started from kbase.us.

Resources

Publications:

kbase.us/publications/

Research Highlights:

kbase.us/research-highlights/

Research Narrative Library:

kbase.us/narrative-library/

User Testimonials:

kbase.us/testimonials/

Project Paper (preprint):

Arkin, A.P. et al. The DOE Systems Biology Knowledgebase (KBase). DOI: <http://dx.doi.org/10.1101/096354>.

User Meetings and Workshops

Learn more at KBase user and developer workshops and the annual KBase user meeting, held jointly with the DOE Joint Genome Institute.

Host a KBase webinar or workshop at your institution. Contact us online or on social media to learn more about hosting an event.

Calendar of Events:

kbase.us/events/

Social Media

Facebook:

facebook.com/DOEKBase/

Twitter:

@DOEKBase

News Blog:

<http://kbase.us/news>

Questions and Support

For more information or assistance, visit: kbase.us/contact-us



Office of Science

January 2018