environmental microbiology

Environmental Microbiology (2012) **14**(1), 4–12 doi:10.1111/j.1462-2920.2011.02585.x

Minireview

Beyond the Venn diagram: the hunt for a core microbiome

Ashley Shade and Jo Handelsman*
Department of Molecular, Cellular, and Developmental Biology, Yale University, 219 Prospect Street, New Haven, CT 06520, USA.

Summary

Discovering a core microbiome is important for understanding the stable, consistent components across complex microbial assemblages. A core is typically defined as the suite of members shared among microbial consortia from similar habitats, and is represented by the overlapping areas of circles in Venn diagrams, in which each circle contains the membership of the sample or habitats being compared. Ecological insight into core microbiomes can be enriched by 'omics approaches that assess gene expression, thereby extending the concept of the core beyond taxonomically defined membership to community function and behaviour. Parameters defined by traditional ecology theory, such as composition, phylogeny, persistence and connectivity, will also create a more complex portrait of the core microbiome and advance understanding of the role of key microorganisms and functions within and across ecosystems.

Introduction

A core microbiome is comprised of the members common to two or more microbial assemblages associated with a habitat (Turnbaugh *et al*., 2007; Hamady and Knight, 2009; Table 1). Identifying the core species (or operational taxonomic units, OTUs) is essential to unravelling the ecology of microbial consortia because it has been proposed that these commonly occurring organisms that appear in all assemblages associated with a particular habitat are likely critical to the function of that type of community. Thus, identifying a core is the first step in defining a 'healthy' community and predicting community responses to perturbation. Understanding what members are core will guide the manipulation of communities to achieve desired outcomes. These insights will contribute to addressing two of this decade's grand challenges in microbial ecology: to predict the impact of global change on biogeochemical cycling, and to manage the human microbiota to enhance human wellness.

Determining whether the human microbiome has a core is a goal of the US National Institutes of Health Human Microbiome Project (Turnbaugh *et al*., 2007). As a result, many studies aim to discover and compare cores associated with mammalian hosts in states of health and disease (see Kuczynski *et al*., 2010 and references therein). Collectively, these studies provide preliminary evidence of cores, and the resulting enthusiasm has produced almost as many reviews, commentary and opinion articles as primary research (e.g. Tschop *et al*., 2009; Turnbaugh and Gordon, 2009; Vael and Desager, 2009; Bäckhed, 2010; Benson *et al*., 2010; Elli *et al*., 2010; Kuczynski *et al*., 2010; Neu *et al*., 2010; Tilg, 2010).

The concept of a core is not restricted to host-associated microbiomes. Environmental microbiologists have long asked similar questions about free-living microbial assemblages. The phrase 'core microbiome' may be used to describe members shared across soils, lakes or wastewater treatment systems. The human microbiome projects have re-ignited popularity of the classic question: which functions do these core microorganisms contribute to the community?

In 2009, Hamady and Knight presented ideas about the nature of a core across individual human beings: the core may be *substantial* (one in which a large proportion of microbial taxa are shared); *minimal, non-existent,* or a *gradient* (for instance, along some dietary or other environmental continuum). They also suggest that some cores are shared only among *subpopulations* of hosts, rather than all individuals of the host species (Hamady and Knight, 2009). These thoughtful and timely ideas provide a starting point for exploring large microbiome data sets.

Despite the renewed focus on cores, what constitutes a core remains elusive. A typical approach is to report the number of species found across localities from a similar habitat based on a presence/absence data set. This

Received 22 April, 2011; accepted 5 August, 2011. *For correspondence. E-mail jo.handelsman@yale.edu; Tel. (+1) 203 432 9119; Fax (+1) 203 432 6161.

Table 1. Terms relevant to core microbiomes.

