Title

Defining and quantifying the core microbiome: Challenges and prospects

Journal

Proceedings of the National Academy of Sciences of the United States of America, 118(51)

ISSN

0027-8424

Authors

Neu, Alexander T
Allen, Eric E
Roy, Kaustuv

Publication Date

2021-12-21

DOI

10.1073/pnas.2104429118

Peer reviewed

PERSPECTIVE

Defining and quantifying the core microbiome: Challenges and prospects

Alexander T. Neu[a,1], Eric E. Allen[b,c], and Kaustuv Roy[a,1]

Edited by Edward F. DeLong, University of Hawai'i at Mānoa, Honolulu, HI, and approved October 27, 2021 (received for review April 7, 2021)

The term "core microbiome" has become widely used in microbial ecology over the last decade. Broadly, the core microbiome refers to any set of microbial taxa, or the genomic and functional attributes associated with those taxa, that are characteristic of a host or environment of interest. Most commonly, core microbiomes are measured as the microbial taxa shared among two or more samples from a particular host or environment. Despite the popularity of this term and its growing use, there is little consensus about how a core microbiome should be quantified in practice. Here, we present a brief history of the core microbiome concept and use a representative sample of the literature to review the different metrics commonly used for quantifying the core. Empirical analyses have used a wide range of metrics for quantifying the core microbiome, including arbitrary occurrence and abundance cutoff values, with the focal taxonomic level of the core ranging from phyla to amplicon sequence variants. However, many of these metrics are susceptible to sampling and other biases. Developing a standardized set of metrics for quantifying the core that accounts for such biases is necessary for testing specific hypotheses about the functional and ecological roles of core microbiomes.

core microbiome | microbiota | microbial ecology | 16S ribosomal RNA gene

The search for the core microbiome has become widespread within the field of microbial ecology. In general, a core microbiome can be defined as any set of microbial taxa, as well as the associated genomic or functional attributes characteristic of a specific host or environment (1–3). This has led to a number of studies focused on the genes (e.g., ref. 4), functional pathways (e.g., ref. 5), and metabolic profiles (e.g., ref. 6) common to microbial communities in a number of environments. Most commonly, however, the search for the core microbiome involves determining which taxa, if any, are shared among two or more microbial communities in a given host species or environment (7). These shared taxa are hypothesized to represent the most ecologically and functionally important microbial associates of that host or environment under the conditions sampled. In fact, it has been suggested that identifying core microbiome components may assist in addressing topics ranging from the maintenance of human oral and gut health (8, 9) to the responses of organisms to anthropogenic climate change (10, 11). The potential utility of these core taxa has led researchers to identify core microbiomes in a wide range of environments and hosts, from the skins of frogs (12) to the Baltic Sea (13) to activated sludge (14), resulting in a rapid increase in the number of studies that include a core microbiome component over the past decade (Fig. 1).

Although analyses of the core taxonomic microbiome have provided a number of insights into the microbial ecology of a multitude of environments and hosts (8, 15–17), they often vary in their criteria for quantifying the core. In general, this involves determining the proportion of samples that share a set of microbial taxa, the relative abundances of shared taxa across samples or hosts, or a combination of the two. The taxonomic level used to define the core can also vary, as a core microbiome may be determined at the level of amplicon sequence

[a]Section of Ecology, Behavior, and Evolution, Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093; [b]Section of Molecular Biology, Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093; and [c]Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093

Author contributions: A.T.N., E.E.A., and K.R. designed research; A.T.N. performed research; A.T.N. analyzed data; and A.T.N., E.E.A., and K.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: aneu@ucsd.edu or kroy@ucsd.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2104429118/-/DCSupplemental.

Published December 3, 2021.

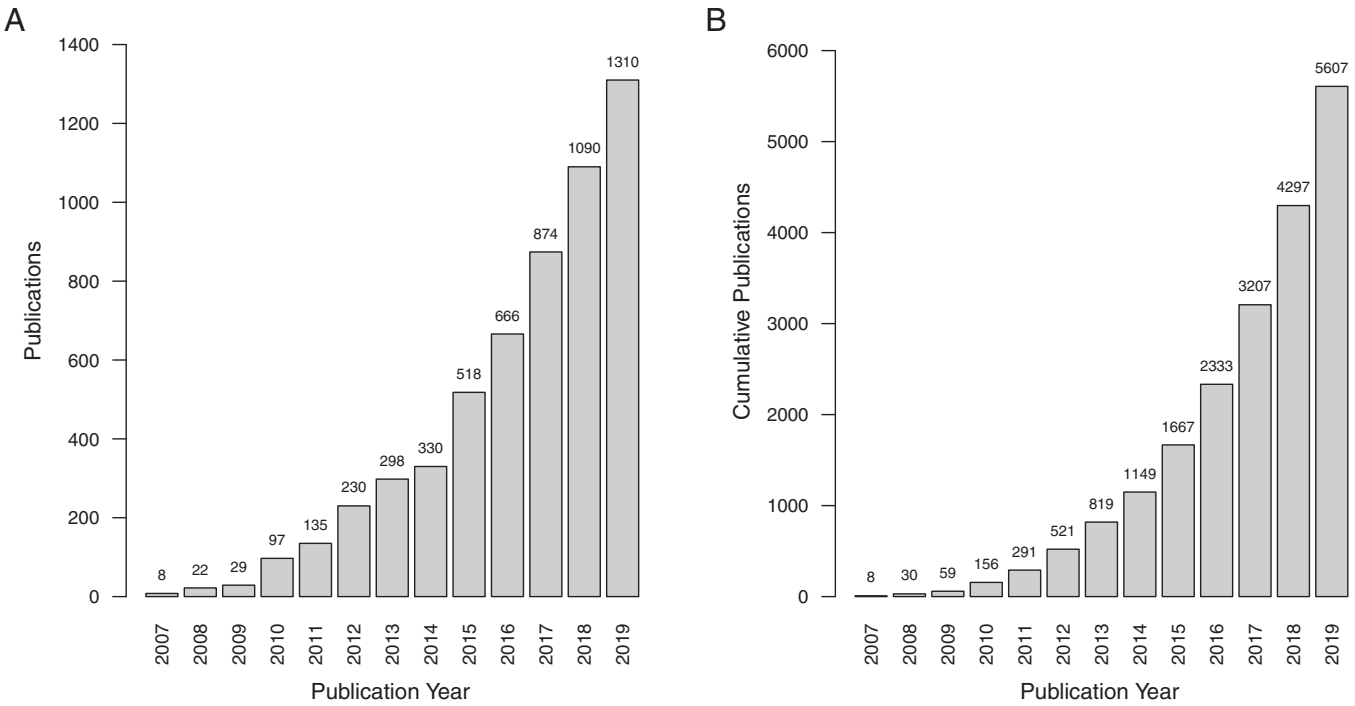

Fig. 1. Number of publications per year (*A*) and cumulative publications (*B*) in Google Scholar including the terms "core microbiome" or "core microbiota" from the introduction of the term in 2007 to 2019 (search conducted on December 22, 2020).

variants (ASVs, sequences clustered at 100% sequence similarity), phyla, or anywhere in between. Furthermore, the spatial and temporal scales over which a core microbiome is quantified are also variable, ranging from samples collected at a single site to a global sampling of a particular host taxon, and from samples collected at a single time point to multiple collections spanning days to years.

