

Research article

Open Access

Algebraic comparison of metabolic networks, phylogenetic inference, and metabolic innovation

Christian V Forst¹, Christoph Flamm^{*2}, Ivo L Hofacker² and Peter F Stadler^{2,3,4}

Address: ¹Bioscience Division, Los Alamos National Laboratory, Mailstop M888, P.O. Box 1663, Los Alamos, NM 87545, USA, ²Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria, ³Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany and ⁴Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

Email: Christian V Forst - chris@lanl.gov; Christoph Flamm* - xtof@tbi.univie.ac.at; Ivo L Hofacker - ivo@tbi.univie.ac.at; Peter F Stadler - peter.stadler@bioinf.uni-leipzig.de

* Corresponding author

Published: 14 February 2006

Received: 01 August 2005

BMC Bioinformatics 2006, 7:67 doi:10.1186/1471-2105-7-67

Accepted: 14 February 2006

This article is available from: <http://www.biomedcentral.com/1471-2105/7/67>

© 2006 Forst et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Comparison of metabolic networks is typically performed based on the organisms' enzyme contents. This approach disregards functional replacements as well as orthologies that are misannotated. Direct comparison of the structure of metabolic networks can circumvent these problems.

Results: Metabolic networks are naturally represented as directed hypergraphs in such a way that metabolites are nodes and enzyme-catalyzed reactions form (hyper)edges. The familiar operations from set algebra (union, intersection, and difference) form a natural basis for both the pairwise comparison of networks and identification of distinct metabolic features of a set of algorithms. We report here on an implementation of this approach and its application to the prokaryotes.

Conclusion: We demonstrate that metabolic networks contain valuable phylogenetic information by comparing phylogenies obtained from network comparisons with 16S RNA phylogenies. The algebraic approach to metabolic networks is suitable to study metabolic innovations in two sets of organisms, free living microbes and *Pyrococci*, as well as obligate intracellular pathogens.

Background

The metabolic networks of a wide variety of organisms, in particular prokaryotes, have been reconstructed by means of a combination of genomic annotations with biochemical and physiological data, see e.g. [1]. These networks are compiled in databases, in particular in the KEGG resource [2].

Large scale bacterial phylogenies that are based on single genes are notoriously plagued by gene transfer, gene

duplication, gene deletion, and functional replacement of genes. The same holds for various approaches towards utilizing gene content for phylogenetic purposes, discussed e.g. by [3-8]. A recent article by [9] addressed this issue by considering the presence or absence of 64 individual sub-pathways that were identified based on the COG division [10] of the National Center for Biotechnology Information. A related approach, based on comparison of individual pathways was discussed by [11-13] and [14]. The pathways necessary for such approaches can be derived

from a given metabolic network either "by hand" or using automated procedures such as metabolic flux analysis, see e.g. [15-19].

Instead of attempting to first reconstruct individual pathways, we take here a more global view by grounding our analysis in the direct comparison of the metabolic networks. While the application of generic graph distances or similarity measures (see e.g. [20]) is certainly appealing, they cannot be used in a straightforward manner for metabolic networks. The reason is that chemical reaction networks do not have a simple representation as graphs, at least not when metabolites are represented as nodes and reactions as edges. Instead, a metabolic network is naturally described by a directed hypergraph [21], or, equivalently, by a directed bipartite graph, in which metabolites and reactions (or, equivalently, the enzymes that catalyze the reactions), are represented by two different types of vertices. More global comparison of metabolic networks, in terms of various network indices and networks motifs, can be found in [22].

This contribution is organized as follows: In the next section we summarize an algebraic approach to comparison and manipulation of chemical reaction networks that is motivated by set theory. We briefly describe the C library that implements this approach. We then demonstrate that the symmetric difference of two metabolic networks can be used to derive a distance measure that is suitable for reconstructing phylogenetic relationship from metabolic network data. More interestingly, however, the same approach can be used directly to extract those subnetworks of the metabolism that are innovations in the particular subtree of the phylogeny. We illustrate our approach using pathogenic procaroyotes as an example.

The algebra of directed hypergraphs

A metabolic network is defined by its metabolites and the system of reactions that inter-converts them. We denote the set of metabolites by X . A chemical reaction E can be described as a pair of multisets (E^-, E^+) , where $E^- \subseteq X$ is the set of educts in the reaction and $E^+ \subseteq X$ is the set of reaction products. Slightly more generally, we can replace the multisets by ordinary sets and instead define the multiplicities of product and educt metabolites by means of the stoichiometric coefficients $n_{x,E}^+$ and $n_{x,E}^-$ of the products and educts, respectively. A metabolic network is thus a pair (X, \mathcal{E}) where \mathcal{E} is a set of reactions. Such a structure is known as a directed hypergraph $\mathfrak{M}(X, \mathcal{E})$, see e.g. [21]. The *stoichiometric matrix* S of the network has the entries

$$S_{xE} = n_{x,E}^+ - n_{x,E}^- \quad (1)$$

where $x \in X$ is a metabolite, and E is a reaction. For completeness, we remark that the set $E^c = E^+ \cap E^-$ are the catalysts of the reaction E . Furthermore, a reaction is autocatalytic if $n_{x,E}^+ - n_{x,E}^- \neq 0$ for some $x \in E^c$. By abuse of notation we write $E = E^+ \cup E^-$ for the set of metabolites involved in the reaction E . Furthermore, we write $\text{supp } \mathcal{E} = \cup \{E | E \in \mathcal{E}\}$ for the set metabolites that actually take part in the reactions. We call a network $\mathfrak{M}(X, \mathcal{E})$ *clean* if $X = \text{supp } \mathcal{E}$ and define the *clean up operator* as $\lfloor \mathfrak{M} \rfloor = (\text{supp } \mathcal{E}, \mathcal{E})$. Furthermore, for a given set \mathcal{E} of reactions and set A metabolites we define

$$\mathcal{E}[A] = \{E \in \mathcal{E} \mid (E^+ \cup E^-) \subseteq A\} \quad (2)$$

The restriction of a network $\mathfrak{M}(X, \mathcal{E})$ to a set A of metabolites is defined as the clean network

$$\mathfrak{M}[A] = \lfloor (A, \mathcal{E}[A]) \rfloor. \quad (3)$$

For short we write $\mathfrak{M}[\mathcal{E}] = \mathfrak{M}[\text{supp } \mathcal{E}]$ for the restriction with respect to a set of reactions. The number of reactions in a network \mathfrak{M} will be denoted by $\|\mathfrak{M}\|$.

