Local piecewise alignment of metabolic pathways

M. Llabrés, M. Tudurí

Department of Mathematics and Computer Science
Computational Biology and Bioinformatics Research Group
University of the Balearic Islands, Spain

Abstract

Comparative analysis of metabolic networks in multiple species yields important information on their evolution and potential pharmacological targets. We present in this work a method to obtain a piecewise local alignment of metabolic networks modeled as Petri nets. We use the MPath2PN tool for the automatic translation of the metabolic data into Petri nets. From the Petri net model, a piece-pairwise alignment is performed based on the network structure similarity.

Metabolic Pathways and Petri Nets

Metabolic Pathway: network of chemical reactions, catalyzed by enzymes, where some molecules (reactants or substrates) are transformed into others (products). Enzymes are not consumed in the reactions. The product of a reaction is the substrate of the next one.
Petri Net: a graphical and mathematical modeling tool which consists of places, transitions, and arcs that connect them. Input arcs connect places with transitions (input places), while output arcs start at a transition and end at a place (output place). Places can contain tokens to describe the dynamics of the net.
Metabolic Pathways as Petri Nets:

- molecules $\equiv$ places
- chemical reactions $\equiv$ transitions
- reactants or substrates $\equiv$ input places
- products $\equiv$ output places
- number of tokens in each place $\equiv$ amount of substance associated with that place

Alignment algorithm overview

Let N be a Petri net (PN) modeling a metabolic pathway. Let R be its set of transitions.

- We consider all possible sequences of transitions $R_1 R_2 ... R_k$ such that the product of transition R_i is the substrate of transition R_{i+1}, and we consider the enzyme-paths that are $w = w_1 w_2 ... w_k$, where each w_i is the associated enzyme of R_i.
- For every pair of Petri nets N_1, N_2, we compute the similarity score of the sequence alignment of every pair of enzyme-paths (w^1, w^2) with w^1 an enzyme-path in N_1 and w^2 an enzyme-path in N_2.
- We consider then the set of all pairwise scores:

$$\Sigma = \{\sigma(w_i^1, w_j^2) \text{ for every pair } (w_i^1, w_j^2) \in N_1 \times N_2\}$$

and we construct the maximum weighted bipartite graph from Σ. From this maximum weighted bipartite graph we obtain a piece-pairwise alignment of the two Petri nets N_1 and N_2.

Methods

All Petri nets modeling metabolic pathways are obtained from the KEGG database (`http://www.genome.jp/kegg/`) using the MPath2PN translator tool.

Enzyme-path calculation: for every Petri net, the set W of subpaths from N is obtained using a backtracking algorithm that operates on the set of input places to obtain the longest possible transition sequence in the following sense: every pair of obtained transition sequences are compared and when one sequence is included into another, only the longest one is considered.

Enzyme-path similarity score: we first calculate an enzyme similarity score matrix by considering the hierarchical enzyme similarity, that is, the number of common most significant EC digits of the enzymes over 4. The values of hierarchical similarity are 0, when the enzymes have their first digit different; 0.25 when the first digit is identical and the second digit is different; 0.5 when the first two digits are identical and the third digit is different; 0.75 when the first three digits are identical and the last one is different; and 1 for two identical EC numbers.

Methods

Enzyme-path similarity score: then, a BioPython implementation of the Smith-Waterman algorithm based on the enzyme similarity score matrix calculated before is used to obtain the pairwise sequence alignment of every pair of enzyme-paths $(w_i^1, w_j^2) \in N_1 \times N_2$, obtained in the first step.

Piece-pairwise local alignment: from the set of pairwise enzyme-paths alignment score, a maximum weighted matching algorithm is considered to obtain the maximum weighted bipartite graph. Such a bipartite graph matches an enzyme-path in N_1 to its most similar enzyme-path in N_2 and thus, we obtain a piece-pairwise alignment of N_1 and N_2. Moreover, from this maximum weighted bipartite graph we define a global score which is the sum of the scores of the aligned enzyme-paths over the maximum number of enzyme-paths in N_1 and N_2

$$S(N_1, N_2) = \frac{\sum \sigma(w_i^1, w_j^2)}{max\{|W_1|, |W_2|\}}$$

where (w_i^1, w_j^2) belongs to the bipartite graph and $|W_k|$ is the cardinal number of the set of subpaths in N_k for $k = 1, 2$.

