



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

1. Introduction

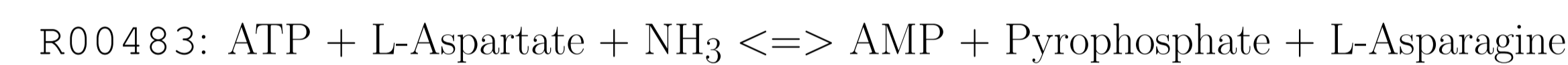
Metabolic networks are generally represented by directed graphs where nodes represent metabolites and an edge corresponds to the presence of a reaction between any two metabolites.

Limitations of directed graphs

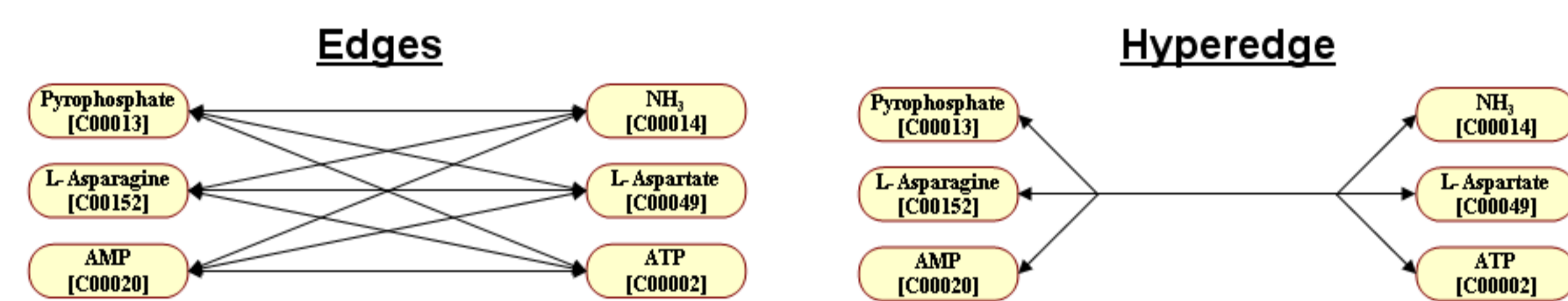
- Cannot capture relationship between more than two metabolites in a reaction
- Ignores dependence between the metabolites
- Not ideal for representing multiple connections (reactions) between metabolites
- Do not provide intuitive approach to analyse metabolic pathways or study evolution of metabolic networks

Hypergraphs

A better solution is to use hypergraphs. Hypergraphs are a generalisation of ordinary graphs where an edge can link more than two vertices. Such an edge is called a *hyperedge*. For example, consider the following reaction.

