

Investigating the microbial community reproducibility: a network-based approach

Matthew Shaun Grainger

February 2024

1 Introduction

Microbial communities, groups of microbial taxa that share and interact within the same environment, underpin all life on Earth (Konopka, 2009; Widder et al., 2016). Together, the microbes within a given community contribute to a range of ecosystem functions. They drive the biogeochemical cycles within Earth's water, soil, and air NEED CITATION. They also directly contribute to disease suppression, soil fertility, plant growth, carbon sequestration, fermentation, and bioremediation NEED CITATION. Because microbial communities have such a range of important functions, there has been considerable interest in controlling them.

There are a number of different ways in which microbial communities can be influenced for the benefit of humans. Probiotics are selections of microbial organisms that are introduced into a target environment, with the intention of directly incorporating them into the resident community NEED CITATION. Prebiotics on the other hand are substances that facilitate the growth of desired microbes, modifying the microbial community more indirectly NEED CITATION. A third approach is via the use of microbial community transplants, which involves the introduction of an entire existing microbial community into a target environment. This has the advantage of preserving microbial interactions and community stability NEED CITATION. All three of these methods rely on the fact that different microbes, and different entire microbial communities, have been linked to different qualities of ecosystem function. NEED CITATION. They also all rely upon the mechanisms of microbial community assembly.

Microbial community assembly needs to be reproducible if microbial communities are to be reliably controlled and applied. For probiotics and whole-community transplants to succeed, the introduced microbes need to be able to assemble into a community within the target environment. For prebiotics to succeed, the addition of a chosen substance into the target environment needs to change the composition and function of the target community. If they are to be reliably used, these methods need to yield the same results each time that

they are employed. For example, applying a given probiotic to a field needs to impact the yield of a target crop in the same way each time that it is used.

Microbial communities are incredibly complex, and they are dynamic across both space and time. Different communities can be found in different habitats across both the macroscale and the microscale NEED CITATION. The microbial community within a given space will also change over time, due to changes in the environment, and stochastic events NEED CITATION. A given microbial community has a huge diversity of microbes, so there may be variation due to chance NEED CITATION. Given all of the factors involved in the determination of microbial communities, it is important to know whether they can be reliably reproduced.

Here, I investigated the reproducibility of microbial communities from tree holes.

2 Methods

2.1 Data

Bacterial amplicon sequence variant (ASVs) abundance table sourced from Pascual-García et al. (2023). In brief, this table contains the abundances of ASVs from 275 bacterial communities within the rainwater pools of beech tree roots (*Fagus sylvatica*). Each community was grown in a sterile beech leaf medium, and incubated at 22°C under static conditions for 1 week so that they could reach stationary phase. The communities were then cryopreserved. Following cryopreservation, each of the 275 communities were revived four times (to provide four replicates), and incubated under static conditions at 22°C for 7 days. Samples of the 275 stationary state communities were taken prior to cryopreservation, and samples of all of the replicates of these 275 communities were taken at the end of the 7 day growth period that followed their revival. The ASV abundance table contains the abundances of NUMBER ASVs within each of the 275 initial samples and in each of the 1100 final samples.

The metadata table that accompanies the ASV abundance table was also sourced from Pascual-García et al. (2023). This metadata table contains information upon both whether each sample is from before cryopreservation or after it, as well as information upon the community class of each sample. In short, these community classes are clusters of communities based upon Jensen-Shannon divergence. They were determined via Partition Around Medoids clustering to reach the classification that showed the highest quality, according to the Calinski-Harabasz index. The bacterial communities clustered in compositional space according to their collection location and date, which means that the different community classes reflect the slightly different environmental conditions from which different tree hole communities originated. In my investigation, therefore, I treat community classes as a proxy for the environment.

2.2 Inferring interactions between ASVs

Split the ASV table and metadata table by pre-cryopreservation (starting) samples and post-cryopreservation (final) samples, converted them into the correct format, and inputted them separately into FlashWeave. FlashWeave is a method for inferring direct interactions between bacteria in heterogeneous systems, via a local-to-global learning approach. It determines which ASVs are directly interacting within a network, and excludes indirect connections between a pair of ASVs that are due to shared connections with other ASVs. FlashWeave has been compared to other state-of-the-art methods for inferring microbial interactions, such as SparCC and SpiecEasi, where it has shown both an increased prediction performance on a variety of synthetic data sets, and an improved reconstruction of expert-curated interactions from the TARA Oceans project (Tackmann et al., 2019). Here, I applied two versions of FlashWeave to both the ASV table from the starting samples and the ASV table from the final samples. The first version of FlashWeave that I applied was FlashWeave-S, intended for homogeneous data that is largely unaffected by different treatments. For this version of FlashWeave, I supplied no metadata table, with the intention of ignoring the presence of community classes. The second version of FlashWeave that I applied to the tables was FlashWeaveHE-S, which is intended to be applied to heterogeneous data with moderate numbers of samples (hundreds to thousands). The community class of each sample was used as the only meta variable. Both versions of FlashWeave output a table of all the direct interactions that are recovered from between all possible pairs of ASVs. This means that ASVs that do not have any inferred direct interactions are excluded from the FlashWeave output tables. For FlashWeaveHE-S, the community classes were included in the table of interactions as if they were ASVs.

