Mobile microbiome networks across spatiotemporal gradients

Rongling Wu (✉ rwu@phs.psu.edu)

  Penn State College of Medicine   https://orcid.org/0000-0002-2334-6421

Libo Jiang

  Beijing Forestry University

Christopher Griffin

  Pennsylvania State University University Park : Penn State

Mobile microbiome networks across spatiotemporal gradients

Libo Jiang[1], Christopher H. Griffin[3], and Rongling Wu[1,2]*

[1]Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

[1,2]Center for Statistical Genetics, Departments of Public Health Sciences and Statistics, The Pennsylvania State University, Hershey, PA 17033, USA

[3]Applied Research Laboratory, The Pennsylvania State University, University Park, PA 16802, USA
griffinch@ieee.org

*Corresponding author: rwu@phs.psu.edu (R.W.)

Abstract

Background: The microbiome, a community of microbes that co-reside in biotic or abiotic environments, underlies biogeochemical cycling, plant and animal development, and human health. Increasing evidence shows that much of the role the microbiome plays is executed through complex interactions among microbes. Thus, network reconstruction has been increasingly used as a tool to disentangle internal workings within microbial communities.

Results: We developed a general framework for recovering microbial interaction networks from any design of microbial experiment. This framework represents a quasi-dynamic game model (qdGM) derived from the seamless integration of evolutionary game theory and allometric scaling laws. The qdGM can not only characterize how individual microbes act singly, but also reveal how different microbes interact with each other to govern microbial community assembly. Beyond existing approaches that can only identify a single overall microbial network from a number of samples, our framework can track and visualize how interaction networks vary from sample to sample and covert sample-specific (personalized) networks into context-specific networks. More importantly, this framework can reconstruct such mobile microbial networks from steady-state data, facilitating the widespread use of network tools to understand the impact of the microbiome on natural processes.

Conclusions: As proof of concept, we used the new framework to analyze human gut microbiota data and interspecific animal-associated microbiota data. Mobile networks reconstructed from each dataset can characterize previously unknown mechanisms that drive the change of microbial interaction architecture and organization along spatiotemporal gradients. This framework provides a tool to generate process-specific microbiome networks that can be readily translated into various biotechnological applications and evolutionary studies.

Keywords: Microbiome, biogeographic ecology theory, evolutionary game theory, microbial network, ordinary differential equation

Background

Microbial communities are an important determinant of a wide range of natural processes, from biogeochemical cycling to plant and animal development to human health [1-4]. In

recent years, stimulated by recent advances in DNA sequencing that make it possible to map the composition of microbial communities with high resolution, a surge of interest has been devoted to understand the ecological mechanisms that govern microbial community assembly and function [5]. Surveys of microbiome composition across various ecological settings from the soil to the human body have consistently revealed that microbes are organized into functional and stable communities through fundamental ecological principles [6-8]. Several studies have begun to reconstruct microbial interaction networks [7-11], but existing approaches, mostly borrowed from biophysical fields, have significant drawbacks in integrating the unique ecological structure of microbiomes and revealing their systematic, mechanistic impact on natural processes.

Increasing evidence shows that microbial community composition changes along environmental gradients [12,13] and from healthy to diseased states [4,14]. Most existing approaches can only infer a single aggregate network from a number of samples, failing to characterize how microbial interactions vary according to geographic, biotic, and developmental contexts and further chart causal links between these interactions and biogeochemical cycling or host phenotypes. Fully informative networks are defined as those filled by an omnidirectional set of bidirectional, signed, and weighted interactions, and they can better reveal the mechanistic workings of microbial assembly. However, reconstructing such networks critically relies on the availability of temporal or perturbed abundance data [15]. These data are difficult to collect in practice; for example, it is ethically and logistically inappropriate to repeatedly sample gut microbiota from human hosts. Xiao et al.'s [16] developed a Jacobian matrix-based approach for reconstructing informative networks from steady-state data, but it is unclear to what extent this approach can be scaled up to predict network properties of large microbial communities across a range of environmental and biological changes.

