

Science Focus Areas use KBase to share new tools and data

Adam P. Arkin¹, **Robert Cottingham**³, **Chris Henry**², Elisha Wood-Charlson^{1*}, Benjamin Allen³, Jason Baumohl¹, Kathleen Beilsmith², David Dakota Blair⁴, Jay Bolton¹, Shane Canon¹, Stephen Chan¹, John-Marc Chandonia¹, Dylan Chivian¹, Zachary Crockett³, Paramvir Dehal¹, Ellen Dow¹, Meghan Drake³, Janaka N. Edirisinghe², José P. Faria², Jason Fillman¹, Tianhao Gu², AJ Ireland¹, Marcin P. Joachimiak¹, Sean Jungbluth¹, Roy Kamimura¹, Keith Keller¹, Vivek Kumar⁵, Sunita Kumari⁵, Miriam Land³, Sebastian Le Bras¹, Zhenyuan Lu⁵, Filipe Lui², Dan Murphy-Olson², Erik Pearson¹, Gavin Price¹, Priya Ranjan³, William Riehl¹, Boris Sadkhin², Samuel Seaver², Alan Seleman², Gwyneth Terry¹, Charles Trenholm¹, Sumin Wang¹, Doreen Ware⁵, Pamela Weisenhorn², Ziming Yang⁴, Shinjae Yoo⁴, Qizhi Zhang²

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

<http://kbase.us>

Project Goals: 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 data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.

Science Focus Area (SFA) research occurs at the cutting edge of the Office of Science's Biological and Environmental Research (BER) strategic goals. However, dissemination of knowledge and the tools capable of extending that knowledge to other biological systems beyond traditional scientific publishing is still challenging. KBase, as a free, open source platform designed to support broad biological systems-level analyses and sharing, provides a unique opportunity for BER programs to expand their impact to the 20,000+ KBase users world-wide.

Here, we describe several SFA-KBase collaborative projects, ranging from efforts to explore strain variation in microbes, protein-ligand binding, or quantitative Stable Isotope Probing (qSIP); improving functional annotation or viral ecogenomics analyses; and engineering microbial communities or community function(s) to support resilience in a time of change. Completed projects will highlight contributions and impact, and newer collaborations will showcase what is to come. SFA projects are also included as KBase Collaborations under the KBase User Working Group website: <https://www.kbase.us/research/user-working-groups/>.

SFA-KBase Collaborative projects (alphabetical):

Bacterial-Fungal Interactions (LANL): Add novel taxonomic annotation tools for bacterial and fungal sequences in KBase (released in KBase: GOTTCHA2, Centrifuge). More information: <https://www.kbase.us/research/chain-sfa/>.

Biofuels (LLNL): Improve on existing functional annotation tools by allowing for multiple annotations to be uploaded, compared, and merged, in support of creating more robust metabolic models (released in KBase: import, compare, and merge metabolic annotations). Coming soon: probabilistic annotation and ensemble metabolic modeling. More information: <https://www.kbase.us/research/stuart-sfa/>.

ENIGMA (LBNL): Establish reference-based metagenome workflow to enable detection and annotation of strain variants (released in KBase: Meta-Decoder Call Variants, Strain Finder v1, Fama Profiling). Coming soon: Long read assembly tools for microbial isolates and metagenomes. More information: <https://www.kbase.us/research/adams-sfa/>.

Microbes Persist (LLNL): Integrate the tools that enable work on viruses (released in KBase: vConTACT2, VirSorter, VirMatcher). Create functionality to detect and fix errors in assembled contigs (in development). Coming soon: Workflow to analyze quantitative stable isotope probe (qSIP) data and integration of *microTrait* to enable user-defined trait-matrix generation from a set of target genomes present in the same environment for functional comparison. More information: <https://www.kbase.us/research/peff-ridge-sfa/>.

Persistence Control (PNNL): Coming soon: Improved protein annotation using machine learning, multi-omics data integration, and structural prediction. More information: <https://www.kbase.us/research/egbert-sfa/>.

Plant Microbe Interfaces (ORNL): Coming soon: Design and omics exploration of synthetic microbial communities. More information: <https://www.kbase.us/research/doktycz-sfa/>.

Protein Structure (ORNL): Establish the first protein structure importer and ligand-binding tools in KBase (released in KBase: CCMPredPy Contact Mapping, AutoDock Vina Protein-Ligand Docking, Import Protein Structure from PDB). More information: <https://www.kbase.us/research/mitchell-sfa/>. To extend this SFA effort, KBase is collaborating with PDB to further improve our protein structure workflows and datatypes in KBase. More information: <https://www.kbase.us/research/protein-data-bank/>.

Soil Microbiome (PNNL): Coming soon: Omics-enabled global gapfilling (OMEGGA) for phenotype-consistent metabolic network reconstruction of microorganisms and communities. More information: <https://www.kbase.us/research/hofmockel-sfa/>.

KBase encourages new projects and collaborations! We provide information on the benefits of collaborating with KBase, and outline how to get started.

This work is supported as part of the [BER Genomic Science Program](#). The DOE Systems Biology Knowledgebase (KBase) is funded by 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.