BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 16 2012, pages 2195–2197
doi:10.1093/bioinformatics/bts323

Systems biology

Advance Access publication June 4, 2012

NetCmpt: a network-based tool for calculating the metabolic competition between bacterial species

Anat Kreimer[1,*,†], Adi Doron-Faigenboim[2,†], Elhanan Borenstein[3,4,5] and Shiri Freilich[4,*]

[1]Department of Biomedical Informatics, Columbia University, New York, NY 10032, USA, [2]Institute of Plant Science, Agricultural Research Organization, Bet Dagan, Israel, [3]Department of Genome Sciences, [4]Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195 and [5]Santa Fe Institute, Santa Fe, NM 87501, USA

Associate Editor: Jonathan Wren

ABSTRACT

Summary: NetCmpt is a tool for calculating the competitive potential between pairs of bacterial species. The score describes the effective metabolic overlap (EMO) between two species, derived from analyzing the topology of the corresponding metabolic models. NetCmpt is based on the EMO algorithm, developed and validated in previous studies. It takes as input lists of species-specific enzymatic reactions (EC numbers) and generates a matrix of the potential competition scores between all pairwise combinations.

Availability and implementation: NetCmpt is provided as both a web tool and a software package, designed for the use of non-computational biologists. The NetCmpt web tool, software, examples, and documentation are freely available online at http://app.agri.gov.il/shiri/NetComp.php.

Contact: anat.kreimer@gmail.com; shiri.freilich@gmail.com

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on February 22, 2012; revised on May 8, 2012; accepted on May 28, 2012

1 INTRODUCTION

The availability of many fully sequenced genomes together with several generic metabolic schemes allows the systematic reconstruction of the metabolic networks of hundreds of species across the tree of life (Kanehisa and Goto, 2000). Analyses of the structure and topology of such genome-scale metabolic networks enabled, despite ignoring thermodynamic properties, various systematic, cross-species, studies of lifestyle and evolution in the microbial world (Freilich *et al.*, 2008; Kreimer *et al.*, 2008; Papp *et al.*, 2011). Beyond studying various aspects of genome evolution, such metabolic-driven approaches allow further processing genomic information into ecological information by using the topological structure of the network for predicting the biochemical composition of species' environment. (Borenstein *et al.*, 2008; Handorf *et al.*, 2008). Such 'Reverse Ecology' approaches predict the set of nutrients a species extract from its environment, providing a proxy for its habitat. The recently published NetSeed web tool allows to freely calculate the environments of the selected input species (Carr and Borenstein, 2012).

On top of predicting the species'-specific biochemical habitat, the systematic inference of biochemical habitats together with the availability of a large collection of species-specific metabolic networks, enabled the conductance of several studies indicating that computational, systems-biology, approaches can be applied to study microbial ecology (Borenstein and Feldman, 2009; Freilich *et al.*, 2010a, b; Janga and Babu, 2008; Klitgord and Segre, 2011). In particular, network-driven approaches allow studying the complex set of interactions microbes form with other species thriving in similar habitats (Freilich *et al.*, 2010a, b, 2011). Notably, within a community of bacterial species sharing limited resources, competitive and cooperative interactions are to a large extent derived by metabolism. Accordingly, a growing number of studies testify for the ability of metabolic-driven computational approaches to describe the metabolic interaction between two species (Klitgord and Segre, 2010, 2011; Stolyar *et al.*, 2007; Wintermute and Silver, 2010). The effective metabolic overlap (EMO) score was introduced by Freilich *et al.* (2010a, b), providing a systematic approach for the quantification of the level of competition between bacterial species through the processing of genomic-driven, metabolic information. Several lines of evidence point at the ecological relevance of the EMO score as a tool for predicting the metabolic competitive potential between bacterial species [(Freilich *et al.*, 2010a, b) and Supplementary Note 1]. Based on the EMO algorithm, here we provide NetCmpt—a publically available, easy to use, tool, for computing the competitive potential between selected input species. Notably, the EMO score is solely aimed at the metabolic dimension while putting aside other aspects and mechanisms of inter-species competition.

The rapid increase in the number of metagenomic projects and whole-community analysis techniques highlights the need for accessible tools allowing delineating the web of interactions in a given environment (Chaffron *et al.*, 2010; Dutilh *et al.*, 2009; Pignatelli *et al.*, 2008). NetCmpt, the first free tool for calculating inter-species competition, is available online as a web tool and software package at http://app.agri.gov.il/shiri/NetComp.php. The relevance of the NetCmpt tool for analyzing data derived from metagenomic data was verified through simulations using incomplete datasets (Supplementary Note 2).

