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

Anat Kreimer1,*, Adi Doron2,#, Elhanan Borenstein3, Shiri Freilich2,*
1Department of Biomedical Informatics, Columbia University, New York 10032, New York, USA
2Newe Yaar Research Center, Agricultural Research Organization, Ministry of Agriculture, P.O. Box 1021, Ramat Yishay 30095, Israel
3Department of Genome Sciences, University of Washington, Seattle, Washington, U.S.A.

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: 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: shiri.freilich@gmail.com

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., 2010; 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., 2010; Freilich, et al., 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; Klitgord and Segre, 2011; Stolyar, et al., 2007; Wintermute and Silver, 2010). The Effective Metabolic Overlap (EMO) score was introduced by Freilich et al. (Freilich, et al., 2010), 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., 2010) and Supplementary Note 1). Based on the EMO algorithm, here we provide NetCmpt – a publicly 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 data sets (Supplementary Note 2).

2 DESIGN AND IMPLEMENTATION

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 (Figure 1). Based on the above, we use the Effective Metabolic Overlap (EMO) score for describing the pair-
wise 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 the optimal environment of each species individually (e.g., metabolites $\alpha$ and $\beta$ for species A and metabolite $\beta$ for species B in Figure 1), the pair-specific environment is then composed solely of the species-specific, unshared, metabolites (metabolite $\alpha$ in Figure 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 than 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).

**Construction of species-specific Metabolic Networks**

For each input species, the computation of its EMO score requires information on usage and construction of input file can be found in Supplementary Note 3.

![Fig. 1](image_url). An illustrative figure of the Effective Metabolic Overlap (EMO) algorithm implemented in the NetCmpt tool. In this example, the pair-specific environment of species A and B is composed of metabolite $\alpha$. 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.

**Computing Effective Metabolic Overlap (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 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 (Figure 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.

**Software availability and usage**

NetCmpt is available as a web tool and a software package at [http://app.agri.gov.il/shiri/NetComp.php](http://app.agri.gov.il/shiri/NetComp.php). Extended information on usage and construction of input file can be found in Supplementary Note 3.
ACKNOWLEDGEMENTS
We thank Eytan Ruppin for development of concepts and ideas.

REFERENCES