159.   The DOE Systems Biology Knowledgebase: Plant Science Domain

Doreen Ware[1,2]* (ware@cshl.edu), David Weston[3], Jason Baumohl[4], Aaron Best[5], Jared Bischof[6], Ben Bowen[4], Tom Brettin[6], Tom Brown[6], Shane Canon[4], Stephen Chan[4], John- Marc Chandonia[4], Dylan Chivian[4], Ric Colasanti[6], Neal Conrad[6], Brian Davison[4], Matt DeJongh[5], Paramvir Dehal[4], Narayan Desai[6], Scott Devoid[6], Terry Disz[6], Meghan Drake[3], Janaka Edirisinghe[6], Gang Fang[7], José Pedro Lopes Faria[6], Mark Gerstein[7], Elizabeth M. Glass[6], Annette Greiner[4], Dan Gunter[4], James Gurtowski[2], Nomi Harris[4], Travis Harrison[6], Fei He[1], Matt Henderson[4], Chris Henry[6], Adina Howe[6], Marcin Joachimiak[4], Kevin Keegan[6], Keith Keller[4], Guruprasad Kora[3], Sunita Kumari[2], Miriam Land[3], Folker Meyer[6], Steve Moulton[3], Pavel Novichkov[4], Taeyun Oh[8], Gary Olsen[9], Bob Olson[5], Dan Olson[5], Ross Overbeek[5], Tobias Paczian[5], Bruce Parrello[5], Shiran Pasternak[2], Sarah Poon[4], Gavin Price[4], Srivdya Ramakrishnan[2], Priya Ranjan[3], Bill Riehl[4], Pamela Ronald[8], Michael Schatz[2], Lynn Schriml[10], Sam Seaver[6], Michael W. Sneddon[4], Roman Sutormin[4], Mustafa Syed[3], James Thomason[2], Nathan Tintle[5], Will Trimble[6], Daifeng Wang[7], Andreas Wilke[6], Fangfang Xia[6], Shinjae Yoo[1], Dantong Yu[1], **Robert Cottingham**[3], **Sergei Maslov**[1], **Rick Stevens**[6], **Adam P. Arkin**[4]

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

DOE Systems Biology Knowledgebase (KBase) has two central goals. The scientific goal is to produce predictive models, reference datasets, and analytical tools and to demonstrate their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance, and use of predictive models and methods.

The plant team is currently focused on reconstruction and modeling of genotype-to- phenotype relationships in plant species relevant for DOE mission. Our workflows accessible via narrative and command line interfaces provide interactive, data-driven analysis and exploration across multiple experiments and diverse data-types. KBase allows our users to process next generation sequencing data to identify novel genomic variation and to quantify genome-wide expression levels. Users can process

expression data to calculate co-expression networks and to identify and annotate functional modules within those networks. Furthermore, we provide computational tools to carry out Genome-Wide Association analysis to identify SNPs and candidate genes significantly correlated with plant phenotypes. Predicted genotype-to-phenotype relationships can be validated by a variety of public and user-generated networks metabolic models in KBase.