



DOE Systems Biology

Knowledgebase

KBA

**Integration and
modeling for
predictive biology**

SE

Sunita Kumari

Ware lab

Cold Spring Harbor Lab

September 13, 2012

KBbase – The new Systems Biology Knowledgebase [kbase.us]

- Collaborative effort designed to accelerate our understanding of microbes, microbial communities, and plants.
- Community-driven, extensible and scalable open-source software framework and application system.
- Offer free and open access to data models and simulations, which will enable scientists and researchers to build new knowledge, test hypotheses, design experiments, and share their findings to accelerate the use of predictive modeling in biological research

What can KBase do?

- Combine heterogeneous data types
- Offer standardized access to bioinformatic and modeling analyses
- Use evidence-supported annotations of genome structure and genetic function
- Discover new associations and network structures in community and molecular networks
- Map genotype to complex organismal traits
- Design and refine experiments using models of metabolism, regulation and community function
- Enable sharing of data, hypotheses, and newly-generated knowledge

The KBase Team:



Collaborators:

- Cold Spring Harbor Laboratory
- University of California, Davis
- Hope College in Michigan
- University of Illinois at Urbana-Champaign
- Yale University
- The Joint Genome Institute
- InterBRC Knowledgebase
- Several University Knowledgebase Projects

→ **New Partners and Collaborators from Users and Stakeholders**

Funding: This work is supported by the U.S. Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

KBase Plants Overall goal

- Meaningful integration of the data to extract knowledge from the wealth of high-throughput data in plant biology.
 - Establish a framework for deploying capability that will allow interactive, data-driven analysis and exploration across multiple-omics experiments.
 - Provide researchers access to comprehensive datasets from high-throughput experiments together with relevant analytical tools and resources.
 - Provide a platform for researchers to analyze their own experimental data and have these results incorporated into the data exploration framework.

KBase-Plants targeted genomes

- Click to edit Master text styles

- Second level

Poplar

- Third level

- Fourth level

- Fifth level



Arabidopsis

Sorghum

Chlamydomonas

Brachypodium



Miscanthus



Switchgrass

Kbase Plants Network

- Develop co-expression analysis resource
 - Arabidopsis and Poplar microarray data as a proof of concept using weighted correlation network analysis (WGCNA – R software package) to explore the expression of genes across experimental conditions, grouped by tissue types, and different stages of plant development.
- Prototyping with Arabidopsis Experiment metadata, curation of a subset of GEO datasets
 - based on PO anatomy and PO growth and development for co-expression network (for spatial and temporal expression)
 - based on different array designs (treatment/condition specific) using EO controlled vocabularies.

Challenges/gaps

- PO temporal and EO terms.
- Need to develop the infrastructure to support metadata and support the integration across experiments, and organisms.
- Need new controlled vocabularies especially that describe bioenergy related processes.