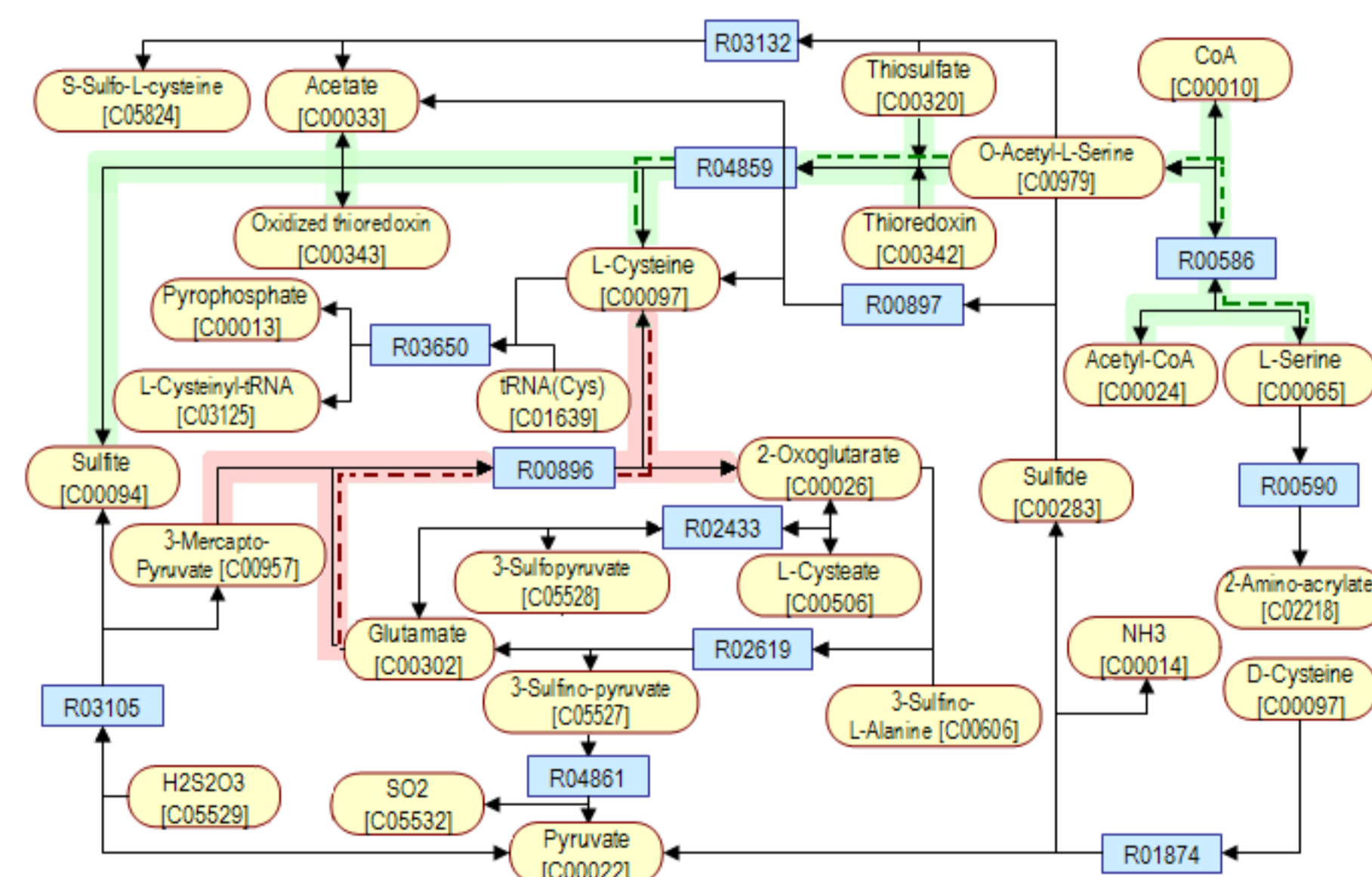


This reaction is shown below as a set of ordinary edges and a hyperedge. The hyperedge clearly captures the relationship between the multiple metabolites.



2. Hypergraph Representation of Metabolic Networks

The cysteine metabolism network from KEGG (Kanehisa et. al, 2006) in *P. fluorescens* PfO-1 is represented below as a hypergraph. Compounds such as ATP, AMP, H₂O not shown. Two possible pathways for cysteine biosynthesis are also marked with dashed lines.



Representing reactions as hyperedges makes it possible to consider them as single entities. This allows studying evolution in terms of loss or gain of reactions.

Hypergraphs also provide an intuitive approach to finding pathways between two metabolites

A pathway is said to exist between two metabolites if there is a sequence of distinct reactions (or hyperedges) connecting the metabolites.