Term	Definition	Examples	Reference
Biome	The world's major ecosystems, defined by temperature gradients in latitude and altitude, precipitation and seasonality.	Subtropical, Mediterranean, Polar	Walter and Box (1976)
Microbiome	An assemblage of microorganisms existing in or associated with a habitat; includes active and interacting member as well transient or inactive members. Requires qualifiers for locality and habitat of interest.	Human microbiome, Earth microbiome, Lake Erie microbiome, Soil microbiome	Lederberg and McCray (2001)
Core microbiome	Organisms common across microbiomes hypothesized to play a key role in ecosystem function within a habitat.	To be determined	Turnbaugh *et al*. (2007)
Habitat	The physical and chemical parameters of an environmental area that determine niche spaces.	Termite hindgut: physical structure, anaerobic conditions, acidity and cellulose availability permit *Spirochaeta* abundance. Siberian tundra: subarctic temperatures, short summers and low solar energy permit boreal forest predominance.	Whittaker *et al.* (1973)
Ecosystem	The interactions and dynamics of physical, chemical and biological components of a locality.	Cornfield, Temperate forest, Permafrost, Tidal marsh, Mouse oral cavity	Odum (1953)
Locality	A spatially defined environmental area.	New York City, Lake Michigan, Soil core from a pea field, Vagina of a human subject	Andrewartha and Birch (1984)
Microbial community	Microorganisms that are co-existing and interacting with flanking microbiome members and/or the environment.	Algal mat, Biofilm, Dental plaque	Little *et al.* (2008)
Niche space	The activity range of a population along the physical and chemical dimensions of a habitat.	Anoxic aquatic or sediment niche spaces for sulfate reducers and methanogens	Hutchinson (1957)
Population	All the organisms belonging to the same species/operational taxonomic unit that live in a locality. For microbes, the species definition will vary by the **genes** of interest; a strain could be a population, as well as mutant and wild-type organisms.	Grizzly bear in Alaska *Sulfolobus islandicus* strain in a geothermal pool Wild-type *Enterococcus faecalis*	Waples and Gaggiotti (2006)
Connectivity	The amount or proportion of interactions within a system. Within a microbiome, this could include interactions among taxa (biotic interactions) or with environmental factors (abiotic) within the locality.	Quorum sensing, Predator-prey dynamics, Resource competition	Gardner and Ashby (1970)
Observation	The sampling unit, includes metadata of locality, habitat, time and/or experimental condition.	Meta-transcriptome of one location in an acid mine drainage site, collected at one time point. 16S tag-sequences from the right palm of one human subject.	
Persistent	Organisms that are consistently detected within a locality through time.	*Pseudomonas aeruginosa* in Cystic Fibrosis patients Firmicutes in infant guts	

results in the visual of the Venn diagram, where circles denote the different microbiomes, and their area of overlap represents the core. Though the Venn diagram is a reasonable first exploration, its conceptual basis omits ecological characteristics that might suggest a more nuanced understanding that connects *who* is present with *what they are doing* in a type of ecosystem.

Here we present a conceptual framework for identifying core microbiomes. Beginning with the Venn diagram representing membership, layers of complexity are added by incorporating composition, phylogeny, persistence and connectivity. The goal is to consider various aspects of core microbiome ecology and bring to it a conceptual framework from traditional ecology. Finally, the examples present a springboard for discussion and future research.

Defining a core microbiome

Using the common definition of the core leads to a data set presented as an OTU table. The OTU table has each OTU ('species' or unit of interest) in a row and community observations (e.g. sampling units; see Table 1) in columns. The table is filled with numerical information about OTU occurrences. The OTU table need not be restricted to 16S sequence-based metagenomic data; it can include data from various 'omics approaches, such as abundance of proteins or metabolites, intensity of hybridization to nucleic acid probes on a chip, or presence/absence of clusters of functional genes.

We use *microbiome* to describe the assemblage of microorganisms, active or inactive, associated with a habitat (Table 1). The *habitat* is the abiotic components that determine biological *niche spaces* of a defined environmental area, or *locality*. A typical *observation* would be a suite of microbial-associated data (for instance, 16S, transcriptome, metagenome) from one time point at a locality, but the relevant *core microbiome* comparison is usually across similar habitats. A *microbial community* is the interacting subset of the microbiome that is an active part of the *ecosystem*, which includes the interactive dynamics of abiotic and biotic components in a habitat.

Membership: shared presence

An analysis of membership considers the shared taxa (or genes) that are present across two or more microbiomes. This is based on a presence/absence data set in which shared OTUs are tallied (Fig. 1A). The visual of a Venn diagram indicates the amount of overlap, but is not effective for presentation of more than four categorical groups. Sørenson's index, which accounts for the number of shared and unique OTUs per group, also describes a membership-based core microbiome.

