NetSeed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment

Rogan Carr[1] and Elhanan Borenstein[1,2,3,*]

[1]Department of Genome Sciences, [2]Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195 and [3]Santa Fe Institute, Santa Fe, NM 87501, USA

Associate Editor: Olga Troyanskaya

ABSTRACT

Summary: NetSeed is a web tool and Perl module for analyzing the topology of metabolic networks and calculating the set of exogenously acquired compounds. NetSeed is based on the seed detection algorithm, developed and validated in previous studies.

Availability: The NetSeed web-based tool, open-source Perl module, examples and documentation are freely available online at: http://depts.washington.edu/elbogs/NetSeed.

Contact: elbo@uw.edu

1 INTRODUCTION

Genome-scale, *in silico* metabolic models provide powerful tools for studying the capacity, function and dynamics of microbial metabolism. Specifically, network-based models have been shown to capture essential features of microbial metabolism revealed through careful analysis of their structure and topology (Alon, 2003). Such topology-based analysis has been used to characterize various functional and evolutionary properties including scaling (Jeong *et al.*, 2000), modularity (Guimera and Amaral, 2005), essentiality (Palumbo *et al.*, 2005), transcriptional regulation of metabolism (Patil and Nielsen, 2005) and adaptation (Kreimer *et al.*, 2008).

However, as organisms evolve and adapt to their environments, the structure of their metabolic networks clearly contains information not only about their metabolic capacity but also about their habitat. The topology of the metabolic network of an organism can therefore be used to characterize its environment and to obtain insights into its ecology. This 'Reverse Ecology' approach was introduced by Borenstein *et al.* (2008), who proposed an algorithm for analyzing the topology of metabolic networks of microbial species and predicting the set of nutrients they acquire from the environment (termed 'seed set'). This framework allows for the quantification of an organism's metabolic dependence on its environment and enables the transformation of high-throughput genomic data into large-scale ecological data.

Formally, the seed set of a directed network, such as a metabolic network, is defined as the minimal subset of nodes necessary to access every other node in the network. Accordingly, for a metabolic network, the seed set corresponds to the minimal subset of the compounds in the network that cannot be synthesized from other compounds in the network and whose existence permits the production of all other compounds in the network. These seed metabolites therefore represent externally acquired compounds and can be viewed as the metabolic interface between an organism and its environment. Borenstein *et al.* (2008) validated these inferred metabolic environments and demonstrated that they serve as successful proxies for organisms' biochemical environments and capture information about their habitats. They further used this approach to study the evolution of the interface between organisms and their environments across the microbial tree of life and identified various determinants of this interface.

The seed set analysis technique has since been used in several computational studies to elucidate microbial metabolism and ecology. Specifically, this approach has been applied to quantitatively characterize the ecological strategies of a large array of microbial species (Freilich *et al.*, 2009), to quantify environmental robustness in bacteria (Freilich *et al.*, 2010a) and to calculate the metabolic overlap between organisms in microbial communities (Freilich *et al.*, 2010b). Cottret *et al.* (2010) implemented the seed detection algorithm to study the metabolic relationship between coresident endocytobionts. Borenstein and Feldman (2009) further extended the seed set concept to predict the interaction between host and parasites on a large scale. The concept of the seed set has also been proposed as a biotechnology and environmental engineering framework for developing media to isolate specific microorganisms (Röling *et al.*, 2010) and for predicting potential syntrophic relationships and selecting taxa for gnotobiotic mice colonization studies (Hansen *et al.*, 2011). Janga and Babu (2008) further promoted the use of the seed set framework for studying the association of species in similar habitats and for developing drugs to target parasitic organisms without harming the host.

As the number of available genome-scale metabolic models is steadily increasing, laying the foundation for extensive reverse-ecology studies (Papp *et al.*, 2011), we have developed NetSeed, a web-based tool to calculate and visualize metabolic network seed sets. NetSeed will make the calculation of the seed sets of metabolic networks simple and widely available. NetSeed is available online with additional documentation and examples at http://depts.washington.edu/elbogs/NetSeed. NetSeed is also available as an open-source (GPL) Perl module, allowing the seed detection algorithm to be included in custom software.

