



Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil



Honggui Wang^{a, b}, Zhong Wei^c, Lijuan Mei^b, Jingxin Gu^b, Suisui Yin^b,
 Karoline Faust^{d, e, f, 1}, Jeroen Raes^{d, e, f}, Ye Deng^g, Yulong Wang^a,
 Qirong Shen^{c, 1}, Shixue Yin^{b, *, 1}

^a Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, China

^b College of Environmental Science and Engineering, Yangzhou University, Yangzhou 225127, China

^c College of Resources and Environmental Science, Nanjing Agricultural University, Nanjing, 210095, China

^d Department of Structural Biology, Flemish Institute for Biotechnology (VIB), Belgium

^e VIB Centre for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium

^f Laboratory of Microbiology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

^g Institute of Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

ARTICLE INFO

Article history:

Received 23 September 2016

Received in revised form

26 November 2016

Accepted 30 November 2016

Available online 6 December 2016

Keywords:

Soil microbial community

Interaction network inference

Ecologically meaningful interactions

Paddy soil

ABSTRACT

High-throughput sequencing technologies have recently made it possible to interrogate the phylogenetic diversity of soils at considerable depth. This ability has led to the development of many computational tools to infer interaction networks from environmental samples. Although such tools have widely been used, they have more often served as a visual means to compare microbial communities across environmental gradients than as a means to appreciate microbial interactions associated with certain ecological processes. Previous studies have often regarded a subnetwork (module) as a functional unit but its functionality in ecological context has never been evidenced. To make better use of these tools in appreciating microbial interactions, we propose the combinational use of different inference tools. This ensemble approach permits the use of more independent predictors and the removal of tool-specific predictions in order to increase prediction accuracy. The purpose of the present study is to identify ecologically meaningful bacterial associations using multi-tool approach. Soil samples were collected in time series from experimental paddy rice plots. Bacterial communities were characterized by high-throughput tag sequencing of 16S rRNA gene fragments. We used three tools, Co-occurrence Network inference (CoNet), Molecular Ecological Network Analysis (MENA) and extended Local Similarity Analysis (eLSA), to infer networks from abundance profiles, partitioned the networks into modules, screened for the modules with $\geq 50\%$ of genus-/species-level nodes, captured the modules that were derived from different tools and shared $\geq 50\%$ of order-level nodes (tool-agreed modules) and tested their robustness against the changes in the tool parameters. By these procedures, two three-tool-agreed modules were found. One represented a guild that is phenotypically associated with aerobic respiration and fermentation and the other represented a guild phenotypically associated with metal/sulphur cycles, all of which are essential processes of water-submerged paddy soils that are mediated by bacteria. These data suggested that the linked members in a module were functionally associated taxa that work together to achieve a distinct function or an ecological process, and thus were ecologically meaningful to the environment. We selected three linked species from a three-tool-agreed module and validated their interactions using co-culture methods. Results showed that the interaction type between *Janthinobacterium lividum* and *Leuconostoc lactis* in the two-species mixture was validated to be ambivalent, positive for one partner and negative for the other. However, this type of interaction was not retained when a third party *Lactococcus piscium* was introduced, signifying the complexity of multi-species interactions. Validation results suggested that the selected species were interacting partners in laboratory but the validated interaction types were different from those inferred. By multi-tool approach, we also

* Corresponding author. Huayang West Road 196, College of Environmental Science and Engineering, Yangzhou University, Yangzhou 225127, Jiangsu Province, China.

E-mail address: sxyin@yzu.edu.cn (S. Yin).

¹ These authors contributed equally to this work.

found that highly linked nodes, which are often referred to as “keystone species” and are frequently interpreted as the species playing important roles in soils, are tool dependent. Among top ten highly linked nodes, only four are conserved across three tools. These results suggest more research is required on the ecological significance of degree-based identification of keystone species. Overall, the present study highlights the potential utility of combined use of inference tools to identify ecologically meaningful bacterial associations in soils and other environmental samples. It is interesting to see what type of ecologically meaningful bacterial associations can be found in other soils.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