Membership cores have been observed in human-associated consortia, but with variable representation. In the gut, there is often a small number of shared phylotypes among individuals, suggesting a minimal membership core (e.g. Tap *et al*., 2009; Qin *et al*., 2010). Conversely, there was a large overlap among oral microbiota phylotypes across communities from three individuals (Zaura *et al*., 2009). Instead of a membership core observed across individuals, 'clusters' and 'enterotype' cores were discovered in vaginal and gut microbiomes, supporting the Hamady and Knight subpopulation hypothesis (Hamady and Knight, 2009; Arumugam *et al*., 2011; Ravel *et al*., 2011). A metagenomic study of the core genes between lean and obese twins revealed a high degree of overlap in genes hypothesized to be important for gut-adapted life style, such as carbohydrate metabolism (Turnbaugh *et al*., 2009). These shared genes were described as 'core' functions, while those genes not shared were 'variable' and used to identify pathways uniquely associated with obesity. Notably, sequencing depth could change the interpretation of the proportion of core phylotypes or genes, as Qin *et al*. showed that 3× sequencing depth revealed a 25% larger core than did 1× coverage. Together, these studies and others suggest that a membership-based core depends on environment, sequencing efforts and host population.

Composition: dominance, rarity and the middle ground

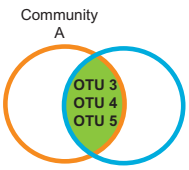
Venn analysis ignores composition, which accounts for the relative abundance of each OTU. Many multivariate

Fig. 1. Defining the core microbiome. A and B represent microbial communities to be compared using the OTU table of occurrences.
A. A core based on shared membership. Input OTU table is presence/absence, and occurrences of shared presence are tallied across communities of interest.
B. A core based on shared composition. In this example, only OTUs that are both shared and in similar proportion are counted toward a core.
C. A core incorporating phylogenetic information. Related OTUs are counted as one unit toward the core, as for OTUs 1 and 2 within lineage ii. The phylogenetic level can be adjusted up or down.
D. A core microbiome including OTUs that are persistent in a series. OTUs that are consistently observed within a community are counted towards a core. (Here, shown as consecutive time points times 1, 2 and 3 for each community A and B).
E. A core including only OTUs interacting (or presumed to be) with other members of its community. Hypotheses of significant OTU interactions are derived from a networks analysis of occurrences in a series. In the network, dashed lines represent putative negative correlations (e.g. species competition or predation), and solid lines with arrows represent putative positive interactions (e.g. both OTUs responding to an increase in a shared resource).

A. Membership:
shared OTU occurences across communities
1 = present, 0 = below detection

	Occurences in community **A**	Occurences in community **B**	Shared occurences **A & B**
OTU 1	1	0	
OTU 2	0	1	
OTU 3	1	1	X
OTU 4	1	1	X
OTU 5	1	1	X

List of observed OTUs

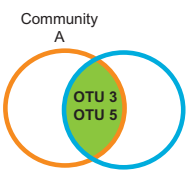


Core = 3 shared OTUs

B. Composition:
similar OTU abundances across communities

	Abundances community **A**	Abundances community **B**	Similar abundances **A & B**
OTU 1	0.4	0	
OTU 2	0	0.1	
OTU 3	0.1	0.1	X
OTU 4	0.2	0.5	
OTU 5	0.3	0.3	X

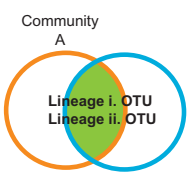
List of observed OTUs



Core = 2 shared OTUs

C. Phylogeny:
shared OTU lineages across communities

	Abundances community **A**	Abundances community **B**	Similar abundances **A & B**
OTU 1	0.4	0	X
OTU 2	0	0.1	
OTU 3	0.1	0.1	X
OTU 4	0	0.8	
OTU 5	0.5	0	

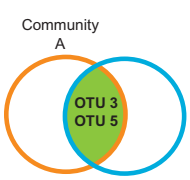


Core = 2 shared OTUs

D. Persistence:
OTUs shared over time within and across communities

	Time series community **A**			Time series community **B**			Shared persistence **A & B**
	time 1	*time 2*	*time 3*	*time 1*	*time 2*	*time 3*	
OTU 1	0.4	0	0	0	0	0	
OTU 2	0	0	0	0.2	0.2	0	
OTU 3	0.1	0.1	0.2	0.1	0.1	0.1	X
OTU 4	0.2	0	0	0.2	0.1	0.1	
OTU 5	0.3	0.9	0.8	0.7	0.6	0.8	X