Here, we explore the diversity of methods that have been used to quantify core microbiomes, determine which practices are most widely used in the literature, benefits and challenges of this methodological variability, as well as implications for understanding the ecological and evolutionary processes that produce and maintain core microbiomes. To do this, we searched Google Scholar using the terms "core microbiome" or "core microbiota" and limited the search to studies published between 2008 and 2019 (search conducted on March 18, 2020). We selected the first 200 primary research articles (excluding reviews and computational methods), which quantified a core microbiome of bacterial and archaeal taxa using 16S rRNA gene amplicon sequencing, a common method for determining the diversity and composition of prokaryotic communities. We then supplemented these articles with those previously downloaded into our own personal reference libraries, using the same date range and methodological criteria. This resulted in a representative sample of 224 studies published between 2008 and 2019 that involve an analysis of the core microbiome (Fig. 2*A* and Dataset S1). These studies are distributed across a variety of different plant and animal hosts, environments (e.g., soil, seawater), and industrial processes (e.g., wastewater treatment, industrial fermentation), which allowed us to further explore the methodologies for quantifying core microbiomes across subfields of microbial ecology (Fig. 2*B*). Although some studies within our sample set also determine a core fungal microbiome or core functional genes (Dataset S1), we have limited our analysis of the literature to core prokaryotic taxa, as these comprise the majority of published empirical studies. However, many of the major points discussed below are also applicable to fungal taxa or functional core microbiomes. We also note that most of the studies referenced here focus on the "core microbiota," which includes the microbial taxa within a particular environment, rather than the "core microbiome," which includes the structure and function of the community as well as the abiotic conditions in their particular environment (18). While such distinction is useful, the published literature predominately uses the term "microbiome" for both types of analyses (Dataset S1) and our use of the term core microbiome here follows that practice.

History and Evolution of the Core Microbiome Concept

One of the original goals of the Human Microbiome Project was to identify the core microbiome or "… whatever factors are common among the microbiomes of all or the vast majority of humans" (1). Thus, the term core microbiome is, by design, incredibly broad, recognizing that there may be multiple types of characteristics shared among different human microbiomes. These include individual genes, metabolic pathways, microbial taxa, as well as any other mechanisms that may play a role in host–microbiome interactions. Early studies commonly used a taxonomic approach, based on 16S rRNA gene-sequencing data, to quantify a core microbiome in healthy humans, as well as in people with specific conditions, such as obesity (4) and inflammatory bowel disease (19). However, these studies varied greatly in their ability to identify a set of core microbial taxa consistently associated with humans. In the gut, for example, few taxa were shared among individuals, and in some cases no overlapping operational taxonomic units (OTUs), or clusters of sequences with a level of shared nucleotide identity, were found (4). In contrast, in the oral microbiome, a number of OTUs were

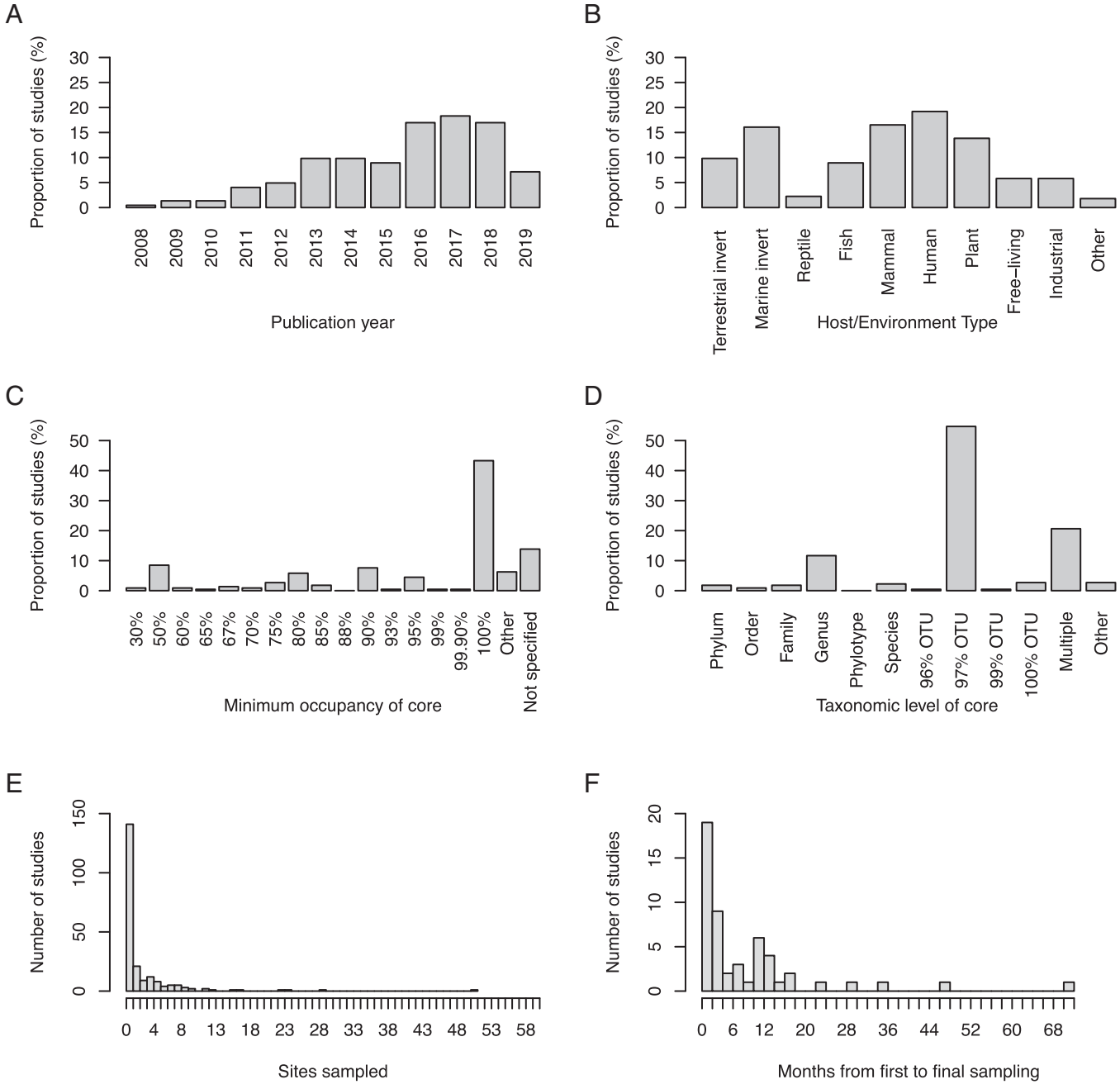

Fig. 2. Barplots of the proportion of studies in our representative dataset ($n = 224$) from each publication year (*A*) and environment (*B*), as well as the minimum occupancy value (*C*; those with no explicitly defined occupancy value are counted as "Other") and taxonomic level (*D*) used to determine the core microbiome. Histograms illustrate the number of sites sampled in each study (*E*) and the number of months from first sampling to the completion of the project for those studies which included a temporal component (*F*; $n = 51$).

identified as members of the taxonomic core of healthy individuals (8, 20). The heterogeneity of the results led to multiple different hypotheses about the core microbiome, including that 1) cores may exist within certain human populations but not globally, 2) a core may be discernible only at a higher taxonomic level such as genera, or that 3) the core may be wholly functional, comprised of functional gene clusters rather than individual taxa (outlined in ref. 2). While these hypotheses are not mutually exclusive, they do provide different insights into the nature and functioning of the human core microbiome. Furthermore, these early studies of human microbiomes provided a framework for subsequent analyses of core microbiomes in a variety of nonhuman hosts and environments.

