CAMP: a computational system for Comparative Analysis of Metabolic Pathways

Chun-Yu Chen[1], Chuan-Hsiung Chang[2]

Keywords: pathway comparison, metabolism, algorithm

1 Introduction.

We present a systematic method for comparing the metabolic pathways based on KEGG (Kyoto Encyclopedia of Genes and Genomes) [1] reference pathway data. Our comparison of over 100 metabolic pathways simultaneously can help researchers to figure out what happened in the course of the evolution and adaptation in each well-defined metabolic pathway. The comparison results can be used to improve the genome annotation results by either identifying missing genes, finding alternative routes or reducing annotation errors. CAMP (Comparative Analysis of Metabolic Pathways) will be freely available for academic or nonprofit use at http://gel.ym.edu.tw/camp/.

2 Materials and Methods.

The KEGG provides metabolic pathway information for each completely sequenced organism in xml file format [1]. Each file contains organism-specific metabolic pathway information. We have used these xml files as our input source, and the algorithm of our computational system outputs the percentage of similarity and/or difference between any two metabolic pathways. We have stored the pathway information as a graph $G_X = (V, E_X)$ in the adjacency list [2] for organism X. V is the common set of vertices representing the compounds involved in reference pathway G. E_A is a set of edges representing all the possible metabolic reactions in organism A referenced to pathway G. So G_A and G_B represent the information of pathway G in organism A and B, respectively. The value of denominator here is the number of relations belongs to G_B, and, the value of similarity is the number of relations shared in both G_A and G_B. The percentage score (similarity over denominator) is returned at the end. Therefore, the critical point of this comparison is to figure out how to define the relation. Our current work indicates this relation can be appropriately determined by three parameters: length of reaction paths, whether to include those relations not existed in both organisms for similarity comparison, and type of comparisons either functional or topological. The last one is illustrated in Figure 1.

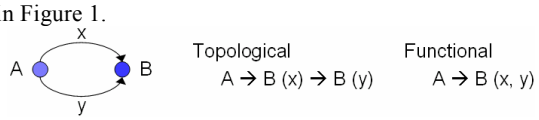

Figure 1: The difference between topological and functional comparison in the adjacency list. Where A and B represent the compounds (vertices), x and y represent the reactions (edges).

3 Results and Conclusion.

[1] Bioinformatics Program, Institute of Health Informatics and Decision Making, National Yang-Ming University, Taipei, Taiwan, 11221. E-mail: cyc@gel.ym.edu.tw

[2] Institute of Bioinformatics, National Yang-Ming University, Taipei, Taiwan, 11221. E-mail: cchang@ym.edu.tw

The metabolic pathway comparison has been implemented by Java with well-organized source code and comments, as well as JavaDoc API. A web-based query page and a Java applet are provided for displaying the comparison results. In the example shown below, we have compared the carbohydrate-related metabolic pathways of different Vibrio species. Our comparison results are consistent with the phylogeny analysis results currently available, i.e. the *Vibrio vulnificus* (VV) YJ016 is much similar to *Vibrio vulnificus* CMCP6, than with *Vibrio parahaemolyticus* (VP) and *Vibrio cholerae* (VC). This result is shown in Figure 2.

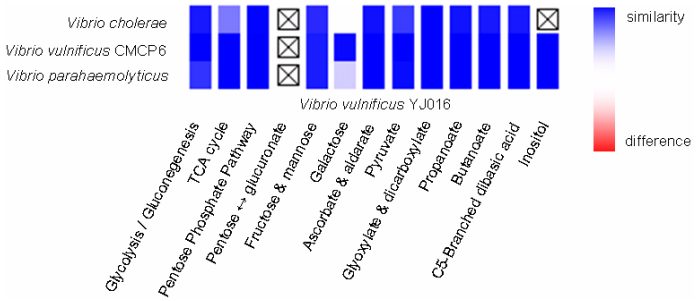

Figure 2: Comparisons of carbohydrate-related metabolic pathways in Vibrio species.

To know exactly what the differences are and why these Vibrio species are so different in the galactose metabolism pathway, the system will response with three pathway views. The first two can display those enzymes only specifically present in VC when referenced to those in VV YJ016, and vice versa. The last view shows the enzymes that are shared among these organisms. All the metabolic pathways of the target organism can also be compared with those of many other organisms as shown in the example of Figure 3. The output results can not only be presented along with the taxonomy information of these organisms, but also be sorted or clustered on both axes. Therefore, our CAMP system can be used to thoroughly compare metabolic pathways in many different ways with annotated genomic information.

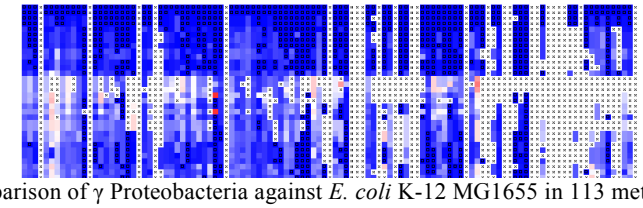

Figure 3: Comparison of γ Proteobacteria against *E. coli* K-12 MG1655 in 113 metabolic pathways.

4 Acknowledgement.

This work is supported by a grant (NSC 92-3112-B-010-020) from the National Science Council of the R.O.C.

5 References.

[1] Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Research, 2004, Vol. 32, Database issue D277-D280

[2] Hopcroft, J. E. and Tarjan, R. E. 1973. "Algorithm 447: Efficient Algorithms for Graph Manipulation," *Communications of the ACM*, 16, 372-378.