

Rahnuma: Hypergraph based tool for metabolic pathway prediction and network comparison

Aziz Mithani¹, Gail M. Preston^{2,*} and Jotun Hein¹

¹Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, UK

²Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

Associate Editor Prof. Thomos Lengauer

ABSTRACT

Summary: We present a tool called Rahnuma for prediction and analysis of metabolic pathways and comparison of metabolic networks. Rahnuma represents metabolic networks as hypergraphs and computes all possible pathways between two or more metabolites. It provides an intuitive way to answer biological questions focusing on differences between organisms or the evolution of different species by allowing pathway based metabolic network comparisons at an organism as well as at a phylogenetic level.

Availability: Rahnuma is available online at <http://portal.stats.ox.ac.uk:8080/rahnuma/>.

Contact: gail.preston@plants.ox.ac.uk

Supplementary Information: Supplementary information is available at the journal's website.

and as a result dependence between metabolites is lost. Rahnuma takes reaction direction into account while predicting pathways between metabolites by using directed hyperedges.

Rahnuma enables users to perform pathway analysis and comparisons for single organisms or groups of organisms, and in the context of a user-defined phylogeny. Rahnuma also allows identification of differences between networks by comparing predicted pathways between two or more metabolites, in contrast to Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2006), which only highlights differences in the traditional pathway maps in terms of presence and absence of reactions. Besides this, Rahnuma provides a simple web interface to perform various analyses, allowing a wide range of users to submit jobs tailored to their specific needs.

1 INTRODUCTION

Recent advances in technology and increasing use of computational techniques to study biological systems have led to the development of a number of computational tools to compute pathways between metabolites and to compare different metabolic networks. Some of the commonly used tools include PathComp (Kanehisa *et al.*, 2006), Pathway Tools (Karp *et al.*, 2002), Pathway Analyst (Pireddu *et al.*, November 2005) and UM-BBD Pathway Prediction System (Ellis *et al.*, 2006). Currently available tools, however, do not allow researchers to study pathway specific differences between multiple organisms or to study the evolution of different species by performing pathway analysis in a phylogenetic context.

Here we describe a tool called *Rahnuma* for prediction and analysis of metabolic pathways and comparison of metabolic networks. The name Rahnuma comes from the Urdu language and literally means "someone who guides through the path". Rahnuma represents metabolic networks as hypergraphs, rather than the commonly used graph representation. A hypergraph is a generalization of an ordinary graph where an edge, called a *hyperedge*, can connect more than two vertices (Yeung *et al.*, 2007). The vertices in the hypergraph are the compounds and the hyperedges are the reactions connecting the compounds. Since a reaction is treated as a single entity in a hypergraph, it can be used to capture relationships between any number of metabolites involved in a reaction unlike ordinary graphs where each edge is independent

2 PATHWAY PREDICTION IN RAHNUMA

Rahnuma computes pathways between individual metabolites or a group of metabolites using depth first traversal of the hypergraph representing the metabolic network. A *pathway* is said to exist between any two metabolites if there is a connected sequence of distinct reactions (or hyperedges) between the two metabolites such that the product of one reaction acts as a substrate in the next reaction. A valid pathway also requires each compound to be present only once in the pathway. Moreover, if a metabolite is acting as a substrate in any of the reactions already present in the pathway then it cannot be used as an intermediary point in the pathway. Conditions like this can easily be handled using hypergraph representation as it treats reactions as complete entities, unlike ordinary graphs where all the connections are independent. Further details including specific algorithmic details and a worked example of pathway prediction are given in the Supplementary Material (Section S1). Rahnuma is able to predict pathways between metabolites that are not identified by PathComp, a graph-based tool available through KEGG (Supplementary Material Section S2.1).

3 DESCRIPTION OF THE TOOL

Rahnuma is written in Java and uses a MySQL database to store data from KEGG. The current dataset includes all KEGG pathway maps relating to metabolism and includes manually curated carbon and nitrogen connections for 32 pathway maps relating to nutrient assimilation and energy metabolism (see Supplementary Table S1). Rahnuma translates KEGG data into hypergraphs allowing it to capture the relationship between multiple metabolites involved in

*to whom correspondence should be addressed

a reaction. The main reasons behind selecting KEGG as the data source for Rahnuma were the versatility of data present in KEGG and the fact that KEGG data is updated on a regular basis.

Rahnuma consists of three main modules, namely Network Analysis, Pathway Analysis and Comparative Analysis. These are described in the following sections. Rahnuma allows the user to choose between descriptive or tabular output format and to specify if either textual or HTML output is required. In the case of HTML output, the entities (organisms, metabolites, reactions, enzymes etc.) are hot-linked to respective entries in KEGG allowing the user to exploit KEGG's integrative capabilities with other databases.