List of observed OTUs



Core = 2 shared OTUs

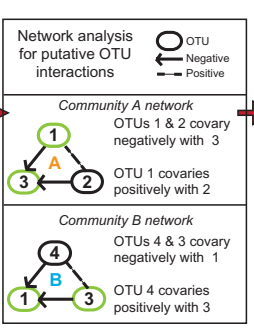
E. Connectivity:
OTUs co-varying within a community and shared across communities

1.

	Time series community **A**			Time series community **B**		
	time 1	*time 2*	*time 3*	*time 1*	*time 2*	*time 3*
OTU 1	0.1	0.2	0.3	0.3	0.2	0.1
OTU 2	0.1	0.2	0.3	0	0.2	0
OTU 3	0.3	0.2	0.1	0.1	0.2	0.3
OTU 4	0.1	0	0	0.1	0.2	0.3
OTU 5	0.4	0.4	0.3	0.5	0.2	0.3

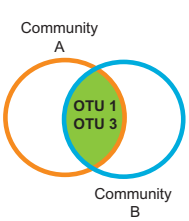
List of observed OTUs

2.



3.

	Covarying & shared **A & B**
OTU 1	X
OTU 2	
OTU 3	X
OTU 4	
OTU 5	



Core = 2 shared OTUs

analyses for exploring microbial communities (multidimensional scaling or distance-based analyses) weigh more abundant OTUs more heavily, resulting in patterns driven by the most abundant OTUs. In contrast, the Venn analysis weights all observed OTUs equally, regardless of their representation in the community.

Accounting for composition could provide key ecological information about a core microbiome (Fig. 1B). The core analysis can be restricted to an OTU subset within a specified abundance range; this eliminates OTUs that have vastly different representation in the community. Numerically dominant organisms drive community-level analyses, but it may be informative to highlight the contribution of rare members by removing dominant members. For instance, in an analysis of freshwater bacteria, members of rare OTUs were disproportionately physiologically active compared with common OTUs, as assessed by the level of 16S RNA gene transcripts (Jones and Lennon, 2010). Notably, the 'common' OTUs, those that are neither dominant nor rare and fall consistently in the middle, are often overlooked. However, these organisms may collectively serve as a supporting backbone, relevant for consistent ecosystem function. Identification of the compositional core provides a foundation for understanding the unique contributions of dominant, rare and common OTUs.

To consider composition in an analysis of core, rank abundance curves (species abundance distribution) can be created for each microbiome. Informed 'cut-offs' can be made to separate the most and least abundant OTUs. The traditional ecology literature suggests that the most abundant 10% should be considered dominant, and the least abundant 65% considered rare (Ugland and Gray, 1982). However, microbial rank abundance curves often include a very small number of dominant members and a long tail of many rare members (Sogin *et al*., 2006). This results in a highly 'skewed' rank abundance distribution that may require different definitions for dominant, common and rare OTUs. The theory and application of species abundance distributions are actively discussed in the classical ecology literature (e.g. McGill *et al*., 2007; Dornelas *et al*., 2009; Henderson and Magurran, 2010), which can inform concepts of dominance and rarity, and provide guidance for definitions in microbial ecology.

Phylogenetic and functional redundancy

Characterizing the relationship of composition and function is a central challenge in microbial ecology. Function and composition are linked in some systems, but not in others. Furthermore, some functions are restricted to certain taxa (e.g. sulfate reduction), but other functions are widespread across diverse groups (e.g. photosynthesis). A microbiome may contain both phylogenetic and functional redundancy. Phylogenetic redundancy occurs when multiple OTUs from the same lineage are present in a microbiome, while functional redundancy occurs when multiple OTUs perform the same action (e.g. nitrogen fixation) within a microbiome. Relationships within and among lineages can be hypothesized by constructing phylogenetic trees from 'omics data sets. Phylogenetic and functional redundancies can overlap when OTUs from the same lineage also have similar functional capabilities. The question for each type, however, is similar to what was proposed for calculating diversity (Martin, 2002): should redundant OTUs count equally with unique OTUs in the definition of the core (Fig. 1C)?

