Improving Draft Metabolic Models in KBase: Tools for Importing, Comparing and Merging Metabolic Annotations

Patrik D'haeseleer[1] (dhaeseleer2@llnl.gov), Jeffrey A. Kimbrel[1], Janaka N. Edirisinghe[2], Felipe Liu[2], James Jeffryes[2], Ali Navid[1], Chris Henry[2], and Rhona K. Stuart[1]

[1]Lawrence Livermore National Laboratory, Livermore CA, USA
[2]Argonne National Laboratory, Lemont IL, USA

https://bio-sfa.llnl.gov/

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multi-trophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

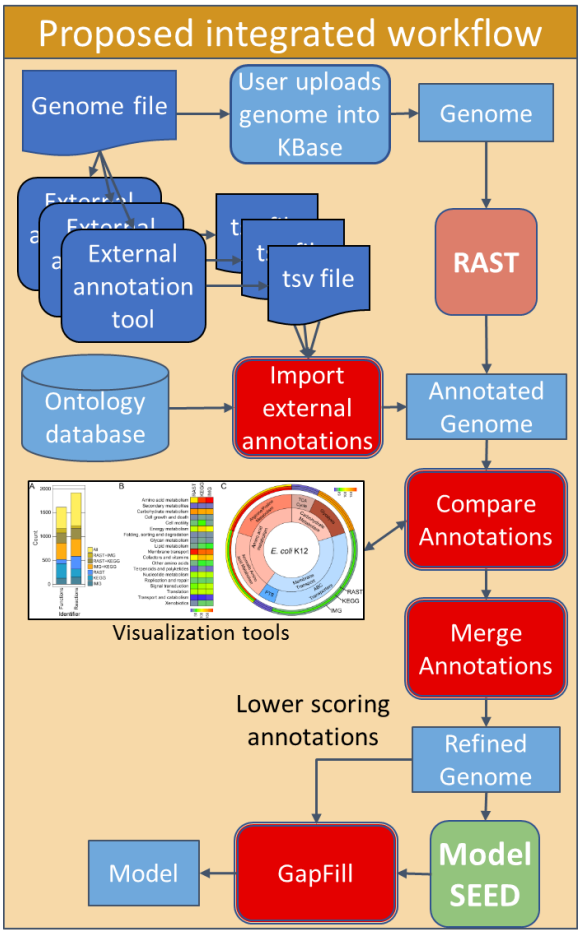
Metabolic pathway analysis, and especially metabolic modeling, is one of the cornerstones of modern Systems Biology, because it allows us to go straight from sequence data to gene functions to an understanding of how whole biological systems function. Metabolic modeling is a critical part of our LLNL Biofuels SFA, which investigates metabolic interactions in bioenergy-relevant microbial communities. Accurate metabolic models require well-annotated genomes. Unfortunately, assigning functional annotations to genes is an imperfect science, and annotated genomes typically contain 30-50% of genes with little or no functional annotation, severely limiting our knowledge of the "parts lists" that the organisms have at their disposal.

In previous research, we have shown that single metabolic annotation tools such as RAST or KEGG tend to be incomplete and inconsistent, and that merging annotation from multiple sources can drastically increase the number of genes and metabolic reactions included in metabolic models (Griesemer *et al*, BMC Genomics 2018). Our comprehensive approach added on average 40% more reactions, 3-8 times more substrate-specific transporters, and 37% more metabolic genes, compared to annotation using only a single tool. These results are even more pronounced for pathways outside of the core carbon metabolism, and for bacterial species that are phylogenetically distant from well-studied model organisms.

The DOE Systems Biology Knowledgebase (KBase) contains a suite of powerful Apps for building genome-scale metabolic models. ModelSEED, originally developed by our collaborator

Chris Henry, is the central Flux Balance Analysis model building App in KBase, and one of the most popular metabolic modeling tools for generating draft models because of its accessibility, ease of use and quality. However, it currently only supports metabolic annotations produced by the Annotate Microbial Genome App, based on RAST (Rapid Annotations using Subsystems Technology). This means that so far it has been impossible for researchers who may prefer to use other high quality annotation tools such as KEGG or even JGI's IMG platform to import their annotations into KBase, let alone merge annotations from multiple sources.

We are currently developing a set of KBase Apps to allow users to upload functional annotations from popular third-party annotation tools, compare and merge them, and use them for metabolic modeling. (1) An Import App (close to completion) will allow user to upload a simple tab-separated file with annotation data in the form of EC numbers, KEGG and MetaCyc reactions identifiers. (2) A Compare App will allow the user to compare metabolic annotations from different sources, by mapping all of them to the ModelSEED reaction database. (3) The Merge App will provide the user with a simple yet flexible scoring mechanism to select a preferred set of annotations from among the full set of functional identifiers mapped to each gene in the genome. (4) Finally, we will also assist the KBase metabolic modeling team to make modifications in the existing ModelSEED App and the Gap Filling tool, to enable users to build models from the merged highest-confidence annotations, and prioritize the remaining lower-scoring annotations for gapfilling.



This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL Biofuels SFA, FWP SCW1039.