MetNetAligner: a web service tool for metabolic network alignments

Qiong Cheng, Robert Harrison, Alexander Zelikovsky *

Department of Computer Science, Georgia State University, Atlanta, GA 30303

Received on December 30, 2008; revised on March 10, 2009; accepted on April 20, 2009

Associate Editor: Limsoon Wong

ABSTRACT

Summary: The accumulation of high-throughput genomic, proteomic, and metabolical data allows for increasingly accurate modeling and reconstruction of metabolic networks. Alignment of the reconstructed networks can help to catch model inconsistencies and infer missing elements. In this note we present the web service tool MetNetAligner which aligns metabolic networks, taking in account the similarity of network topology and the enzymes’ functions. It can be used for predicting unknown pathways, comparing and finding conserved patterns, and resolving ambiguous identification of enzymes. The tool supports several alignment options including allowing or forbidding enzyme deletion and insertion. It is based on a novel scoring scheme which measures enzyme-to-enzyme functional similarity and a fast algorithm which efficiently finds optimal mappings from a directed graph with restricted cyclic structure to an arbitrary directed graph.

Availability: MetNetAligner is available as web-server at: http://alla.cs.gsu.edu:8080/MinePW/pages/gmapping/GMMain.html. Contact: alexz@cs.gsu.edu; csczqcox@cs.gsu.edu;

INTRODUCTION

Metabolism is a vital cellular process whose understanding is critical to human disease studies and drug discovery. The accumulation of high-throughput genomic, proteomic and metabolical data allows for increasingly accurate modeling and reconstruction of metabolic networks. Comparison among the reconstructed networks can catch model inconsistencies and infer missing elements. With the growth of identified metabolic networks, computational tools are necessary for the comparison. Network alignment is convenient for comparing and exploring metabolic networks – it can be used for predicting unknown or alternative pathways and pathway holes as well as resolving ambiguities and finding inconsistencies in existing pathway descriptions.

A metabolic pathway/network can be represented as a directed graph which vertices are enzymes. Each of its edges connects enzymes catalyzing consecutive reactions. Pinter et al. (2005) formulated the network alignment as a labeled subtree homeomorphism problem – given a vertex labeled pattern tree \( P \) (representing an unknown pathway) and text graph \( T \) (representing a known pathway), find the minimum cost transformation of \( P \) into subtrees of \( T \) by edge subdividing with degree-two vertices. Pinter et al. (2005) gave an efficient algorithm for the case when \( T \) is a tree. Yang et al. (2007) allowed to delete pattern nodes and gave an efficient algorithm for path matching and an exponential algorithm for arbitrary pattern and text graphs.

The MetNetAligner web service tool relies on a fast dynamic programming based algorithm for metabolic network alignment. The algorithm finds an optimal alignment between arbitrary pattern and text graphs allowing enzyme deletion and insertion as well as matching similar enzymes (see Cheng et al. (2008) for details). The algorithm is efficient for arbitrary text graphs and pattern graphs with a restricted cyclic structure. Note that mimicking evolutionary machinery of gene duplication (Sharan et al. (2006)), similarly to Yang et al. (2007) our algorithm allows to map different pattern enzymes into the same text enzyme.

MetNetAligner provides simple and intuitive web-interfaces and several services such as pathway retrieval, visualization and upload services. Below, we briefly describe our alignment algorithm, present the web service tool, and give usage examples. The algorithm validation and discussion are in supplementary materials.

METABOLIC NETWORK ALIGNMENT

We distinguish two kinds of vertex deletions: (i) bypass deletion corresponding to the replacement of a few consecutively acting enzymes with a single multifunctional enzyme or enzyme using an alternative catalysis and (ii) strong deletion symbolizing the matching of a proper connected subgraph of the pattern network. Thus, a bypass deletion of a vertex \( v \) with a single incoming and a single outgoing edge is the replacing of a \( u-v-w \)-path with a single \( (u,w) \)-edge formally represented as mapping \( v \) into \( b \). A strong deletion of a vertex \( v \) is removing of arbitrary vertex \( v \) together with all its incoming and outgoing edges which is formally represented as mapping into \( d \).

