Rahnuma: Hypergraph based Tool for Analysis and Comparison of Metabolic Pathways

Aziz Mithani, Arantza Rico†, Rachel Jones†, Gail Preston† and Jotun Hein
Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK; Email: mithani@stats.ox.ac.uk
†Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

Introduction

Metabolic networks such as one shown in Figure **??** are generally represented by directed graphs where nodes represent metabolites and an edge corresponds to the presence of a reaction between any two metabolites. There are, however, various problems associated with this representation.

- A reaction may have more than one substrate which react to give one or more products, a scenario which cannot be captured by ordinary graphs.
- By treating each metabolite separately in a reaction dependence between the metabolites is lost.
- There may be more than one reaction involving any two metabolites. Ordinary graphs are not ideal for representing multiple connections between metabolites.

A better solution is to use hypergraphs, where an edge can link more than two vertices. This allows enumeration of multiple reactions connecting metabolites and pathway prediction between metabolites or groups of metabolites.

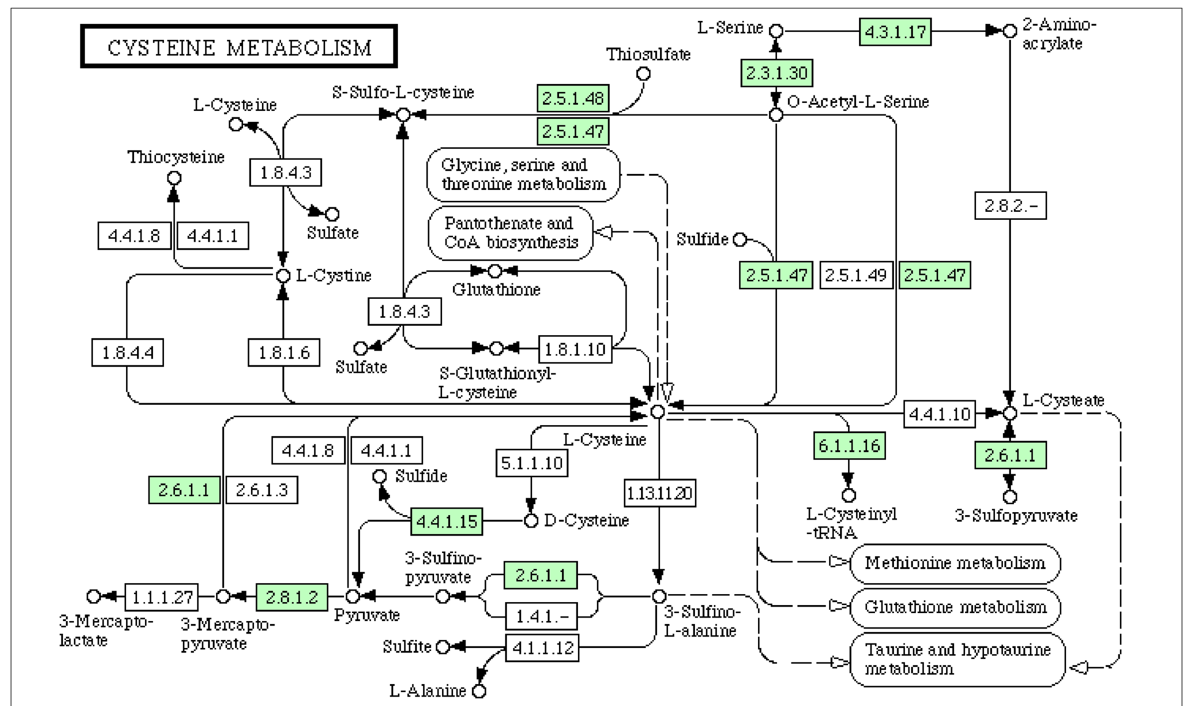

Figure 1: A standard representation of the cysteine metabolism network in *P. fluorescens* PfO-1 from KEGG[1].