3. Rahnuma: Tool for Metabolic Pathway Analysis and Network Comparison

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

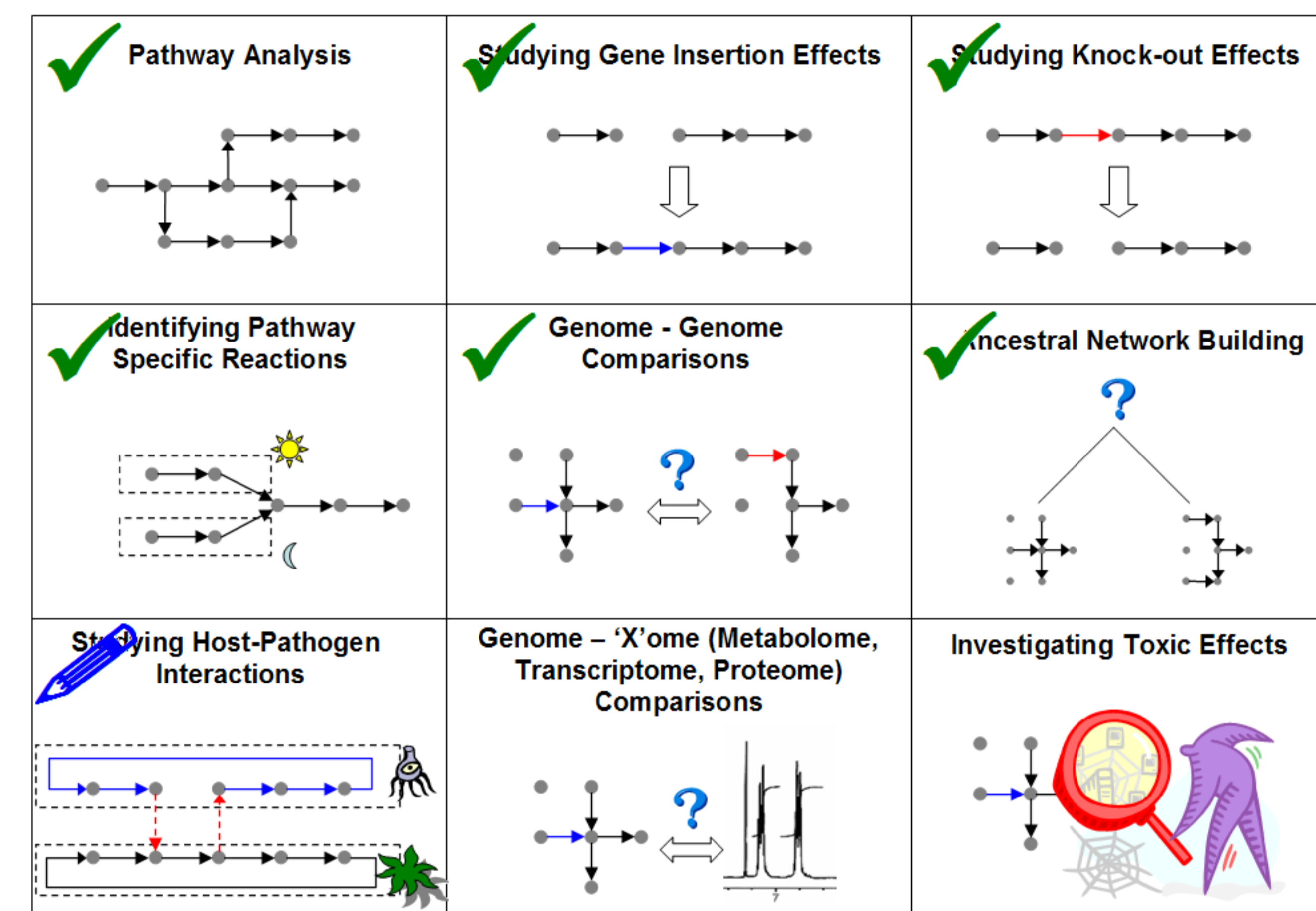
Rahnuma – someone who guides through the path (Urdu language)

Modules

1. Pathway prediction and analysis
2. Network comparison
3. Ancestral network building

Application Areas

Different studies involving metabolic networks are shown below. Rahnuma is capable of performing six of these analyses marked with green and seventh one is under development.