In order to compare networks in a systematic way, we need to be able to determine the differences and the commonalities of two networks. Inspired by the usual algebra of sets, which is based on the operations "union", "intersection", and "difference", we introduce analogous mathematical constructions for chemical reaction networks. Fig. 1 illustrates the basic operations of this "network algebra" which are formally defined below.

In the following let $\mathfrak{M}'(X', \mathcal{E}')$ and $\mathfrak{M}''(X'', \mathcal{E}'')$ be two networks. Of course we have $\mathfrak{M}' = \mathfrak{M}''$ iff $X' = X''$ and $\mathcal{E}' = \mathcal{E}''$. The empty network will be denoted by \emptyset .

Union

The union $\mathfrak{M} = \mathfrak{M}' \cup \mathfrak{M}''$ is defined as the network $(X' \cup X'', \mathcal{E}' \cup \mathcal{E}'')$. Note that \mathfrak{M} is clean if both \mathfrak{M}' and \mathfrak{M}'' are clean.

Intersection

The intersection $\mathfrak{M} = \mathfrak{M}' \cap \mathfrak{M}''$ is defined as the clean network

$$\mathfrak{M} = \lfloor (X' \cap X'', \mathcal{E}' \cap \mathcal{E}'') \rfloor \quad (4)$$

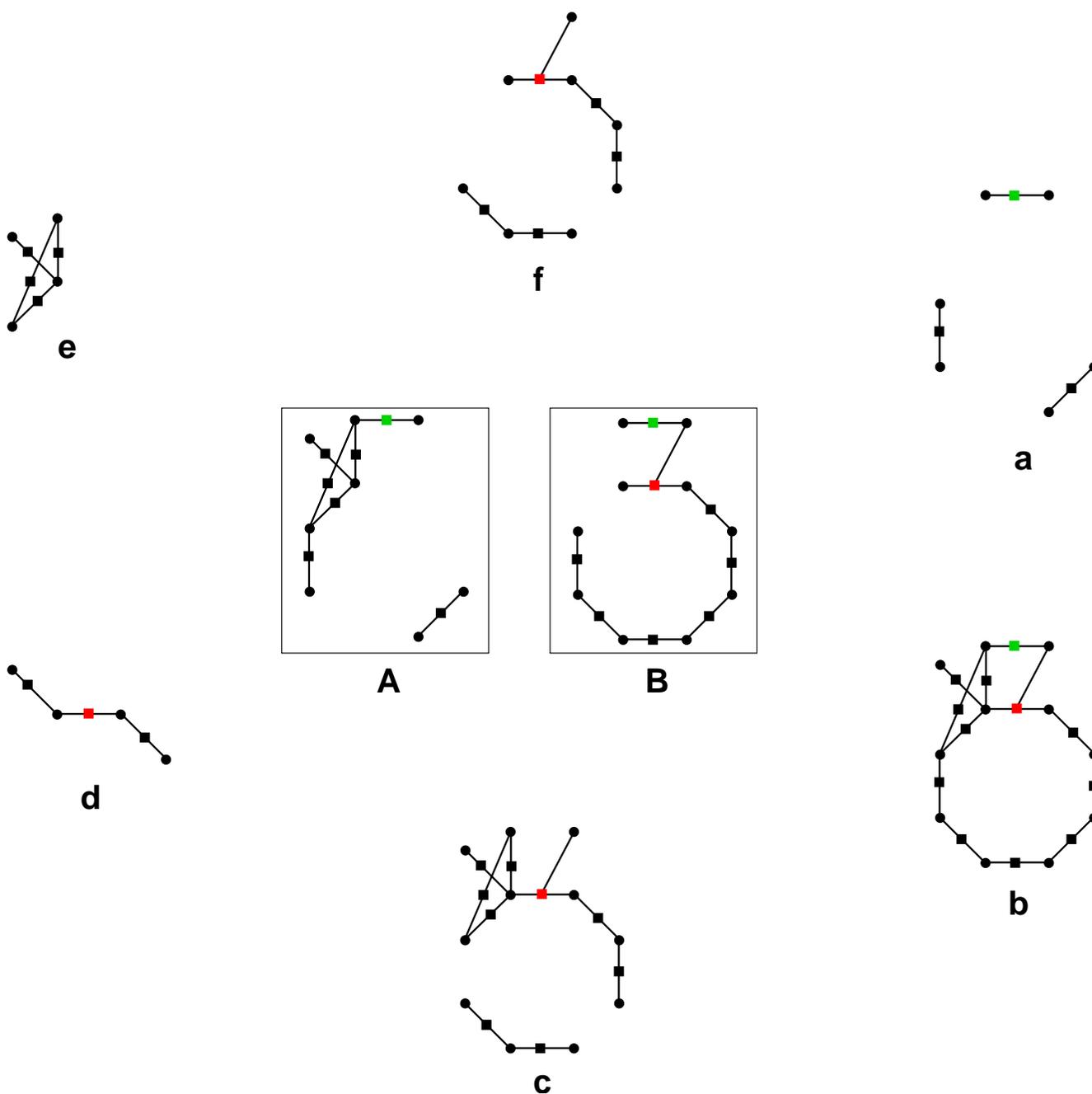


Figure 1

Graphical representation of the basic binary operations of the network algebra. Diagrams (A) and (B) summarize the citric-acid cycle of *P. horikoshii* and *H. pylori* [31]. Hypergraphs can always be drawn as bipartite graphs with one class of vertices representing the hypergraph vertices (chemical species, ●), while the other class of vertices encodes the hyperedges (chemical reactions, ■). Each reaction is connected by (directed) arrows from its educts and to its products. For clarity of presentation we have omitted the direction of the arrows (most reactions are reversible) as well as small molecules such as CO_2 and H_2O here. Furthermore, two reactions are marked in color, namely the ones catalyzed by citrate synthase in red, and pyruvate dehydrogenase in green. The results of the basic operations are as follows: (a) Intersection $A \cap B$; (b) Union $A \cup B$; (c) Symmetric Difference $A \Delta B$; (d) Strict Symmetric Difference $A \diamond B$; (e) Difference $A \setminus B$; (f) Difference $B \setminus A$.