Hypergraph

Hypergraphs are a generalisation of ordinary graphs where an edge (reaction) may connect any number of nodes or vertices (metabolites) as shown in Figure **??**. This makes them ideal for representing the relationships between multiple metabolites in a reaction.

> A hypergraph is a pair $\mathscr{H} = (V, E)$ where $V = \{V_1, V_2, \ldots, V_n\}$ is the set of vertices and $E = \{E_1, E_2, \ldots, E_m\}$ is the set of non-empty subsets of V called hyperedges such that $E_i = V_i^S \cup V_i^P$ where $V_i^S, V_i^P \subseteq V$ are the sets of substrates and products respectively involved in the reaction corresponding to edge E_i.

Hypergraphs provide an intuitive approach to finding pathways between two metabolites. A metabolite is reachable from another metabolite if there is a sequence of distinct reactions (or hyperedges) connecting the metabolites.

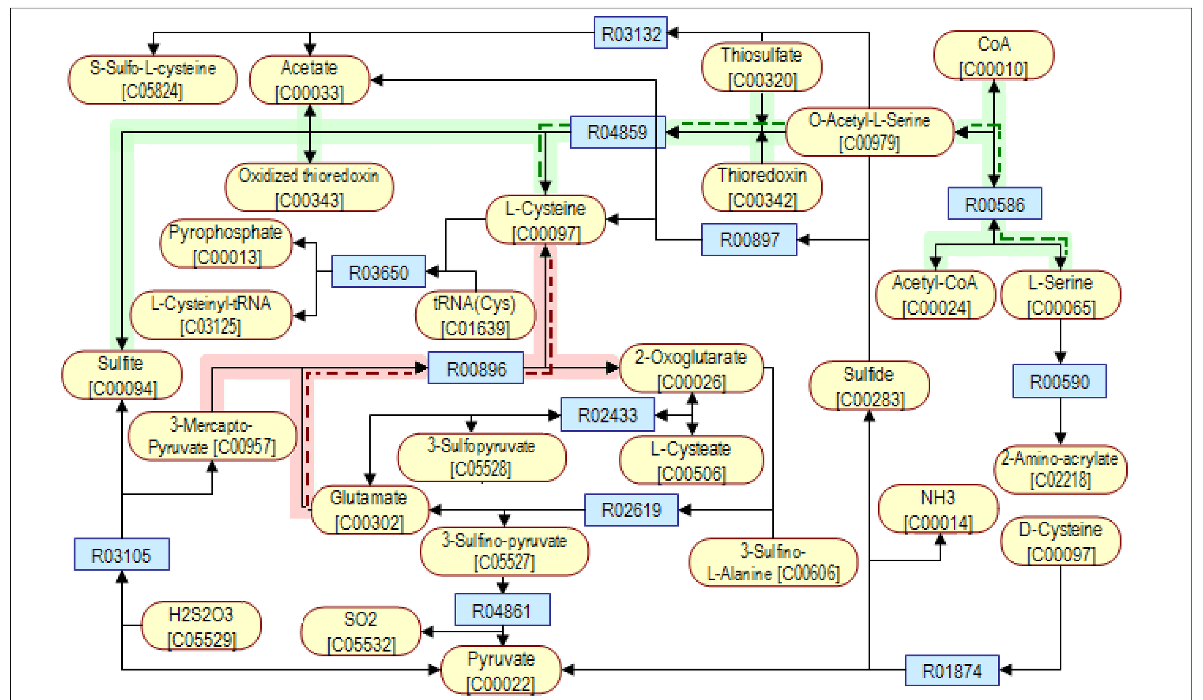

Figure 2: Hypergraph representation of network shown in Figure **??** with two possible cysteine synthesis pathways marked. Compounds such as ATP, AMP, H_2O not shown.

Methodology

Rahnuma is a pathway analysis and comparison tool that represents metabolic networks as hypergraphs.

Pathway Analysis Pathways are computed by performing a depth-first traversal of the hypergraph in either of the following two modes.

