Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Abundance-occupancy distributions to prioritize plant core microbiome membership

Ashley Shade[1,2,3,4,5] and Nejc Stopnisek[1,3,4]

Core microbiome members are consistent features of a dataset that are hypothesized to reflect underlying functional relationships with the host. A review of the recent plant-microbiome literature reveals a variety of study-specific approaches used to define the core, which presents a challenge to building a general plant-microbiome framework. Abundance-occupancy distributions, used in macroecology to describe changes in community diversity over space, offer an ecological approach for prioritizing core membership for both spatial and temporal studies. Additionally, neutral models fit to the abundance-occupancy distributions can provide insights into deterministically selected core members. We provide examples and code to systematically explore a core plant microbiome from abundance-occupancy distributions. Though we focus on examples from and discussions relevant to the plant microbiome, the abundance-occupancy method can be widely and generally applied to prioritize core membership for any microbiome.

Addresses
[1] Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing MI 48824, United States
[2] Department of Plant Soil and Microbial Sciences, Michigan State University, East Lansing MI 48824, United States
[3] Plant Resilience Institute, Michigan State University, East Lansing MI 48824, United States
[4] Great Lakes Bioenergy Research Center, Michigan State University, East Lansing MI 48824, United States
[5] Program in Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing, United States

Corresponding author: Shade, Ashley (shadeash@msu.edu)

Current Opinion in Microbiology 2019, **49**:50–58

This review comes from a themed issue on **Environmental microbiology**

Edited by **Roeland Berendsen** and **Klaus Schlaeppi**

For a complete overview see the Issue and the Editorial

Available online 10th November 2019

https://doi.org/10.1016/j.mib.2019.09.008

Introduction

Core microbiomes are shared features of microbial communities (e.g. Refs. [1–3]). Starting with a 2011 report of a core microbiome 'similar in composition to animal gut microfaunas' among carnivorous pitcher plants [4], the scientific interest in identifying core plant microbiomes increased scientific interests after a seminal study of *Arabidopsis* roots [5••]. Since then, the concept of a core microbiome has been applied to many plant species grown in numerous contexts, and as defined in a variety of biologically or statistically motivated ways (Table 1, Supporting Table S1 and references therein). Because of their consistency, core microbiome members are inferred to have some importance for host fitness (e.g. Ref. [6]) and therefore potential for managing plant microbiomes towards desired outcomes [7–12]. However, every study design is different, and there are almost as many different ways to define a core microbiome as there are studies. Therefore, a remaining challenge for microbiome research in general, and for plant microbiomes in particular, is to synergize across different studies towards a universal understanding of the core microbiome that can be used to support microbiome prediction and management, across plant species and contexts, and on a changing planet.

There are different motivations for wanting to define a core microbiome (Supporting Table S1). Some studies simply want to know which taxa are most abundant on a host. This core definition can be informed by ranked abundance curves. However, many microbial strains that have functional consequences for plants often are rare (e.g. many plant pathogens), and abundance-defined cores miss important members that are not dominant in the community. Other studies want to know which taxa are always detected on a host or in a condition (e.g. presence/absence). This core definition can be informed by a Venn diagram. However, the preponderance of rare taxa in microbiome datasets will inflate the number of taxa detected in particular growth conditions or genotypes. Therefore, a detection-defined core membership may include many unimportant members that are transient to the conditions of interest. Finally, some studies define core microbiomes by enrichment in particular conditions (e.g. Ref. [13]). Therefore, core members are those that are consistently detected in higher abundance in the focal condition than comparative conditions. This core definition incorporates aspects of detection and abundance and can provide ecological insights to prioritize members that are selected by the host or adapted to the host environment.

There are some intuitive commonalities among the approaches to define a core that have been recently applied to plant microbiomes: they often include some threshold of taxon proportional contribution and/or a

Table 1

Representative studies that have defined core plant microbiomes by abundance, occupancy, and both. Please see supporting Table S1 for a full list of studies and their details of plant species, compartment, and core microbiome definitions

	Plant	Study
Abundance	31 plant species	[55]
	Rice (*Oryza sativa*)	[21]
	Poplar tree (*Populus deltoides*)	[56•]
Occupancy	Maize (*Zea mays*)	[22]
	30 angiosperm species	[57]
	Sugarcane (*Saccharum* spp.)	[58]
Combined	Switchgrass (*Panicum virgatum*) and miscanthus (*Miscanthus* × *giganteus*)	[20]
	Maize (*Zea mays*)	[59]
	Canola (*Brassica napus*)	[60]

minimal detection across samples. In ecology, these intuitive properties are called abundance and occupancy, respectively. Abundance-occupancy distributions offer a consistent, ecologically-motivated approach for prioritizing core membership. In macroecology, species abundance-occupancy distributions are often applied to explore large-scale patterns in taxon distributions ([14], Figure 1). A taxon's mean relative abundance is log10 transformed, and then plotted against the proportion of discrete samples in which it occurs (with occupancy of 1 to be found in all samples). Together, abundance and occupancy provide rich information for interpreting diversity patterns at both population and community levels.

The abundance-occupancy distribution can be used to generate a supported hypothesis of core microbiome membership from any type of community dataset, and to then prioritize those core taxa for subsequent follow-up studies to test their functional importance. Here, we discuss first how the abundance-occupancy distribution may be similarly applied and interpreted to prioritize core membership over space and time (Figure 1). We inform a threshold for core membership by quantifying the contribution of top-ranked members (in occupancy) to beta diversity. We then discuss the use of the neutral model of microbial community assembly to discriminate taxa by community assembly processes underlying their distributions. Fitting a neutral model through the abundance-occupancy data can be useful for interpreting and discussing core membership as well as for hypothesis building. Finally, we end with caveats and limitations.