High-throughput sequencing technologies have recently made it possible to interrogate phylogenetic diversity in environmental samples at considerable depth of coverage. This ability has led to the development of many computational tools to infer interaction networks from complex microbial communities, on the basis of abundance profiles. Among the long list are Co-occurrence Network inference (CoNet) (Faust et al., 2012), Molecular Ecological Network Analysis (MENA) (Deng et al., 2012), extended Local Similarity Analysis (eLSA) (Xia et al., 2013), Sparse Correlations for Compositional data (SparCC) (Friedman and Alm, 2012), Learning Interactions from Microbial Time Series (LIMITS) (Fisher and Mehta, 2014), Correlation inference for Compositional data through Lasso (CCLasso) (Fang et al., 2015), Microbe-microbe Interaction Networks (MInt) (Biswas et al., 2015), Regularized Estimation of the Basis Covariance (REBACCA) (Ban et al., 2015) and Sparse Inverse Covariance Estimation for Ecological Association Inference (SPIEC-EASI) (Kurtz et al., 2015). Some of these tools have been evaluated using synthetic data (Weiss et al., 2016) or using data on plankton interactions reported from the literature (Lima-Mendez et al., 2015). Using these tools, microbial interactions in various environmental samples have been explored, such as marine water (Steele et al., 2011; Gilbert et al., 2012), fresh water (Eiler et al., 2012; Kara et al., 2013), the human microbiome (Qin et al., 2010; Arumugam et al., 2011; Duran-Pinedo et al., 2011; Faust et al., 2012; Greenblum et al., 2012; Endesfelder et al., 2014) and soils (Janssen, 2006; Zhou et al., 2011; Barberan et al., 2012; Lu et al., 2013; Hoppe et al., 2014). Networks inferred from each set of samples exhibited numerous links between numerous taxa. However, these links are often difficult to understand from biological and/or ecological perspectives. Previous studies have demonstrated that most networks inferred from environmental samples exhibit non-random patterns, but the patterns may or may not represent biological interactions. For example, two taxa may respond to soil temperature in a similar manner, and consequently result in similar abundance profiles, without being interaction partners. It is not clear whether the linked members are functionally associated taxa that work together to achieve a distinct function or an ecological process in the given environment. Indeed, interpreting such networks from biological/ecological perspectives is difficult, partially because of the complexity of microbial communities and partially because network inference tools cannot distinguish between true ecological interactions and other non-random processes (Faust and Raes, 2012). Consequently, abundance-based network inference tools have served more often as a powerful visual means to compare microbial communities across environmental gradients than as a means to appreciate microbial interactions associated with certain ecological processes.

To make better use of inference tools to appreciate microbial interactions in environmental samples, we propose the

combinational use of different tools to capture tool-agreed networks. Different tools usually use different algorithms, methods to treat similarity scores and approaches to filter noises (false links). Some tools are designed to address particular computational questions. For example, eLSA (Ruan et al., 2006) and MInt (Biswas et al., 2015) are particularly suited to capture the time-lag relationships between two nodes (OTUs, taxa or species) in time-series data. MENA uses random matrix theory to determine thresholds on similarity scores (Zhou et al., 2010), whereas other tools determine thresholds by using permutation tests. CoNet is designed to use multiple similarity and distance measures (Pearson, Spearman, Bray Curtis, Kullback-Leibler and others) in combination to capture measure-agreed edges (links), and is based on the idea that different similarity measures would agree on true interactions and disagree on false interactions (Faust et al., 2012), whereas other tools use either the Pearson or Spearman correlation method. Owing to the different techniques employed, different tools often produce different networks, thus making it difficult for users to appreciate which result represents biological interactions in a given environment. Furthermore, it has been demonstrated that no single inference method performs optimally across all datasets, and the prediction accuracy can be significantly improved with multiple inference methods (Marbach et al., 2012). A recent evaluation also concluded that combining tools improves their performance (Weiss et al., 2016). Therefore, tool-agreed networks would be of higher value to capture ecologically meaningful interactions than tool-specific networks. Unfortunately, previous studies on soil samples have often relied on a single tool. If taxa in such tool-agreed networks provide sufficient phenotypic information, this information can be used to interpret the functionality of networks and relate it to a certain soil process. This approach allows exploration of whether a network represents a functional guild associated with a distinct soil process. In addition, highly linked (high degree) taxa in inferred networks have often been referred to as “keystone species” and interpreted as species playing important roles in communities. However, the degree-based keystone species are frequently claimed based on single inference method only. Whether such species are tool-dependent has not been investigated. This kind of information helps us further understand the degree-based keystone species in environmental samples.

In the present work, we used water-submerged paddy soils as a model ecosystem. Rice paddy is an ecosystem full of aerobic, microaerobic and anaerobic zones and interfaces in micro- and macro-scales. Such ecosystem supports the life of bacteria with diverse physiologies. It is also known that biogeochemical variables fluctuates over time (Ponnampertuma, 1972), which drive the fluctuations of bacterial abundances over time, and thus provide an opportunity for us to sample soils with sufficient abundance variations that are necessary for inferring reliable network. We used three tools, CoNet, MENA and eLSA, to infer networks on the basis of abundance profiles, partitioned them into modules (sub-networks, each comprising a group of nodes more densely connected to each other than to nodes outside the group), screened for those

Download English Version:

<https://daneshyari.com/en/article/5516518>

Download Persian Version:

<https://daneshyari.com/article/5516518>

[Daneshyari.com](https://daneshyari.com)