2 DESIGN AND IMPLEMENTATION

NetSeed determines the seed set of a metabolic network as described in Borenstein *et al.* (2008). Here, the network is represented as

---

*To whom correspondence should be addressed.
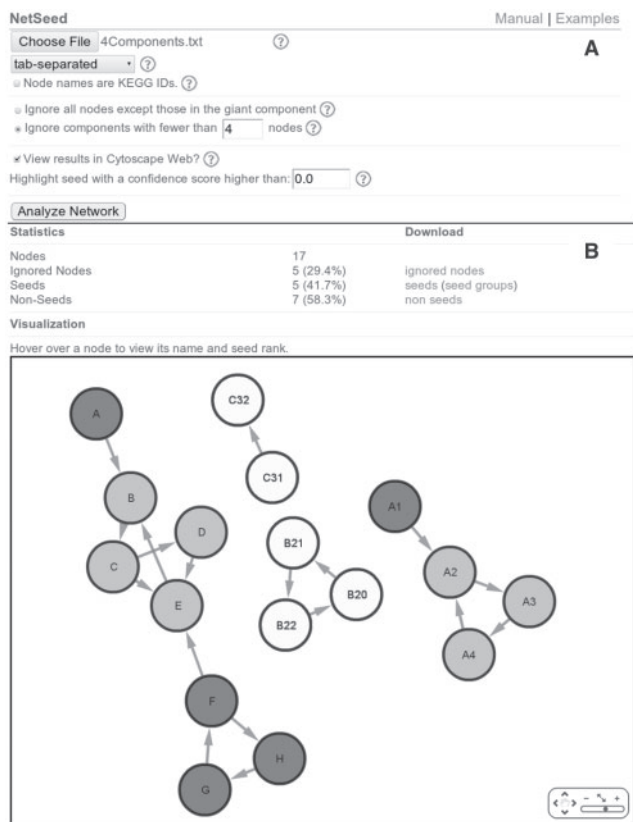
Fig. 1. The NetSeed web tool. (**A**) The user first specifies the network to analyze and various algorithm parameters. (**B**) The result page gives statistics and links to the results. The network is shown in CytoscapeWeb, with seeds highlighted in dark grey and ignored nodes in white.

a directed graph, with nodes as metabolites and edges as reactions linking substrates to products. The seed set of the network is calculated in two steps. First, the network is decomposed into its strongly connected components (SCCs) using Kosaraju's algorithm (Aho *et al.*, 1974). SCCs of a directed network are defined as maximal subsets of the network where there exists a path from each node to every other node in the subset. Due to this interconnectivity, SCCs have the property that either all or none of their nodes are potential seeds. The second step then determines which SCCs are sources, having no incoming edges. The nodes in these source SCCs are the potential candidates of the seed set of the network. Seed candidates that belong to the same SCC (seed group) have the property that the presence of one in the seed set will generate all the other in the group and therefore cannot be distinguished in terms of their seed status. To quantify this relationship, each seed in the network is given a 'confidence level' corresponding to the inverse of the size of their seed group. See Borenstein *et al.* (2008) for more details.

The NetSeed web tool (Fig. 1A) is a Perl CGI built on top of the NetSeed Perl module, which performs the seed set calculation. The user first selects a network (several formats are supported) and

determines whether to restrict the calculation by component size (e.g. only the giant component or those of a minimum size) and whether to view the seeds in Cytoscape Web (Lopes *et al.*, 2010). If Cytoscape Web is enabled, the user may specify the minimum confidence level for labeling seeds; this will not affect the analysis.

When user then selects 'Analyze Network', the network file is uploaded and the seed set calculated. The results page (Fig. 1B) displays statistics on the analysis and links to files containing a list of seeds with their confidence levels, seed groups, nodes ignored by the calculation (as specified by the user) and non-seed nodes. If requested, the network is rendered in Cytoscape Web with the highlighted seed set. If the node names correspond to KEGG IDs, links to the KEGG database are generated. Detailed examples and an in-depth manual for both the NetSeed web tool and Perl module are available on the NetSeed website.

ACKNOWLEDGEMENTS

REFERENCES

Aho,A. *et al.* (1974) *The Design and Analysis of Computer Algorithms.* Addison-Wesley, Reading.

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

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

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

Cottret,L. *et al.* (2010) Graph-based analysis of the metabolic exchanges between two co-resident intracellular symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with their insect host, *Homalodisca coagulata*. *PLoS Comput. Biol.*, **6**, e1000904.

Freilich,S. *et al.* (2009) Metabolic-network-driven analysis of bacterial ecological strategies. *Genome Biol.*, **10**, R61.

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.

Guimera,R., Amaral,L.A.N. (2005) Functional cartography of complex metabolic networks. *Nature.* **433**, 895–900.

Hansen,E.E. *et al.* (2011) Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proc. Natl Acad. Sci. USA*, **108**, 4599–4606.

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.

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

Lopes,C.T. *et al.* (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.

Palumbo,M. *et al.* (2005) Functional essentiality from topology features in metabolic networks: A case study in yeast. *FEBS Lett.*, **579**, 4642–4646.

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

Patil,K.R. and Nielsen,J. (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl Acad. Sci. USA*, **102**, 2685–2689.

Röling,W.F.M. *et al.* (2010) Systems approaches to microbial communities and their functioning. *Curr. Opin. Biotechnol.*, **21**, 532–538.