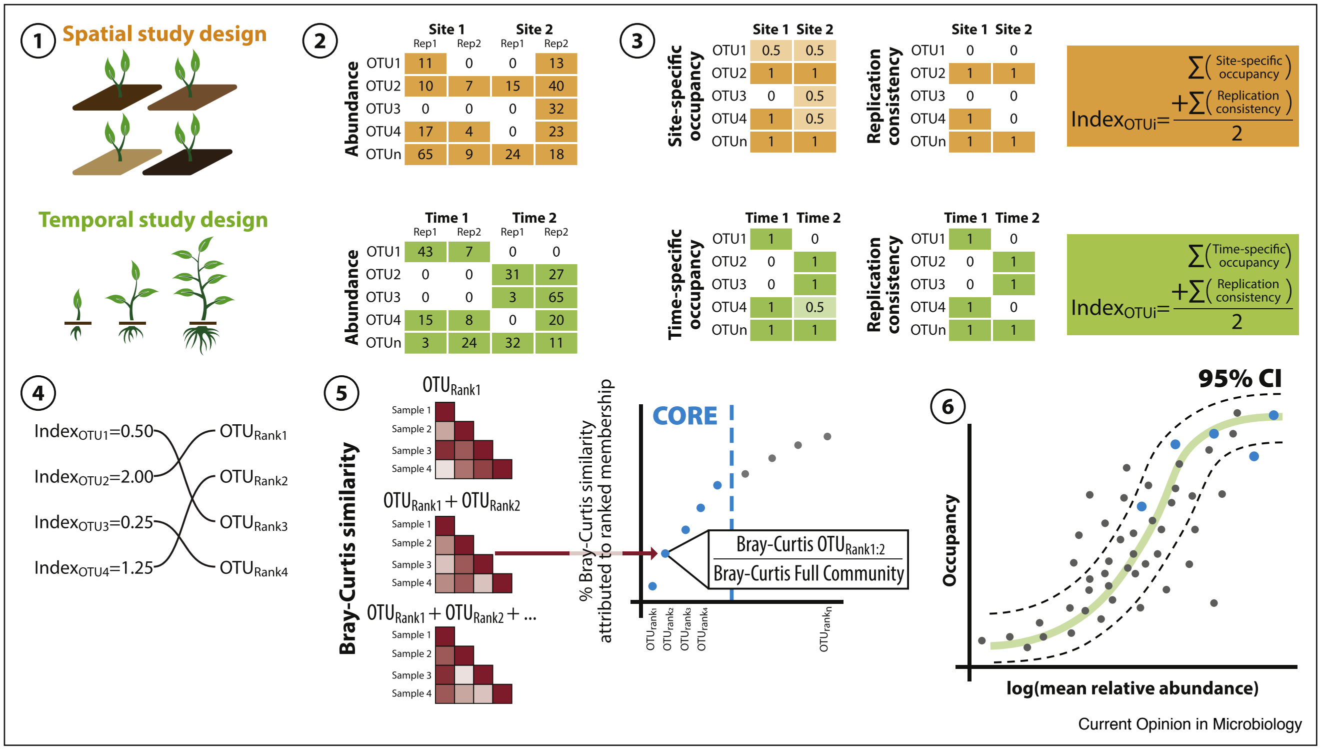
Prioritizing a core microbiome over space

In macroecology, abundance-occupancy distributions are commonly used to think about changes in diversity over space [15•,16,17], though it has not often been applied to think about 'core' taxa, a concept more actively applied and studied in microbiome ecology [5••,18]. In microbiome ecology, there are several study types that have an explicit spatial component and allow for direct transfer of the abundance-occupancy method to the dataset. For plant-microbiome studies in particular, spatial components are inherent because of the high environmental heterogeneity that is sometimes host-imposed (e.g. soils chemistry, root zones) [9,19]. In some plant-microbiome experimental designs, efforts are taken to minimize pot-to-pot, row-to-row or field-to-field variability so that interesting attributes to the factor of interest can be isolated (e.g. plant genotype, management) [20•]. In other studies, characterizing plant-microbiome changes to spatial environmental gradients proffer a biogeographic question in which space itself is the factor of interest [21••,22]. Core microbiomes can be informed from any of these spatial study designs using abundance-occupancy distributions.

The spatial study design, however, is important for the calculation of taxon occupancy. The simplest way to calculate occupancy would be to have all discrete samples contribute equally to the total, expressed as a proportion of 1 (or a percentage out of 100%). This approach is most conservative and will restrict the core to only those taxa that are detected in every sample, and therefore will also be biased towards more abundant taxa (see neutral model discussion, below). However, with replication over space, it is sometimes ecologically relevant to consider the replicates collectively and to eliminate taxa that are inconsistent. Alternatively, occupancy could be viewed as detection within a location/treatment, such that as long as the taxon was represented in each location (though not necessarily in all replicates within that location), it was counted as occurring there. This approach is less conservative and so there may be false positive taxa that, upon functional testing, are later dropped from the core set. However, this approach is also more likely to include taxa that are in, on average, medium to low abundance, depending on the structure of the community.