Mode	Description
Connection based	• Requires specification of a file containing connections between substrates and products of each reaction • At each step of the pathway calculation, all the reactions in which the metabolite being considered is a substrate are processed • For each reaction all outgoing connections from the metabolite are considered
Reaction based	• Explores all possible connections between substrates and products of a reaction • Allows specification of two optional files containing (i) list of metabolites to be avoided, e.g. ATP, AMP etc. (ii) element(s) being tracked, e.g. carbon or nitrogen. • Connections to compounds in the avoid list or not having specified element(s) are ignored

In addition to the maximum length of the pathway to be reported, parameters may be set to include following functionalities.

Type	Annotation	*in silico* Experiment
P	Predicted Reaction	Gene Insertion
KO	Invalid Reaction	Knock-out

Network Comparison Pathway computation can be extended to compare two networks in order to identify pathways or reactions that are missing in one or the other network. Network comparison can be performed in following two modes.

Mode	Description
Full	All reactions present in the network are considered
Pathway-based	Only Reactions involved in the pathways from a metabolite to one or more metabolites are considered

Results

Pathway Analysis Minimum predicted pathway lengths to ammonia from selected amino acids are shown below along with the experimental results[2].

	P. fluorescens PfO-1		*P. syringae* DC3000	
	Min Length	Biolog	Min Length	Biolog
Alanine	3	+	4	−/w
Arginine	1	+	2	+
Cysteine	4	−/w	5	−/w
Glutamine	1	+	1	+
Phenylalanine	1	+	3	w
Proline	3	+	4	+
Serine	1	+	1	+
Tryptophan*	2	−/w	2	−/w

*The tryptophan assimilation pathway involves reaction `R02722`, a tryptophan synthesis reaction, catalysed by the enzyme tryptophan synthase `(EC: 4.2.1.20)`. However, due to reversible nature of the reaction it may proceed in the reverse direction.

The general result is summarised below.

> Amino acids that are rapidly assimilated as carbon or nitrogen sources are predicted to have short pathways between the amino acid and the TCA cycle or ammonia.

Network Comparison Results are summarised in Figure **??** which compares the number of reactions in different *Pseudomonas* strains.

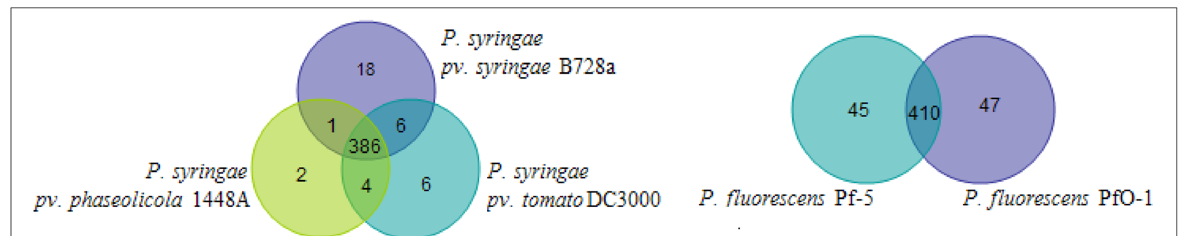

Figure 3: Number of reactions in different strains of *P. syringae* (left) and *P. fluorescens* (right) based on full network comparison.

Case Studies

Case Study 1: Predicting Amino Acid Assimilation Pathways

Observation: Alanine is rapidly assimilated as a carbon and nitrogen source in *P. fluorescens* PfO-1.

Hypothesis: Alanine has short pathways to the TCA cycle and ammonia.

Analysis: We first used KEGG's PathComp tool to calculate pathways from alanine to the TCA cycle and ammonia. Pathways were calculated separately for each metabolite in the TCA cycle. Results are shown below where n denotes the pathway length.

	TCA cycle			Ammonia		
	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$
P. fluorescens PfO-1	✗	✗	✗	✗	✗	✗
Reference Network	✗	✓	✓	✗	✓	✓

We then used Rahnuma to compute the pathways using its inherent functionality of specifying multiple destinations. Results are given below with a subset of reactions shown in Figure **??**. None of these reactions were picked by the PathComp tool.

