

KBase's Narrative Interface: A User Interface for Creating Reproducible Systems Biology Workflows

Nomi Harris*¹ (nlharris@lbl.gov), William J. Riehl¹, Erik Pearson¹, Adam P. Arkin¹, Chris Henry², Bob Cottingham³, and the KBase Team at the following institutions:

¹Lawrence Berkeley National Laboratory, Berkeley, CA; ²Argonne National Laboratory, Argonne, IL; ³Oak Ridge National Laboratory, Oak Ridge, TN; ⁴Brookhaven National Laboratory, Upton, NY; ⁵ Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

<http://kbase.us>

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a free, open-source software and data platform that enables researchers to collaboratively generate, test, compare, and share hypotheses about biological functions; analyze their own data along with public and collaborator data; and combine experimental evidence and conclusions to model plant and microbial physiology and community dynamics. KBase's analytical capabilities currently include (meta)genome assembly, annotation, comparative genomics, transcriptomics, and metabolic modeling. Its web-based user interface supports building, sharing, and publishing reproducible, annotated analysis workflows with integrated data. Additionally, KBase has a software development kit that enables the community to add functionality to the system.

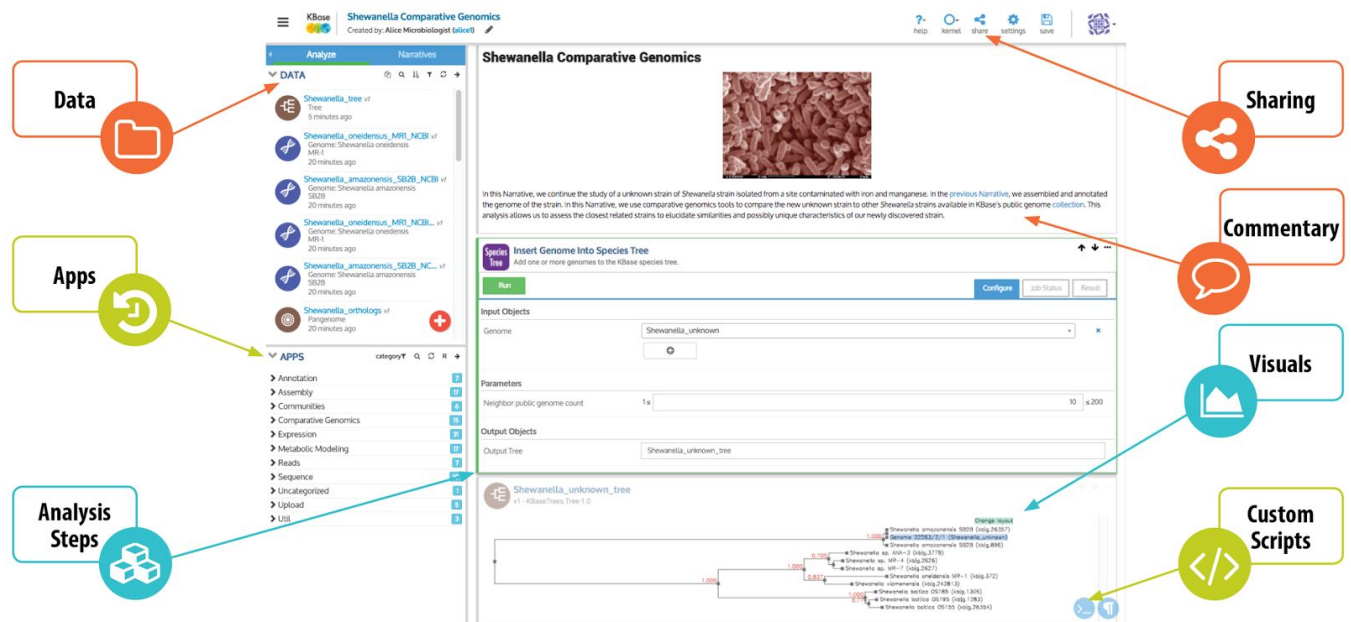


Figure 1. A KBase Narrative is an interactive, dynamic, and persistent document created by users that promotes open, reproducible, and collaborative science.

KBase's Narrative User Interface (Fig. 1) enables users to (i) upload their private data; (ii) search and retrieve extensive public reference data; (iii) access data shared with them by others; (iv) share their own data with others; (v) select and run applications on their data; (vi) view and analyze the results from those applications; and (vii) record their thoughts and interpretations along with the analysis steps.

All of these activities take place within a point-and-click “notebook” environment—the Narrative interface is built on top of the Jupyter Notebook platform. When a user begins a new computational experiment in KBase, they create a new “notebook” to hold this experiment, which we call a “Narrative”. Every action performed by a user shows up as a new Narrative “cell”. App cells show the chosen input parameters for the application and the results of the analysis. Apps interoperate seamlessly to enable a range of scientific workflows. In addition to running apps, users can create and run blocks of code within a Narrative using “code cells”, which allows them to, for example, run large-scale studies in KBase (e.g., building thousands of models at once). Users can also leverage the flexibility of code cells to add custom analysis steps that are not yet available as KBase apps. Markdown cells allow users to add formatted text and figures to a Narrative to annotate and describe the thought process behind the scientific workflow being crafted.

A finished Narrative represents a complete record of everything the authors did to complete their computational experiment. Once a Narrative has been shared (or made public), other users can copy the Narrative and rerun it on their own data, or modify it to suit their scientific needs. Thus, public Narratives serve as resources for the user community by capturing valuable data sets, associated computational analyses, and scientific context describing the rationale behind a scientific study. This recording of a user's KBase activities within a sharable Narrative is a central pillar of KBase's support for reproducible transparent research, simplifying the re-purposing, re-application, and extension of scientific techniques.

The kbase.us website has a number of resources to introduce you to the Narrative Interface. The [Narrative Interface guide](#) provides a detailed explanation of the features of the interface. Video tutorials, such as “[KBase Quickstart](#)”, are available on our [YouTube channel](#). [Narrative tutorials](#), such as “[Assembly and Annotation of Prokaryotic Genomes](#)”, are interactive tutorials that you can copy and re-run on the example data or your own data.

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.