The DOE Systems Biology Knowledgebase (KBase): Progress towards a system for collaborative and reproducible inference and modeling of biological function

Adam P. Arkin*1 (aparkin@lbl.gov), Jason Baumohl1, Aaron Best2, Jared Bischof2, Ben Bowen1, Tom Brettin2, Tom Brown2, Shane Canon1, Stephen Chan1, John-Marc Chandonia1, Dylan Chivian1, Ric Colasanti2, Neal Conrad2, Brian Davison3, Matt DeJongh5, Paramvir Dehal1, Narayan Desai2, Scott Devoid2, Terry Disz2, Meghan Drake3, Janaka Edirisinghe2, Gang Fang7, José Pedro Lopes Faria2, Mark Gerstein7, Elizabeth M. Glass5, Annette Greiner1, Dan Gunter1, James Gurtowski5, Nomi Harris1, Travis Harrison2, Fei He4, Matt Henderson1, Chris Henry2, Adina Howe2, Marcin Joachimiak1, Kevin Keegan2, Keith Keller1, Guruprasad Kora3, Sunita Kumari3, Miriam Land3, Folker Meyer2, Steve Moulton2, Pavel Novichkov1, Taeyun Oh8, Gary Olsen9, Bob Olson2, Dan Olson2, Ross Overbeek2, Tobias Paczian2, Bruce Parrello7, Shiran Pasternak5, Sarah Poon1, Gavin Price1, Srividya Ramakrishnan5, Priya Ranjan2, Bill Riehle, Pamela Ronald8, Michael Schatz5, Lynn Schriml10, Sam Seaver2, Michael W. Sneddon1, Roman Sutormin1, Mustafa Syed1, James Thomason5, Nathan Tintle6, Will Trimble2, Daifeng Wang7, Doreen Ware5, David Weston2, Andreas Wilke2, Fangfang Xia2, Shinjae Yoo1, Dantong Yu4, Bob Cottingham3, Sergei Maslov4, Rick Stevens2

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, 6Hope College, Holland, MI, 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 computational 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 including (meta)genomic assembly, annotation, network inference and modeling, thereby 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.

Systems biology is driven by the ever-increasing wealth of data resulting from new generations of genomics-based technologies. With the success of genome sequencing, biology began to generate and accumulate data at an exponential rate. In addition to the massive stream of sequencing data, each type of technology that researchers use to analyze a sequenced organism adds another layer of complexity to the challenge of understanding how different biological components work together to form a functional living system. Achieving this systems-level understanding of biology will enable researchers to predict and ultimately design how biological systems will function under certain conditions. A collaborative computational environment is needed to bring researchers together so they can share and integrate large, heterogeneous datasets and readily use this information to develop predictive models that drive scientific discovery.

The advancement of systems biology relies not only on sharing the results of projects through traditional