Let graphs \( P = (V_P, E_P) \) and \( T = (V_T, E_T) \) represent a pattern and text metabolic networks, respectively. Let \( f : P \to T \) map every vertex in \( V_P \) to \( V_T \cup \{b, d\} \). A pair of pattern vertices \( u, v \in V_P \) is called contracted if \( u \) and \( v \) are mapped into two different text vertices and either \( (u,v) \in V_P \) or there exists a \( u-v \)-path in \( P \) whose all intermediate vertices are bypass deleted. Let \( E'_P \) be the set of all vertex pairs in \( P \) contracted by \( f \) (note that \( E_P \subseteq E'_P \)). The mapping \( f \) is called a network alignment if for each contracted pair \((u,v) \in E'_P\) there exists a \( f(u)-f(v) \)-path in the text \( T \). The number of insertions, i.e., the minimum number of intermediate vertices in such path is denoted \( e(f(u), f(v)) \).

The alignment should be penalized (i) for mismatches between aligned enzymes, (ii) for strong deletions, (iii) for bypass deletions,
and (iv) for insertions. Thus we obtain the following cost function.\[
\text{cost}(f) = \sum_{u \in V_P} \Delta(u, f(u)) + \lambda \sum_{(u,v) \in E_P^f} \sigma(f(u), f(v)),
\]
where $\Delta(u, f(u))$ is the penalty of mismatch between enzymes corresponding to pattern vertex $u$ and text vertex $f(u)$. $\Delta(u, d)$ and $\Delta(u, b)$ are penalties for strong and bypass deletions, respectively, and $\lambda$ is the penalty for a single enzyme insertion.

We proposed an enzyme-to-enzyme dissimilarity score $\Delta$ based on 4-level EC encoding $d_1$:$d_2$:$d_3$:$d_4$, where $d_4$ is the number of enzymes. The numeric values are usually positive integers and equal 0 (or -) only if the corresponding subclass is unknown. If $d_1$'s or $d_2$'s of two enzymes are different and non-zero, then $\Delta = 0$ (see Cheng et al. (2007)). Notice that if two enzymes are different in the corresponding position but one of them is 0, then we do not increase lack of information.

Given two metabolic networks $P$ and $T$, the problem is to find the optimal (minimum-cost) network alignment from $P$ to $T$. Our dynamic programming algorithm (Cheng et al. (2008)) aligns multi-source trees to arbitrary networks in time $O(|V_P|^2 |E_P| \log |V_T|)$. When the pattern has a limited cycle structure, i.e., it is sufficient to delete a small number of vertices to break all undirected cycles, then our algorithm can find the optimal alignment in polynomial time.

### THE WEB-SERVER

**Data Source.** The MetNetAligner web service is applicable to pathways from Bio-Cyc database (Karp et al. (2007)). The metabolic networks of five organisms (E. coli, S. cerevisiae, B. subtilis, Halobacterium sp. NRC-1, and T. thermophilius) are readily available for aligning. The files with metabolic pathways of other organisms can be obtained from Bio-Cyc and extracted using our supporting tools.

**Implementation and features.** MetNetAligner employs Tomcat/Apache as the application server and uses MySQL as the data storage server. Our tool is portable to several operating systems and compatible to database servers. A web browser with Java SE 6 installed is required for graph visualization.

The work with MetNetAligner is divided into three phases: first, a set of parameters must be specified that describes the alignment mode, parameters and scoring schemes. Second, a user chooses organisms and pathways for pattern and text from the existing database. Enzyme-enzyme model of pathway and alignment results will be graphically displayed (see Fig. 1). The graph layout is based on a force-directed method (Barnes et al. (1986)). A user can customize the visualization by dragging vertices with the mouse. Labels with a light gray background represent the pattern vertices and those with a white background represent text vertices. Third, for the batch modes, the best K alignment results are sorted by P-value or alignment score. The options, parameters and modes are specified as follows.

**Alignment options.**
- Allow/forbid enzyme deletion and insertion;
- EC-notation based dissimilarity score from Cheng et al. (2007); EC-common-upper-class based dissimilarity score from Tohsato et al. (2006).
- Models of random graph for P-value computation: edge reshuffling and vertex label reshuffling;

**Alignment parameters.**
- Deletion penalties for pattern or text vertices;
- Single enzyme insertion penalty $\lambda$

**Supported batch modes for pathway alignments.**
- Alignment of two pathways. The alignment result is visualized (see an example on Fig. 1). The visualization can be customized in a drag and drop manner and uploaded to the web tool.
- Alignments from a single pattern pathway to all pathways in the text organism. The best K alignments (K is user-specified) are displayed sorted according to P-value or alignment score.
- Alignments from every pathway in the pattern organism to every pathway in the text organism.

In conclusion, MetNetAligner is a web service tool for aligning and visualizing metabolic networks allowing that can be used for finding conserved patterns and resolving ambiguous enzymes.

**Acknowledgments.** Q.C. was supported by Georgia State University Molecular Basis of Disease Fellowship.

**REFERENCES**