	TCA cycle			Ammonia		
	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$
P. fluorescens PfO-1	✗	✓	✓	✗	✗	✓
Reference Network	✗	✓	✓	✓	✓	✓

Pathway predictions based on the KEGG reference network suggest that *P. fluorescens* PfO-1 could have an unidentified aminotransferase corresponding to reaction `R00258` which would lead to a shorter pathway of length 2 from alanine to ammonia (Figure **??**).

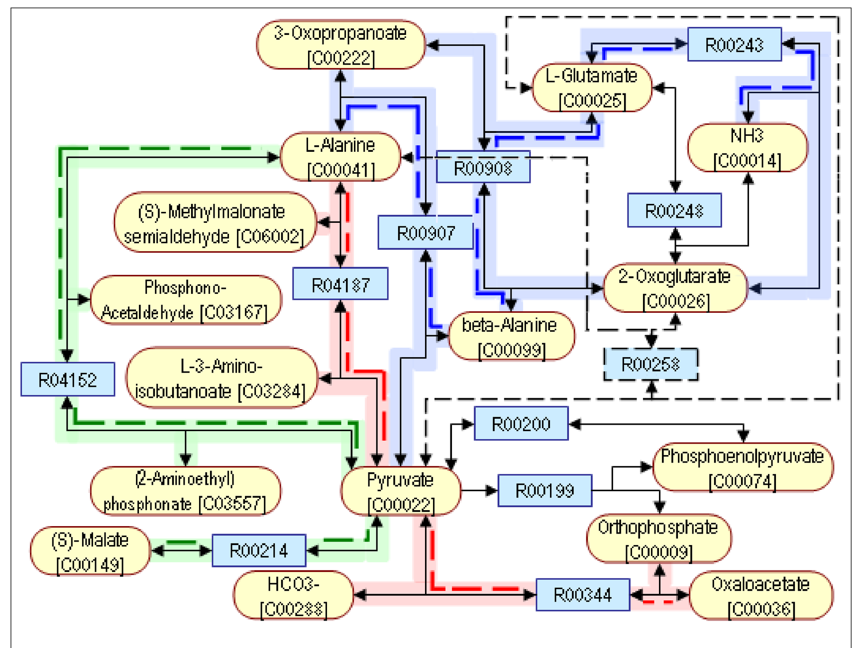

Figure 4: A subset of reactions reported by Rahnuma to be possibly involved in alanine assimilation pathways in *P. fluorescens* PfO-1. Reactions involved in two possible pathways to the TCA cycle (red and green) and one pathway to ammonia (blue) are highlighted with possible routes for carbon and nitrogen flow marked as dashed lines. Also shown in dashed line (black) is a predicted reaction corresponding to an aminotransferase which might be present in nitrogen assimilation pathway.

Case Study 2: Pathway-Based Network Comparison

Task: Compare the metabolic networks of *P. fluorescens* PfO-1 and *P. syringae* pv. tomato DC3000 to identify reactions involved in amino acid assimilation that are missing in one of the two networks.

Method: We performed pathway-based comparisons of metabolic networks of *P. fluorescens* PfO-1 and *P. syringae* pv. tomato DC3000 allowing pathways with length up to 6 reactions. Longer pathways are biologically unlikely based on current knowledge of amino acid assimilation in bacteria and include reactions that may not be involved in the assimilation pathways *in vivo*.

Results: The results for network comparison are summarised in Figure **??**.

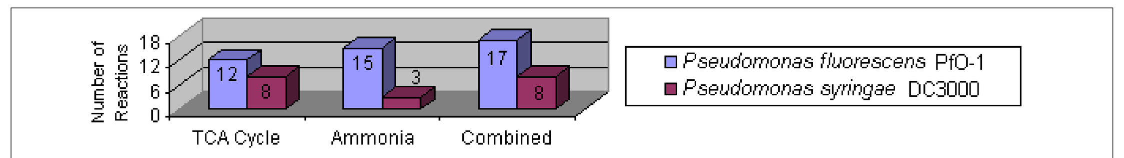

Figure 5: Number of reactions present in one but missing in the other network.