Redundancy is important for defining and interpreting a core. For example, disturbance ecology posits that a robust community contains multiple species that serve similar roles (e.g. the 'insurance hypothesis'; Yachi and Loreau, 1999). If this principle applies to microbial communities, redundant members may buffer disturbance response. For example, redundancy may enable recovery of community function following application of an antibiotic if only not all functionally redundant members are sensitive to the antibiotic (Dethlefsen *et al*., 2008; Antonopoulos *et al*., 2009).

Many 'omics studies provide the taxonomic identities of OTUs, and often can be analysed to identify phylogenetic signals in microbiomes (e.g. Blomberg *et al*., 2003). A first step is to determine whether the data set contains a phylogenetic signal that indicates an environmental filter. An environmental filter is a chemical or physical constraint that restricts community membership to those organisms that can are well-adapted to environmental conditions. Examples of extreme environmental filters are high pressure near hydrothermal vents, high pH in certain insect guts, or high temperatures in hot springs. Environmental filters indicated by phylogenetic signals suggest concurrence of phylogenetic and functional redundancy, which, in turn, can reveal environmental conditions that drive microbial community structure across habitats and thereby elucidate possible cross-system predictors of composition and structure.

Phylogenetic signals can be identified by analyses of phylogenetic species variability or phylogenetic species evenness (Helmus *et al*., 2007). A key assumption in these tests is that closely related species respond similarly to environmental factors (Blomberg *et al*., 2003). If a microbiome is comprised of more closely related taxa than expected, this signal suggests an environmental filter, as demonstrated in aquatic, soil and sediment ecosystems (Horner-Devine and Bohannan, 2006; Newton *et al*., 2007).

More information about a microbiome will provide a basis for hypotheses about functional redundancy. In the absence of sufficient information about functional roles

of taxa, phylogenetic signals may predict functional roles of OTUs and whether the core microbiome contains redundancy.

Persistence: dynamic microbiomes

Microbiomes are dynamic in time and space. Temporal and spatial sampling will identify OTUs that are consistently detected, or persistent, in a microbiome (Fig. 1D). A transient OTU could be a methodological error or a non-native 'tourist' species. Work in freshwater systems has shown that individual OTU dynamics do not necessarily reflect that of the whole microbiome, and that OTUs clustered by similar response patterns reveal the complexity of differential responses to environmental changes (Shade *et al*., 2010a). For host-associated microbial communities, the temporal core may be defined as those organisms that are present across developmental stages of the host, which is necessarily linked to time, but presents other experimental and biological challenges to researcher and microbe. Therefore, accounting for variation in persistence may inform understanding of the core.

After a series of observations is collected, the definition of a core may be restricted to persistent OTUs that are detected across all observations, or to a subset of occurrence patterns. For instance, a recent meta-analysis to uncover a core gut microbiome included OTUs detected in all samples taken from the same person at four time points (Sekelja *et al*., 2011). An alternative approach to defining persistence is to weight OTUs by the number of observations in which they occur, rather than using arbitrary cut-offs. This is important in interpreting time series data from environmental microbial communities, as demonstrated by an 11-month study of marine microbes in the Western English Channel (Gilbert *et al*., 2009). Though only 0.5% of observed OTUs were detected at every time point, this represented half of the generated reads and revealed that that the persistent core was comprised of abundant community members. Finally, persistence and transience could be considered similarly to concepts of dominance and rarity. Rank abundance curves in space to assess dominance may be mathematically compatible with those in time to assess the persistence (Magurran, 2008). Applying models to species persistence distributions (rather than species abundance distributions) would provide a quantitative framework for identifying a core microbiome of routinely occurring organisms.

Connectivity: interacting microbes

The concept of the core microbiome could also include information about OTU interactions, putative or proven. Many microbiologists and ecologists agree that a community is comprised of *interacting* species within a locality (e.g. Little *et al*., 2008; Siepielski and McPeek, 2010). This qualification is important, as any DNA-based sequence analysis detects OTUs that may be inactive, deceased or transient. Thus, to identify a core microbiome by connectivity, taxa are omitted if there is no indication of biological interactions (Fig. 1E). A core definition based on connectivity would include OTUs that have hypothetical interactions within a microbiome (uncovering the microbial community within a locality), but also are shared across microbiome (uncovering the core microbiome across similar habitats). One caveat to this is that some predictions about the lack of connectivity are flawed because lack of knowledge; i.e., there may be a biological connection not detected by the methods applied to the ecosystem.

