Review

CellPress

Omics: Fulfilling the Promise

Functional and phylogenetic assembly of microbial communities in the human microbiome

Afrah Shafquat[1], Regina Joice[1], Sheri L. Simmons[2], and Curtis Huttenhower[1,3]

[1] Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave, Boston, MA 02115, USA
[2] Bay Paul Center, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA
[3] The Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), 301 Binney Street, Cambridge, MA 02142, USA

Microbial communities associated with the human body, that is, the human microbiome, are complex ecologies critical for normal development and health. The taxonomic and phylogenetic composition of these communities tends to significantly differ among individuals, precluding the definition of a simple, shared set of 'core' microbes. Here, we review recent evidence and ecological theory supporting the assembly of host-associated microbial communities in terms of functional traits rather than specific organisms. That is, distinct microbial species may be responsible for specific host-associated functions and phenotypes in distinct hosts. We discuss how ecological processes (selective and stochastic forces) governing the assembly of metazoan communities can be adapted to describe microbial ecologies in host-associated environments, resulting in both niche-specific and 'core' metabolic and other pathways maintained throughout the human microbiome. The extent to which phylogeny and functional traits are linked in host-associated microbes, as opposed to unlinked by mechanisms, such as lateral transfer, remains to be determined. However, the definition of these functional assembly rules within microbial communities using controlled model systems and integrative 'omics' represents a fruitful opportunity for molecular systems ecology.

Microbial communities within the human microbiome

Microbial communities associated with the human body are complex ecological systems governed by many of the same processes that shape plant and animal communities. Because the composition of the human-associated microbiome is critical for normal immune development and health [1,2], there is a pressing need to develop robust models of microbiome structure and function grounded in ecological theory [3,4]. Therapies such as fecal transplants have tremendous potential to alter the microbiome to improve health [5], but these are currently blunt instruments that modify community structure in an untargeted manner. In order to understand and modulate the structure of the human microbiome – and of other microbial communities – we must better characterize the fundamental ecological processes that underlie microbiome assembly. That is, what properties (organisms, metabolic functions, regulatory circuits, etc.) must be present in 'normal' host-associated microbial communities, and how are they established and maintained?

This question was one of the early motivating factors for studies such as the Human Microbiome Project (HMP) [6], which hypothesized that essential microbial species might be distributed 'universally' among healthy human beings [7,8]. Such a 'core' microbiome, accompanied by some level of inter-individual variation, would be analogous to the core- and pan-genomes present in many microbial species [9]. A (near-)universally present, minimal set of microbial species in healthy individuals would have provided a very clear target for microbiological characterization and identification of biomarkers to distinguish between healthy and dysbiotic individuals. However, early results showed that inter-individual variability even among closely related

Glossary

Ecological drift or stochastic forces: change in population structure over time due to random factors, i.e., neutral evolution not due to selective pressure.
Functional ecology: the study of organisms within one or more habitats based on their functional traits (biochemical capabilities or larger-scale phenotypes) and the effects these have on selective pressures and community steady states.
Metacommunity: collection of populations that contribute to a particular ecology, typically connected by genetic or organismal flow of varying rates. For example, the metacommunity of a particular individual's skin microbiome might include other individuals' skin (from which organismal flow might vary depending on contact) and other body sites within that host (from which organismal flow might vary depending on similarity of the habitat).
Niche-based selection: selective pressure based specifically on occupation of a particular habitat, such that organisms with appropriate functional roles ultimately occupy that niche over time.
Selective pressure: processes that result in natural selection operating on a population, changing the composition of genetic material and/or of clades in the community. Examples include nutrient availability, diet, or climate.
Steady state: composition of features (organisms, genetic material, or compounds) within a population that does not change (or changes only slightly) over time.

Corresponding author: Huttenhower, C. (chuttenh@hsph.harvard.edu).
Keywords: human microbiome; microbial ecology; metagenomics; functional ecology.

0966-842X/$ – see front matter
 http://dx.doi.org/10.1016/j.tim.2014.01.011

individuals precluded the existence of a meaningful core microbiome in terms of microbial species [7,10]. Although the same few phyla are typically present within each body site niche of the human microbiome, their relative abundances can vary by tenfold or more among individuals, and it is not uncommon for even closely matched individuals to share 50% or less of their microbial genera [7]. This result suggested that if a core microbiome exists, it is not defined by the presence of specific microbial taxa.

