

168. Command Line Environment in the DOE Systems Biology Knowledgebase

Bob Olson¹, James Thomason², Thomas Brettin*¹ (brettin@cels.anl.gov), Jason Baumohl⁴, Aaron Best³, Jared Bischof¹, Ben Bowen⁴, Tom Brown¹, Shane Canon⁴, Stephen Chan⁴, John-Marc Chandonia⁴, Dylan Chivian⁴, Ric Colasanti¹, Neal Conrad¹, Brian Davison⁵, Matt DeJongh³, Paramvir Dehal⁴, Narayan Desai¹, Scott Devoid¹, Terry Disz¹, Meghan Drake⁵, Janaka Edirisinghe¹, Gang Fang⁷, José Pedro Lopes Faria¹, Mark Gerstein⁷, Elizabeth M. Glass¹, Annette Greiner⁴, Dan Gunter⁴, James Gurtowski², Nomi Harris⁴, Travis Harrison¹, Fei He⁶, Matt Henderson⁴, Chris Henry¹, Adina Howe¹, Marcin Joachimiak⁴, Kevin Keegan¹, Keith Keller², Guruprasad Kora⁵, Sunita Kumari², Miriam Land⁵, Folker Meyer¹, Steve Moulton⁵, Pavel Novichkov⁴, Taeyun Oh⁸, Gary Olsen⁹, Dan Olson¹, Ross Overbeek¹, Tobias Paczian¹, Bruce Parrello¹, Shiran Pasternak², Sarah Poon⁴, Gavin Price⁴, Srivdya Ramakrishnan², Priya Ranjan⁵, Bill Riehl⁴, Pamela Ronald⁸, Michael Schatz², Lynn Schriml¹⁰, Sam Seaver¹, Michael W. Sneddon⁴, Roman Sutormin⁴, Mustafa Syed⁵, James Thomason², Nathan Tintle³, Will Trimble¹, Daifeng Wang⁷, Doreen Ware², David Weston⁵, Andreas Wilke¹, Fangfang Xia¹, Shinjae Yoo⁶, Dantong Yu⁶, **Robert Cottingham**⁵, **Sergei Maslov**⁶, **Rick Stevens**¹, **Adam P. Arkin**⁴

¹Argonne National Laboratory, Argonne, IL, ²Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ³Hope College, Holland, MI, ⁴Lawrence Berkeley National Laboratory, Berkeley, CA, ⁵Oak Ridge National Laboratory, Oak Ridge, TN, ⁶Brookhaven National Laboratory, Upton, NY, ⁷Yale University, New Haven, CT, ⁸University of California, Davis, CA, ⁹University of Illinois at Champaign-Urbana, Champaign, IL, ¹⁰University of Maryland, College Park, MD

<http://kbase.us>

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

KBase provides a variety of user experiences targeted to meet the needs of a diverse user- base in the systems biology community. A rich programmatic API is provided in wide variety of languages for maximum performance and flexibility for computationally advanced users. The graphical Narrative interface provides an intuitive visual experience to facilitate hands-on interactive systems biology. Between these is the command line interface, providing a powerful but simple command-driven interface capable of high- throughput interaction with KBase tools and data. The online IRIS web application (<http://iris.kbase.us>) permits users to run commands in KBase from any computer with web access without installing any local software. It offers a wide variety of online tutorials, scripting, and data visualization features. The installable KBase client for Mac and Ubuntu Linux enables users to run KBase commands from a shell on their own computers, supporting seamless integration with locally installed data and tools.

The installable client also includes API libraries in numerous programming languages (e.g., Java, JavaScript, Perl, Python, JSON RPC), enabling users to program directly against the KBase API from their own computers and workstations. In this way, the existing computational infrastructure in any academic lab can seamlessly integrate KBase tools and data into local data analysis pipelines.

The KBase command line environment now offers a powerful suite of over 800 distinct commands that encompass genome assembly, annotation, metabolic modeling, expression data analysis, phenotype analysis, and GWAS. Since the initial release in February 2013, the KBase command line environments have collectively been applied by over 300 unique users to: (i) annotate or import over 34K genomes, (ii) construct over 40K metabolic models, and (iii) run over 110K FBA simulations. In this poster, we will highlight all command-line clients and tools available in KBase, with example workflows for genome assembly, annotation, metabolic modeling, and phenotype reconciliation.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.