A systematic exploration of the core is advised to determine the ecologically-supported core membership and

Figure 1



Abundance

	Site 1 Rep1	Site 1 Rep2	Site 2 Rep1	Site 2 Rep2
OTU1	11	0	0	13
OTU2	10	7	15	40
OTU3	0	0	0	32
OTU4	17	4	0	23
OTUn	65	9	24	18

Site-specific occupancy

	Site 1	Site 2
OTU1	0.5	0.5
OTU2	1	1
OTU3	0	0.5
OTU4	1	0.5
OTUn	1	1

Replication consistency

	Site 1	Site 2
OTU1	0	0
OTU2	1	1
OTU3	0	0
OTU4	1	0
OTUn	1	1

$$\text{Index}_{OTUi} = \frac{\sum(\text{Site-specific occupancy}) + \sum(\text{Replication consistency})}{2}$$

Abundance

	Time 1 Rep1	Time 1 Rep2	Time 2 Rep1	Time 2 Rep2
OTU1	43	7	0	0
OTU2	0	0	31	27
OTU3	0	0	3	65
OTU4	15	8	0	20
OTUn	3	24	32	11

Time-specific occupancy

	Time 1	Time 2
OTU1	1	0
OTU2	0	1
OTU3	0	1
OTU4	1	0.5
OTUn	1	1

Replication consistency

	Time 1	Time 2
OTU1	1	0
OTU2	0	1
OTU3	0	1
OTU4	1	0
OTUn	1	1

$$\text{Index}_{OTUi} = \frac{\sum(\text{Time-specific occupancy}) + \sum(\text{Replication consistency})}{2}$$

$$\frac{\text{Bray-Curtis OTU}_{Rank1:2}}{\text{Bray-Curtis Full Community}}$$

Current Opinion in Microbiology

A generalized method to use abundance-occupancy distributions to prioritize core microbiomes over space and time. **(a)** Spatial (orange) and temporal (green) study designs have different considerations and priorities for determining occupancy. **(b)** A taxon table of taxa (operational taxonomic units, OTUs) in rows and samples (communities, including replication) in columns. Presence/absence is shown by shading of the cells, and taxon counts (abundances) are the cell values. **(c)** Occupancy is determined according to study design, considering detection across site or over time, plus replication consistency. Spatial and temporal methods proceed similarly after the indices for ranking are calculated. **(d)** Taxa are ranked by index. **(e)** Bray-Curtis similarity is calculated for the whole dataset, and then also for only the top-ranked taxa. The contribution of the top-ranked taxa is divided by the total Bray-Curtis similarity to calculate a percent contribution of the prospective core set to beta diversity. The next-ranked taxon is added consecutively to find the point in the ranking at which adding one more taxon offers diminishing returns on explanatory value for beta diversity. **(f)** A hypothetical example of the neutral model of community assembly fit to an abundance-occupancy distribution. Each point is a different taxon, and the core taxa are blue large points while non-core taxa are small black points. Each point is a taxon plotted by its mean log10 relative abundance and occupancy, the solid green line is the neutral model, and the dashed gray lines are 95% confidence intervals around the model fit. The points that fall outside the 95% model confidence are inferred to be deterministically, rather than neutrally, selected. Points above the model are selected by the (plant) environment, and points below the model are dispersal limited.

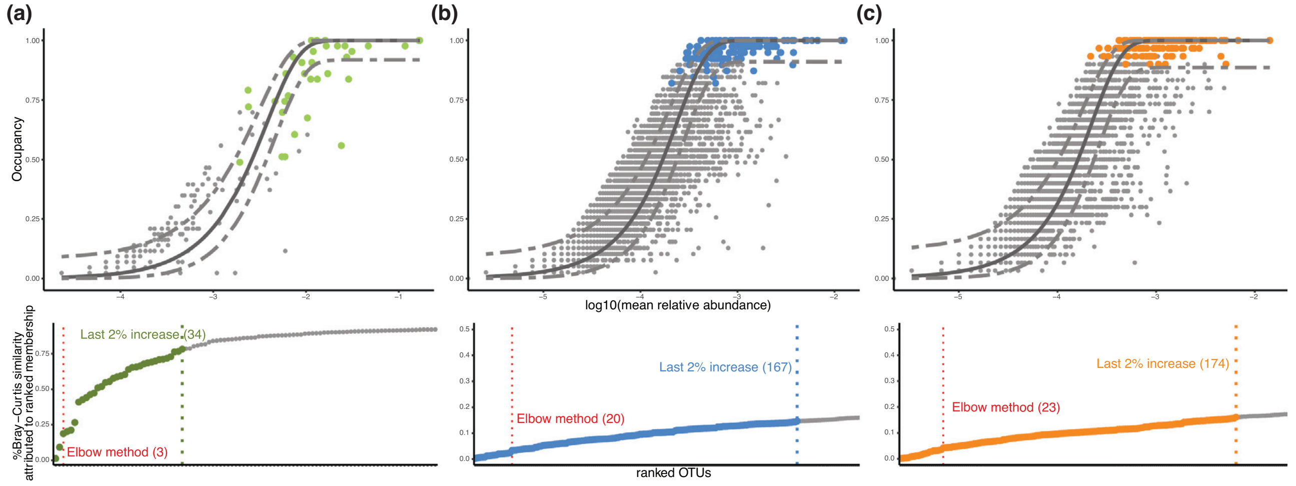
threshold [23], though it should be combined with ecological insight from the experimental design and research question posed. We recommend an iterative exploration to quantify the explanatory value of the core membership for beta diversity (Box 1, Figure 1–5). This process includes ranking taxa by occupancy, consecutively determining the collective contribution of ranked taxa to the beta diversity, and applying a threshold that suggests diminishing returns when adding the next ranked-taxon.

