OXFORD

Data and text mining

MetaCoMET: a web platform for discovery and visualization of the core microbiome

Yi Wang[1,2,3], Ling Xu[2,3], Yong Q. Gu[1] and Devin Coleman-Derr[2,3,]*

[1]USDA-ARS, Western Regional Research Center, Crop Improvement and Genetics Research Unit, Albany, CA 94710, USA, [2]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA and [3]USDA-ARS, Plant Gene Expression Center, Albany, CA 94710, USA

*To whom correspondence should be addressed.
Associate Editor: Jonathan Wren

Abstract

Motivation: A key component of the analysis of microbiome datasets is the identification of OTUs shared between multiple experimental conditions, commonly referred to as the core microbiome.

Results: We present a web platform named MetaCoMET that enables the discovery and visualization of the core microbiome and provides a comparison of the relative abundance and diversity patterns between subsets of samples within a microbiome dataset. MetaCoMET provides an efficient and interactive graphical interface for analyzing each subset defined by the union or disjunction of groups within the Venn diagram, and includes a graphical taxonomy summary, alpha diversity metrics, Principal Coordinate analysis, abundance-based heatmaps, and a chart indicating the geographic distribution of each sample.

Availability and Implementation: MetaCoMET is a user-friendly and efficient web platform freely accessible at http://probes.pw.usda.gov/MetaCoMET or http://aegilops.wheat.ucdavis.edu/MetaCoMET.

Contact: devin.coleman-derr@ars.usda.gov

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

1 Introduction

A common goal in amplicon-based microbiome analysis is the identification of operational taxonomic units (OTUs) that are either shared between or distinct among multiple experimental conditions (treatments, environments, time points, etc.). The term core microbiome has been used to describe these shared microbes (Hamady and Knight, 2009; Turnbaugh, *et al.*, 2007) and is typically defined as the set of OTUs common among the microbial communities from different but related conditions or habitats. Commonly occurring components within complex microbial assemblages are of interest for a variety of reasons, and may well be critical to the general function of those communities (Shade and Handelsman, 2012). Although some web tools, like METAREP, SILVAngs and VAMPS provide analysis and visualization functions for microbiome datasets, they provide limited tools for the identification of the core microbiome and these tools focus primarily on the presence or absence of OTUs within the data.

Here, we present a web platform named MetaCoMET that enables the easy discovery and visualization of the core microbiome using one of three alternative methods and the comparison of relative abundance and diversity patterns for subsets of the data as defined by the core analysis.

2 Methods

As an extension and application of the ideas presented by Shade and Handelsman (Shade and Handelsman, 2012), we employ three standard methods (membership, composition, persistence) that can be used to define the core microbiome.

The Membership method is the simplest and most commonly used method for determining the core microbiome using an OTU

table, and is based on the presence and absence of OTUs among the different microbiomes being compared (Supplementary Fig. S1A). If an OTU is observed at a level greater than the user-specified absolute or relative abundance threshold within any of the samples representing a specific group, it is counted as being a member of that group. This method is helpful for cases when the minimal membership core is expected to be small, or when the presence of a microbe alone may indicate a shift in microbiome function, for instance in the detection of microbe-mediated disease.

The Composition method places OTUs within the Venn diagram based on relative abundance, rather than presence/absence (Supplementary Fig. S1B). In this method, each OTU is automatically assigned membership to the group for which it has the highest relative abundance, and membership in other groups is assigned if the relative abundance of that OTU is similar as defined by a user-specified threshold. This method ensures that every OTU will be included in the final analysis, and will only appear in a single overlap within the Venn diagram. This method of analysis may be helpful when there is a large or complete overlap in OTUs that are observed across different groups, but also substantial differences in the relative abundance of these OTUs in the groups being compared that could drive important functional differences.

The Persistence method enables the identification of OTUs that are consistently detected within a microbiome experiment (Shade and Handelsman, 2012). In this method, we consider that the OTU should be present within a user-specified fraction of the samples within the group, in order to be counted as belonging to that group (Supplementary Fig. S1C). This method can be useful for tracking OTUs that are stable across experimental replication, or across time in longitudinal studies.