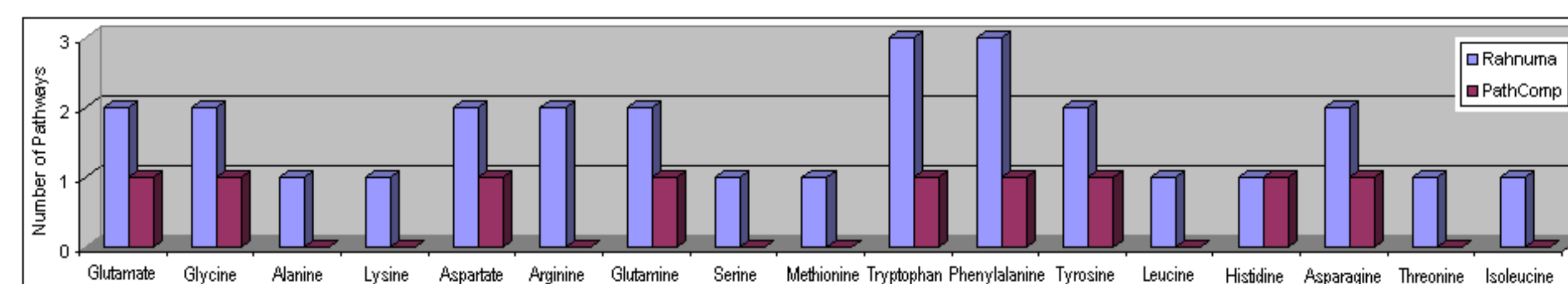


Implementation

- Dataset extracted from KEGG using custom scripts
- 1293 reactions involved in nutrient assimilation and related pathways manually annotated
- Written in Java with web interface in Java Server Pages
- Uses advanced data structures to efficiently compute the pathways between the metabolites or group of metabolites
- Explores all reactions between the metabolites during pathway prediction and, therefore, reports all possible pathways between metabolites

Performance

The number of assimilation pathways of unit length to ammonia in KEGG's reference network reported by KEGG's PathComp and Rahnuma for selected amino acids are shown below.



4. Case Study 1: Predicting Amino Acid Assimilation Pathways

Minimum predicted pathway lengths to ammonia from selected amino acids are shown below along with the experimental results (Rico and Preston, accepted).

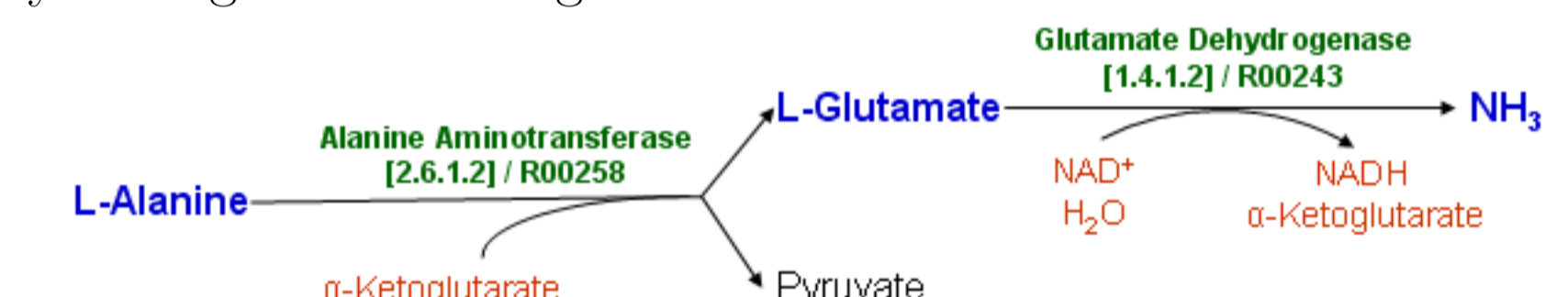
	Reference Network	<i>P. fluorescens</i> PfO-1		<i>P. syringae</i> DC3000	
	Min Length	Min Length	Biolog	Min Length	Biolog
Alanine	1	3	+	4	-/w
Arginine	1	1	+	2	+
Cysteine	2	4	-/w	5	-/w
Glutamine	1	1	+	1	+
Phenylalanine	1	1	+	3	w
Serine	1	1	+	1	+
Tryptophan*	1	2	-/w	2	-/w

*Tryptophan 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.

