A method for customized cross-species metabolic pathway comparison

Eric Lo[2] wxluo@gel.ym.edu.tw
Takuji Yamada[1] takuji@kuicr.kyoto-u.ac.jp
Michihiro Tanaka[1] mtanaka@kuicr.kyoto-u.ac.jp
Masahiro Hattori[1] hattori@kuicr.kyoto-u.ac.jp
Susumu Goto[1] goto@kuicr.kyoto-u.ac.jp
Chuan-Hsiung Chang[3] cchang@ym.edu.tw
Minoru Kanehisa[1] kanehisa@kuicr.kyoto-u.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan
[2] Bioinformatics Program, Institute of Genetics, National Yang-Ming University, Taipei 112, Taiwan, ROC
[3] Institute of Bioinformatics, National Yang-Ming University, Taipei 112, Taiwan, ROC

Keywords: metabolic pathway, alternative pathway

1 Introduction

In the post-human genomic era, more and more other genomes have been sequenced. Comparative genomic analysis gives us an opportunity to find out the similarities and differences among different organisms. Moreover, comparative analysis of metabolic pathway can give us important information on their evolution and potential pharmacological targets. In this study, we propose a cross-species pathway comparison method on customized starting and ending points of pathways. We use both the reconstructed pathway information of different microorganisms from KEGG [2] and the Pathcomp [5], which can generate possible pathways between two compounds, to generate all possible paths in specific organism for decomposition of pathway. With this method, we can find out organism-specific alternative pathway and the similarity between the species compared.

2 Method

2.1 Pathway and similarity definition

In this study, pathway is defined as a directed graph which consists of the set of vertexes V and the set of edges E and is denoted by G(V, E), where a vertex represents a metabolite, an edge represents a reaction and direction of edge represents the direction of reaction. $G_{A(C\rightarrow D)}$ and $G_{B(C\rightarrow D)}$ represent a pathway from compound C to compound D in organism A and organism B, respectively, Path(G) stands for the set of paths in the graph G, || stands for the cardinality of the set. The similarity of G_A and G_B from compound C to compound D can be formularized as

$$Similarity(G_{A(C\rightarrow D)}, G_{B(C\rightarrow D)}) = \frac{\left|Path\left(G_{A(C\rightarrow D)}\right) \cap Path\left(G_{B(C\rightarrow D)}\right)\right|}{\left|Path\left(G_{A(C\rightarrow D)}\right) \cup Path\left(G_{B(C\rightarrow D)}\right)\right|} \quad (1)$$

2.2 Decomposition of the metabolic pathway to all possible paths

First, the reaction information and compound information are gathered from KGML [6] and LIGAND database [1] in KEGG, respectively. We constructed the graph from specific starting compound to another specific ending compound by using the information from the reaction section of KGML. Then, Pathcomp was used to find out all the possible paths in the graph.

2.3 Cross-species comparison

Given S_A as the set of all possible paths in species A and S_B as the set of all possible paths in species B, we do all against all comparison between S_A and S_B. We use two different methods to compare the metabolic pathway differences among various species.

First, given two path strings S_1, S_2 from species A and species B, respectively. The program compares S_1 and S_2 by using exact string matching with two different criteria: "functionally matched" means we only consider whether the function of reaction is the same, in other words, only compare compound sequence. On the other hand, both compound sequence and reaction sequence are considered in "topologically matched". If there is no any difference between two strings, we consider the path common in the two species compared.

Second, we use dynamic programming algorithm to do global alignment of one reaction sequence from species A and another one from species B. The information content [4] of reaction classification (RC) number [3] is used as the score. It is easier to derive common reaction mechanism in pathway from RC number than EC number. In addition, we also check log odds ratio to ensure the matching is not randomly matched.

3 Result

The result shown in Fig. 1 is the comparison of *Escherichia coli* (eco), *Saccharomyces cerevisiae* (sce) and *Salmonella typhi* (sty) from alpha-D-glucose-6-phosphate to pyruvate, the path length set to equal or less than 10. Functionally matched exact string matching and the number in the graph show the number of path found. The result shows the similarity among three species compared. With the data, we found that *E. coli* and *S. typhi* are very similar (similarity(G_{eco},G_{sty}) = 0.62) to each other but they are not similar to yeast (similarity(G_{eco}, G_{sce}) = 0.06, similarity(G_{sty}, G_{sce}) = 0.07). With the different match criteria used, we also found that the common paths in topologically matched are less than functionally matched (not shown). This is reasonable because topologically matched method is stricter than functionally matched method. In addition, common paths and organism-specific paths can also be shown in the output (Fig. 2). We can find out organism-specific paths and the paths are the potential candidates of alternative pathway in the specific organism. To validate the result, our program interface can also output the compound structure and the mapping of the reaction to known pathway. Validation of the results by using thermodynamic property of reactions is planned for our future work.

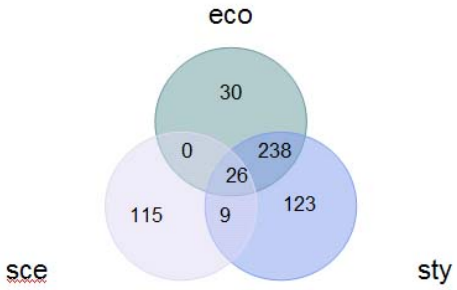

Fig. 1. Venn's diagram shown the similarity among three species compared

Salmonella specific path:

sty Unique path :

*7 C00668 <R01786> C00267 <R00878> C00095 <R00867> C05345 <R04779> C05378 <R01070> C00111 <R01016> C00546 <R00203> C00022 [show reaction chain] [reaction map to pathway]

Fig. 2. The result of organism-specific paths (partial)

References

[1] 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.

[2] Kanehisa, M. and Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28: 27-30, 2000.

[3] Kotera, M., Okuno, Y., Hattori, M., Goto, S. and Kanehisa, M., Computational Assignment of the EC numbers for Genomic-scale Analysis of Enzymatic Reactions. *J. Am. Chem. Soc.*, in press.

[4] Tohsato, Y., Matsuda H., Hashimoto, A., A Multiple Alignment Algorithm for Metabolic Pathway Analysis Using Enzyme Hierarchy. *Proc. 8th International Conference on Intelligent Systems for Molecular Biology* 376-383, 2000

[5] http://www.genome.jp/kegg/tool/pathcomp.html

[6] http://www.genome.jp/kegg/xml/