High-quality Genome-scale Metabolic Reconstructions for Multi-scale Microbial Communities

Laura Kutt* (laura.kutt@ext.uni.lu), Stefania Magnusdottir, Almut Heinken, Eugen Bauer, Dimitry A. Ravcheev, Ronan M.T. Flemingand **Ines Thiele**

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg.

http://stanford.edu/group/SOL/multiscale/

Project Goals: There are an increasing number of automatically reconstructed genome scale metabolic models available for various organisms from databases, such as KBase; however, these models require further microbe specific refinement. The refinement process is time-consuming and laborious. This project introduces a pipeline that reduces many errors that are associated with automatic reconstructions. The reconstructions are modified to include correct reaction directionalities and to allow feasible ATP production. Further refinement focuses on manual curation of the models against current available knowledge. It involves extensive literature search for carbon sources and fermentation products, and comparative genomic approach for central metabolism, respiration, and amino acid biosynthesis pathways. The inclusion of these pathways increases the predictive power of the models. High quality metabolic models can be then applied to genetic engineering of specific microbes to increase the yield or degradation of compound of interest, or investigate microbe-microbe relationships in complex microbial communities.

In silico genome scale metabolic models are often used in genetic engineering to optimize the microbe metabolism of for a specific task, for example for degradation of environmental pollutants in the case of oil spills. This type of optimization is usually done by flux balance analysis, which allows evaluation of gene deletions in the context of metabolic system. Often the elimination of a specific metabolic enzyme does not lead to the increased yield of the compound of interest due to compensatory effects in metabolic systems. In silico models enable the production of possible solutions, which can be tested in laboratory. However, the predictions depend on the quality of the genome-scale metabolic models used. The Department of Energy Systems Biology Knowledgebase (KBase) provides tools for automatic generation of these genome scale metabolic models. However, these models need substantial improvement to provide reliable microbe specific predictions. Our pipeline identifies the main errors associated with the draft models, and undertakes manual curation of species-specific pathways. The pipeline first checks the metabolic networks for reaction directionalities, internal flux loops and futile cycles, which are common problems in the automatic draft models, for example due to high number of transport reactions. The second stage involves extensive manual curation of various metabolic pathways such as fermentation products, carbon sources, amino acid biosynthesis, central metabolism, and respiratory pathways. This refinement is necessary due to high variability between different genera of bacteria, for example, uptake of carbon sources is highly species specific. This pipeline also checks for anaerobic growth, and the increased biomass production in the presence of oxygen. Anaerobic growth should be enabled for microbes that thrive in anaerobic environments. This pipeline fulfils a need for faster reconstruction of high quality models, especially in the light of increasing availability of (meta-) genomic data. We also developed a computational toolbox for modeling of multi-scale microbial communities as well as the tailoring of the microbial community models based on meta-omics data. The reconstruction and modeling software is implemented in Matlab (Mathworks, Inc., Natic, MA, USA) and shall form an extension to The COnstraint-Based Reconstruction and Analysis (COBRA) Toolbox [1].

References

[1] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.

This work was in part supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.