An alternate hypothesis is that the core human microbiome is defined by a balance of metabolic and other functional capabilities within each body site or microhabitat. That is, the core microbiome maintained in healthy humans may consist of common gene families and pathways rather than common organisms. Theoretical work in metazoan ecology suggests that if selection within each niche (see Glossary) acts primarily on functions that are not phylogenetically unique, the result over evolutionary time is the accumulation of sets of functionally equivalent species [11]. Selection on function rather than phylogenetic identity within niches is plausible in microbial populations given high rates of lateral gene transfer, which means that not all functional traits are stably associated with particular phylogenetic markers [12]. If this was the case, a 'functional' core microbiome would act at the gene and pathway level somewhat like the human genome does for small polymorphisms, with less abundant (but highly phenotypically significant) variation occurring outside of the core and basic, common processes maintained within it. Identifying a functional core for the microbiome and other microbial communities has the potential to identify metabolic pathways critical for each environment that are maintained despite stochastic variation in the presence or absence of particular taxa. These core pathways may provide new potential targets for molecular therapies and community engineering.

Sequencing-based techniques for investigating microbial communities are currently quite cost-effective, and consist at a high level of (i) targeted amplicon-based techniques (e.g., 16S rRNA gene sequencing) and (ii) shotgun metagenomic or metatranscriptomic approaches, both reviewed in depth elsewhere [13–15]. Both can be applied to the human microbiome and other communities to study different microbial features (i.e., phylogenetic vs. genomic composition) with various tradeoffs with respect to cost, efficiency, and data quality and biases. For example, primers targeting conserved regions of 16S rRNA can lead to underrepresentation of clades with variation at the primer site. For example, the commonly used 967F primer can underrepresent clades such as the *Bacteroidetes* [16], 357F/926R (included in HMP protocols [17]) will deplete *Bifidobacterium* [18], and 515F/806R can poorly amplify *Propionibacterium*, one of the most common skin microorganisms [13]. Conversely, shotgun metagenomics can be heavily influenced by human genomic 'contamination' in environments with a high eukaryotic cell fraction, for example, saliva [19]. However, used in tandem, current technologies can provide a rich picture of microbial community phylogenetic and functional structure in order to determine how these two features interact to assemble a core human microbiome. In this review, we discuss the functional organization of human-associated microbial communities and the underlying ecological processes that affect it to drive microbiome assembly.

Assembly of habitat-specific microbial communities

To understand the processes involved in the generation and maintenance of a core human microbiome, it is important to identify the key ecological processes underlying patterns in microbial distributions. Microbial communities are shaped by a combination of selective and stochastic forces qualitatively similar to, but quantitatively differing from, those in metazoan systems: niche-based selection, dispersal, ecological drift, and mutation [11]. The balance between these forces varies over space and time and can be difficult to determine based on species abundance distributions alone, particularly when diversity is high [20]. Indeed, most studies on the roles of selection and stochasticity on microbial community assembly based on such observations suggest that both factors play significant roles in structuring communities [21,22]. This is due in part to the large differences in temporal and spatial scales relative to metazoan communities, and it requires microbial ecology to account, for example, for more stochasticity than may be apparent in larger-scale systems.

The neutral theory of community assembly has thus been particularly successful in explaining microbial ecologies [23], although interestingly it was originally developed to explain tropical trees sharing niches (slow-growing shade-tolerant and fast-growing sun-tolerant species) [24]. In its classical form it posits that steady-state species abundance distributions are based only on ecological drift, dispersal, and the diversity of organisms available for colonization (referred to as the metacommunity). Because it assumes that species are effectively functionally equivalent, it can act as a null model in contrast to selective pressure or niche specialization. However, when combined with a model of specialization and extended to evolutionary time, neutral theory suggests that sets of functionally equivalent species can evolve within the niches that are most prevalent [25]. That is, the dynamics within each niche remain neutral, but selection occurs to set the boundaries of each species set. The necessary condition for this process to occur is the absence of factors that promote competitive exclusion between functionally similar species. This pattern of sets of functionally similar species ('emergent neutrality') can appear via several different possible pathways [26,27] and is more likely to occur in species-rich communities [28]. Finally, these models typically only describe communities at steady state, such that species sets can also follow each other in successional processes; for example, any of a set of sun-tolerant, fast-growing trees could colonize a vacant spot in a forest followed by any of a set of slow-growing, shade-tolerant trees.

Microbial communities, particularly those associated with the dynamic environment of a host, are typically a good fit to such a model. They are often quite diverse, with a relatively small number of high-abundance members coupled with a long tail of low abundance members [16]. High rates of genetic exchange between spatially and phylogenetically associated microbes provide a mechanism