Rahnuma provides users with the option to use all pathway maps, curated pathway maps, or to select a subset of pathway maps to perform analyses. The user can also specify a list of metabolites to be ignored, thereby allowing the user to exclude current metabolites such as O₂, H₂O and CO₂ from pathway predictions and neighborhood analyses.

3.1 Network Analysis

The Network Analysis module builds metabolic networks over a phylogeny and reports the reactions present at various levels of the phylogeny. Reaction equality across multiple networks is established using KEGG reaction Ids. The networks can be built in three modes:

1. **Union** Combines reactions from all the networks below the current level of the phylogeny.
2. **Intersection** Considers only those reactions that are present in all the networks present below the current level.
3. **Reaction Neighborhood** Considers reactions for which the proportion of neighbors (Yeung *et al.*, 2007) present in at least one of the networks is greater than the specified cut-off.

An example of ancestral network building using Rahnuma under different modes is shown in Supplementary Figure 1.

3.2 Pathway Analysis

The Pathway Analysis module allows the user to perform three types of analyses (further explained in Supplementary Section S1):

1. **Pathway prediction** Allows the user to perform simple pathway prediction between two or more metabolites.
2. **Reactions exclusive to a pathway** Identifies reactions that are involved in pathways from only one of the start metabolites.
3. **Reactions acting as bridges in a pathway** Identifies reactions that if deleted from the network will result in all pathways being removed between the specified metabolites.

Analyses can be performed in either organism or phylogeny mode using manually curated connections (where available), KEGG RPairs, or by pairing substrates and products involved in reactions. The organism mode requires the user to select a reference network built using KEGG's reference pathway, or to specify one or more organisms. The phylogeny mode asks the user to specify a phylogeny on which the selected analysis should be performed. When building a single network in organism mode or building networks over a phylogeny, Rahnuma provides flexibility over how networks should be built (Section 3.1). An example of pathways

formed by merging multiple networks is shown in Supplementary Figure 2 and the performance of the tool when multiple networks are combined is discussed in Supplementary Section S2.2.

When predicting pathways, Rahnuma allows the user to specify more than one start/end metabolite. This not only eliminates the need of submitting multiple jobs with same parameters but also allows the user to study a functionality such as the tricarboxylic acid (TCA) cycle or urea cycle as a whole. Details of different pathway prediction parameters are available on Rahnuma's website.

3.3 Comparative Analysis

The Comparative Analysis module allows the user to compare two metabolic networks. Two types of analyses are included:

1. **Standard comparison** Performs comparative analysis over a phylogeny or between two groups of organisms.
2. **All but one comparison** Identifies pathways or reactions that are present in only one organism in a group but absent in all other organisms and vice versa.

For both comparisons, the user can select between two comparison modes (i) full network and (ii) pathway dependent. Full network comparison takes into account all reactions present in the networks whereas pathway dependent mode builds upon the pathway analysis module described in Section 3.2 and performs comparisons based on reactions involved in specified pathways. An example demonstrating the comparative analysis feature in Rahnuma is given in the Supplementary Material (Section S3).

4 CONCLUSION

Rahnuma is a versatile tool that can be used for many different applications including predicting catabolic or biosynthetic pathways, identifying knock-out targets that may compromise pathway function, and comparing pathways and reactions at different levels of a phylogeny. It can also be used to compare metabolic networks in distantly related, but metabolically linked organisms such as hosts and symbionts. The unique features of this tool, which allow the users to answer specific biological questions, distinguish it from currently available tools and are likely to prove useful to both experimental biologists and bioinformaticians, who will be able to use Rahnuma to test and develop hypotheses and to explore and interpret experimental results.

ACKNOWLEDGMENTS

This work was supported by the grants from the Higher Education Commission, Government of Pakistan (A.M.), The Royal Society (G.P.) and the Biotechnology and Biological Sciences Research Council grant BB/E007872/1 (G.P.).

REFERENCES

- Ellis,L.B.M. *et al.* (2006) The University of Minnesota Biocatalysis/Biodegradation Database: The First Decade, *Nucleic Acids Res.*, **34**, D517–D521.
- Kanehisa,M. *et al.* (2006) From genomics to chemical genomics: new developments in KEGG, *Nucleic Acids Res.*, **34**, D354–D357.
- Karp,P. *et al.* (2002) The Pathway Tools Software, *Bioinformatics*, **18**, S225–232.
- Pireddu,L. *et al.* (2005) Pathway Analyst - Automated Metabolic Pathway Prediction, *Proceedings of the IEEE Symposium CIBCB*.
- Yeung,M. *et al.* (2007) Estimation of the number of extreme pathways for metabolic networks., *BMC Bioinformatics*, **8**, 363.