Note that $(\mathcal{E}' \cap \mathcal{E}'') [X' \cap X''] = \mathcal{E}' \cap \mathcal{E}''$.

Difference

The *difference* $\mathfrak{M} = \mathfrak{M}' \setminus \mathfrak{M}''$ is defined as the clean network

$$\mathfrak{M} = \lfloor (\text{supp}(\mathcal{E}' \setminus \mathcal{E}''), \mathcal{E}' \setminus \mathcal{E}'') \rfloor \quad (5)$$

The difference network contains all reactions occurring in \mathfrak{M}' but not \mathfrak{M}'' , and all metabolites occurring in the remaining reactions.

The *strict difference* $\mathfrak{M} = \mathfrak{M}' \setminus \setminus \mathfrak{M}''$ is the clean network

$$\mathfrak{M} = \lfloor (X' \setminus X'', (\mathcal{E}' \setminus \mathcal{E}''), [X' \setminus X'']) \rfloor \quad (6)$$

The new network contains only those metabolites occurring in \mathfrak{M}' but not \mathfrak{M}'' , and only those reactions from \mathfrak{M}' that can be performed with the remaining metabolites. Thus, we have $\|\mathfrak{M}' \setminus \setminus \mathfrak{M}''\| \leq \|\mathfrak{M}' \setminus \mathfrak{M}''\|$.

Symmetric difference

The *symmetric difference* $\mathfrak{M} = \mathfrak{M}' \Delta \mathfrak{M}''$ is defined as the clean network $\mathfrak{M} = \lfloor (\mathfrak{M}' \cup \mathfrak{M}'') \setminus (\mathfrak{M}' \cap \mathfrak{M}'') \rfloor$.

Strict symmetric difference

The *strict symmetric difference* $\mathfrak{M} = \mathfrak{M}' \diamond \mathfrak{M}''$ is $\mathfrak{M} = \lfloor (\mathfrak{M}' \cup \mathfrak{M}'') \setminus \setminus (\mathfrak{M}' \cap \mathfrak{M}'') \rfloor$.

The Vienna Reaction Network Library Vienna-RNL implements these basic set-theoretic operations on chemical reaction networks. It is available under the GNU Public License from [32] and as additional file: 1. The library is written in platform independent ANSI C and provides basic data structures for chemical reactions and their networks, IO routines for reading and writing and various formats, as well as set operations such as the union, intersection, or difference of two chemical reaction networks. It is intended for the use in conjunction with the user's own C programs or PERL scripts.

The library contains IO-Routines for reading SBML [23], an XML based dialect for the standardized representation of systems biology models, writing of SBML is currently being implemented. The capability of reading and writing SBML will make the functionality of the Vienna Reaction Network Library accessible to about 80 other software systems which support SBML [33].

Phylogenies from networks

Datasets were retrieved from the KEGG database on metabolic networks [2], which holds genomic and network data of about 20 Archaea, 200 Bacteria and 20 Eucarya, where in particular the data of many Eukaryotes are incomplete. In a preparatory step we decomposed the individual KEGG-pathways into their chemical reactions and combined these to a complete network for each organism.

The simplest approach to inferring phylogenetic relationships from metabolic networks is to use a distance measure d on the set of reaction networks. We use here

$$d(\mathfrak{M}', \mathfrak{M}'') = \frac{\|\mathfrak{M}' \Delta \mathfrak{M}''\|}{\|\mathfrak{M}'\| + \|\mathfrak{M}''\| - \|\mathfrak{M}' \cap \mathfrak{M}''\|} = \frac{\|\mathfrak{M}' \Delta \mathfrak{M}''\|}{\|\mathfrak{M}' \cup \mathfrak{M}''\|} \quad (7)$$

Alternatively, the strict symmetric difference $\mathfrak{M}' \diamond \mathfrak{M}''$ could be used to define a difference measure. Furthermore, other normalizations of the difference measure could be used. By calculating tree distances between the 16S rRNA phylogenies and the network phylogenies, using the treedist program from the phylib package [24], we have observed, however, that equ.(7) performs best with respect to reproducing trusted 16S RNA phylogenies. Tree distances between the 16S rRNA phylogeny and the network phylogeny calculated with the symmetric difference are 2 and 0.023, using symmetric difference and branch score distance measure, respectively. Compared with tree distances between the 16S rRNA phylogeny and the network phylogeny utilizing the strict symmetric difference yields a much larger tree distance of 10 and 0.093, respectively.

Distance-based network phylogenies are computed using the Fitch algorithm [25] implemented in the phylib package as well as using the splits-decomposition algorithm from the SplitsTree package [26].

An example comprising a selection of bacterial and archaeal metabolic networks is shown in Fig. 2, see Table 1 for the list of species used. The phylogeny inferred from the metabolic networks conforms almost perfectly with the maximum parsimony tree computed from the 16S rRNA sequences of the same organisms. The rRNA sequences were aligned using clustalx. The minor discrepancies are due to poorly resolved nodes as can be seen in the split-decomposition network below. The congruence of rRNA trees and network-based phylogenies demonstrates that the structure of metabolic networks encodes phylogenetic information. Metabolic networks thus constitute a source of phylogenetic signal that is completely independent of sequence information.