After applying the methods described above for core microbiome discovery, MetaCoMET generates a new OTU table for each overlapping region of the Venn diagram, and allows for interactive visual exploration of these OTU tables through commonly used statistical techniques. Example analyses that help illustrate the utility of MetaCoMET are available for data from a recent study of the rice microbiome (Edwards *et al.*, 2015).

3 Results

As input, the MetaCoMET analysis accepts an OTU table in BIOM (The Biological Observation Matrix) format (both versions 1.0 or 2.0 supported) containing community abundance data (McDonald *et al.*, 2012). The OTU table can be filtered by OTU or sample abundance, using either an absolute or relative threshold, which allows the user the option of removing low-abundance taxa and samples. Additionally, MetaCoMET allows for prefiltering by OTU taxonomic terms, which enables the user to focus on or exclude OTUs matching specific key words in downstream analysis. MetaCoMET also allows users to normalize their data to account for differences in read depths across their samples or to prefilter their data based on phylogenetic relatedness to collapse closely related OTUs into new lineages.

MetaCoMET has an efficient and interactive graphical interface that will generate a Venn diagram viewer (Supplementary Fig. S2) using jvenn (Bardou *et al.*, 2014) for comparing between two to six user-defined groups within their microbiome data. In addition, MetaCoMET provides functionality for analyzing both the overall dataset, as well as each subset defined by the union or disjunction of groups within the Venn diagram. This functionality utilizes preexisting tools to create a graphical taxonomy summary (Ondov *et al.*, 2011) (Supplementary Fig. S3), abundance-based heatmaps (Skuta

et al., 2014) (Supplementary Fig. S4), Principal Coordinate analysis (Vazquez-Baeza *et al.*, 2013) (Supplementary Fig. S5A), alpha diversity metrics (Caporaso *et al.*, 2010) (Supplementary Fig. S5B), a chart of the geographic distributions of each sample (Supplementary Fig. S6), and a built-in search for identifying keywords within the OTU table (Supplementary Fig. S7). MetaCoMET also integrates the recently publish tool Phinch (Bik, 2014) in the analysis result, adding alternative visualization features including bubble charts, Sankey diagrams and donut partitions (Supplementary Fig. S8), and has been optimized and validated for use with the Google Chrome browser.

4 Conclusion

The interactive Venn diagram viewer in MetaCoMET provides a high-level summary of shared and unique OTUs, as well as the most commonly needed visualization features for comparative exploration of the data. As input, MetaCoMET only requires users to upload a standard OTU table in BIOM file format and to define group membership, and importantly allows the core microbiome analysis to be performed quickly and without the need to install programs locally. As such, MetaCoMET provides researchers with a powerful tool with low overhead for analyzing and interpreting the biological meaning of microbiome data and enhancing ecological understanding.

References

Bardou,P. *et al.* (2014) jvenn: an interactive Venn diagram viewer. *BMC Bioinf.*, **15**, 293.

Bik,H.M. (2014) Phinch: an interactive, exploratory data visualization framework for –Omic datasets. *bioRxiv*.

Caporaso,J.G. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**, 335–336.

Edwards,J. *et al.* (2015) Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, E911–E920.

Hamady,M. and Knight,R. (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res.*, **19**, 1141–1152.

McDonald,D. *et al.* (2012) The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome–ome. *GigaScience*, **1**, 7.

Ondov,B.D. *et al.* (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinf.*, **12**, 385.

Shade,A. and Handelsman,J. (2012) Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.*, **14**, 4–12.

Skuta,C. *et al.* (2014) InCHlib – interactive cluster heatmap for web applications. *J. Cheminf.*, **6**, 44.

Turnbaugh,P.J. *et al.* (2007) The human microbiome project. *Nature*, **449**, 804–810.

Vazquez-Baeza,Y. *et al.* (2013) EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience*, **2**, 16.