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

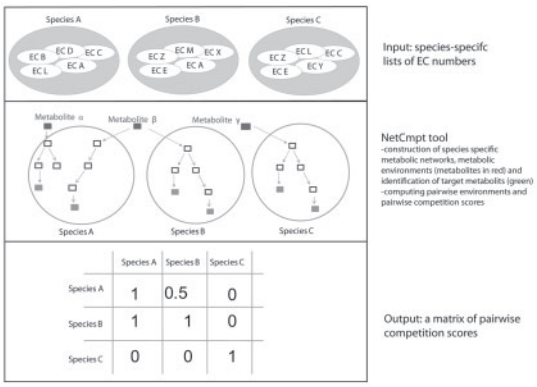

Fig. 1. An illustrative figure of the EMO algorithm implemented in the NetCmpt tool. In this example, the pair-specific environment of species A and B is composed of metabolite α. EMO score denotes 1 minus the fraction of produced essential metabolites. Species B cannot produce any of its target metabolites leading to an EMO score 1. Species A can produce half of its target metabolites leading to an EMO score 0.5. The metabolic environment Species C is composed solely of non-overlapping metabolites, hence species C is not metabolically affected by the presence of the other species

2 DESIGN AND IMPLEMENTATION

2.1 General description of the EMO algorithm and its implantation

For each input species (represented by a list of EC reactions), we automatically reconstruct its metabolic network, its optimal metabolic environment and a list of essential metabolites it has to produce in order to grow (Fig. 1). Based on the above, we use the EMO score for describing the pairwise competitive potential between all pairwise combinations formed between input species. Briefly, for each species-pair we simulate a pair-specific optimal environment composed of all non-overlapping metabolites of a given combination. That is, following computing the optimal environment of each species individually (e.g. metabolites α and β for species A and metabolite β for species B in Fig. 1), the pair-specific environment is then composed solely of the species-specific, unshared, metabolites (metabolite α in Fig. 1). Within this constructed, pair-specific environment, we compute for each pair member the fraction of produced target metabolites. Note that this asymmetric procedure may provide different competition relations between A→B and B→A, as one would intuitively expect. EMO score denotes 1 minus the fraction of produced essential metabolites. Competition score (EMO score) 1 indicates that two species compete on the same resources; Competition score (EMO score) 0 indicates that two species utilize different metabolites for growth. Overall, the EMO score ranges from 0 to 1 and provides an estimate to the metabolic consequences of the co-presence of species B and A. Notably, optimal environment (as used here) allow revealing the competitive potential between bacterial pairs, while different settings, beyond the scope of the EMO score, can reveal other types of interactions (Freilich *et al*., 2011).

2.2 Construction of species-specific metabolic networks

For each input species, the computation of its EMO score requires information on its metabolic capacities, provided as a list of metabolic reactions (EC numbers). The metabolic enzymes are then mapped to reactions as in Freilich *et al*. (2010a, b). The reaction representation allows mapping the enzymes into a generic metabolic network, constructed as in Ma and Zeng (2003) where nodes represent reactions and edges represent common metabolites. Notably, reactions are directional, as inferred from the metabolic scheme used in Freilich *et al*. (2010a, b).

The species-specific network is then derived from the generic network, based on the species-specific reaction content. Let $E_1 = \{e_1^1, e_2^1, \ldots, e_n^1\}$ denote the set of enzymes that catalyze reaction R_1, and $E_2 = \{e_1^2, e_2^2, \ldots, e_m^2\}$ denote the set of enzymes that catalyze reaction R_2. Then, for example, if the species is represented by two enzymes $\{e_3^1, e_2^2\}$ one from E1 and the other from E_2 and a product of R_1 is a substrate of R_2, the species directed metabolic network would be $R_1 \rightarrow R_2$. Edges in the network are considered directed at the reaction level. Notably, frequent metabolites (that is participating in more than 10 reactions) as well as water, protons and electron components are removed from the generic network (Kharchenko *et al*., 2005; Raymond and Segre, 2006). This graph-based representation of metabolic reactions is a common and well established tool in analyzing and studying metabolic networks (Alon, 2003; Feist *et al*., 2009; Jeong *et al*., 2000; Oberhardt *et al*., 2009).

2.3 Constructing a list of target metabolites and externally consumed metabolites (simulated metabolic environments)

To construct species-specific target metabolite lists, the set of metabolites that each species produces is intersected with a generic list of 'target metabolites' that are likely to be essential for growth in most species (Becker and Palsson, 2005; Feist *et al*., 2007; Oh *et al*., 2007). Metabolic growth environments are inferred using the seed algorithm developed by (Borenstein *et al*., 2008) and implemented in the NetCmpt code according to (Carr and Borenstein, 2012). This algorithm predicts the 'seed'—a set of all exogenously acquired compounds, given the metabolic network. We now provide the species-specific seed sets as part of the tool's output.