Networks analyses can generate hypotheses about interacting OTUs. Networks are gaining popularity in the microbial ecology literature (Chaffron *et al*., 2010; Freilich *et al*., 2010; Shade *et al*., 2010b), and they are standard in microbial ecology workflows such as QIIME (Caporaso *et al*., 2010). There are many options for networks analysis, including some borrowed from other disciplines. For example, self-organizing maps from neural network analyses were applied to an acid-mine drainage microbial community to understand whether strong environmental constraints limited nucleotide composition (Dick *et al*., 2009). There are fewer networks analyses designed specifically for microbial communities, but one example is the local similarity analysis (LSA), created to understand long-term dynamics of marine bacterial communities (Ruan *et al*., 2006). Two strengths of LSA are: (i) it uses a correlation statistic that accommodates non-linear relationships between OTUs, and (ii) a time component can be incorporated. These are important for uncovering relationships that are not necessarily direct or instantaneous, such as, complementary oscillating dynamics between two populations that compete for a shared resource. The development of rigorous network analyses tailored to the questions of microbial ecology and capable of handling large 'omics microbial data sets will become essential for understanding the role of connectivity among members of core microbiomes. The number or strength of interactions within a microbial community may be more informative than the identity of OTUs for predicting and managing core microbiomes across systems. Classic work by Gardner and Ashby (1970) suggested that the proportion of interactions within a system modulate stability, but as the system becomes larger (more possible interactions), it may become suddenly unstable after a 'critical' proportion of interactions are exceeded. This would be a fascinating concept to test with microbiomes for different levels of complexity, from insect gut or acid mine drainage communities to diverse mammalian gut or tropical soil communities.

Caveats and discussion points

There are many caveats attendant to all of these definitions of the core microbiome. One source of substantial flexibility is the OTU definition. There are various alignment and clustering algorithms for sequence-based OTU definitions, each with different sources of bias and tolerances to method error (Quince *et al*., 2011). In many cases, the number of shared OTUs depends on the degree of sequence identity or taxonomic level chosen. For example, all bacterial communities would share one core member if defined at the domain level and, conversely, no two communities would share any core members if shared core members were defined by sequence identity of 99.9%. In an analysis of 14 environmental habitats, there was scant evidence for a core at 98%, 95%, 92% and 89% sequence identity in the 16S rRNA genes (Nemergut *et al*., 2010). In this comparative study, only 16% of OTUs were present in more than one habitat at the order level.

New methods of defining OTUs may uncover core microbiomes that were previously undetected. In a recent meta-analysis that revealed a core microbiome of gut habitats, 16S sequences were first translated into pentamer frequencies to reduce complexity, and then these frequencies were analysed by a principal components method and clustered in ordination space to define OTUs (Sekelja *et al*., 2011). By comparing this method with a 16S rRNA gene phylogenetic analysis, the authors found they could reconstruct expected diversity (Rudi *et al*., 2006). Their results suggest that using multiple methods for defining OTUs is useful in exploration of the core microbiome. A significant challenge is deciding which statistical method is most appropriate and biologically meaningful.

There are additional caveats to uncovering cores. The assignment of each OTU to a taxon introduces bias, but databases become more comprehensive with each sequencing initiative, bolstered by the efforts of the projects to target phylogenetically underrepresented microbes for sequencing (e.g. GEBA; http://www.jgi.doe.gov/programs/GEBA/index.html). Microbial profiling technology will affect description of the core. As an example from PCR-based sequencing, the read length, choice of primers and sequencing region, and depth of sequencing for each observation will influence perception of the core (e.g. Schloss, 2010; Wu *et al*., 2010). Finally, replication is imperative (Prosser, 2010). If an OTU is not detected consistently across replicates, it should not be included in the core. Care should be taken to best understand each source of bias, and to interpret results accordingly.