Predictive analysis based on KEGG's reference network suggests the presence of reaction R00258 in *P. fluorescens* PfO-1 corresponding to an as yet unidentified aminotransferase. This would add following pathway of length 2 for nitrogen assimilation.



5. Case Study 2: Studying Evolution of Metabolic Networks

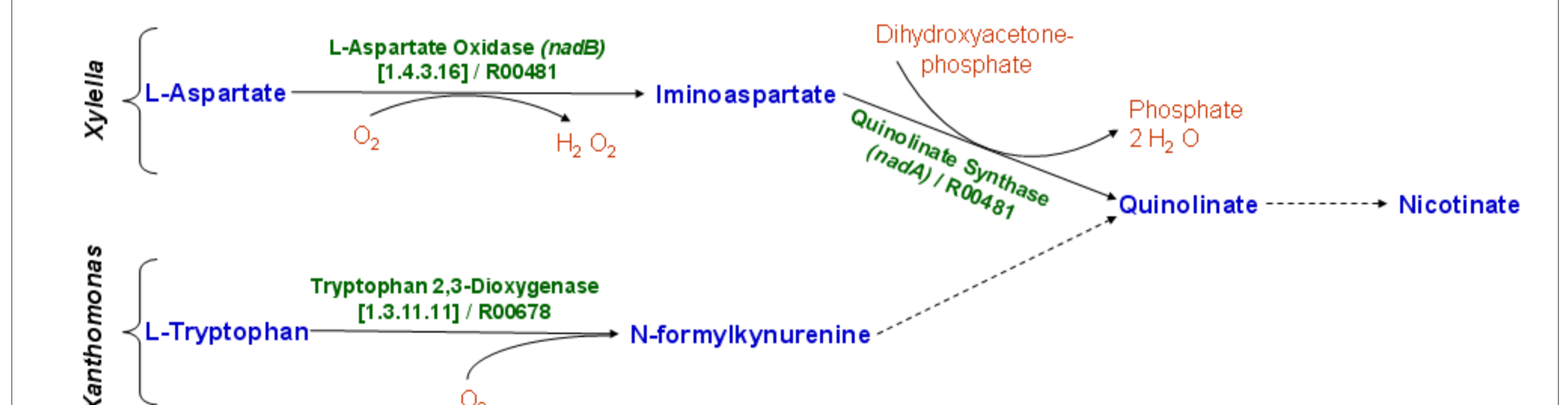
To study a common hypothesis that *Xylella* has undergone metabolic reduction during evolution from *Xanthomonas* to specialise for growth in plant vessels, we compared the metabolic networks of these organisms.

Full network comparison between the two sets of organisms highlights following differences.

Species	Extra Reactions
<i>Xylella</i>	2
<i>Xanthomonas</i>	139

The two reactions absent from *Xanthomonas* but present in *Xylella* are R00357 and R01257 corresponding to enzymes *nadB* and *yoyH* respectively.

Example: Analysis of *nadB* gene



Observation: The *nadB* gene is present and conserved in many γ -proteobacteria, however the *nadB* gene of *Xylella* is most similar to *nadB* from α -proteobacteria *Rizobium*.

Possible Explanation: The *nadB* gene may have been lost and re-gained in *Xylella* perhaps through an ancestral strain that contained both *nadB* genes at some point.

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 phenotypes: An integrated approach to study nutritional specialization of *Pseudomonas syringae* in the plant apoplast (accepted)

