

The KBase Software Development Kit Makes KBase an Extensible Community Resource

Paramvir Dehal*¹ (psdehal@lbl.gov), Shane Canon¹, **Adam P. Arkin¹**, **Christopher Henry²**, **Bob Cottingham*³**, Nomi Harris¹ 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.

KBase was designed to be an extensible community resource. This extensibility is supported by the [KBase Software Development Kit \(SDK\)](#), which enables external developers to incorporate third-party open source tools as KBase apps. Such third-party apps are beginning to appear in the system, and the pace of development is increasing.

The KBase SDK is a set of command-line tools and a web interface that enable any developer to build, test, register, and deploy new or existing software as KBase apps. All SDK apps and any underlying tools that are registered in the KBase system must adhere to a standard open-source license (opensource.org/licenses). Information about the app developer is maintained in the documentation for that app so credit can be given to the contributor. Data provenance, job management, usage logging, and app versioning are handled automatically by the platform, allowing developers to wrap new scientific tools quickly with minimal KBase-specific training. Other existing platforms offer similar support for third-party development, but KBase’s data model provides the additional benefit of improving interoperability of third-party applications by imposing a single data format and specification on all data types consumed or produced by each app.

A number of external developers have used the SDK to add new KBase apps. These include:

- “Classify PhenotypeSet with Decision Tree” (see Fig. 1), which builds a decision tree in order to classify results of a phenotype set to understand the effect the media compounds have on any defined phenotype (Mikaela Cashman, University of Nebraska)

- SBMLTools, tools to import and integrate published metabolic models (Felipe Liu and Sonia Rocha, University of Minho)
- optStoic (in development), which optimizes metabolic networks for the production of chemical products and biofuels (Costas Maranas and Anupam Chowdhury, Pennsylvania State University)

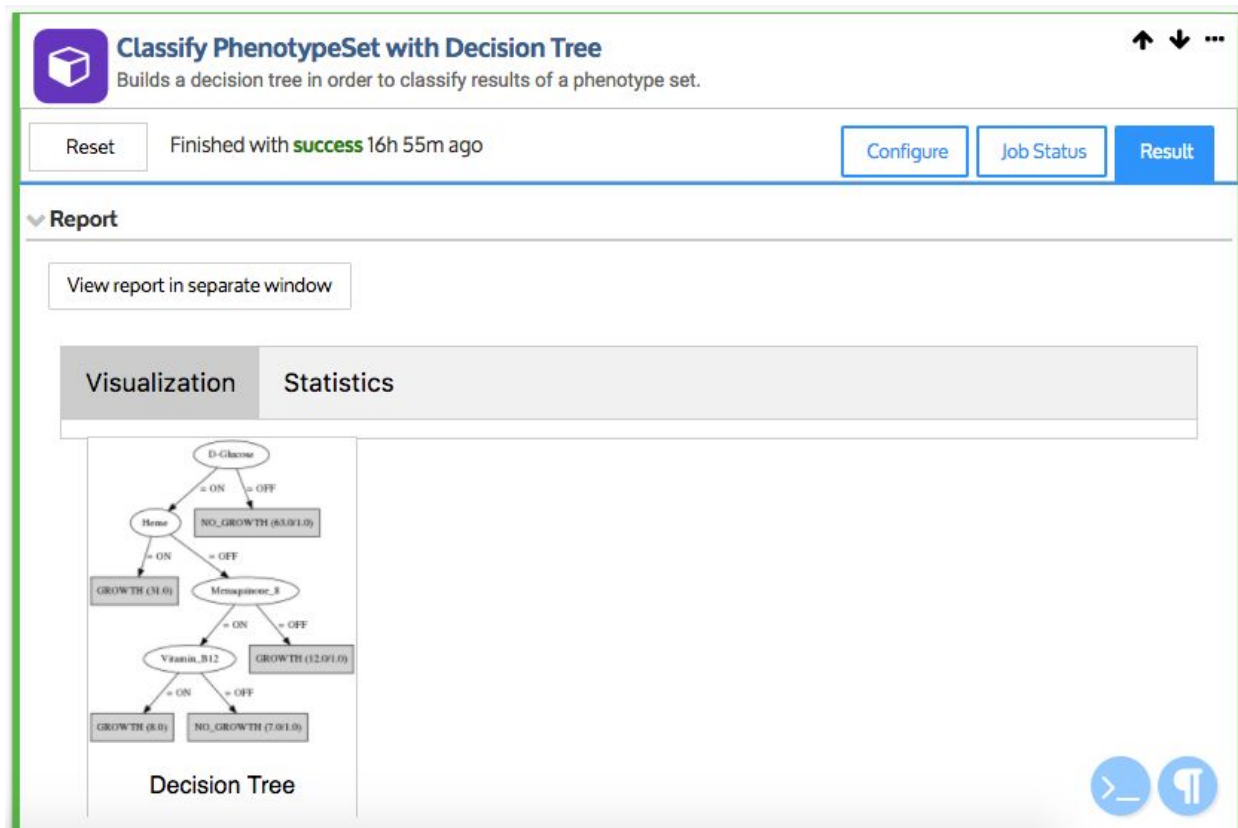


Figure 1. The app called “Classify PhenotypeSet with Decision Tree” was added by third-party developer Mikaela Cashman using the KBase SDK.

These third-party apps, along with many others wrapped by KBase-internal and third-party developers with the SDK, can be found in the KBase App Catalog (<https://narrative.kbase.us/#appcatalog>). More information about the KBase SDK is available at https://github.com/kbase/kb_sdk/blob/master/README.md.

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.