The first comprehensive overview of the conceptual basis for the core microbiome, by Shade and Handelsman (7), revealed that the search for a core microbiome was still largely in its discovery phase, relying heavily on Venn diagrams to identify taxa (usually OTUs) shared among samples. They further outlined additional approaches for future research, such as inclusion of relative abundances of taxa in identifying cores and inclusion of a "persistent" or "dynamic" core microbiome measured across timescales (detailed below). These advances have led to

Table 1. Pros, cons, and suggestions for improvement for the currently available methods of quantifying core microbiomes

Metric type	Pros	Cons	Suggestions for improvement
Occurrence only	Computationally simple	No abundance information	Maximizing sequencing depth and replicate sampling
	Commonly used in the literature	Arbitrary cutoffs	Using multiple occurrence cutoffs
	Includes rare taxa	Can be heavily impacted by sampling coverage and sequencing depth	Using a range-through approach in conjunction with deeper sequencing
Relative abundance only	Computationally simple	Impacted by sequencing depth and inadequate spatial and temporal sampling coverage	Increasing geographic and temporal sampling, especially for widespread and/or lower abundance taxa
	Incorporates abundance information to identify taxa likely to be of functional importance	Arbitrary cutoffs	Ensuring uniform sequencing depth across samples
		Affected by rarefaction and related methods for sample standardization	
Abundance–occurrence	Based on macroecological theory	Often use arbitrary cutoffs of abundance and occurrence	Need to better establish macroecological relationships for microbial taxa
	New methods (including code) are being developed in this space	Currently assumes that macroecological relationships in microbial taxa are similar to those in plants and animals	Constraining analyses by phylogenetic or functional groups
	Can potentially differentiate the stochastic from the deterministically selected core		Ensuring that scale of spatial and temporal sampling is adequate to reliably capture macroecological relationships

discoveries, such as the stability of particular microbial strains in the guts of individual humans (21), and that certain microbial taxa persist across major environmental transitions, such as the saltwater-to-freshwater migration in Atlantic salmon (22). Identification of core microbial taxa can also lead to better understanding of their functional roles through targeted culturing or other 'omics approaches (such as metagenomics or metatranscriptomics), as exemplified in the honey bee system (23, 24). On the other hand, a few studies across phylogenetically diverse hosts, ranging from caterpillars to crustaceans, have found that the putative core microbiome was comprised of transient taxa repeatedly acquired from the environment rather than obligate associates of the host (25, 26). Finally, some authors have proposed moving away from a taxonomic approach altogether to an entirely functional definition of the core microbiome, where the core "… can be defined as a whole set of microbial vehicles, including replicators coding for essential functions for holobiont fitness" (27). One proposed solution to this issue is to specify the type of core microbiome of interest, such as a temporal core, ecological core, or functional core (3).

Current Methods for Identifying the Taxonomic Core Microbiome

Core microbiomes are typically quantified using one of three methods: 1) the occurrence of microbial taxa across multiple samples of the same host/environment, 2) the relative abundances of microbial taxa across such samples, or 3) some combination of the two. These metrics can be readily estimated from a standard OTU table containing the number of sequence counts of each OTU (or other units of analysis) detected in each sample. However, as discussed below, these methods have a number of pros and cons (Table 1) and their applications have been highly variable, making it difficult to compare the sizes and compositions of core microbiomes across different hosts and environments.

Occurrence Only. In our dataset, 75.4% of core microbiomes were quantified based solely on the occurrence of a taxon (OTU, genus, etc.) within a proportion of the samples collected (Dataset S1). Occurrence data are readily available, easily comparable, and provide information about the spatial or temporal span over which the microbe and host/environment interact. The proportions of sites, samples, or time points over which a microbe must occur to be considered core, however, is always at the discretion of the authors. In the most liberal cases from our dataset, a 30% occurrence standard was used, meaning that any OTU detected in at least 30% of samples was considered a core member (17, 28). Others have used cutoffs between 50 and 99.9%, depending on the study system and number of total samples (Fig. 2*C*). Most commonly, though, studies required a taxon be observed in 100% of samples to be considered core (Fig. 2*C*). This approach aims to identify the obligate, or seemingly obligate, relationships between host and microbe. A potential criticism, however, is that a 100% occurrence requirement is likely to miss lower abundance taxa that have fallen below the threshold of detection in one or more samples, but are of functional and ecological importance (e.g., ref. 29). The probability of this occurring depends, in part, on the sequencing depth of the samples, which is discussed in greater detail below. One proposed solution is to quantify the core microbiome using multiple occurrence cutoffs and evaluate how these different values affect the composition of the core. For example, in *Xestospongia* sponges, changing the occurrence cutoff has a relatively small impact on the composition of the

core (30). However, in other cases different occurrence cutoffs can also result in substantially different numbers of core taxa (e.g., refs. 4 and 31).

Relative Abundance Only. Nine studies (4%) within our sample set quantified the core microbiome using only relative abundance criteria (Dataset S1). In certain cases, though, occurrence data may have been used but not explicitly stated in the text. In some of these studies, the most abundant set of taxa in each sample type are identified as core members (32, 33). In others, taxa considered to be core were those that were preferentially enriched in the host of interest relative to the surrounding environment (16, 34–36). This method assumes that a high, or at least significantly increased, relative abundance in a particular host or environment is evidence of a stronger and more stable association with that system. A criticism of this approach is that some low-abundance taxa may still play important functional roles within the host (37, 38). Furthermore, relative abundances of some taxa may change substantially over time (e.g., ref. 39), so identifying core microbiomes using the abundance distribution of a single temporal snapshot may miss taxa that may be more common at other time points. One way to test for such variability is to use the relationship between the mean and variance of relative abundance of each OTU (e.g., ref. 40), a trend known as Taylor's power law (41). While this law has wide applications in ecology and beyond, in the present context, if the temporal or spatial variance of an OTU is greater than the mean, then the taxon should be considered less stable and its inclusion in the core microbiome potentially problematic. Such analyses, however, require replicate samples over time/space, something that is currently lacking in many core microbiome analyses.