An alternative approach to defining the core is based on the set of OTUs that are consistently absent rather than present in samples of the microbiome or habitat. Detection technology will always limit the interpretation of data about absence because absence cannot be differentiated from an OTU that is present but below the limit of detection. However, community saturation will improve with technical advances, and confidence in the ability to exhaustively sample less-complex microbiomes may facilitate evaluation of shared absences. Coupling functional assays (functional metagenomics and chip-based analyses), analyses of processes (transcriptomics, proteomics, metabolomics) and 'omics-enabled profiling (shot-gun metagenomic or 16S rRNA gene sequencing) will deepen our understanding of the significance of both shared present and absent OTUs.

Concluding remarks

Ultimately, the appropriate definition of a core microbiome depends on the ecological question addressed. Thus, application of multiple definitions of the core microbiome will enhance ecological understanding. A Venn diagram is only the beginning.

Acknowledgements

A.S. is a Gordon and Betty Moore Foundation Fellow of the Life Sciences Research Foundation. Argonne National Laboratory provided travel support to A.S. to present these ideas at the 2010 Soils Workshop. This work was supported in part by USDA Grant No. 2006-35319-17466. We thank Jon Holt for comments on a previous version of this manuscript.

References

Andrewartha, H.G., and Birch, L.C. (1984) *Ecological Web: More on the Distribution and Abundance of Animals*. Chicago, IL, USA: The University of Chicago Press.

Antonopoulos, D.A., Huse, S.M., Morrison, H.G., Schmidt, T.M., Sogin, M.L., and Young, V.B. (2009) Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infect Immun* **77:** 2367–2375.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., *et al*. (2011) Enterotypes of the human gut microbiome. *Nature* **473:** 174–180.

Bäckhed, F. (2010) 99th Dahlem Conference on Infection, Inflammation and Chronic Inflammatory Disorders: the normal gut microbiota in health and disease. *Clin Exp Immunol* **160:** 80–84.

Benson, A., Kelly, S., Legge, R., Ma, F., Low, S., Kim, J., *et al*. (2010) Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci USA* **107:** 18933–18938.

Blomberg, S.P., Garland, T., and Ives, A.R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57:** 717–745.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Chaffron, S., Rehrauer, H., Pernthaler, J., and von Mering, C. (2010) A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res* **20:** 947–959.

Dethlefsen, L., Huse, S., Sogin, M.L., and Relman, D.A. (2008) The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6:** e280.

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Yelton, A.P., and Banfield, J.F. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10:** R85.

Dornelas, M., Moonen, A.C., Magurran, A.E., and Barberi, P. (2009) Species abundance distributions reveal environmental heterogeneity in modified landscapes. *J Appl Ecol* **46:** 666–672.

Elli, M., Colombo, O., and Tagliabue, A. (2010) A common core microbiota between obese individuals and their lean relatives? Evaluation of the predisposition to obesity on the basis of the fecal microflora profile. *Med Hypotheses* **75:** 350–352.

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., and Ruppin, E. (2010) The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res* **38:** 3857–3868.

Gardner, M.R., and Ashby, W.R. (1970) Connectance of large dynamic (cybernetic) systems: critical values for stability. *Nature* **228:** 784.

Gilbert, J., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., *et al*. (2009) The seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11:** 3132–3139.

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

Helmus, M.R., Bland, T.J., Williams, C.K., and Ives, A.R. (2007) Phylogenetic measures of biodiversity. *Am Nat* **169:** E68–E83.

Henderson, P.A., and Magurran, A.E. (2010) Linking species abundance distributions in numerical abundance and biomass through simple assumptions about community structure. *Proc R Soc B Biol Sci* **277:** 1561–1570.

Horner-Devine, M.C., and Bohannan, B.J.M. (2006) Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87:** S100–S108.

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harb Symp Quant Biol* **22:** 415–427.

Jones, S.E., and Lennon, J.T. (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci USA* **107:** 5881–5886.

Kuczynski, J., Costello, E., Nemergut, D., Zaneveld, J., Lauber, C., Knights, D., *et al*. (2010) Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol* **11:** 210.

Lederberg, J., and McCray, A.T. (2001) 'Ome Sweet 'Omics – a genealogical treasury of words. *Scientist* **15:** 8.

Little, A., Robinson, C., Peterson, S., Raffa, K., and Handelsman, J. (2008) Rules of engagement: interspecies interactions that regulate microbial communities. *Ann Rev Microbiol* **62:** 375–401.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., *et al*. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol Lett* **10:** 995–1015.