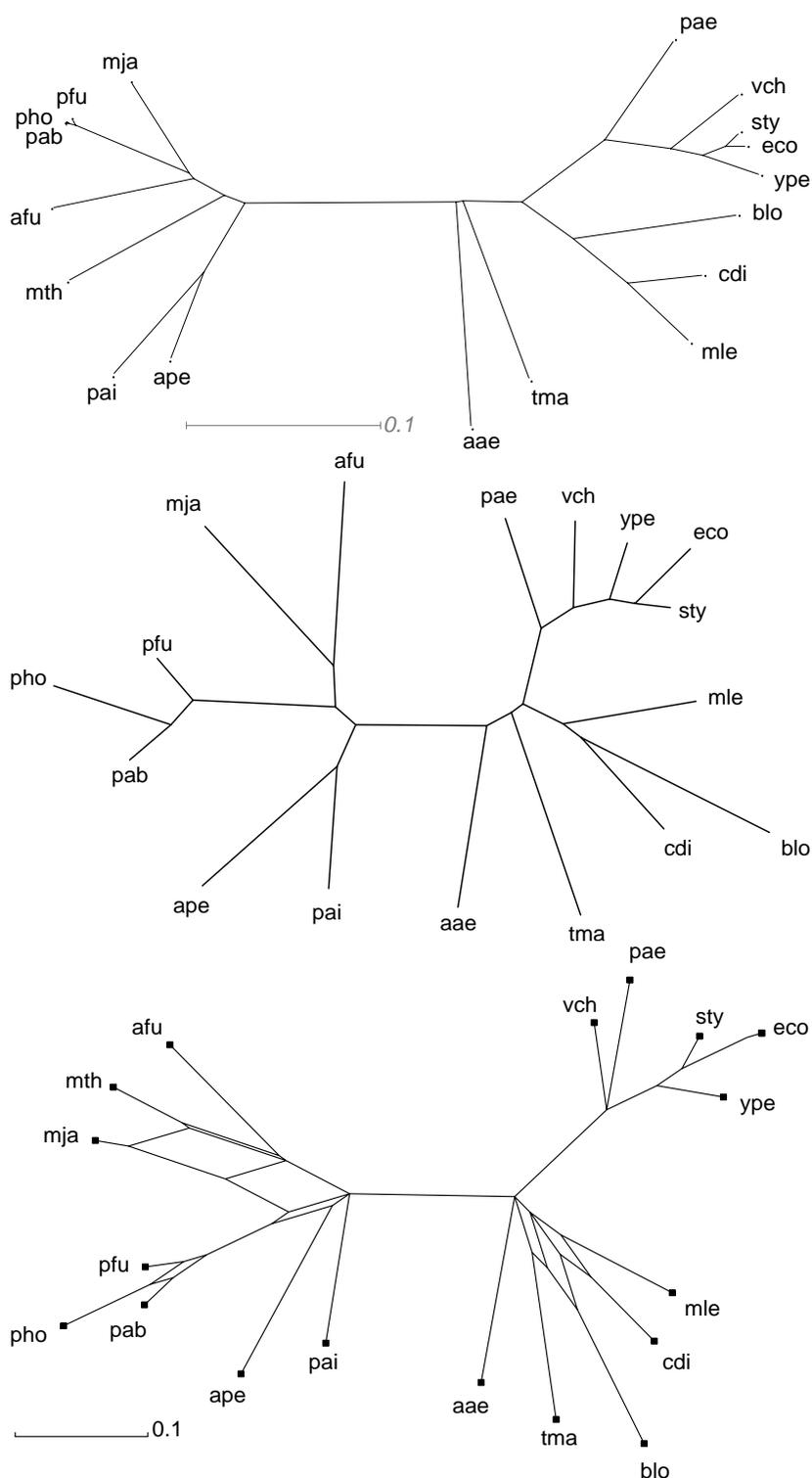


Figure 2

Unrooted phylogenies. (top) Maximum parsimony tree of 16S rRNA sequences. (center) Phylogenetic tree calculated from metabolic network data using the Fitch algorithm for distance matrices. (bottom) Phylogenetic tree calculated from metabolic network data using Splits decomposition with the Fitch-Margoliash power 2 fit for distance matrices. Species abbreviations are collected in Table 1.

Table 1: Metabolic networks used in this study.

Domain		Species	KEGG Id	Genomic Sequence	
Bacteria	Proteobacteria	Gamma	Escherichia coli K-12 MG1655	eco	U00096
			Buchnera aphidicola	buc	BA_000003
			Salmonella typhi CT18	sty	NC_003198
			Yersinia pestis CO92	ype	NC_003143
			Vibrio cholerae	vch	NC_002505
			Pseudomonas aeruginosa	pae	NC_002516
		Rickettsiales	Rickettsia prowazekii	rpr	NC_000963
			Wolbachia endosymbiont	wol	NC_002978
	Firmicutes	Mollicutes	Mycoplasma genitalium	mge	L43967
			Mycoplasma pneumoniae	mpn	NC_000912
			Ureaplasma urealyticum	uur	NC_002162
	Spirochaetes		Borrelia burgdorferi	bbu	AE000783
			Treponema pallidum	tpa	NC_000919
			Treponema denticola	tde	NC_002967
	Actinobacteria		Mycobacterium leprae	mle	NC_002677
		Bifidobacterium longum	blo	NC_004307	
		Corynebacterium diphtheriae	cdi	NC_002935	
Hyperthermophilic bacteria		Aquifex aeolicus	aae	AE000657	
		Thermotoga maritima	tma	AE000512	
Archaea	Euryarchaeota	Methanocaldococcus jannaschii	mja	NC_000909	
		Methanothermobacter thermoautotrophicus	mth	NC_000916	
		Archaeoglobus fulgidus	afu	NC_000917	
		Pyrococcus horikoshii	pho	BA000001	
		Pyrococcus abyssi	pab	NC_000865	
		Pyrococcus furiosus	pfu	NC_003413	
	Crenarchaeota		Aeropyrum pernix	ape	BA000002
			Pyrobaculum aerophilum	pai	NC_003364