Here, we develop a general framework for reconstructing fully informative microbial networks using static data collected from any design of experiment. We introduce ecology theory [18] to define new concepts that describe the ecological relationship of an individual microbe (defined as the niche) to all microbes (defined as the habitat). We integrate these concepts and evolutionary game theory [19] to derive a quasi-dynamic game model (qdGM). By modeling the niche-habitat relationship across samples, the qdGM can quantify how

microbes vary independently and how they interact with each other over natural processes. Variable selection equips qdGM with a computational capacity to reconstruct high-dimensional but sparse microbial networks. Taken together, not relying on temporal or perturbed data, our framework can reconstruct mobile microbial networks that topologically vary among hosts, environments, and times.

The qdGM formulation

Networks are snapshots of biological processes and particularly central to the functionality of microbes in ecological communities across time and space. Thus, the reconstruction of dynamic networks is of paramount importance to better understand and reveal general principles that mediate microbe-environment (biotic and abiotic) relationships. Not relying on temporal data, the qdGM can infer dynamic microbial networks. To formulate this theory, it is useful to consider a series of *n* samples, each representing an independent microbial community, which are drawn to reflect a spectrum of ecological or health processes. In ecology, a habitat is defined as the sum of all resources and environmental conditions in a community all species can jointly exploit to live and grow [20], whereas a niche defined as the subset of those factors that supports a specific species' survival and growth [21]. Being part of the habitat, different niches may have the same or different types and numbers of environmental dimensions (including biotic and abiotic) [22], which establishes the material, energy, or signal basis for various species to interact with each other through cooperation or competition. In this regard, the pattern of how each niche (part) scales with the habitat (whole) determines the type and strength of interspecific interactions. We argue that the allometric scaling of a niche with the habitat as a part-whole relationship can be explained by resource allocation theory [23] that obeys the power law [24].

Under natural conditions, modeling the niche-habitat relationship via its underlying environmental components is impossible given the confounding effects of multiple environmental factors. Since the number of microbes that inhabit the community virtually reflects its carrying capacity to supply essential resources and conditions that maintain these microbes' survival and growth [25,26], we coin the total amount of abundance of all microbes in a sample as the habitat index (HI) of this sample and the abundance level of a specific microbe in the sample as the niche index (NI) of this microbe within this sample. Let

M_{ji} denote the abundance level (or NI) of a microbe j ($i = 1, .., m$) from sample i ($i = 1, .., n$) and $H_i = \sum_{j=1}^{m} M_{ji}$ denote the HI of sample i. Thus, modeling how M_{ji} scales with W_i can facilitate the characterization of how microbes diversify and interact with each other across samples. The concepts of HI and NI, analogous to the environmental index defined to quantify the overall quality of site in terms of the accumulative growth of all plants [27,28], can reflect the overall adaptation of all microbes or a specific microbe to the microbial assembly environment.

The microbes, co-inhabiting the same environment, promote or inhibit each other in a complex network. This process acts like a game in which each player may choose to compete or cooperate with its opponents to maximize its payoff. Such strategic choices include a rational judgement based on a player's accrued knowledge of the environment affected by other players [29]. Evolutionary game theory models the dynamic change of strategy frequency in the population as a result of evolutionary pressure [19]. In an evolving population, any strategy used by an individual to maximize its payoff would be constrained by strategies of other individuals that also strive to maximize their own payoffs and, ultimately, this process through natural selection, would optimize the structure and organization of the population, making it reach maximum payoff [19]. We integrate such evolutionary game theory and concepts of NI and HI through Lokta-Volterra prey-predator equations [30] to develop a system of ordinary differential equations (ODEs), expressed as

$$\frac{dM_{ji}}{dW_i} = g_j\big(M_{ji}(W_i):\Theta_j\big) + \sum_{j'=1,j'\neq j}^{m} g_{jj'}\big(M_{j'i}(W_i):\Theta_{jj'}\big), j = 1, \dots, m; i = 1, \dots, n \qquad (1)$$

which quantifies and dissects the scaling relationship of NI with HI into two distinct components, the endogenous component $g_j(\cdot)$ that describes how a specific microbe j is distributed across samples if it is assumed to be in isolation and the exogenous component $g_{jj'}(\cdot)$ that specifies how the across-sample distribution of the focal microbe is affected by other microbes j' ($j' \neq j$). Equation (1) quantifies how microbes exist independently and dependently with others across samples, presenting a system of ordinary differential equations (ODEs) with the time derivative replaced by the HI derivative, which are called the qdODEs. Based on these qdODEs, we formulate the qdGM as an alternative theory to time-

based evolutionary game theory. The endogenous and endogenous components of qdODEs are described by ODE parameters Θ_j and $\Theta_{jj'}$, respectively, which need to be estimated from microbial observations.