Magurran, A. (2008) Diversity Over Time. *Folia Geobot* **43:** 319–327.

Martin, A.P. (2002) Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl Environ Microbiol* **68:** 3673–3682.

Nemergut, D., Costello, E., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S., *et al*. (2010) Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13:** 135–144.

Neu, J., Lorca, G., Kingma, S., and Triplett, E. (2010) The intestinal microbiome: relationship to type 1 diabetes. *Endocrinol Metab Clin North Am* **39:** 563–571.

Newton, R.J., Jones, S.E., Helmus, M.R., and McMahon, K.D. (2007) Phylogenetic ecology of the freshwater *Actinobacteria acI* lineage. *Appl Environ Microbiol* **73:** 7169–7176.

Odum, E.P. (1953) *Fundamentals of Ecology*. Philadelphia, PA, USA: Saunders.

Prosser, J.I. (2010) Replicate or lie. *Environ Microbiol* **12:** 1806–1810.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., *et al*. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59–65.

Quince, C., Lazen, A., Davenport, R.J., and Turnbaugh, P.J. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12:** 38.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S.K., McCulle, S.L., *et al*. (2011) Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci USA* **108:** 4680–4687.

Ruan, Q.S., Dutta, D., Schwalbach, M.S., Steele, J.A., Fuhrman, J.A., and Sun, F.Z. (2006) Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22:** 2532–2538.

Rudi, K., Zimonja, M., and Naes, T. (2006) Alignment-independent bilinear multivariate modelling (AIBIMM) for global analyses of 16S rRNA gene phylogeny. *Int J Syst Evol Microbiol* **56:** 1565–1575.

Schloss, P. (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **6:** e1000844.

Sekelja, M., Berget, I., Næs, T., and Rudi, K. (2011) Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *ISME J* **5:** 519–531.

Shade, A., Chiu, C.Y., and McMahon, K.D. (2010a) Seasonal and episodic lake mixing stimulate differential planktonic bacterial dynamics. *Microb Ecol* **59:** 546–554.

Shade, A., Chiu, C.Y., and McMahon, K.D. (2010b) Differential bacterial dynamics promote emergent community robustness to lake mixing: an epilimnion to hypolimnion transplant experiment. *Environ Microbiol* **12:** 455–466.

Siepielski, A.M., and McPeek, M.A. (2010) On the evidence for species coexistence: a critique of the coexistence program. *Ecology* **91:** 3153–3164.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-P., *et al*. (2009) Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* **11:** 2574–2584.

Tilg, H. (2010) Obesity, metabolic syndrome, and microbiota multiple interactions. *J Clin Gastroenterol* **44:** S16–S18.

Tschop, M.H., Hugenholtz, P., and Karp, C.L. (2009) Getting to the core of the gut microbiome. *Nat Biotechnol* **27:** 344–346.

Turnbaugh, P.J., and Gordon, J.I. (2009) The core gut microbiome, energy balance and obesity. *J Physiol (Lond)* **587:** 4153–4158.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007) The Human Microbiome Project. *Nature* **449:** 804–810.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., *et al*. (2009) A core gut microbiome in obese and lean twins. *Nature* **457:** 480–484.

Ugland, K.I., and Gray, J.S. (1982) Lognormal distributions and the concept of community equilibrium. *Oikos* **39:** 171–178.

Vael, C., and Desager, K. (2009) The importance of the development of the intestinal microbiota in infancy. *Curr Opin Pediatr* **21:** 794–800.

Walter, H., and Box, E. (1976) Global Classification of natural terrestrial ecosystems. *Vegetatio* **32:** 75–81.

Waples, R.S., and Gaggiotti, O. (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol Ecol* **15:** 1419–1439.

Whittaker, R.H., Levin, S.A., and Root, R.B. (1973) Niche, habitat, and ecotope. *Am Nat* **107:** 321–338.

Wu, G.D., Lewis, J.D., Hoffmann, C., Chen, Y.Y., Knight, R., Bittinger, K., *et al*. (2010) Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol* **10:** 206.

Yachi, S., and Loreau, M. (1999) Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc Natl Acad Sci USA* **96:** 1463–1468.

Zaura, E., Keijser, B.J.F., Huse, S.M., and Crielaard, W. (2009) Defining the healthy 'core microbiome' of oral microbial communities. *BMC Microbiol* **9:** 259.