The use of distance measures reduces the available information on the network structure already in the first step. We therefore complement distance based phylogenetic analysis with parsimony methods using reaction content: For a given set of organisms, we first calculated the union \mathbf{U} of their metabolic networks. For each organism we then constructed a reaction profile reflecting presence or absence of each reaction in \mathbf{U} in the metabolic network of the respective organism. This approach thus is reduced to reconstructing phylogenies from character-tables that represent the presence/absence of particular reactions in the reaction network. It should be noted that this is similar, but not quite the same, as using the presence or absence of orthologous enzymes (see e.g. [3-8]). The main difference is that the network based approach tolerates functional replacements that may occur, e.g., through horizontal gene transfer [9].

Metabolic innovations

The algebraic approach to metabolic network evolution can also be used in a straightforward way to trace the history of metabolic innovations. To this end, consider a (trusted) unrooted phylogenetic tree T in which each leaf of T is labeled with the metabolic network \mathbf{m}_k of the corresponding taxon k . Each edge e of T defines a split, i.e., a bipartition $\sigma_e = \{U_e, \hat{U}_e\}$ of the set of taxa. Here we regard splits as directed. Note that mathematically we can define innovations at each split in both directions. One of the two subsets U or \hat{U} , however, contains the ancestral state, hence only one direction makes biological sense: this is the one where the ancestral state (root of the tree) is located in the sub-set \hat{U} . This knowledge has to be provided externally. Consider an (arbitrary) directed split $\sigma =$

(U, \hat{U}) on the given set of taxa, i.e., a pair of sets of taxa (U, \hat{U}) such that $U \cap \hat{U} = \emptyset$, and $U \cup \hat{U} = \Theta$. We define the differential metabolic network

$$\mathcal{D}(\sigma) = \left(\bigcup_{k \in U} \mathfrak{m}_k \right) \setminus \left(\bigcup_{k \in \hat{U}} \mathfrak{m}_k \right) \quad (8)$$

The network $\mathcal{D}(\sigma)$ describes the *metabolic innovations* in U relative to the "background" \hat{U} .

As discussed in the previous section, network phylogenies are rather sensitive with respect to life-style and environmental constraints. The organisms whose metabolic networks have been utilized to compute the 16s rRNA tree shown in Fig. 2 are able to freely live in the environment with a reasonably large capacity for adaptation.

As a first example we analyzed the unique metabolic network from the *Pyrococcus* genus. Figure 3 shows the network phylogeny from Fig. 2 with the *Pyrococcus spp.* clade highlighted. The resulting differential network indicates reactions present in *Pyrococcus spp.* but absent in all other organisms of the phylogeny (Figure 3). For example, reaction R01087 is catalyzed by Maleate cis-trans-isomerase which is utilized in maleate assimilating and high-temperature bacteria. A second sub-network involving both ADP-forming acetate and propanoate CoA ligases is potentially used in the organisms to convert between acetate and propanoate and their corresponding CoA forms.

As a second example we analyzed our set of reference organism (Figure 2) with obligatory intracellular pathogens. Figure 4 shows the phylogeny with the selected pathogens (dashed oval). Interestingly, *Mollicutes*, such as *Mycoplasmae* and *Spirochaetes*, such as *Treponema* are grouped together. They all possess a minimal gene-set, and thus a highly optimized and host-dependent metabolic network. Surprisingly, this set of organisms has specific reactions that are absent in the remaining organisms of the phylogeny. Figure 5 shows the corresponding differential network which consists of five sub-network. The two largest networks involve sugar-conversions and parts of glycolysis. Smaller networks correspond to formylation of tetrahydrofolate as well as cholin and carnitine pathways.

Discussion

The Vienna Reaction Network Library introduced above treats chemical reaction networks, and metabolic networks in particular, as directed hypergraphs. A framework borrowed from set algebra provides natural definitions of

unions, intersections, and differences that can be used to compare the metabolic networks of different organisms. We have demonstrated that metabolic networks convey phylogenetic information and can indeed be used to infer phylogenetic relationships of free-living organisms in a way that is similar to gene-content based approaches. In contrast to the latter, however, metabolic network based phylogenies are less sensitive to the effects of functional replacement, e.g., through horizontal gene transfer.

Differences of metabolic networks among subtrees of a trusted phylogeny, or more generally, along any split of interest in a set of organisms can be computed directly, making it easy to study metabolic innovations in particular clades. A first application of our network phylogeny analysis involved three members of the *Pyrococcus spp.* clade. The metabolic reactions resulting from the split between the *Pyrococci* and the remaining organisms involve the maleate cis-trans-isomerase reaction, ADP-forming acetate and propanoate CoA ligase reactions as well as beta-D-Glucose:NAD(P)⁺ 1-oxoreductase.

Our second example considers a class of intra-cellular pathogens that includes *Mycoplasmae*, *Ureaplasmae*, and *Spirochete*. Their restricted repertoire of metabolic reactions reflect the specialized life-style. Many metabolic pathways are not required in such a rich environment and have been lost in the course of evolution. On the other hand, constructing a network phylogeny including these microbes, we observe metabolic reactions assembling an unconnected network that is present in this set of intracellular pathogens and absent in the remaining organisms. Such reactions include phosphorylation and conversions of sugars and derivatives, deaminating lyase reactions, and reactions involving carnitine, choline and tetrahydrofolate.

At present, metabolic network data are compiled by a multitude of methods, and at least in part are constructed by genomic similarity with other organisms. Strictly speaking, therefore, we cannot view metabolic network data such as those compiled in the KEGG database as independent from genomic data. With the recent advances of experimental techniques in metabolomics (see e.g. [27-29]), however, the situation is rapidly improving.