Given the high environmental heterogeneity across soils, another spatial consideration is the range of the biogeographic area included in the study. A global study should expect to harbor more diversity and fewer taxa with an occupancy of 1 than a greenhouse study using a standard potting soil mix. Therefore, while there will not be a one-size-fits-all core inclusion threshold, use and systematic exploration of a core set from abundance-occupancy distributions can provide a consistent approach for prioritizing ecologically important taxa over space and time.

Prioritizing a core microbiome over time

Longitudinal studies in microbiome ecology are becoming more common, and are of particular interest for plants-microbiome because of known changes in phytohormone and root exudate production over plant development [24–26]. Another interesting temporal component is how plants with perennial or biannual life

Box 1 Quantifying the contribution of the core membership to beta diversity



Current Opinion in Microbiology

One way to systematically explore the abundance and occupancy inclusion thresholds used to define the core microbiome is to ask how well the resulting core membership reflects the overarching patterns of the full dataset, or beta diversity. Here, we offer one method and R code to systematically explore the core inclusion criteria. We emphasize that the calculation can be applied to different study designs: a replicated time series (Figure i-A, switchgrass phyllosphere [20]); a genotype-defined core (Figure i-B, monkey flower ecotype recruitment in a reciprocal transplant experiment [48]); and a site-defined core (Figure i-C, a biogeography study of common bean planted across U.S. growing regions, unpublished data from the authors). In each case, abundance-occupancy distributions are fit to the neutral model. The solid line show the model fit with dashed lines showing the 95% confidence intervals. Points highlighted in green, blue or orange are those found to belong to the spatial, genotypic or temporal plant core, respectively. The R code for the neutral model is from Ref. [32].

Taxa are first ranked by occupancy according to the study design, and there is an option to also include a weight for taxon abundance. Then, the contribution of a subset of core taxa to beta diversity is quantified by calculating the proportion of total community resemblance that is attributable to the core subset. We use Bray-Curtis (Eq. (1), see also Ref. [40]), but other resemblances are interchangeable with the note that different resemblances emphasize different aspects of beta diversity and should be considered carefully.

Finally, to identify a point in which an increase in the core inclusion threshold will offer marginal return in explanatory value, we offer two methods: a more stringent 'elbow' approach (first-order difference, see explanation below) and a final percent increase in beta-diversity (we recommend 2% or more).

The cumulative explanatory value with adding the next-ranked taxon into the core (e.g. as in Figure 1 part 5) are shown for each of the datasets in Figure i. The red line distinguishes core taxa by the first-order difference and the green, blue or orange line by the last 2% increase in explanatory value by Bray-Curtis similarity. In Equation 1, C is the ontribution of a set of ranked taxa (by occupancy/abundance) to the total Bray-Curtis similarity, and BC is Bray-Curtis, designated as including all OTUs or only a hypothesized core set.

Equation 1

$$C = 1 - \frac{BC_{OTU_{core}}}{BC_{OTU_{all}}}$$

The method to find the 'elbow' of a graph is based on the first-order difference (from numerical differentiation). In this method, a score is assigned to each cutoff point by separating the curve into two parts and calculating the difference in the average rates of change for both of these parts. Then, the elbow point is selected to be the cutoff point that maximizes this score.

The script for the analysis, including elbow and percent increase methods for selecting a core, can be found at https://github.com/ShadeLab/PAPER_Shade_CurrOpinMicro.

histories may differently recruit and maintain core microbiome members from plants with annual strategies [27]. In addition, some plant compartments, like epiphytic phyllosphere surfaces, may be expected to experience greater environmental fluctuations over time (seasons) than other compartments, like root endophytes [21••]. However, there is much work to be done to understand how the microbiota that live in these compartments change temporally, and, because destructive sampling is typically required, there can be logistical challenges in executing temporal replication within a study's design. Therefore, quantitative temporal studies of core plant-microbiota are at their onset and the field is relatively open.

Abundance-occupancy distributions have been applied in other systems to identify core microbiota over time. 'Persistent' and 'transient' were used to categorize microbiome members by occupancy bins over a daily time series in the human gut [28]. A neutral model was applied to abundance-occupancy distributions that were calculated from data collected over discrete time points in zebrafish development [29]. Given the need to understand microbiome changes over plant development or with seasonal changes, these approaches are conceptually aligned with goals in plant-microbiome research.

As for calculating occupancy over space, there are considerations for collapsing the occupancy axis to include only the most ecologically relevant series for the question rather than the entire sample set. This is, arguably, more important for defining core microbiome over time because occupancy is indicative of persistence: which taxa are most consistently detected in association with the plant, even as the environment changes over time. Persistence is a key ecological consideration for the maintenance of system stability, which is often a translational outcome desired in understanding core microbiota. In addition, the temporal dynamics of core plant microbiota offer important baseline series against which to compare disturbances (e.g. for crops, growing seasons with flood, drought, or other management changes).

We have recently used abundance-occupancy distributions to prioritize a core in the perennial switchgrass phyllosphere (epiphytes), grown in the field for up to two seasons in a replicated block design [20•]. We discuss this work in detail to show how the occupancy calculation can be ecologically informed. Because we were interested in the seasonal trends in microbiome structure, replicated plots were crucial for identifying consistent core members, and occupancy was conservatively defined to include taxa detected in all replicate fields at a given time point. Defining the core set as having an occupancy of 1 at any time point accounted for the seasonal aspect of the dataset, and then we were able to observe which of those members were also persistent over the growing season and across two years. We found that much of the core membership had high mean abundance and also were persistent seasonally, and that it was the collective dynamics of these core members that defined the temporal patterns in beta-diversity for the study (Box 1). In summary, setting a core microbiome occupancy threshold of 1 for a longitudinal study is likely overly conservative when environmental changes occur, but a replicated time series can offer support for prioritizing core taxa that are key at particular times and under particular seasonal conditions.