2.3 Detecting bacterial communities

Used functionInk to identify communities of bacteria within each network (one network for each application of FlashWeave). This involved converting the interaction tables into the correct format, including via the addition of a 'Type' column. This modified interaction table was then inputted into the detailed pipeline of functionInk, specifying that the interactions were weighted, undirected, and typed. This pipeline identifies ASVs that are structurally equivalent, and groups them into clusters that represent communities. For a pair of ASVs to be structurally equivalent, they need to share the same types of links (interactions) with the same neighbours (Pascual-García and Bell, 2020). Here, different types of links were based upon both whether the interaction was positive or negative, and on whether it was between two ASVs, between a community class and an ASV, or between two community classes. This type was stored in the 'Type' column. There was also an additional type for if there were discrepancies between the networks in terms of whether an interaction was positive or negative.

The functionInk pipeline outputs the cluster that each ASV was grouped

into, as well as a measure of the partition densities of the network. If a network has a higher external partition density, it mainly contains guilds, and if a network has a higher internal partition density, it mainly contains modules. Guilds are communities of bacteria that are clustered due to their shared interactions with bacteria outside of the guild; a high external partition density indicates that there are many links between members of a guild and members of another community. On the other hand, modules are communities of bacteria that interact predominantly with other members of the module, rather than to external communities; a high internal partition density indicates that there are many links between members of the same module (Pascual-García and Bell, 2020). Here, I ran the functionInk pipeline for each network until the step at which the maximum internal partition density was reached.

2.4 Network comparison

Calculated various metrics to describe the four networks. These included the number of ASVs within each network, the number of ASVs shared between each pair of networks, the number of positive and negative interactions within each network, the number of positive and negative interactions shared between each pair of networks, and the number of clusters within each network. Additionally used Cohen’s Kappa to measure the agreement between each pair of networks in terms of which ASVs were present, the sign of the interaction (positive or negative) between pairs of ASVs that were found within both networks, whether a given pair of ASVs that were found within both networks were found within the same cluster, and whether a given pair of ASVs that were positively correlated within both networks were found within the same cluster.

2.5 Identifying the core community

Removed the weightings from the interactions within the interaction tables for the starting sample and final sample networks that were produced by FlashWeaveHE-S, such that they were described as either '1' for positive or '-1' for negative. Combined these interaction tables, added a 'Type' column, and inputted the resulting table into the detailed pipeline of functionInk. Added an additional column for the network(s) within which each interaction was found, and used this to visualise the network in Cytoscape. Similarly added a column for the network(s) in which each ASV was found, as well as information upon whether it was an ASV or a community class node, to the functionInk output file that provides information upon the cluster that each ASV is found in. This was used as the network table for Cytoscape.

Used the network visualisation to identify clusters of bacteria with large numbers of interactions.

3 Results

4 Discussion and Conclusion

improve crop yield, produce pharmaceuticals, food and beverage production, treat wastewater, treat industrial waste, biological control of pests and diseases also contribute to human health, agriculture, and industry. They are involved in soil fertility, plant growth, carbon sequestration, bioremediation, waste treatment, food and drink manufacturing, and disease suppression. NEED CITATION

Needs to include: - Methods used in microbial community ecology - Mention of all the sequencing information available - Microbial communities are dynamic across both space and time... Lots of factors influence them - Mechanisms driving microbial community assembly - Microbial communities consist of bacteria and fungi etc... Here, we focus upon bacteria. - Different microbial communities have been linked to different ecosystem functions. - Microbial community transplants are of increasing interest as a solution to problems in human health, agriculture, and waste remediation. - If microbial community transplants are to be widely utilised, their results need to be reproducible. - Alternative ways of characterising microbial communities (other than co-occurrence networks) - Alternative ways of co-occurrence (SparCC, SpiecEasi etc) - Explain pipeline - how does co-occurrence reflect microbial communities

References

- Konopka, A. (2009). What is microbial community ecology? *The ISME Journal*, 3(11):1223–1230.
- Pascual-García, A., Rivett, D., Jones, M. L., and Bell, T. (2023). Replaying the tape of ecology to domesticate wild microbiota. *bioRxiv*.
- Pascual-García, A. and Bell, T. (2020). functionink: An efficient method to detect functional groups in multidimensional networks reveals the hidden structure of ecological communities. *Methods in Ecology and Evolution*, 11.
- Tackmann, J., Matias Rodrigues, J. F., and von Mering, C. (2019). Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Systems*, 9(3):286–296.e8.
- Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., Brown, S. P., Momeni, B., Shou, W., Kettle, H., Flint, H. J., Haas, A. F., Laroche, B., Kreft, J.-U., Rainey, P. B., Freilich, S., Schuster, S., Milferstedt, K., van der Meer, J. R., Großkopf, T., Huisman, J., Free, A., Picioreanu, C., Quince, C., Klapper, I., Labarthe, S., Smets, B. F., Wang, H., Soyer, O. S., and Fellows, I. N. I. (2016). Challenges in microbial ecology:

building predictive understanding of community function and dynamics. *The ISME Journal*, 10(11):2557–2568.