Conclusion

Our comparative approach to metabolic network analysis, which focuses on individual reactions rather than on aggregate features such as pathways, simplifies the identification of metabolic innovations and, in particular, facilitates the recognition of organisms as potential biological threat agents based on their metabolic repertoire. Furthermore, the ability to easily identify differences in metabolic

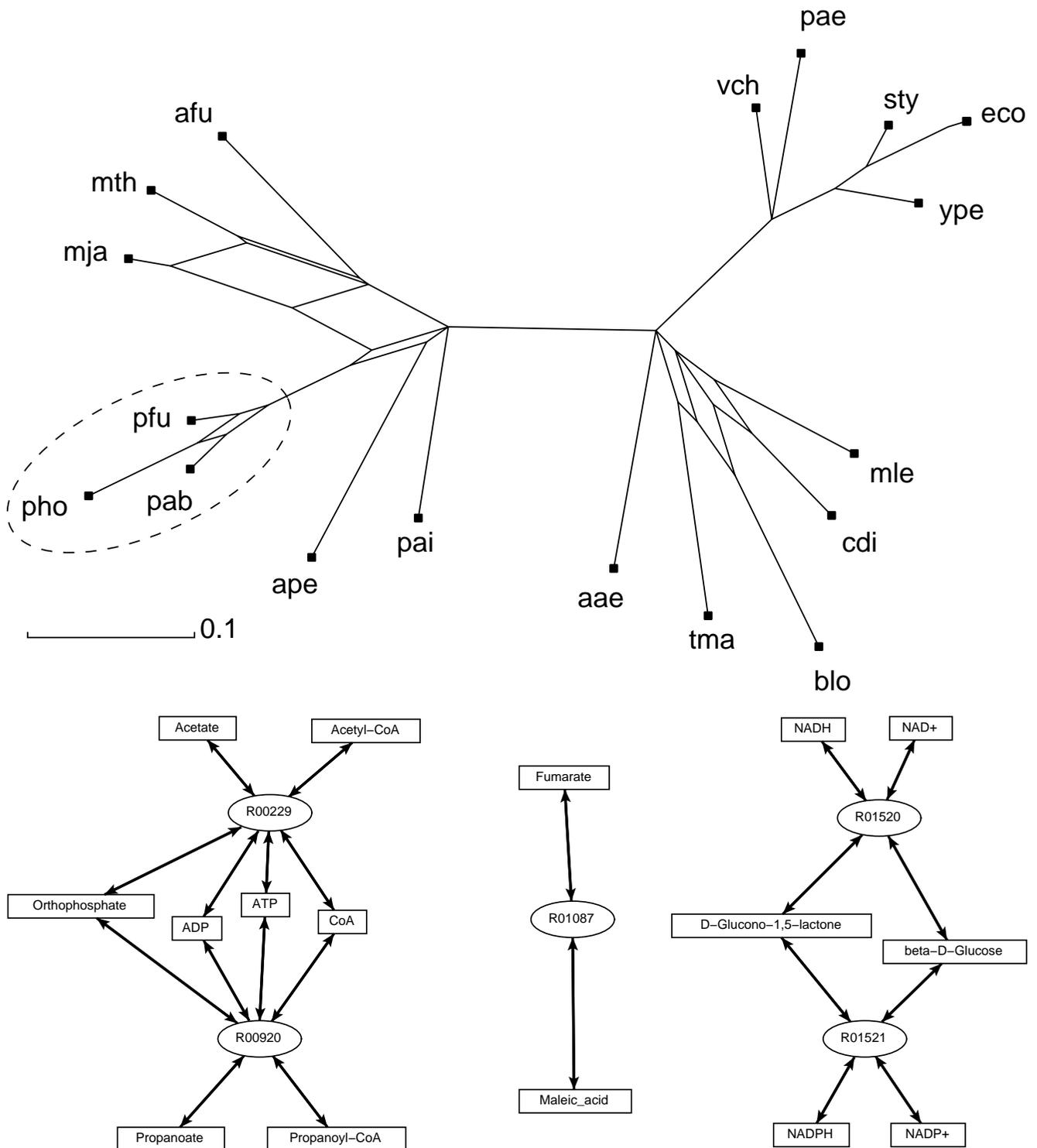


Figure 3
 (top) The *Pyrococcus* spp. clade has been selected (dashed oval) for differential network analysis. (bottom) Differential metabolic network. Numbers in the ovals refer to reaction ids in the KEGG database.