Neutral model to inform core taxa that are deterministically assembled

The general shape of the abundance-occupancy distribution is expected to be an 'S', with the most abundant taxa exhibiting highest occupancies and the rare taxa exhibiting lowest occupancies (Figure 1). This is the 'neutral' expectation of the data that assumes unlimited dispersal of microbial cells and no fitness differences among taxa. This neutral model was discussed in detail and brought to the attention of many microbial ecologists by Sloan and Curtis in their theoretical explorations and applications of abundance-occupancy distributions from wastewater treatment microbial communities [30,31].

The neutral model is used as a null hypothesis of community assembly, and is often tested in parallel with deterministic models to ask which model best describes the overall community assembly. However, the neutral model can offer an interesting extension to prioritizing core microbiota because it can discriminate taxa that are expected to be deterministically selected from taxa that are expected to be neutral (e.g. occupancy as expected, given mean abundance). Taxa that fall outside of the confidence intervals meet a deterministic expectation (Figure 1). In particular, taxa that fall above the neutral model prediction are in higher occupancy than expected by their mean abundance in the dataset, and taxa that fall below the prediction are in higher abundance than expected by their occupancy. For plant microbiomes, we may interpret this to form a hypothesis that, in the former case, the plant environment selects for these taxa, and, in the latter case, that these taxa are dispersal limited. Parallel interpretation was previously applied in the developing zebra gut, human lungs and skin, as well as from a wider range of hosts and environments [32••,33,34,35,36,37,38]. Among the core set of taxa selected by abundance and occupancy, those that are deterministically selected above the model fit (selected by the plant environment) can be given heighten priority for robust colonization efficiency and/or the potential to have interactions with the plant. Therefore, we recommend exploring the assembly mechanisms of core taxa to glean deeper ecological insights into their distributions and to guide functional follow-up studies.

Considerations and limitations

We have focused on the application of abundance-occupancy distributions to detect core microbiomes from marker gene amplicon datasets, such as the 16S rRNA gene or the ITS region. The motivation for this focus is because of the continuing utility of marker genes for large-scale and cross-system analyses to inform generalities in the core plant microbiome (Box 2). However, a consideration when using this approach concerns

Box 2 A case for continued analyses of marker genes for prioritizing the core

High-throughput sequencing is a standard method used to assess cultivation-independent microbial community structure based on phylogenetically-informative marker genes (e.g. 16S rRNA, ITS or 18S rRNA genes). Because of multiplexing and high-yield sequence output, marker gene sequencing studies are increasing in the complexity of their designs and the number of samples included for comparison. Microbiome comparisons are being conducted at unprecedented spatial and temporal scales, and including hundreds to thousands of genotypes with genome-wide association studies [49]. To this end, marker gene datasets, and especially 16S rRNA gene amplicon datasets, remain the most numerous and broad as far as inclusivity to various plant species, genotypes, soils, conditions, and ecosystems. While these cannot speak directly to microbial functionality *in situ* [50], there is evidence that particular microbial traits can be very phylogenetically conserved [51]. Regardless, marker gene studies remain the most useful for large-scale analyses and meta-analyses. Meta-analyses of datasets combined across different studies offer a key strategy for gaining macroecological insights into the causes and consequences of plant-microbiomes diversity.

Recent works suggest that there are a suite of functional genes commonly harbored by plant microbiomes that are of importance for plant-microbe interactions [52,53••,54]. Therefore, we suggest using abundance-occupancy distributions from marker gene studies to prioritize members that are likely to have functional importance. An advantage is that marker gene studies often offer deeper observation of taxa that are in low abundance (rare) than what can be covered by metagenomes, allowing for a window into identifying core taxa that are persistent but not highly abundant.

After using a marker gene study to prioritize a core, it is important then to follow-up with studies to confirm their functionality. These experiments include isolation and construction of synthetic ‘core’ communities, and also direct plant-microbe experiments to quantify any benefit to the host in both unstressed and stressed conditions. Comparative genomics on sets of consistently defined core members will provide insights into their common functional genes and interaction mechanisms. Taken together, marker gene analysis remains a first step in interrogation of the core plant microbiome.

informed interpretation of the core given inclusion criteria without functional data. For example, taxa that lay just outside of abundance-occupancy thresholds for core inclusion may later prove to be functionally important. Researchers will be most confident in the inclusion of core taxa that have the highest abundance or occupancy, simply because there are more robust patterns apparent of the taxa that are observed more frequently. Taxa that conditionally contribute to the system may not meet the inclusion criteria for ‘core’ based on their abundance-occupancy, but we suggest using complementary approaches to identify these responsive taxa [39,40], especially for longitudinal studies or after stress.

