The 2019 KBase Fungal Biochemistry Curation Jamboree: Insights and Lessons Learned

Jeremy Zucker,[1]* (jeremy.zucker@pnnl.gov), Neeraj Kumar,[1] Janaka Eirisinghe,[3] Louise Glass[2], Kyle Pomraning, Scott Baker, Erin Bredeweg, and **Chris Henry[3]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]University of California, Berkeley, Berkeley, CA; [3]Argonne National Laboratory, Lemont, IL

https://kbase.us

Project Goals: Short statement of goals. (Limit to 1000 characters)

Fungal genome-scale metabolic models are an efficient way of predicting phenotypes across various environmental conditions. These metabolic models are a key tool in understanding fungal-bacterial and plant-fungal community behavior.

However, automating the construction of high-quality fungal models has been a challenge. Recently, Kbase has introduced a fully automated fungal model construction pipeline. However, the underlying biochemistry data that is derived from published fungal models still needs to be significantly improved to build high-quality models.

To achieve this task, KBase organized a fungal biochemistry curation jamboree at PNNL. They reached out to the fungal community in the DOE space who were willing to contribute to curation of the fungal biochemistry. KBase has also developed an Escher map-based curation environment where curators can visualize biochemistry in the form of biochemical pathways and have the ability to curate reactions and assign gene families to those reactions. Based on feedback from the curation jamboree, KBase, has now set up standards in their curation process such as linking of literature citations and evidence-codes justifying the curation events.

Abstract

Recently, KBase has introduced a methodology to construct genome-scale fungal models in an automated fashion based on a set of reactions that are derived from 14 published fungal metabolic models. As the basis for the method, they produced a fungal model template that encompasses the biochemistry data from the published fungal models and the structural annotations from the associated fungal genomes.

KBase's approach uses structural annotations of any user-submitted fungal genome and computes a set of orthologous proteins against the curated fungal template in order to assert the presence or absence of specific biochemical reactions and pathways. These orthologous families are then curated and mapped to biochemistry by expert curators. The related biochemistry data is then propagated to construct a new draft metabolic model. Once the draft metabolic models are derived, additional reactions are added based on available functional annotations. This method is deployed in the Department of Energy Systems Biology Knowledgebase (KBase) (https://narrative.kbase.us/) as an app called "Build Fungal Model". This method is able to produce a draft fungal metabolic model in an around one hour.

KBase recognized that the core component of a pipeline for producing high-quality draft fungal models is a well curated biochemistry. However, the underlying biochemistry data that is derived from the published models needs to be significantly improved and reconciled onto a controlled

vocabulary in order to avoid redundancy and the incorrect representation of fungal biochemistry. To accomplish this, they reached out to fungal experts in DOE space and hosted a “Fungal Curation Jamboree” at Pacific Northwest National Laboratory”. Parallel to this effort they built a fungal curation environment based on Escher maps (https://escher.github.io/#/) where the biochemistry data and the gene associations can be visualized in the form of biochemical pathways. This allows the curators to quickly make the curation decisions on reactions and associate gene families to those reactions. This talk will focus on the outcomes of the Fungal Biochemistry Curation Jamboree and the lessons learned.