Example: Reactions present in *P. syringae* pv. tomato DC3000 but missing (either absent or unidentified) in *P. fluorescens* PfO-1 that are present in pathways between amino acids and ammonia are listed below.

Rxn Id	Reaction	Amino Acids
R00565	L-Ornithine + Guanidinoacetate <=> L-Arginine + Glycine	L-Arginine, L-Aspartate, L-Glutamate L-Glutamine, L-Isoleucine, L-Leucine L-Phenylalanine, L-Tyrosine, L-Valine γ-Aminobutanoate
R01731	L-Aspartate + Prephenate <=> Oxaloacetate + L-Arogenate	L-Asparagine, L-Aspartate
R01989	L-Ornithine + γ-Guanidinobutanoate <=> L-Arginine + γ-Aminobutanoate	(same as R00565)

Network comparison can be extended to study a wide range of scenarios:

- genome-'X'ome (X: metabolome, transcriptome, proteome) comparisons
- predicting knock-out/insertion effects
- understanding host-pathogen interactions
- investigating toxic effects

Case Study 3: Rahnuma vs. KEGG's PathComp

Objective: Compare the performance of Rahnuma with that of KEGG's PathComp tool.

Analysis Besides allowing network comparison, Rahnuma provides edge over PathComp in many areas. Some of these are mentioned below.

Pathway Prediction As shown in Case Study 1, PathComp fails to identify alanine assimilation pathways which are picked by Rahnuma. Some other cases where PathComp fails to identify pathways such as arginine, lysine, methionine and serine can be seen in Figure **??**.

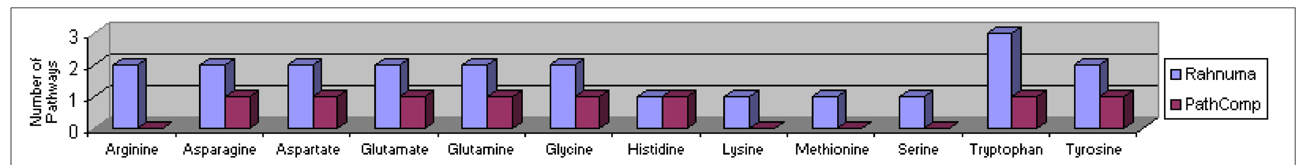

Figure 6: Number of assimilation pathways of unit length to ammonia in KEGG's reference network reported by PathComp and Rahnuma for selected amino acids.

Multiple Pathways Rahnuma explores all reactions between the metabolites and, therefore, reports all possible pathways from a metabolite to another metabolite or group of metabolites (Figure **??**). PathComp, on the other hand, works only on binary relationship between the metabolites. For example, Rahnuma reports the following nitrogen assimilation pathways of unit length from glycine.

```
1.  Glycine <R00364> Ammonia  (Enzyme:  Glycine:ferricytochrome-c oxidoreductase)
2.  Glycine <R00366> Ammonia  (Enzyme:  Aminoacetic acid:oxygen oxidoreductase)
```

Invalid Pathways PathComp, unlike Rahnuma, treats all reactions as reversible and, as a result, also reports pathways that are biologically incorrect. An example of invalid pathway reported by PathComp is the following.

```
γ-Aminobutanoate (GABA) <R00261> L-Glutamate <R00248> Ammonia
```

`R00261` is an irreversible reaction that only breaks glutamate into GABA and CO_2.

References:

1. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. *From genomics to chemical genomics: new developments in KEGG.* Nucleic Acids Res. 34, D354-357 (2006).
2. Rico, A., and Preston, G. M. Apoplast phenoarrays: An integrated approach to study nutritional specialization of Pseudomonas syringae in the plant apoplast (submitted)

Acknowledgements:

THE ROYAL SOCIETY
bbsrc biotechnology and biological sciences research council
Higher Education Commission, Government of Pakistan
St. Anne's College, University of Oxford