Abundance–Occurrence. While occurrence and abundance have each been used individually to quantify the core microbiome, combining them can provide a more conservative approach. This paired method was used in 11.6% of studies in our dataset (Dataset S1). Many studies using this method set a minimum relative abundance threshold under which a taxon is disqualified from core membership. Then, the remaining taxa are assessed by the number of sites/samples in which they occur, again using an assigned cutoff value. In the studies sampled here, the minimum relative abundance threshold for an OTU to be considered for core membership ranged from 0.001 (42) to 4.5% (43), while occurrence cutoffs ranged from 50 to 100% across host species, geographic sites, or treatments (Dataset S1).

Abundance–occurrence relationships have a conceptual basis in the macroecological literature, as it has been shown in larger eukaryotes (44, 45), and more recently in marine bacterioplankton (13), that the local relative abundance of a species is generally positively correlated with the number of sites that species occupies within a region. However, whether this relationship is universally applicable across environments and microbial taxa is still unknown. Recently, though, there have been a few attempts to use the abundance–occurrence relationship to determine the compositions of core microbiomes. For example, Li et al. (46) designed a simple model that visualizes taxa that lie above user-specified abundance and occurrence cutoffs values, using a binomial distribution to account for differences in sequencing depth. Shade and Stopnisek (47) developed a method using abundance–occurrence relationships and contribution to Bray–Curtis similarity between samples in order to identify potential members of the core microbiome. They further compared the abundance–occurrence distributions of these taxa to a null model (48) in order to determine which core members may be deterministically selected by the host/environment (47).

While such use of abundance–occurrence relationships has the potential to better constrain the make-up of the core microbiome, it is important to keep in mind the underlying assumptions of such models. The first assumption is that this relationship is likely to be phylogenetically and functionally constrained. Abundance–occurrence relationships in plants and animals have been shown to be constrained by phylogeny or functional guilds (49) and different clades of microbial taxa also show different patterns of abundance and occurrence across spatial scales (50, 51). Thus, using the abundance–occurrence relationship across the entire microbiome to identify the core, as is the current practice, is likely to be problematic. Instead, it would be better to use this approach to identify core taxa at the level of individual clades and then aggregate across the whole community. Another critical assumption here is that any observed abundance–occurrence relationship of microbes using a small number of samples or populations is truly characteristic of that taxon. In larger eukaryotic communities, the abundance of a species is usually nonuniform across its geographic range (52), and often tends to peak near the center of the distribution (53). Whether such an "abundant center" pattern is also true of microbial taxa is currently unknown, but abundances of individual microbial taxa are also known to change along spatial gradients (51, 54). Thus, abundance–occurrence models are likely to be most useful for constraining the make-up of core microbiomes when they are based on large, representative samples across the geographic distributions of individual taxa, again, something that is currently rare.

The Role of Sequencing Depth. The number of microbial sequences obtained from each sample, or the sequencing depth, remains an important consideration in microbiome studies. Multiple studies of environmental microbiomes using massively deep sequencing efforts have shown that rare taxa, often undetected by shallower sequencing, make up a substantial portion of many microbial communities (55, 56). This suggests that maximizing sequencing depth is important for quantifying core microbiomes, especially when using strict occurrence cutoffs. This is also true of functional cores determined using other 'omics datasets, as shallower sequencing is likely to miss low-abundance gene clusters and provide incomplete coverage of the whole community (e.g., ref. 57).

In addition to overall sequencing depth, a related issue is accounting for differences in sequencing depths across a set of samples when quantifying core microbiomes. Some laboratory-based methods can help to control for sequencing depth and include standardizing the mass or volume of samples prior to DNA extraction and pooling libraries in equal molarity before sequencing. Despite these efforts, though, variation in sequencing depths across different samples in the same study can be substantial, and the use of statistical standardization methods to account for such variations has been the subject of much discussion and debate. Some have suggested that data should be rarefied (58, 59) to a common sampling depth, typically to the level of the sample with fewest sequences, while others argue that such rarefaction is "inadmissible" and favor approaches that transform or scale sequence counts (60, 61). While these

debates have led to the development of new statistical tools to account for sequencing depth, they have primarily focused on analyses of the alpha (e.g., refs. 62 and 63) or beta (e.g., refs. 64 and 65) diversities. The performance of such approaches in the context of quantifying the core microbiome remains poorly explored. Within our dataset, multiple studies rarefied sequence data before determining a core (e.g., refs. 12 and 66), while others did not rarefy their data at all (e.g., refs. 42 and 67). A recent study found that, all else being constant, the number and identities of core microbiome members can change after rarefying data to four different sequencing depths (68). This suggests that using rarefied data may provide an incomplete or potentially inaccurate picture of the core microbiome, as rarefying can remove tens or hundreds of thousands of sequences from individual samples, especially when variation in sequencing depth across samples is large.

In addition to taxon sampling, sequencing depth is likely to also have an impact on spatial occurrences of individual taxa (69), and thus affect occurrence-based measures of the core microbiome as well as the use of abundance–occurrence models discussed above. A potential solution here, borrowed from the macroecological literature, is to use a range-through approach where a taxon is presumed to be present everywhere within its geographical range limits, even though it may be missing from some samples within that range (69). For example, in samples collected along a latitudinal transect a taxon would be counted as present in all sites between its northernmost and southernmost point of detection, even if it is not explicitly sampled there (e.g., ref. 51). While some of the observed absences are likely to be real, driven by local environmental conditions or other factors, the range-through metric can provide a hypothesis about the occurrences of individual taxa in a specific sample, which can then be tested by repeated sampling or deeper sequencing. Another approach would be to sequence multiple technical replicates from the same sample to determine the likelihood of a true negative result, as has been done in human genotyping data (70), athough this creates potential tradeoffs with additional sequencing costs or decreased sequencing depth per individual sample.

Unspecified Definition. We found that more than 18% of the sampled studies did not include a methodology for quantifying the core microbiome in their *Methods* sections. In some cases, the criteria used for determining the core were provided in the *Results* or *Discussion*, while in other cases those criteria were unclear or not explicitly stated at all. This practice has also been mentioned by others (47) and should be strongly discouraged, as it creates a challenge for those aiming to replicate a study or to undertake comparative or metaanalyses of core microbiomes across studies.

Effects of Spatial, Temporal, and Phylogenetic Scales

Spatial Scale of Sampling. Regardless of whether the core microbiome is quantified based solely on occurrences or a combination of occurrence and abundance, the choice of spatial scale over which the host is sampled can have a strong influence on the composition of the core. For example, the diversity, composition, and function of the core microbiome of two host populations that are geographically close to each other is likely to be very different from that of two host populations separated by hundreds of kilometers. In our dataset, 67% of studies used samples collected at a single site (Fig. 2*E*), though in many of these cases, a core was determined between experimental treatments, disease treatments, time points, etc. (Dataset S1). The other 33% of studies analyzed samples collected from at least two distinct sites, as defined by the authors (Fig. 2*E*). However, these studies vary greatly in the spatial scale over which these locations were sampled, ranging from 2 locations within the same bay (71) to 23 collections on 6 continents (72) (Fig. 2*E*).

