Functional Genomics Based Approach For Reconstruction of Genome Scale Metabolic Network Models

Edward Vitkin & Tomer Shlomi
Computer Science, Technion – Israel Institute of Technology

Intro to METABOLISM

Metabolism is the set of chemical reactions that happen in living organisms to maintain life
- Understanding of metabolism is important for diverse areas in Medicine and Bio-Industry.
- The metabolic network is described using the stoichiometric matrix (S) and modeled as a hyper-graph, with metabolites as vertices and reactions as directed hyper-edges. The stoichiometric coefficient of metabolite i in the reaction j is encoded by non-zero entry at Sij. Vector (v) of fluxes through the model reactions has to fulfill the equation: Sv=0, where x is the vector of metabolite concentrations and t is the time.
- Two major challenges in metabolic reconstruction are:
  i. Filling of network gaps (gap-filling)
  ii. Assignment of genes to reactions (gene-assignment)

• Flux Balance Analysis (FBA) is the major computational approach used in gap-filling algorithms. FBA is a LP-based framework seeking optimal reaction flux vector under various constraints:

  $\max \{v_{i max}\} \\ s.t.: \ S \cdot v = 0 \ \ (1)$
  $v_l \leq v \leq v_u \ \ (2)$

In this FBA example the optimization function is maximum organism growth rate, and the set of constraints contains only the most typical ones:
2. Boundaries of reactions rates.

Reaction Data

Universal Reaction Database:
- All the known reactions gathered from various species

Draft Organism Network:
- Known reactions
- Metabolic content
- Utilized carbon sources

MIRAGE algorithm

Stage I Calculation of functional genomics weights for candidate reactions
Stage II Elimination of candidate reactions according to received weights, maintaining network consistency
Stage III Combination of received N results to the Final Model

Functional Genomics Data

- Phylogenetic weight is calculated based on the normalized maximal Jaccard correlation between reaction phylogenetic profile and the corresponding profiles of its known neighbors.
- Expression weight is calculated based on the normalized average Pearson correlation between gene expression profiles of potential enzyme-coding genes, and the expression profiles of genes of its known neighbors.

Validation

E. Coli iAF1260

Synechocystis PCC6803

• The quality of MIRAGE is marked with a violet star symbol. The quality of several controls are marked with a triangle, bar, and circle. The predictive performance of the functional genomic data (without metabolic flux analysis) is shown by the straight lines. The performance of random predictions of gap-filling reactions is colored blue.
• Combination of functional genomics data with flux analysis outperforms gap-filling using each of them separately.
• Integration of several functional genomics data sources increases the quality of the achieved results.

Application on Cyanobacteria

Model Reconstruction
- Algorithm was applied to reconstruct 36 Cyanobacteria with known genomes.
- Available known reactions described species similarities, while MIRAGE predictions demonstrated their differences.
- Based on sequence similarity, genes were found for many of predicted reactions.

Astatxanthin Production
- Standard metabolic engineering strain design method was applied to predict possible Astatxanthin production
- Astatxanthin is a powerful antioxidant and important food and drug agent
- We predicted positive overproduction for 15 out of 25 Astatxanthin containing bacteria
- Cyanobacteria may utilize up to 40% of carbon producing up to 2.18 umol gDW⁻¹ h⁻¹ of Astatxanthin

Results Summary

- Novel gap-filling algorithm (MIRAGE) was created, which combines accepted flux analysis methods together with available functional genomics data.
- Several speed-up heuristics allowed reconstruction of organism model less than in 1 hour.
- MIRAGE achieves significantly higher model quality comparing to known alternatives.
- We applied MIRAGE to reconstruct 36 cyanobacteria models.
- Standard metabolic engineering strain design method demonstrated promising results for 15 cyanobacteria when applied to predict possible Astatxanthin production.