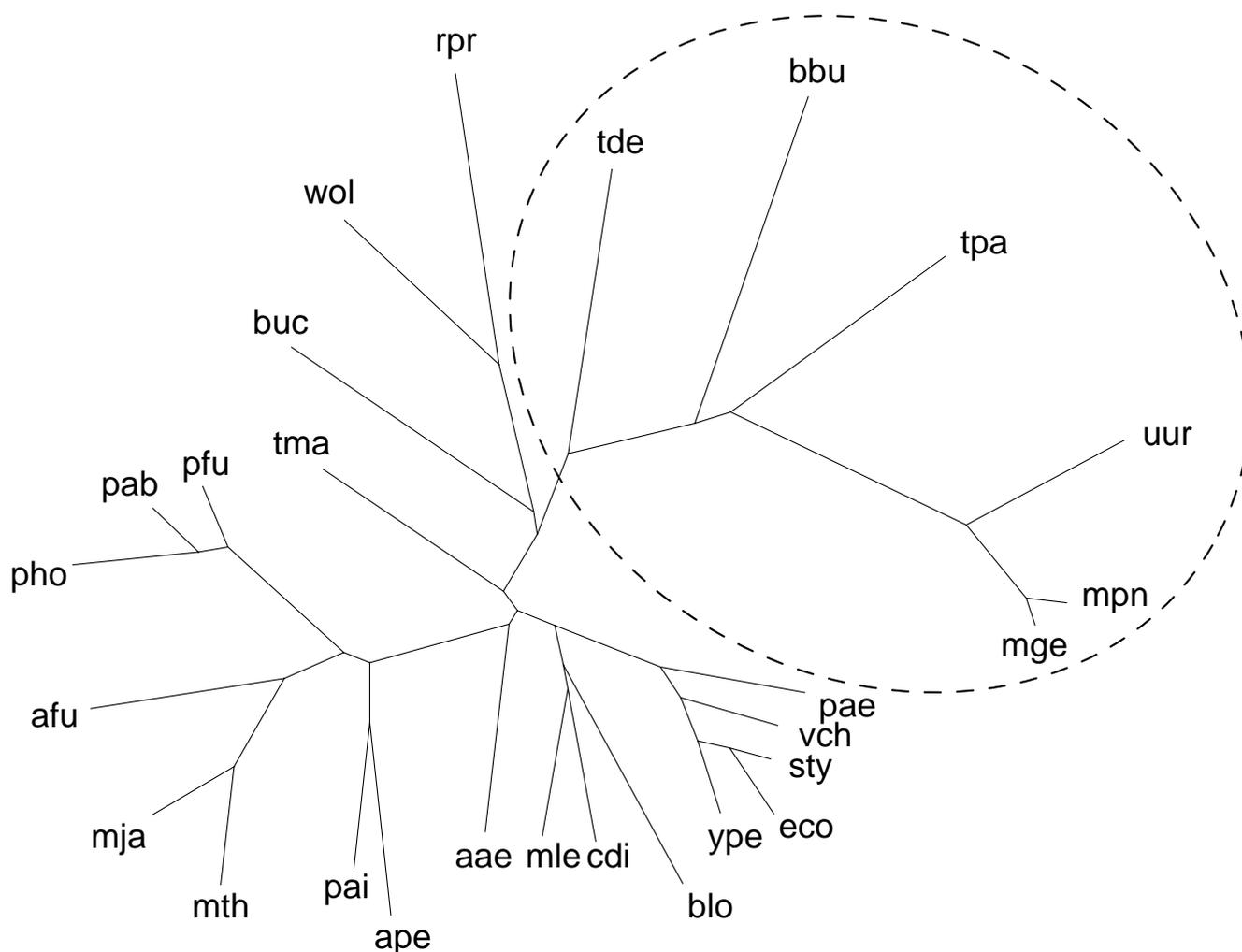


Figure 4

Unrooted network phylogeny using PHYLIP with the Fitch-Margoliash algorithm. A set of obligatory intracellular pathogens has been selected (dashed oval) for differential network analysis (see text).

capacity between pathogens should be useful for devising a refined classification of pathogenicity based on metabolic capabilities.

In this contribution we have restricted ourselves to unweighted networks. Distance measures between networks, however, could be refined by attaching weights to both vertices and (hyper-)edges without requiring significant algorithmic changes. These could reflect, e.g., how essential a reaction or a metabolite is for each organism. With the increasing amount and accuracy of available data it might also be feasible to devise a stochastic model of the evolution of metabolic networks, which could then be

turned into a scoring scheme for a generalized version of (local) graph alignment along the lines of [30].

Authors' contributions

The idea on a metabolic network algebra has been conceived by ILH and PFS during a visit at the Santa Fe Institute. The implementation was substantially improved by CF. CVF developed the distance measurements and applications to microbial metabolic networks. All authors collaborated closely in writing the manuscript.

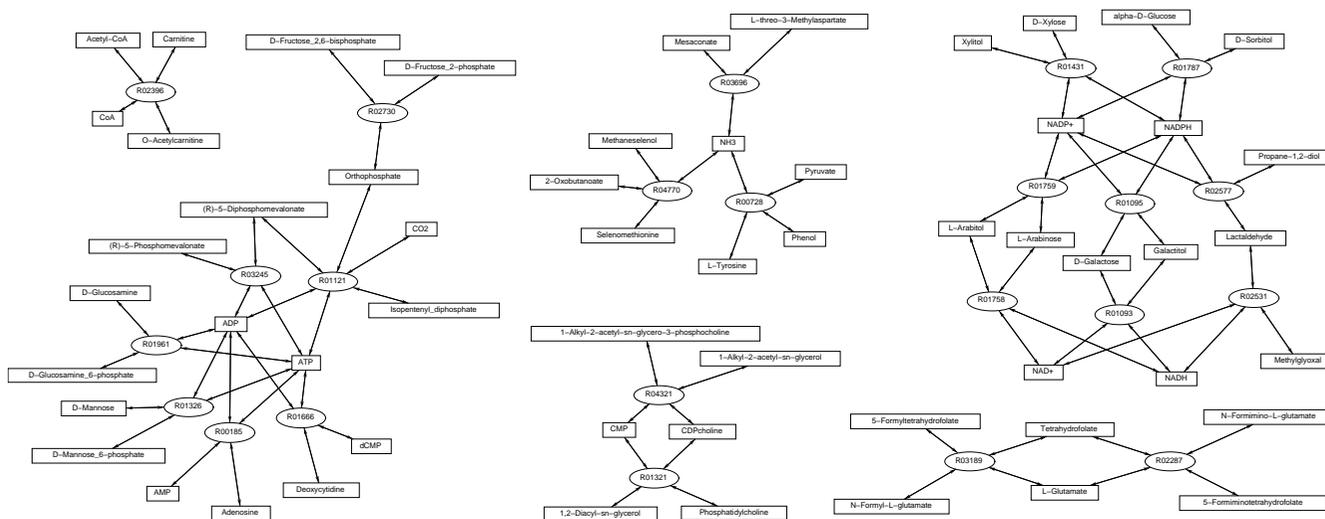


Figure 5
Differential network corresponding to the split shown in Figure 4. These reactions are specializations of the intracellular parasites.