Because of the large variance in spatial scale over which published core microbiomes have been quantified, we have examples of cores that are essentially local (71) to those that are "region-specific" (67), and adapted to a host in a specific environment (e.g., ref. 73), to those that cover the entire geographic distribution of a host. The citrus tree rhizosphere, for example, was sampled from 23 sites spanning 6 continents, 6 climate regimes, and 7 soil types, effectively covering a large proportion of the areas in which citrus is grown as an agricultural product, providing a core citrus rhizosphere microbiome consisting of 132 genera (72).

Industrial and environmental microbiome studies can also utilize large-scale spatial sampling to test for widely distributed core microbiomes. For example, an extensive study of 51 wastewater treatment plants from 5 countries did not identify any core microbial OTUs present across all samples, but showed that anaerobic digester conditions, rather than geographic location, were significant correlated with microbial community composition (74). Similarly, a study of 6 soil types from 24 locations in eastern Europe showed that only two OTUs were shared across all samples, while many taxa appeared to be core to individual soil types (75).

Temporal Scale of Sampling. Of the articles in our sample, 22.3% included some form of temporal replication, ranging from 3 d (28) to 6 y (14). Of these, the majority were sampled over fewer than 2 mo, likely due to the challenges of repeated sampling across a prolonged period of time (Fig. 2*F*). However, such studies provide unique insight into whether core microbiomes persist in a particular location or host system despite potential changes in diet, development, or environment.

Free-living environmental microbiomes are particularly amenable to this type of repeated temporal sampling. Long-term sampling allows for the determination of how microbial communities fluctuate over time (e.g., ref. 76), as well as the taxa that persist through major events, such as hurricanes in the upper troposphere (77) and beach oiling in the Gulf of Mexico (78). Such studies can also reveal the level of stability present in the microbiome of a particular environment, as exemplified by a study of soda lakes in British Columbia, which found a highly persistent core microbiome in these lakes over the course of 4 y (79). Similarly, in long-lived host species such as humans, it is possible to sample the same individual through time to determine the level of community stability within the host. Studies of this type have identified particular OTUs that are able to persist over multiyear periods in the guts and mouths of certain individuals (80, 81).

Long-term studies of hosts with shorter generation times, or where repeated sampling proves difficult, require different approaches. One approach is to sample multiple individuals of an organism across different stages of the life cycle to determine if any core microbial taxa persist throughout. Studies using this method have shown that chickens maintain a large set of core microbial genera throughout their lives (82) and Atlantic

salmon maintain a relatively small core microbiome across the freshwater-to-saltwater transition (83), while parasites experience high levels of microbiome turnover when moving between hosts during their life cycle (84). A second method is to sample individuals of the same species at the same site over multiple time horizons. Although this type of sampling cannot determine which microbial taxa persist over time in a specific individual of a given host, any shared OTUs identified using this approach represent a species or population level core microbiome that is repeatedly acquired from the surrounding environment or vertically transmitted from parent to offspring. For example, monthly sampling of six sponge species over the course of 3 y showed that host species with denser microbial loads harbored larger core communities (85). Similarly, populations of the intertidal bivalve *Donax gouldii*, collected at four time points over 11 y, only had six ASVs present across all time points, indicating a small but persistent core (39).

Taxonomic Resolution of the Core Microbiome. Determining the taxonomic resolution of the core microbiome begins prior to sequencing, as the selection of 16S rRNA gene region and corresponding PCR primers play a significant role in the composition of the microbiome (86). Certain primer pairs may miss particular microbial groups or provide lower resolution due to limited coverage in taxonomic databases. Furthermore, taxonomic assignments provided by one primer set may not directly correspond to those from a different set, making core microbiome comparisons across studies difficult (86). Sequence length also determines the taxonomic level available for analysis, as shorter reads (e.g., 100 base pairs) provide less specific taxonomic information and, therefore, sequences may not be able to be assigned at lower taxonomic levels. Maximizing sequence length by merging paired-end reads, or generating full-length 16S rRNA gene amplicons using long-read technologies (e.g., ref. 87), are potential ways to increase the taxonomic resolution available for further analysis.

Regardless of sequencing strategy, though, the taxonomic level chosen for core microbiome quantification has important implications for the ecological and functional relevance of the core. For example, core microbiomes identified at the phylum level, as in four of the studies sampled here (Fig. 2*D*), may offer limited insights about the specific ecological and functional roles of those microbes. On the other hand, in nearly 50% of the studies in our dataset the core was determined using OTUs clustered at the 97% level, reflecting the popularity of this sequence similarity cutoff in delineating microbial OTUs (Fig. 2*D*). This proportion may be even higher, as bacterial species and phylotypes are often classified at the 97% sequence similarity level but were included as separate groups in these analyses when not explicitly defined. Interestingly, only 10 studies included a more stringent cutoff (99 or 100% OTUs) (Fig. 2*D*) despite the fact that OTUs generated using 100% sequence similarity (also known as ASVs or zero-radius OTUs) have become a commonly used taxonomic descriptor with the increasing use of denoising applications, such as deblur (88) and DADA2 (89). This difference could reflect, in part, a lag in the use of these units for core microbiome study, as they were not available until 2016 or 2017 (*SI Appendix*, Fig. S1).

Alternatively, 97% OTUs, or higher taxonomic levels, such as genera, may be more appropriate for determining the core microbiome since 100% sequence similarity may be too stringent for delineating functional or ecological differences among sequences (47). For example, two sequences with a single base pair variation would be identified as different OTUs using the 100% cutoff, but would be collapsed into a single 97% OTU. Furthermore, certain microbial taxa are known to contain multiple copies of the 16S rRNA gene, which can vary intragenomically by >1% sequence identity (90, 91). Thus, using the 100% OTU cutoff may inflate the size of the core microbiome and increase potential redundancy, particularly if core taxa contain many copies of the 16S rRNA gene (e.g., Gammaproteobacteria) (90). On the other hand, it has recently been shown that as many as 16 unique microbial strains with diverse temperature preferences and carbohydrate utilization profiles can cluster into a single 97% OTU (92). The limited resolution of databases used for taxonomic identification may also decrease the utility of determining the core at lower taxonomic levels, particularly in nonhuman samples, as many microbial groups are still unknown or uncharacterized and therefore cannot provide much information about community function. In more than 20% of cases, though, the authors quantified core microbiomes using multiple different taxonomic criteria. This is done by either 1) clustering OTUs at different levels of sequence similarity (e.g., ref. 93) or 2) using the levels of taxonomic classification produced by a database [e.g., greengenes (94) or Silva (95)]. Multiple classifications have the potential to provide additional information regarding the strength of the association between the microbial taxa and the host/environment. For example, if some functions are conserved at higher taxonomic levels than a microbial genus, they may play a core functional role within a host, but individual OTUs within that genus may have a more limited occurrences and not appear as core members.

Choice of Host Taxa. Nearly 88% of studies in our sample set focused on microbiomes associated with plant and animal hosts (including various microenvironments within humans and other hosts), rather than environmental or industrial samples. Most frequently, core microbiomes were identified for a single host species, but 42 studies attempted to identify such cores using two or more host species of varying phylogenetic relatedness. In some cases, the phylogenetic relationships among the hosts are known, and a shared core microbiome represents a potential cophylogenetic history or conserved functional/ecological role across species. For example, 18 "types" of *Symbiodinium* dinoflagellate spanning at least 8 species were found to share 3 highly abundant microbial OTUs, which were predicted to play various roles in nutrient acquisition and stress tolerance (96). Similar analyses using 36 strains of *Leptocylindrus* diatoms found that only one microbial OTU, a member of the genus *Roseovarius*, was shared by all of the strains (97). Even more broadly, one study tested whether a highly diverse set of sponge hosts (32 species) share a core microbiome, but found no evidence for such conservatism (98).