Additionally, taxa in high occupancy may be contaminants that were introduced during sampling or processing. Proper controls, especially for low biomass samples like the leaf endophyte, flowers, and seeds, are required to discriminate core taxa from contaminants [41,42•]. Also, core taxa in high occupancy should be compared to taxa found in bulk soils. It is expected that bulk soils are a reservoir of plant microbiome diversity, and so detection of potentially core plant taxa in bulk soil does not invalidate their potential importance for the plant. Statistical analyses that discriminate differences in relative abundances of high occupancy taxa between bulk soil and plant environments can provide deeper insight as to their specificity to the plant. Discriminant analyses (e.g. Ref. [43]), indicator taxon analysis [44], and many other statistical methods can be used to assess taxa abundance differences between bulk soils and plants.

The way that the operational taxonomic units are defined will ultimately influence the size of the core. There has been a rise in popularity of defining taxa at 100% sequence identity rather than 97% sequence identity [45,46]. For simplicity, we will refer to these as zero-radius OTUs (zOTUs), but there are other terms applied to these clusters (amplicon sequence variants, exact sequence variants). On one hand, using zOTUs may decrease the number of taxa observed at the full occupancy of 1, but on the other hand, it may also allow for redundant taxa within the core from sequencing errors; these zOTUs that would have otherwise been clustered with the parent sequence at 97% identity. Without compromising on rigorous quality control of the raw sequencing data, we recommend using a lower identity threshold to first cluster (e.g. 97% sequence identity) and for the initial exploration of abundance-occupancy distributions. Then, those core taxa that are found to contribute substantially to beta diversity can be subset and re-clustered at 100% to determine potential strain-level variation and their distributions for each core taxon. The benefit of this approach is that zOTUs that are derived from the same 97% OTU and exhibit identical dynamics can be collapsed together as an ecotype. This does not mean that the collapsed zOTUs are identical, nor confirm the one zOTU is a sequencing error of the parent zOTU, but it does mean that these taxa have redundant ecological patterns in association with the plant. Collapsing them reduces redundancy in the dataset and supports focus on an ecological core membership. Because it has been shown that using zOTUs or OTUs defined at 97% sequence identity often reveal the same overarching patterns in comparative diversity [47], it is expected that either taxonomic definition will adequately inform a core subset that can be prioritized for functional studies.

In our review of the literature for Supporting Table S1, we noticed that some studies did not include the analysis of the core microbiome as part of the description of the statistical methods, often mentioning the inclusion criteria almost as an aside in the results section. Applying abundance-occupancy distributions to discover a core emphasizes the need for careful statistical consideration and transparency about the underlying data structures for the inclusion criteria, and potentially will lead to improved oversight and rigor in the methods. Making the code and statistical workflows available as to exactly how the core was defined (Box 1), and including these details in the statistical methods sections of manuscripts, will increase reproducibility across studies and support general understanding of core microbiomes.

Conclusions

Members of the core microbiome can be discovered and prioritized from amplicon sequencing datasets using the abundance-occupancy distributions borrowed from macroecology. These distributions can be similarly applied to both spatial and temporal study designs, and then combined with neutral models to improve hypotheses about the drivers of core microbiome assembly, to interpret mechanisms that select for core taxa, and to inform discussion and interpretation of the core. Abundance-occupancy distributions allow researchers to discover and prioritize members of the core microbiome based on their underlying ecology. They can offer a systematic first step towards understanding the functional importance of the core for the plant. More generally, this framework from macroecology can offer strategies to predict and understand core microbiomes [15•–17].

Conflict of interest statement

Nothing declared.

Acknowledgements

This material is based upon work supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409, and in part by the Michigan State University Plant Resilience Institute. AS acknowledges support from the USDA National Institute of Food and Agriculture and Michigan State University AgBioResearch.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.mib.2019.09.008.

1. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP *et al.*: **A core gut microbiome in obese and lean twins**. *Nature* 2009, **457**:480-484.

2. Shade A, Handelsman J: **Beyond the Venn diagram: the hunt for a core microbiome**. *Environ Microbiol* 2012, **14**:4-12.

3. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N *et al.*: **Moving pictures of the human microbiome**. *Genome Biol* 2011, **12**:R50.

4. Koopman MM, Carstens BC: **The microbial phyllogeography of the carnivorous plant *Sarracenia alata***. *Microb Ecol* 2011, **61**:750-758.

5. •• Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, del Rio TG *et al.*: **Defining the core Arabidopsis thaliana root microbiome**. *Nature* 2012, **488**:86-90.

This study popularized the conceptual framework of core microbiomes to plants.

6. Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A: **The importance of the microbiome of the plant holobiont**. *New Phytol* 2015, **206**:1196-1206.

7. Orozco-Mosqueda M del C, Rocha-Granados M del C, Glick BR, Santoyo G: **Microbiome engineering to improve biocontrol and plant growth-promoting mechanisms**. *Microbiol Res* 2018, **208**:25-31.

8. Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, Morsy M, Eisen JA, Leach JE, Dangl JL: **Research priorities for harnessing plant microbiomes in sustainable agriculture**. *PLoS Biol* 2017, **15**:e2001793.

9. Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, Fukuda S, Ushio M, Nakaoka S, Onoda Y *et al.*: **Core microbiomes for sustainable agroecosystems**. *Nat Plants* 2018, **4**:247-257.

10. Gopal M, Gupta A: **Microbiome selection could spur next-generation plant breeding strategies**. *Front Microbiol* 2016, **7**:1971.

11. Finkel OM, Castrill G, Herrera Paredes S, Salas González I, Dangl JL: **Understanding and exploiting plant beneficial microbes**. *Curr Opin Plant Biol* 2017, **38**:155-163 http://dx.doi.org/10.1016/j.pbi.2017.04.018.

12. Berg G, Rybakova D, Grube M, Köberl M: **The plant microbiome explored: implications for experimental botany**. *J Exp Bot* 2016, **67**:995-1002.