Additional material

Additional file 1

The Vienna Reaction Network Library Vienna-RNL implements these basic set-theoretic operations on chemical reaction networks. It is available under the GNU Public License from [32]

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-7-67-S1.gz]

Acknowledgements

This work was supported in part by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung, Project No. PI5893, by the German DFG Bioinformatics Initiative BIZ-6/1-2, by the Austrian Gen-AU bioinformatics integration network sponsored by BM-BWK and BM-WA, and by the Laboratory Directed Research and Development program of the Los Alamos National Laboratory, Project No. 20040184ER.

References

1. Becker SA, Palsson BO: **Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation.** *BMC Microbiology* 2005, **5**:8. [epub]
2. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucl Acids Res* 2004, **32**:D277-D280.
3. Fitz-Gibbon ST, House CH: **Whole genome-based phylogenetic analysis of free-living microorganisms.** *Nucleic Acids Res* 1999, **27**:4218-4222.
4. Ma HW, Zeng AP: **Phylogenetic comparison of metabolic capacities of organisms at genome level.** *Mol Phylogenet Evol* 2004, **31**:204-213.
5. Snel B, Bork P, Huynen MA: **Genome phylogeny based on gene content.** *Nat Genet* 1999, **21**:108-110.
6. Snel B, Bork P, Huynen MA: **Genomes in flux: the evolution of archaeal and proteobacterial gene content.** *Genome Res* 2002, **12**:17-25.

7. Wolf YI, Rogozin IB, Grishin NV, Tatusov RL, Koonin EV: **Genome trees constructed using five different approaches suggest new major bacterial clades.** *BMC Evol Biol* 2001, **1**:8. [epub]
8. Yang S, Doolittle RF, Bourne PE: **Phylogeny determined by protein domain content.** *Proc Natl Acad Sci USA* 2005, **102**:373-378.
9. Hong SH, Kim TY, Lee SY: **Phylogenetic analysis based on genome-scale metabolic pathway reaction content.** *Appl Microbiol Biotechnol* 2004, **65**:203-210.
10. Tatusov RL, Koonin EV, Lipman DJ: **A Genomic Perspective on Protein Families.** *Science* 1997, **278**:631-637.
11. Dandekar T, Schuster S, Snel B, Huynen MA, Bork P: **Pathway Alignment: application to the comparative analysis of glycolytic enzymes.** *Biochem J* 1999, **343**:115-124.
12. Forst CV, Schulten K: **Evolution of Metabolism: A New Method for the Comparison of Metabolic Pathways Using Genomic Information.** *J Comp Biol* 1999, **6**:343-360.
13. Forst CV, Schulten K: **Phylogenetic analysis of metabolic pathways.** *J Mol Evol* 2001, **52**:471-489.
14. Heymans M, Singh AK: **Deriving phylogenetic trees from the similarity analysis of metabolic pathways.** *Bioinformatics* 2003, **19**(Suppl 1):i138-i346.
15. Schilling CH, Palsson BØ: **The underlying pathway structure of biochemical reaction networks.** *Proc Natl Acad Sci USA* 1998, **95**:4193-4198.
16. Schilling CH, Letscher D, Palsson BØ: **Theory for the systematic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective.** *J Theor Biol* 2000, **203**:229-248.
17. Schuster S, Fell DA, Dandekar T: **A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks.** *Nature Biotechnol* 2000, **18**:326-332.
18. Xiong M, Zhao J, Xiong H: **Network-based regulatory pathways analysis.** *Bioinformatics* 2004, **20**:2056-2066.
19. Gagneur J, Klamt S: **Computation of elementary modes: a unifying framework and the new binary approach.** *BMC Bioinformatics* 2004, **5**:175. [epub]
20. Bunke H, Shearer K: **On a relation between graph edit distance and maximum common subgraph.** *Pattern Rec Let* 1997, **18**:689-694.
21. Zeigarnik AV: **On Hypercycles and Hypercircuits in Hypergraphs.** In *Discrete Mathematical Chemistry, DIMACS series in discrete mathematics and theoretical computer science Volume 51*. Edited by: Hansen P, Fowler PW, Zheng M. Providence, RI: American Mathematical Society; 2000:377-383.

22. Zhu D, Qin ZS: **Structural comparison of metabolic networks in selected single cell organisms.** *BMC Bioinformatics* 2005, **6**:8. [epub] [Doi:10.1186/1471-2105-6-8]
23. Hucka M, Finney A, Bornstein BJ, Keating SM, Shapiro BE, Matthews J, Kovitz BL, J SM, Funahashi A, Doyle JC, Kitano H: **Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project.** *Syst Biol* 2004, **1**:41-52. [Doi: 10.1049/sb:20045008]
24. Felsenstein J: **Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods.** *Methods Enzymol* 1996, **266**:418-427.
25. Fitch W, Margoliash E: **Construction of phylogenetic trees.** *Science* 1967, **155**:279-284.
26. Huson DH: **SplitsTree: analyzing and visualizing evolutionary data.** *Bioinformatics* 1998, **14**:68-73.
27. Brown SC, Kruppa G, Dasseux JL: **Metabolomics applications of FT-ICR mass spectrometry.** *Mass Spectrom Rev* 2005, **24**:223-231.
28. Griffin JL: **Metabolic profiles to define the genome: can we hear the phenotypes?** *Philos Trans R Soc Lond B Biol Sci* 2004, **359**:857-871.
29. Sumner LW, Mendes P, Dixon RA: **Plant metabolomics: large-scale phytochemistry in the functional genomics era.** *Phytochemistry* 2003, **62**:817-836.
30. Berg J, Lässig M: **Local graph alignment and motif search in biological networks.** *Proc Natl Acad Sci USA* 2004, **101**:14689-14694.
31. Huynen MA, Dandekar T, Bork P: **Variation and evolution of the citric-acid cycle: a genomic perspective.** *Trends Microbiol* 1999, **7**:281-291.
32. [<http://www.tbi.univie.ac.at/software/Vienna-RNL/>].
33. [<http://www.sbml.org/>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