Instead of phylogenetic affinities, some studies have tested whether hosts with similar traits (e.g., mode of digestion, diet) share core microbial taxa. Ruminant mammals, for example, have been found to share a number of core genera across a wide geographic range, which are hypothesized to play a role in digestion and fermentation (15, 99). Similarly, a phylogenetically diverse group of cycad-eating insects were found to share five microbial OTUs in common, at least one of which was predicted to provide functional benefits for digestion of cycad tissue (100).

Synthesis and Future Directions

Determining the most effective way to quantify the core microbiome remains challenging, with some arguing that a taxonomic approach is no longer useful and that a core functional microbiome should be prioritized (27). However, despite the utility in understanding the core functional properties of the microbiome, a taxonomic approach remains highly practical for a few important reasons. First, the per sample cost of 16S rRNA gene sequencing remains much lower than the cost to generate the high-quality metagenomes and metatranscriptomes necessary for detailed functional profiling. Second, this approach is better suited when macroecological relationships, such as abundance–occurrence models of microbial taxa, are used for constraining the composition of the core microbiome. Finally, a taxonomic core microbiome provides a list of potentially ecologically relevant taxa which can be prioritized for targeted culturing and 'omics study. This also allows for the development of testable hypotheses about the roles of these organisms within the microbiome.

However, as shown by this representative sampling of the literature, quantifying a core microbiome is not straightforward. Perhaps more important, though, is the fact that these differences in methodology for quantifying the core change its functional definition. For example, abundance-driven metrics prioritize the most dominant members of the community, or those that have most effectively colonized a particular environment. However, this type of core likely overlooks a number of ecologically and functionally important but low-abundance taxa and may be skewed by high levels of variance. On the other hand, occurrence-based metrics often require taxa to be present in every sample to be counted in the core, which can miss relevant taxa due to inadequate sequencing depth or sampling effects. Relaxing this criterion necessarily requires using an arbitrary cutoff for the number of samples a microbial OTU must be present in to count as part of the core. These different cutoff values include taxa with different levels of association to the host/environment and different ecological roles, thereby changing the way the core is defined. Methods that combine abundance and occurrence have been introduced to overcome some of these challenges (e.g., ref. 46), and are well-supported in the broader ecological literature, but many still require arbitrary occurrence and abundance cutoffs, which vary widely across studies. Such combined methods, which use modeling approaches, represent a potentially significant advancement, but their broad applicability to microbial taxa is not yet certain.

In addition to specific metrics for quantifying a core, the issue of spatial and taxonomic grain at which the core should be determined also remains fluid. As discussed above, the spatial extent of sampling can have a strong impact on the make-up of the core microbiome at the population or species level, especially for hosts that have large geographic distributions. Similarly, determining the taxonomic level at which the core should be quantified also remains a challenge. Although 97% OTUs are the most commonly used taxonomic units in our dataset, cores are also routinely identified using ASVs, genera, or phyla, thus introducing different levels of potential functional and ecological redundancy into the core microbiome and changing its practical definition. One issue that has largely been ignored when quantifying taxonomic cores is the fact that even consistent associations between a host and a group of microbes identified by any of the methods discussed above could still simply reflect repeated acquisitions of microbes from the environment by the host rather than obligate host–microbe relationships where microbes play important functional roles (25). Experiments, quantitative PCR, and microscopy techniques can all be used to test whether a putative core microbiome is a stable component of the host physiology (26) and plays a significant functional role in the host (25), but such approaches remain rare.

In summary, our review of the literature clearly shows that the term "core microbiome" represents different things to different researchers, which makes comparative analyses and metaanalyses of the core microbiome across hosts and environments very difficult, if not impossible. While a single metric is unlikely to capture all the different aspects of core microbiomes, we hope that the information provided is a useful starting point for the development of measures of the core microbiome that are robust to sampling, sequencing depth, and other issues discussed here. Such metrics are necessary not only for testing specific hypotheses about the functional and ecological roles of core microbiomes, but also for understanding the general nature of core microbiomes and the ecological and evolutionary processes that generate and maintain these stable associations between certain microbes and their hosts. Finally, we provide a set of recommendations below that could serve as the starting point for achieving this goal.

First, explicitly define and state the criteria used for determining the core microbiome in the *Methods* section of the manuscript. As noted in this review, these methods provide important context for interpreting the results but are often not adequately described.

Second, when conducting spatial analyses, explicitly distinguish between local, regional, and range-wide cores (e.g., refs. 67 and 73). When conducting temporal analyses, explicitly distinguish between short-term, seasonal, and multiyear cores. They each require their own contextual definitions but provide important information about the potential spatial and temporal stability of any core associations.

Third, sequence as deeply as possible and ensure an adequate number of sequencing replicates to determine the core. Standardizing the size of samples (pre-DNA extraction) or molarity (for library pooling) of samples are potential strategies to achieve uniform sequencing depth. In spatial datasets, it may be possible to compensate for variable sequencing depths using the range-through approach described above.

Fourth, rarefying samples to a common sequencing depth is best avoided while quantifying the core microbiome, especially when variation in sequencing depth across samples is large. Such an approach can lead to underestimates of core size and inaccurate core composition (68).

Fifth, the use of macroecological null models for constraining the make-up of the core microbiome should be based on adequate spatial sampling coverage, especially for widely distributed microbial taxa. Most existing analyses of core microbiomes do not have enough spatial and temporal coverage for computing meaningful macroecological relationships.

Data Availability. All study data are included in the article and supporting information.

Acknowledgments

We thank the members of the K.R. and E.E.A. laboratories for helpful discussions on this topic.