13. Schlaeppi K, Dombrowski N, Oter RG, Ver Loren Van Themaat E, Schulze-Lefert P: **Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives**. *Proc Natl Acad Sci U S A* 2014, **111**:585-592.

14. Gaston KJ, Blackburn TM, Greenwood JJD, Gregory RD, Quinn RM, Lawton JH: **Abundance-occupancy relationships**. *J Appl Ecol* 2000, **37**:39-59.

15. • Shade A, Dunn RR, Blowes SA, Keil P, Bohannan BJM, Herrmann M, Küsel K, Lennon JT, Sanders NJ, Storch D *et al.*: **Macroecology to Unite All Life, Large and Small**. *Trends Ecol Evol* 2018, **33**:731-744 http://dx.doi.org/10.1016/j.tree.2018.08.005.

This work discusses macroecological models and concepts that can be applied generally to microbiome ecology.

16. Lennon JT, Locey KJ: **Macroecology for microbiology**. *Environ Microbiol Rep* 2017 http://dx.doi.org/10.1111/1758-2229.12512.

17. Barberán A, Casamayor EO, Fierer N: **The microbial contribution to macroecology**. *Front Microbiol* 2014 http://dx.doi.org/10.3389/fmicb.2014.00203.

18. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP *et al.*: **A core gut microbiome in obese and lean twins**. *Nature* 2009, **457**:480-484.

19. Kuzyakov Y, Razavi BS: **Rhizosphere size and shape: temporal dynamics and spatial stationarity**. *Soil Biol Biochem* 2019, **135**:343-360.

20. • Grady KL, Sorensen JW, Stopnisek N, Guittar J, Shade A: **Assembly and seasonality of core phyllosphere microbiota on perennial biofuel crops**. *Nat Commun* 2019, **10** http://dx.doi.org/10.1101/446369.

This study presents a replicated time series to identify a seasonally persistent and dynamic core leaf microbiome on a perennial biofuel grass.

21. •• Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V: **Structure, variation, and assembly of the root-associated microbiomes of rice**. *Proc Natl Acad Sci U S A* 2015, **112**:E911-20.

This study investigated colonization and assembly to the rice roots and also included different plant compartments to identify a core microbiome based on differential abundances of taxa.

22. Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, González-Peña A, Peiffer J, Koren O, Shi Q *et al.*: **Large-scale replicated field study of maize rhizosphere identifies heritable microbes**. *Proc Natl Acad Sci U S A* 2018, **115**:7368-7373.

23. Astudillo-García C, Bell JJ, Webster NS, Glasl B, Jompa J, Montoya JM, Taylor MW: **Evaluating the core microbiota in complex communities: a systematic investigation**. *Environ Microbiol* 2017, **19**:1450-1462.

24. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, Cho H, Karaoz U, Loqué D, Bowen BP *et al.*: **Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly**. *Nat Microbiol* 2018, **3**:470.

25. Chaparro JM, Badri DV, Vivanco JM: **Rhizosphere microbiome assemblage is affected by plant development**. *ISME J* 2013, **8**:790-803.

26. Staley C, Ferrieri AP, Tfaily MM, Cui Y, Chu RK, Wang P, Shaw JB, Ansong CK, Brewer H, Norbeck AD *et al.*: **Diurnal cycling of rhizosphere bacterial communities is associated with shifts in carbon metabolism**. *Microbiome* 2017, **5**:65.

27. Dombrowski N, Schlaeppi K, Agler MT, Hacquard S, Kemen E, Garrido-Oter R, Wunder J, Coupland G, Schulze-Lefert P: **Root microbiota dynamics of perennial *Arabis alpina* are dependent on soil residence time but independent of flowering time**. *ISME J* 2017, **11109**:43-55.

28. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N *et al.*: **Moving pictures of the human microbiome**. *Genome Biol* 2011, **12**.

29. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF, Doebeli M, Falkowski PG, Fenchel T *et al.*: **High taxonomic variability despite stable functional structure across microbial communities**. *Nat Ecol Evol* 2016, **1**:0015.

30. Curtis TP, Sloan WT: **Towards the design of diversity: stochastic models for community assembly in wastewater treatment plants**. *Water Sci Technol* 2006, **54**:227-236.

31. Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP: **Modeling taxa-abundance distributions in microbial communities using environmental sequence data**. *Microb Ecol* 2007, **53**:443-455.

32. •• Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannan BJ: **Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development**. *ISME J* 2016, **10**:655-664.

This study developed R code for fitting the neutral model to a developmental series of zebrafish, and it can be generally applied to other microbiomes.

33. Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, Schmidt TM: **Application of a neutral community model to assess structuring of the human lung microbiome**. *mBio* 2015, **6**.

34. Tong X, Leung MHY, Wilkins D, Cheung HHL, Lee PKH: **Neutral processes drive seasonal assembly of the skin mycobiome**. *mSystems* 2019, **4**.

35. Sieber M, Pita L, Weiland-Bräuer N, Dirksen P, Wang J, Mortzfeld B, Franzenburg S, Schmitz RA, Baines JF, Fraune S *et al.*: **Neutrality in the Metaorganism**. *PLoS Biol* 2019, **17**:e3000298.

36. Zeng Q, Rodrigo A: **Neutral models of short-term microbiome dynamics with host subpopulation structure and migration limitation**. *Microbiome* 2018, **6**:80.