Experimental results & Discussion

To test the effectiveness of our alignment method we consider the Glycolysis pathway in the KEGG database in organisms from the three domains of life: Eukaryote, Bacteria and Archaea.

- Eukaryote: Homo sapiens (hsa), Drosophila melanogaster (dme), Mus musculus (mmu) and Saccharomyces cerevisiae (sce).
- Bacteria: Escherichia coli K-12 MG1655 (eco), Helicobacter pylori 26695 (hpy), Salmonella enterica subsp. enterica serovar Typhi CT18 (sty) and Streptococcus pyogenes SF370 (serotype M1) (spy).
- Archaea: Aeropyrum pernix (ape), Thermococcus kodakaraensis (tko) and Pyrobaculum aerophilum (pai).

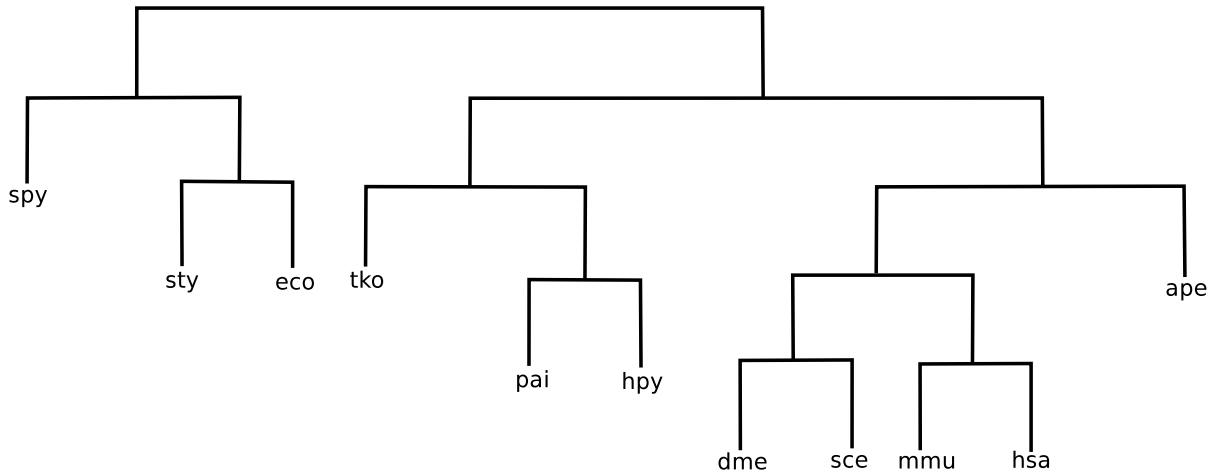
We perform the piece-pairwise alignment explained before between all pairs of the selected glycolysis pathways and we obtain the final similarity scores shown in the following table.

Experimental results & Discussion

	ape	eco	hsa	mmu	sce	sty	dme	hpy	pai	spy	tko
ape	1	0.269	0.685	0.685	0.664	0.269	0.639	0.517	0.40	0.272	0.488
eco	-	1	0.293	0.293	0.253	1	0.252	0.133	0.099	0.899	0.141
hsa	-	-	1	1	0.825	0.293	0.857	0.329	0.250	0.274	0.466
mmu	-	-	-	1	0.825	0.293	0.857	0.329	0.250	0.274	0.466
sce	-	-	-	-	1	0.253	0.926	0.358	0.273	0.237	0.519
sty	-	-	-	-	-	1	0.252	0.133	0.099	0.899	0.141
dme	-	-	-	-	-	-	1	0.347	0.270	0.236	0.511
hpy	-	-	-	-	-	-	-	1	0.727	0.151	0.598
pai	-	-	-	-	-	-	-	-	1	0.114	0.527
spy	-	-	-	-	-	-	-	-	-	1	0.134
tko	-	-	-	-	-	-	-	-		-	1

In order to visualize the obtained results, we also calculate the phylogenetic tree below using average-link clustering.

Experimental results & Discussion



As we can see here, the Eukaryote organisms are clustered together, and we can even appreciate that the Mammals organisms (hsa, mmu) are in the same cluster. There is also another cluster with three of the four Bacteria organisms but, the Archaea organisms are not included in the same cluster.