1 P. J. Turnbaugh *et al.*, The human microbiome project. *Nature* **449**, 804–810 (2007).
2 M. Hamady, R. Knight, Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).
3 A. Risely, Applying the core microbiome to understand host-microbe systems. *J. Anim. Ecol.* **89**, 1549–1558 (2020).
4 P. J. Turnbaugh *et al.*, A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
5 Y. Jiang, X. Xiong, J. Danska, J. Parkinson, Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome* **4**, 2 (2016).
6 I. Stefanini *et al.*, Core microbiota and metabolome of *Vitis vinifera* L. cv. Corvina grapes and musts. *Front. Microbiol.* **8**, 457 (2017).
7 A. Shade, J. Handelsman, Beyond the Venn diagram: The hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12 (2012).
8 E. Zaura, B. J. F. Keijser, S. M. Huse, W. Crielaard, Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol.* **9**, 259 (2009).
9 F. Bäckhed *et al.*, Defining a healthy human gut microbiome: Current concepts, future directions, and clinical applications. *Cell Host Microbe* **12**, 611–622 (2012).
10 D. A. Hutchins *et al.*, Climate change microbiology—Problems and perspectives. *Nat. Rev. Microbiol.* **17**, 391–396 (2019).
11 T. D. Ainsworth, R. D. Gates, Ocean biology. Corals' microbial sentinels. *Science* **352**, 1518–1519 (2016).
12 K. Christian, C. Weitzman, A. Rose, M. Kaestli, K. Gibb, Ecological patterns in the skin microbiota of frogs from tropical Australia. *Ecol. Evol.* **8**, 10510–10519 (2018).
13 M. V. Lindh *et al.*, Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers. *Environ. Microbiol.* **19**, 1222–1236 (2017).
14 A. M. Saunders, M. Albertsen, J. Vollertsen, P. H. Nielsen, The activated sludge ecosystem contains a core community of abundant organisms. *ISME J.* **10**, 11–20 (2016).
15 G. Henderson *et al.*; Global Rumen Census Collaborators, Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* **5**, 14567 (2015).
16 D. S. Lundberg *et al.*, Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**, 86–90 (2012).
17 T. D Ainsworth *et al.*, The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J.* **9**, 2261–2274 (2015).
18 G. Berg *et al.*, Microbiome definition re-visited: Old concepts and new challenges. *Microbiome* **8**, 103 (2020).
19 B. P. Willing *et al.*, A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854.e1 (2010).
20 A. M. Eren, G. G. Borisy, S. M. Huse, J. L. Mark Welch, Oligotyping analysis of the human oral microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2875–E2884 (2014).
21 J. J. Faith *et al.*, The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).
22 C. E. Dehler, C. J. Secombes, S. A. M. Martin, Seawater transfer alters the intestinal microbiota profiles of Atlantic salmon (*Salmo salar* L.). *Sci. Rep.* **7**, 13877 (2017).
23 Z. L. Sabree, A. K. Hansen, N. A. Moran, Independent studies using deep sequencing resolve the same set of core bacterial species dominating gut communities of honey bees. *PLoS One* **7**, e41250 (2012).
24 P. Engel, N. A. Moran, Functional and evolutionary insights into the simple yet specific gut microbiota of the honey bee from metagenomic analysis. *Gut Microbes* **4**, 60–65 (2013).
25 T. J. Hammer, D. H. Janzen, W. Hallwachs, S. P. Jaffe, N. Fierer, Caterpillars lack a resident gut microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9641–9646 (2017).
26 G. G. Martin *et al.*, Absence of a microbiome in the midgut trunk of six representative Crustacea. *J. Crustac. Biol.* **40**, 122–130 (2020).
27 P. Lemanceau, M. Blouin, D. Muller, Y. Moënne-Loccoz, Let the core microbiota be functional. *Trends Plant Sci.* **22**, 583–595 (2017).
28 M. J. Sweet, B. E. Brown, R. P. Dunne, I. Singleton, M. Bulling, Evidence for rapid, tide-related shifts in the microbiome of the coral *Coelastrea aspera*. *Coral Reefs* **36**, 815–828 (2017).
29 S. M. Huse, Y. Ye, Y. Zhou, A. A. Fodor, A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One* **7**, e34242 (2012).
30 C. Astudillo-García *et al.*, Evaluating the core microbiota in complex communities: A systematic investigation. *Environ. Microbiol.* **19**, 1450–1462 (2017).
31 N. Segata *et al.*, The reproductive tracts of two malaria vectors are populated by a core microbiome and by gender- and swarm-enriched microbial biomarkers. *Sci. Rep.* **6**, 24207 (2016).
32 B. Beckers, M. Op De Beeck, N. Weyens, W. Boerjan, J. Vangronsveld, Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome* **5**, 25 (2017).
33 J. Brodie *et al.*, Characterising the microbiome of *Corallina officinalis*, a dominant calcified intertidal red alga. *FEMS Microbiol. Ecol.* **92**, fiw110 (2016).
34 J. Edwards *et al.*, Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E911–E920 (2015).
35 K. Schlaeppi, N. Dombrowski, R. G. Oter, E. Ver Loren van Themaat, P. Schulze-Lefert, Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 585–592 (2014).
36 Y. K. Yeoh *et al.*, The core root microbiome of sugarcanes cultivated under varying nitrogen fertilizer application. *Environ. Microbiol.* **18**, 1338–1351 (2016).
37 S. Banerjee, K. Schlaeppi, M. G. A. van der Heijden, Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microbiol.* **16**, 567–576 (2018).
38 A. Jousset *et al.*, Where less may be more: How the rare biosphere pulls ecosystems strings. *ISME J.* **11**, 853–862 (2017).
39 A. T. Neu, I. V. Hughes, E. E. Allen, K. Roy, Decade-scale stability and change in a marine bivalve microbiome. *Mol. Ecol.* **30**, 1237–1250 (2021).
40 J. Oh, A. L. Byrd, M. Park, H. H. Kong, J. A. Segre; NISC Comparative Sequencing Program, Temporal stability of the human skin microbiome. *Cell* **165**, 854–866 (2016).
41 L. R. Taylor, Aggregation, variance and the mean. *Nature* **189**, 732–735 (1961).
42 R. E. Antwis, J. M. D. Lea, B. Unwin, S. Shultz, Gut microbiome composition is associated with spatial structuring and social interactions in semi-feral Welsh Mountain ponies. *Microbiome* **6**, 207 (2018).
43 M. L. Leis, M. O. Costa, Initial description of the core ocular surface microbiome in dogs: Bacterial community diversity and composition in a defined canine population. *Vet. Ophthalmol.* **22**, 337–344 (2019).
44 I. Hanski, Dynamics of regional distribution: The core and satellite species hypothesis. *Oikos* **38**, 210–221 (1982).
45 K. J. Gaston *et al.*, Abundance-occupancy relationships. *J. Appl. Ecol.* **37**, 39–59 (2000).
46 K. Li, M. Bihan, B. A. Methé, Analyses of the stability and core taxonomic memberships of the human microbiome. *PLoS One* **8**, e63139 (2013).
47 A. Shade, N. Stopnisek, Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr. Opin. Microbiol.* **49**, 50–58 (2019).
48 K. E. Sullam *et al.*, Environmental and ecological factors that shape the gut bacterial communities of fish: A meta-analysis. *Mol. Ecol.* **21**, 3363–3378 (2012).
49 J. H. Brown, *Macroecology* (University of Chicago Press, 1995).
50 A. T. Neu, M. E. Torchin, E. E. Allen, K. Roy, Microbiome divergence of marine gastropod species separated by the Isthmus of Panama. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.07.08.451645. Accessed 11 July 2021.
51 A. T. Neu, E. E. Allen, K. Roy, Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from *Mytilus californianus*, an intertidal foundation host. *J. Biogeogr.* **48**, 2839–2852 (2021).

