Comparative analysis of enzymatic reactions reveals phylogenic evolution of metabolic pathways

Michihiro Tanaka[1]
mtanaka@kuicr.kyoto-u.ac.jp

Naoki Nariai[2]
nariai@ims.u-tokyo.ac.jp

Thomas Handorf[3]
handorf@physik.hu-berlin.de

Oliver Ebenhöh[3]
oliver.ebenhoeh@rz.hu-berlin.de

Satoru Miyano[2]
miyano@ims.u-tokyo.ac.jp

Minoru Kanehisa[1]
kanehisa@kuicr.kyoto-u.ac.jp

Reinhart Heinrich[3]
Reinhart.Heinrich@rz.hu-berlin.de

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University Gokasho, Uji, Kyoto 611, Japan

[2] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[3] Theoretical Biophysics, Institute of Biology, Humboldt University Berlin.

Keywords: metabolic pathway, enzymatic reaction, phylogenic evolution

1 Introduction

Metabolic pathways consist of compound units and reaction units. Representing them as nodes and edges of a graph enables us a statistical analysis of metabolic pathways. Especially, the analysis based on the network topology provides large-scale qualitative knowledge on metabolic pathways [1-2]. On the other hand, quantitative analyses, such as the flux balance analysis (FBA) [3-4], are limited to analyses of small networks, because metabolite concentrations are unknown and the identification of alternative pathways is difficult. In this work, we assigned phylogenic information to each reaction of the KEGG database [5] and reconstructed the metabolic pathway for each organism. We show that lower organisms use more conserved metabolic reactions while higher organisms use more species-specific reactions, which suggests that metabolic pathways have evolved utilizing existing old pathways.

2 Method and Results

In the REACTION section of the KEGG LIGAND database [6], there are about 6,000 entries for substrate-product relationships representing metabolic and other reactions. Using this reaction set, we made the total of 2435 phylogenic profiles based on the information about mapping to each genome in KEGG.

We defined the organism coverage for reactions as follows: The number of organisms that contained enzyme genes catalyzing the reaction is counted and divided by the number of all organisms. The result is shown in the histogram of figure 1. By using this organism coverage, we evaluated the evolutionary conservation of reactions in 8 species: *Synechocystis sp*, *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Dorosophila melanogaster, Arabidopsis thaliana* and *Homo sapiens*.

Also, we defined the reconstruction rate of metabolic networks for each species as follows: The number of each organism's specific reactions whose organism coverage is higher than a predetermined threshold is divided by the number of all reactions in the organism. We set the threshold value ranging from 0% to 100%, and obtained the plot of the reconstruction rate of organisms in Figure 1. We revealed that the lower organism-coverage reactions contribute to reconstructing metabolic pathways in higher organisms, on other hand, the higher organism-coverage reactions contribute more in lower organisms

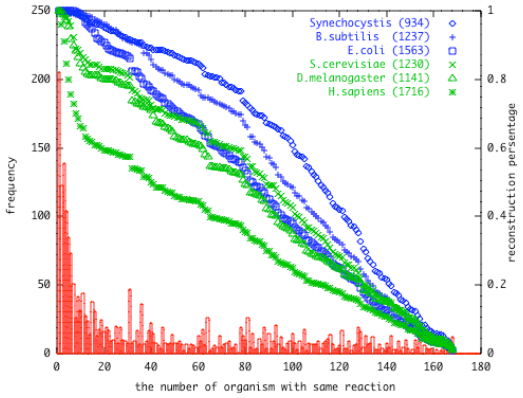

Figure 1: Distribution of the organism coverage for reactions calculated by using 168 organisms and Plot of the reconstruction rate of organism.

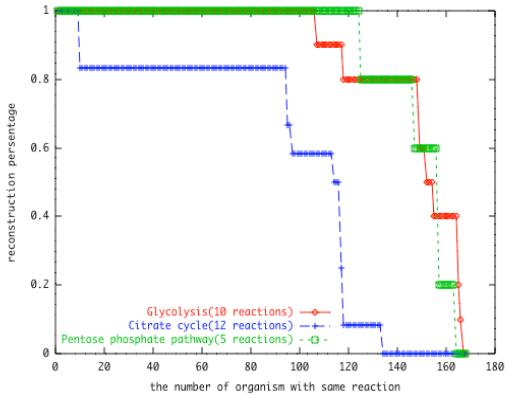

Figure 2: Plot of the reconstruction rate of 3 pathways (Glycolysis, Citrate cycle and Pentose phosphate pathway) in super creature.

3 Discussions

In this work, we found some drop points in reconstructing metabolic pathways for high organism-coverage reactions in the 8 species. Some reactions assigned to this point were localized in androgen and estrogen metabolism. Figure 2 was shown that the citrate cycle pathway incorporated the lower conservative reactions when compared to other two pathways: glycolysis and pentose phosphate pathway. In this study, we confirmed that the variations of reconstructing metabolic pathways corresponds to the phylogenic relationships. In the future work, we would carry out this analysis for more species.

Aknowledgement

This work was supported by the international exchange program between Kyoto University, University of Tokyo, and Humboldt University Berlin.

References

[1] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000)

[2] Guelzim, N., Bottani, S., Bourgine, P. & Kepes, F. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genet*. **31**, 60–63 (2002)

[3] Jin YS., Jeffries TW., Stoichiometric network constraints on xylose metabolism by recombinant Saccharomyces cerevisiae. *Metab Eng*. 6(3)**:** 229-238, 2004

[4] Mahadevan R., Edwards JS., Doyle FJ., Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J. 83(3)**:** 1331- 1340, 2002

[5] Kanehisa M., Goto S., Kawashima S., Okuno Y., Hattori M., The KEGG resource for deciphering the genome, Nucleic Acids Res. 32**:** 277-280, 2004.

[6] Goto, S., Okuno, Y., Hattori, M., Nishioka, T. and Kanehisa, M., LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res*. 30**:** 402-404, 2002