In the Supplementary Text, we provide a detail on how to obtain the maximum likelihood estimates (MLEs) of Θ_j and $\Theta_{jj'}$ by solving qdODEs in equation (1) under the maximum likelihood setting implemented with variable selection. By integrating this system of qdODEs over W_i, we estimate the integrals of the endogenous and exogenous components, $G_j(W_i)$ and $G_{jj'}(W_i)$, respectively, and contextualize these integrals into a graph in which the former represents nodes and the latter stands for edges that link different microbes. Because both $G_j(W_i)$ and $G_{jj'}(W_i)$ are a function of W_i, the microbial networks reconstructed are sample-specific. These so-called mobile networks represent a series of snapshots of natural processes.

Identifying and testing context-specific networks

If the habitats are sampled from different biological, experimental, or environmental contexts, we can convert sample-specific networks into context-specific networks and characterize how interaction architecture changes along spatiotemporal gradients. Assume that n samples contain C contexts, each with n_c samples ($c = 1, \ldots, C$). We extend equation (1)'s qdODEs to include context-specific endogenous and exogenous components of each microbe j, described by Θ_j^c and $\Theta_{jj'}^c$, expressed as

$$\frac{dM_{ji}}{dW_i} = \sum_{c=1}^{C} \left[g_j\left(M_{ji}(W_i)\colon\Theta_j^c\right) + \sum_{j'=1,j'\neq j}^{m} g_{jj'}\left(M_{j'i}(W_i)\colon\Theta_{jj'}^c\right) \right], j = 1, \ldots, m; i = 1, \ldots, n \quad (2)$$

where i denotes the ith sample of context c when this context is considered.

The algorithm described in the Methods can be used to solve the MLEs of Θ_j^c and $\Theta_{jj'}^c$ from which to estimate the endogenous and exogenous components of microbial abundance, $G_j^c(W_i)$ and $G_{jj'}^c(W_i)$, under context c based on the integrals of equation (2). Equation (1) fits the abundance data across all samples, regardless of their origins, which is called a context-agnostic model, whereas equation (2) fits the data across samples separately from different contexts, thus called a context-specific model. In the Methods, we also provide a procedure for testing whether n samples can be classified into C contexts based on the log-

likelihood test of context-agnostic vs. context-specific models. If the context-specific model is accepted, we need to reconstruct microbial networks separately for different contexts. The overall network under context c can be reconstructed by estimating the average values of nodes and edges over n_c habitats, expressed as

$$G_j^c = \frac{\sum_{i=1}^{n_c} G_j^c(W_i)}{n_c}, \tag{3A}$$

$$G_{jj'}^c = \frac{\sum_{i=1}^{n_c} G_{jj'}^c(W_i)}{n_c}. \tag{3B}$$

Thus, we will have C such context-specific microbial networks. We can test and compare how microbial networks change under different treatments, with environmental signals, and over times. All this is of fundamental importance for our detailed understanding of microbial impacts on natural processes.

An approach for testing the relationship between microbial networks and contexts is to draw the HI-varying curve for each microbe and decompose it into its endogenous and exogenous curve components. By comparing these two types of curves, we can assess how the existence of microbes is influenced by others and how this influence is determined by contexts.

