

## Auxotrophy Prediction and Community Metabolic Modeling Reveals Functional Delegation in a Variety of Microbiome Systems

Pamela Weisenhorn\*<sup>2</sup> (pweisenhorn@anl.gov), **Christopher Henry**<sup>2</sup>, Janaka Edirisinghe<sup>2</sup>, **Adam P. Arkin**<sup>1</sup>, **Bob Cottingham**<sup>3</sup> and the KBase Team at the following institutions

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

<http://kbase.us>

**Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a free, open-source software and data platform that enables researchers to collaboratively generate, test, compare, and share hypotheses about biological functions; analyze their own data along with public and collaborator data; and combine experimental evidence and conclusions to model plant and microbial physiology and community dynamics. KBase's analytical capabilities currently include (meta)genome assembly, annotation, comparative genomics, transcriptomics, and metabolic modeling. Its web-based user interface supports building, sharing, and publishing reproducible, annotated analysis workflows with integrated data. Additionally, KBase has a software development kit that enables the community to add functionality to the system.**

KBase includes a pipeline that supports the development and analysis of community metabolic models starting from metagenomic data. These community models are valuable tools for improving our understanding of the molecular interactions and ecological principles that guide the behavior of microbial communities for numerous applications in the environment and industry. This new KBase pipeline integrates 11 analysis steps, including: metagenome assembly; binning of assembled contigs by species; assessment of assembly and genome quality and completeness; genome annotation; species metabolic model reconstruction; merging of models into a community model; mapping RNA-seq reads to individual species; model gapfilling; flux balance analysis; comparison of flux and expression profiles; and simulation of growth phenotypes. All eleven steps are available within a user-friendly point-and-click interface at <http://narrative.kbase.us>.

We applied this pipeline to understand the ecology and trophic interactions occurring within three microbiome-based datasets: (i) a lab-constructed community comprised of the cyanobacterium *Thermosynechococcus elongatus* supporting the heterotrophic bacterium *Meiothermus ruber* (Figure 1a); (ii) 120 naturally occurring soil isolates, many of which are comprised of 2-3 tightly coupled species (Figure 1b); (iii) a 13-species microbiome that performs electrosynthesis (Figure 1c); and (iv) a larger 18-species natural community comprising an epsomitic phototrophic microbial mat in Hot Lake, Washington. We demonstrate how the tools in KBase can integrate community Biolog profiles and transcriptomic profiles (Figure 1d) in our community models of these systems. Further, we show how our platform can predict the trophic interactions that occur among species in each of these systems. Overall, we find significant

trophic interactions in all communities, involving numerous essential metabolites, including amino-acids, vitamins, and cofactors. We also find numerous cases of metabolic handoff, particularly in our lab-constructed, electrosynthetic, and Hot Lake communities, which all include an autotrophic organism that performs carbon fixation functions for all other species in the community. We see how individual species form complex ecological networks of interactions where each species derives chemical energy from the nutrients in the environment, while stability is preserved through essential trophic interactions. We also see how changes in the nutrient content of the environment can disrupt these webs of trophic interactions.

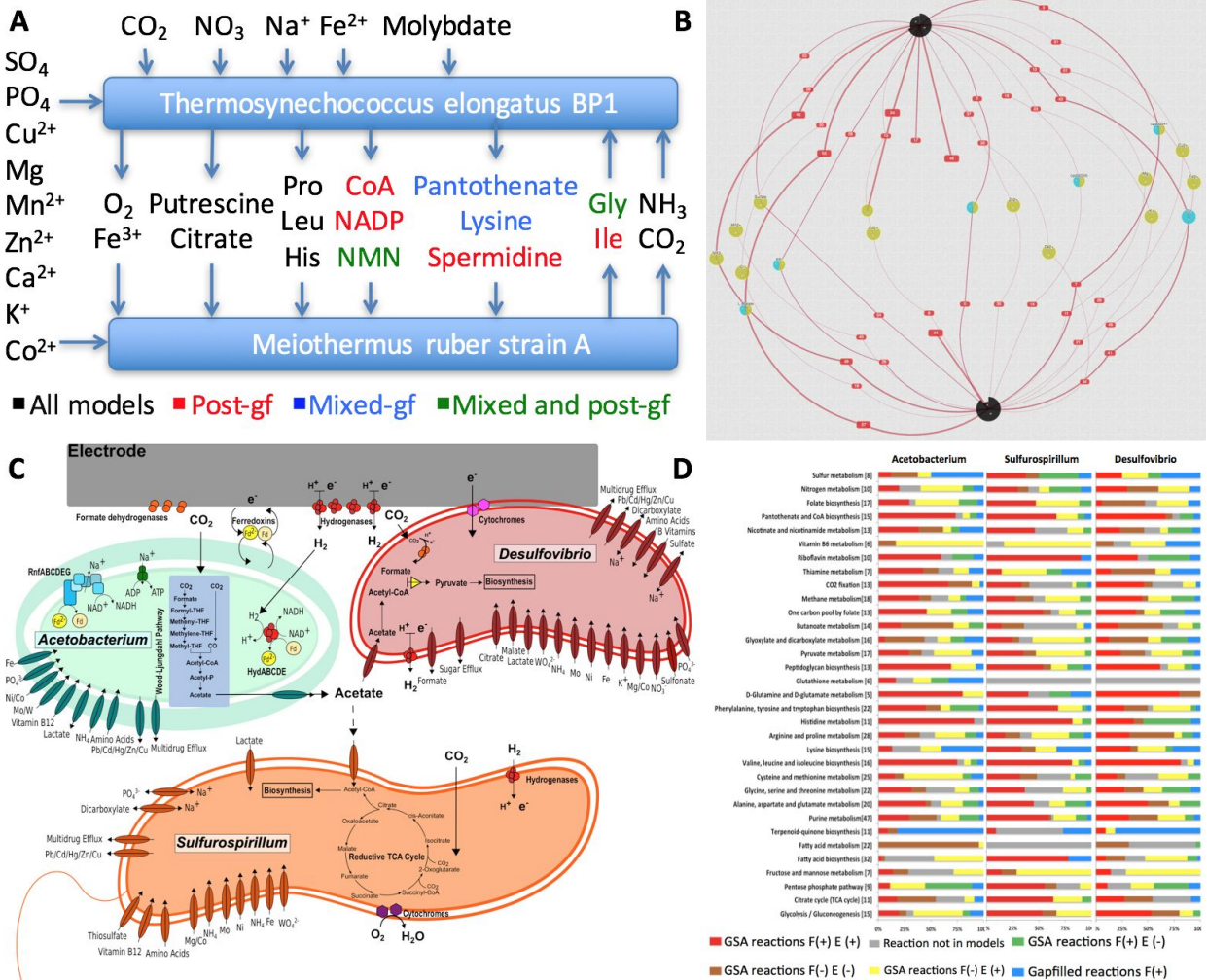


Figure 1. KBBase was applied to predict interspecies interactions (A-C) with validation from transcriptomic data (D). The tools have been applied to a variety of systems, including a 2-species phototrophic consortia (A), numerous multi-species soil isolates (B), and a 13-species electrosynthetic microbiome (C).

KBBase is funded by the Genomic Science program within 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.