37. Li L, Ma ZS: **Testing the neutral theory of biodiversity with human microbiome datasets**. *Sci Rep* 2016, **6**.

38. Ofiţeru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, Sloan WT: **Combined niche and neutral effects in a microbial wastewater treatment community**. *Proc Natl Acad Sci U S A* 2010, **107**:15345-15350.

39. Shade A, Gilbert JA: **Temporal patterns of rarity provide a more complete view of microbial diversity**. *Trends Microbiol* 2015, **23**:335-340.

40. Shade A, Jones SESE, Caporaso JG, Handelsman JJ, Knight R, Fierer N, Gilbert JAAJA, Gregory Caporaso J, Handelsman JJ, Knight R *et al.*: **Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity**. *mBio* 2014, **5**:e01371-14.

41. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, McWeeney S: **Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments**. *mSystems* 2019, **4**:e00290-19.

42. • Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS: **Contamination in low microbial biomass microbiome studies: issues and recommendations**. *Trends Microbiol* 2019, **27**:105-117.

An important read about microbiome controls and quality checks for researchers working with low biomass samples.

43. Anders S, Huber W: **Differential expression analysis for sequence count data**. *Genome Biol* 2010, **11**:R106.

44. De Cáceres M, Legendre P: **Associations between species and groups of sites: indices and statistical inference**. *Ecology* 2009, **90**:3566-3574.

45. Callahan BJ, McMurdie PJ, Holmes SP: **Exact sequence variants should replace operational taxonomic units in marker-gene data analysis**. *ISME J* 2017, **11**:2639-2643 http://dx.doi.org/10.1038/ismej.2017.119.

46. Edgar RC: **Updating the 97% identity threshold for 16S ribosomal RNA OTUs**. *Bioinformatics* 2018, **34**:2371-2375 http://dx.doi.org/10.1093/bioinformatics/bty113.

47. Glassman SI, Martiny JBH: **Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units**. *mSphere* 2018, **3**:e00148-18.

48. Bowsher AW, Kearns PJ, Popovic D, Lowry DB, Shade A: **Locally-adapted Mimulus ecotypes differentially impact rhizosphere bacterial and archaeal communities in an environment-dependent manner**. *bioRxiv* 2019 http://dx.doi.org/10.1101/652883.

49. Beilsmith K, Thoen MPM, Brachi B, Gloss AD, Khan MH, Bergelson J: **Genome-wide association studies on the phyllosphere microbiome: embracing complexity in host–microbe interactions**. *Plant J* 2019, **97**:164-181.

50. Lemanceau P, Blouin M, Muller D, Moënne-Loccoz Y: **Let the Core Microbiota Be Functional**. *Trends Plant Sci* 2017, **22**:583-595 http://dx.doi.org/10.1016/j.tplants.2017.04.008.

51. Martiny AC, Treseder K, Pusch G: **Phylogenetic conservatism of functional traits in microorganisms**. *ISME J* 2013, **7**:830-838 http://dx.doi.org/10.1038/ismej.2012.160.

52. Levy A, Conway JM, Dangl JL, Woyke T: **Elucidating Bacterial Gene Functions in the Plant Microbiome**. *Cell Host Microbe* 2018, **24**:475-485 http://dx.doi.org/10.1016/j.chom.2018.09.005.

53. •• Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F *et al.*: **Genomic features of bacterial adaptation to plants**. *Nat Genet* 2018, **50**:138-150 http://dx.doi.org/10.1038/s41588-017-0012-9.

This study identified a suite of common functional genes harbored by plant-associated microbes, and can inform targeted genome and metagenomes analyses to investigate the functional potential of the core.

54. Bruto M, Prigent-Combaret C, Muller D, Moënne-Loccoz Y: **Analysis of genes contributing to plant-beneficial functions in plant growth-promoting rhizobacteria and related Proteobacteria**. *Sci Rep* 2015, **4**:6261.

55. Yeoh YK, Dennis PG, Paungfoo-Lonhienne C, Weber L, Brackin R, Ragan MA, Schmidt S, Hugenholtz P: **Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence**. *Nat Commun* 2017, **8**:215.

56. • Timm CM, Carter KR, Carrell AA, Jun S-R, Jawdy SS, Vélez JM, Gunter LE, Yang Z, Nookaew I, Engle NL *et al.*: **Abiotic Stresses Shift Belowground Populus-Associated Bacteria Toward a Core Stress Microbiome**. *mSystems* 2018, **3**:e00070-17.

This study used many complementary plant- and microbe- centric approaches to determine the stress microbiome (microbial amplicon, leaf RNAseq and metabolomics, bacterial isolation).

57. Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ: **Assembly and ecological function of the root microbiome across angiosperm plant species**. *Proc Natl Acad Sci U S A* 2018, **115**:E1157-E1165.

58. Hamonts K, Trivedi P, Garg A, Janitz C, Grinyer J, Holford P, Botha FC, Anderson IC, Singh BK: **Field study reveals core plant microbiota and relative importance of their drivers**. *Environ Microbiol* 2018, **20**:124-140.

59. CHEN L, XIN X, ZHANG J, REDMILE-GORDON M, NIE G, WANG Q: **Soil Characteristics Overwhelm Cultivar Effects on the Structure and Assembly of Root-Associated Microbiomes of Modern Maize**. *Pedosphere* 2019, **29**:360-373.

60. Lay CY, Bell TH, Hamel C, Harker KN, Mohr R, Greer CW, Yergeau É, St-Arnaud M: **Canola root-Associated microbiomes in the Canadian Prairies**. *Front Microbiol* 2018:9.