52 J. P. Sexton, P. J. McIntyre, A. L. Angert, K. J. Rice, Evolution and ecology of species range limits. *Annu. Rev. Ecol. Evol. Syst.* **40**, 415–436 (2009).
53 J. H. Brown, D. W. Mehlman, G. C. Stevens, Spatial variation in abundance. *Ecology* **76**, 2028–2043 (1995).
54 M. J. Choudoir, J. R. Doroghazi, D. H. Buckley, Latitude delineates patterns of biogeography in terrestrial Streptomyces. *Environ. Microbiol.* **18**, 4931–4945 (2016).
55 M. L. Sogin *et al.*, Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12115–12120 (2006).
56 S. M. Gibbons *et al.*, Evidence for a persistent microbial seed bank throughout the global ocean. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4651–4655 (2013).
57 J. Pereira-Marques *et al.*, Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front. Microbiol.* **10**, 1277 (2019).
58 S. Weiss *et al.*, Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
59 D. T. McKnight *et al.*, Methods for normalizing microbiome data: An ecological perspective. *Methods Ecol. Evol.* **10**, 389–400 (2019).
60 P. J. McMurdie, S. Holmes, Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
61 G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
62 C.-H. Chiu, A. Chao, Estimating and comparing microbial diversity in the presence of sequencing errors. *PeerJ* **4**, e1634 (2016).
63 A. Alberdi, M. T. P. Gilbert, A guide to the application of Hill numbers to DNA-based diversity analyses. *Mol. Ecol. Resour.* **19**, 804–817 (2019).
64 C. Martino *et al.*, A novel sparse compositional technique reveals microbial perturbations. *mSystems* **4**, e00016–e00019 (2019).
65 J. Liu *et al.*, A semiparametric model for between-subject attributes: Applications to beta-diversity of microbiome data. *Biometrics*, 10.1111/biom.13487 (2021).
66 M. Aira, M. Pérez-Losada, J. Domínguez, Microbiome dynamics during cast ageing in the earthworm *Aporrectodea caliginosa*. *Appl. Soil Ecol.* **139**, 56–63 (2019).
67 I. Burgsdorf *et al.*, Biogeography rather than association with cyanobacteria structures symbiotic microbial communities in the marine sponge *Petrosia ficiformis*. *Front. Microbiol.* **5**, 529 (2014).
68 M. P. Ramakodi, Effect of amplicon sequencing depth in environmental microbiome research. *Curr. Microbiol.* **78**, 1026–1033 (2021).
69 K. Roy, J. D. Witman, "Spatial patterns of species diversity in the shallow marine invertebrates: Patterns, processes, and prospects" in *Marine Macroecology*, J. D. Witman, K. Roy, Eds. (The University of Chicago Press, 2009), pp. 101–121.
70 J. Kim *et al.*, The use of technical replication for detection of low-level somatic mutations in next-generation sequencing. *Nat. Commun.* **10**, 1047 (2019).
71 M. Turon, J. Cáliz, X. Triadó-Margarit, E. O. Casamayor, M. J. Uriz, Sponges and their microbiomes show similar community metrics across impacted and well-preserved reefs. *Front. Microbiol.* **10**, 1961 (2019).
72 J. Xu *et al.*, The structure and function of the global citrus rhizosphere microbiome. *Nat. Commun.* **9**, 4894 (2018).
73 A. Hernandez-Agreda, W. Leggat, P. Bongaerts, T. D. Ainsworth, The microbial signature provides insight into the mechanistic basis of coral success across reef habitats. *MBio* **7**, e00560-e16 (2016).
74 R. Mei *et al.*, Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome. *Water Res.* **124**, 77–84 (2017).
75 E. V. Pershina *et al.*, Investigation of the core microbiome in main soil types from the East European plain. *Sci. Total Environ.* **631-632**, 1421–1430 (2018).
76 J. A. Gilbert *et al.*, The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* **11**, 3132–3139 (2009).
77 N. DeLeon-Rodriguez *et al.*, Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2575–2580 (2013).
78 R. J. Newton *et al.*, Shifts in the microbial community composition of Gulf Coast beaches following beach oiling. *PLoS One* **8**, e74265 (2013).
79 J. K. Zorz *et al.*, A shared core microbiome in soda lakes separated by large distances. *Nat. Commun.* **10**, 4230 (2019).
80 J. G. Caporaso *et al.*, Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
81 I. Martínez, C. E. Muller, J. Walter, Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS One* **8**, e69621 (2013).
82 J. Ding *et al.*, Inheritance and establishment of gut microbiota in chickens. *Front. Microbiol.* **8**, 1967 (2017).
83 K. Rudi *et al.*, Stable core gut microbiota across the freshwater-to-saltwater transition for farmed Atlantic salmon. *Appl. Environ. Microbiol.* **84**, e01974–e17 (2018).
84 F. Jorge, N. M. Dheilly, R. Poulin, Persistence of a core microbiome through the ontogeny of a multi-host parasite. *Front. Microbiol.* **11**, 954 (2020).
85 J. R. Björk, R. B. O'Hara, M. Ribes, R. Coma, J. M. Montoya, The dynamic core microbiome: Structure, dynamics and stability. bioRxiv [Preprint] (2018). https://www.biorxiv.org/content/10.1101/137885v2. Accessed 18 March 2020.
86 I. Abellan-Schneyder *et al.*, Primer, pipelines, parameters: Issues in 16S rRNA gene sequencing. *MSphere* **6**, e01202–e01220 (2021).
87 B. J. Callahan, D. Grinevich, S. Thakur, M. A. Balamotis, T. Ben Yehezkel, Ultra-accurate microbial amplicon sequencing directly from complex samples with synthetic long reads. *Microbiome* **9**, 130 (2021).
88 A. Amir *et al.*, Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* **2**, e00191-e16 (2017).
89 B. J. Callahan *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
90 T. Větrovský, P. Baldrian, The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* **8**, e57923 (2013).
91 T. Coenye, P. Vandamme, Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol. Lett.* **228**, 45–49 (2003).
92 A. B. Chase *et al.*, Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. *MBio* **8**, e01809–e01817 (2017).
93 K. Dougal *et al.*, Identification of a core bacterial community within the large intestine of the horse. *PLoS One* **8**, e77660 (2013).
94 T. Z. DeSantis *et al.*, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
95 C. Quast *et al.*, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
96 C. A. Lawson, J. B. Raina, T. Kahlke, J. R. Seymour, D. J. Suggett, Defining the core microbiome of the symbiotic dinoflagellate, Symbiodinium. *Environ. Microbiol. Rep.* **10**, 7–11 (2018).
97 P. A. Ajani *et al.*, The microbiome of the cosmopolitan diatom leptocylindrus reveals significant spatial and temporal variability. *Front. Microbiol.* **9**, 2758 (2018).
98 S. Schmitt *et al.*, Assessing the complex sponge microbiota: Core, variable and species-specific bacterial communities in marine sponges. *ISME J.* **6**, 564–576 (2012).
99 M. M. O' Donnell, H. M. B. Harris, R. P. Ross, P. W. O'Toole, Core fecal microbiota of domesticated herbivorous ruminant, hindgut fermenters, and monogastric animals. *MicrobiologyOpen* **6**, e509 (2017).
100 S. Salzman, M. Whitaker, N. E. Pierce, Cycad-feeding insects share a core gut microbiome. *Biol. J. Linn. Soc. Lond.* **123**, 728–738 (2018).