Transcriptome Profiling and Genome-Scale Metabolic Modeling in Fungal and Plant genomes using KBase.

Sunita Kumari*[1] (kumari@cshl.edu), Vivek Kumar[1], Sam Seaver[2], Janaka Edirisinghe[2], Priya Ranjan[3], Zhenyuan Lu[1], Doreen Ware[1], Paramvir Dehal[4], Ben Allen[3], Jason Baumohl[4], Jay Bolton[4], Shane Canon[4], Stephen Chan[4], John-Marc Chandonia[4], Dylan Chivian[4], Mikayla Clark[3], Meghan Drake[3], Jose Faria[2], Annette Greiner[4], Tian Gu[2], James Jeffryes[2], Marcin Joachimiak[4], Roy Kamimura[4], Keith Keller[4], Miriam Land[3], Sebastian Le Bras[4], Akiyo Marukawa[4], Sean McCorkle[5], Cheyenne Nelson[4], Dan Murphy-Olson[2], Erik Pearson[4], Gavin Price[4], William Riehl[4], Boris Sadkhin[2], Alan Seleman[2], Gwyenth Terry[4], Elisha Wood-Charlson[4], Shinjae Yoo[5], Qizhi Zhang[2], Diane Zheng[4], **Chris Henry[2], Bob Cottingham[3], Adam P. Arkin[4]**

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4]Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Brookhaven National Laboratory, Upton, 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 DOE computational infrastructure to perform sophisticated systems biology analyses. KBase is freely available and enables scientists to upload their own data, analyze it alongside collaborator and public data, and share workflows and conclusions. Additionally, KBase has a software development kit that enables the community to add functionality to the system.

KBase is an open, web-accessible computational environment for systems biology research focused on microbes, fungal, plants and their communities. It provides a range of integrated biological data types and associated analysis tools (Apps) that include gene expression analysis, metabolic modeling, comparative genomics and functional genomics. The user-friendly KBase Narrative Interface offers the researchers and bioinformaticians a range of analysis tools and data resources that accelerate the pace of functional genomics research by allowing large-scale sample processing, expression-level quantification and integration of gene expression profiles with downstream functional analysis including clustering of expression profiles based on different algorithms, ontology enrichment, metabolic networks and gene regulation.

KBase currently has 78 plant genomes from the JGI Phytozome database, and 134 fungal genomes from the JGI MycoCosm database. KBase has several data resources that originated from the PlantSEED project which combines plant comparative genomics, functional annotation

of enzymes, and reconstruction of plant primary metabolism for individual species. Plant-specific compounds and reactions, collected from public sources such as KEGG, MetaCyc, and AraCyc, have been integrated into PlantSEED and made available in KBase, where they can be used for plant metabolic modeling.

KBase also integrates reactions, biomass formulations, and gene annotations from 12 published genome-scale metabolic models for diverse fungal genomes. These published models are used by a new app that enables users to automatically construct a new genome-scale metabolic model for any fungal genome. These fungal models are an efficient way of predicting phenotypes across various environmental conditions.

KBase is also actively engaged with the external community to help us improve our tools and workflows for functional genomics especially support for gene expression, regulation, and epigenetics in plant science. These capabilities are directly relevant to important DOE research targets such as optimizing biomass production in biofuel feedstocks.