2.4 Computing EMO

To score the metabolic consequences of the presence of species B in the environment of species A on A's metabolic capacities we remove from A's predicted growth environment (i.e. A's seed) all the metabolites consumed by B, simulate growth of species A on this depleted media and quantify the fraction of essential metabolites that A can still produce (Fig. 1). Growth simulations were done using the expansion method (Ebenhoh *et al*., 2004; Handorf *et al*., 2005)—an approach where networks of increasing size are constructed starting from an initial set of externally consumed substrates (computed as described above) by stepwise addition of those reactions whose substrates are produced in the current core network.

2.5 Software availability and usage

NetCmpt is available as a web tool and a software package at http://app.agri.gov.il/shiri/NetComp.php. Extended information on usage and construction of input file can be found in Supplementary Note 3.

ACKNOWLEDGEMENT

We thank Eytan Ruppin for development of concepts and ideas. We thank Nehama Karvasarsky for her help in establishing the website.

Conflict of Interest: none declared.

REFERENCES

Alon,U. (2003) Biological networks: the tinkerer as an engineer. *Science*, **301**, 1866–1867.

Becker,S.A. and Palsson,B.O. (2005) Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.*, **5**, 8.

Borenstein,E. and Feldman,M.W. (2009) Topological signatures of species interactions in metabolic networks. *J. Comput. Biol.*, **16**, 191–200.

Borenstein,E. *et al.* (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc. Natl. Acad. Sci. USA*, **105**, 14482–14487.

Carr,R. and Borenstein,E. (2012) NetSeed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. *Bioinformatics*, **28**, 734–735.

Chaffron,S. *et al.* (2010) A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.*, **20**, 947–959.

Dutilh,B.E. *et al*. (2009) Increasing the coverage of a metapopulation consensus genome by iterative read mapping and assembly. *Bioinformatics*, **25**, 2878–2881.

Ebenhoh,O. *et al*. (2004) Structural analysis of expanding metabolic networks. *Genome Inform.*, **15**, 35–45.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Feist,A.M. *et al.* (2009) Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.*, **7**, 129–143.

Freilich,S. *et al.* (2008) Metabolic innovations towards the human lineage. *BMC Evol. Biol.*, **8**, 247.

Freilich,S. *et al.* (2010a) Decoupling environment-dependent and independent genetic robustness across bacterial species. *PLoS Comput. Biol.*, **6**, e1000690.

Freilich,S. *et al.* (2010b) The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res.*, **38**, 3857–3868.

Freilich,S. *et al.* (2011) Competitive and cooperative metabolic interactions in bacterial communities. *Nat. Commun.*, **2**, 589.

Handorf,T. *et al.* (2008) An environmental perspective on metabolism. *J. Theor. Biol.*, **252**, 530–537.

Handorf,T. *et al*. (2005) Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J. Mol. Evol.*, **61**, 498–512.

Janga,S.C. and Babu,M.M. (2008) Network-based approaches for linking metabolism with environment. *Genome Biol.*, **9**, 239.

Jeong,H. *et al.* (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.

Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

Kharchenko,P. *et al*. (2005) Expression dynamics of a cellular metabolic network. *Mol. Syst. Biol.*, **1**, 0016.

Klitgord,N. and Segre,D. (2010) Environments that induce synthetic microbial ecosystems. *PLoS Comput. Biol.*, **6**, e1001002.

Klitgord,N. and Segre,D. (2011) Ecosystems biology of microbial metabolism. *Curr. Opin. Biotechnol.*, **22**, 541–546.

Kreimer,A. *et al.* (2008) The evolution of modularity in bacterial metabolic networks. *Proc. Natl. Acad. Sci. USA*, **105**, 6976–6981.

Ma,H. and Zeng,A.P. (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, **19**, 270–277.

Oberhardt,M.A. *et al*. (2009) Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.*, **5**, 320.

Oh,Y.K. *et al.* (2007) Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.*, **282**, 28791–28799.

Papp,B. *et al*. (2011) Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.*, **12**, 591–602.

Pignatelli,M. *et al.* (2008) Metagenomics reveals our incomplete knowledge of global diversity. *Bioinformatics*, **24**, 2124–2125.

Raymond,J. and Segre,D. (2006) The effect of oxygen on biochemical networks and the evolution of complex life. *Science*, **311**, 1764–1767.

Stolyar,S. *et al.* (2007) Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol.*, **3**, 92.

Wintermute,E.H. and Silver,P.A. (2010) Emergent cooperation in microbial metabolism. *Mol. Syst. Biol.*, **6**, 407.