Ecological and social dissection of microbial networks

In the context l-specific network, $G_j(l)$ represents the intrinsic capacity with which microbe j grows in isolation, and $G_{jj'}(l)$ reflects the pattern and strength with which microbe j is affected by microbe j' under context l. If $G_{jj'}(l)$ is positive or negative, this suggests that microbe j is promoted or inhibited by microbe j', respectively. The zero value of $G_{jj'}(l)$ suggests that microbe j' is neutral to microbe j. The size of $G_{jj'}(l)$ describes the strength of promotion or inhibition. By comparing $G_{jj'}(l)$ and $G_{j'j}(l)$, we can determine whether and how these two microbes reciprocally trigger impacts on each other. If both are positive or negative and equal in magnitude, this suggests that microbes j and j' have established a relationship defined as *symmetric mutualism* or *symmetric antagonism*. If $G_{jj'}(l)$ and $G_{j'j}(l)$ are positive or negative but not equal in magnitude, these two microbes have a relationship defined by *asymmetric mutualism* or *asymmetric antagonism*, respectively. If $G_{jj'}(l)$ is positive or negative but $G_{j'j}(l)$ is zero, this suggests that microbe j', respectively, exerts

commensalism or *amensalism* toward microbe *j*. If $G_{jj'}(l)$ is positive but $G_{j'j}(l)$ is negative, this suggests that microbe *j'* offers *altruism* toward microbe *j* or microbe *j* triggers *parasitism* or *predation* on microbe *j'*. Taken together, we reconstruct a fully informative network that encapsulate bidirectional, signed, and weighted microbial interactions.

If a microbe has more links than the average of all microbes (i.e., connectivity) within a network, this microbe is regarded as a keystone microbe. A keystone species plays a disproportionately large role in maintaining the structure of an ecological community relative to its abundance [11,31]. Such species affect many other organisms in an ecosystem and help to determine the types and numbers of various other species in the community. The determination of a keystone species is based on the total number of links with it. Yet, a link can be active or outgoing in which case a species affects other species through promotion or inhibition. A link can also be passive or incoming, which means that a species is promoted or inhibited by other species. If a microbe has more outgoing links compared to its incoming links, this microbe is regarded as being "social." If the number of incoming links of a microbe is larger than that of its outgoing links, this microbe is "subordinate." A microbe can be viewed as a "leader," if the number of its outgoing links outperforms the average of all microbes. A "solitary" microbe is one that has fewer links, both outgoing and incoming. The classification of microbes into social, subordinate, leader, and solitary types can help researchers not only better understand the role of each microbe, but also exacerbate the mass or signals that are essential for relevant interaction patterns.

Validating and testing the qdGM

We show how qdGM can be used for network reconstruction in practice. As the proof of concept, we use two examples, human gut microbiota and microbiota residing in bodies of different animal species, although qdGM has a more widespread use in microbial research. We also perform a simulation study to statistically validate its usefulness and utility.

Networking the gut microbiota
Our model was used to analyze a gut microbial abundance dataset containing eight phyla, 23 classes, 28 orders, 50 families, and 101 genera collected from 127 heathy hosts from a

founder, the Hutterites [32]. These hosts include different sexes, different ages, and two treatments, use vs. no use of antibiotic intervention, of whom 93 were measured in winter, 91 in the following summer, and 57 in both. We define a measurement as a sample and, thus, this study contains 184 samples. We first reconstructed 184 sample-specific (i.e., dynamic) networks at the genus level based on the context-agnostic model of equation (1) and then test whether these networks should be classified into season-, sex-, age-, and treatment-dependent microbial networks using the context-specific model of equation (2).

We calculated the log-likelihood ratio of microbial abundance (817.73) under the season-specific and season-agnostic model, which is larger than the critical threshold at the 5% significance level. Despite the existence of a season-specific difference, microbial networks display a *structural* similarity between winter and summer. A total of 101 genera are expected to form 5,050 possible interactions or edges, but networks in both seasons only have a small portion of interactions detected to exist, i.e., 213 (4.2%) for winter and 225 (4.5%) for summer, forming highly sparse microbial networks (Fig. 1**A**). The total number of interactions follows a scale-free network law (Fig. 1**B**; P < 0.01). The networks do not contain mutualism, antagonism, and parasitism, but are dominated by two asymmetric interactions – commensalism (129 or 2.6% for winter 121 or 2.4% for summer) and amensalism (84 or 1.6% for winter 104 or 2.1% for summer). We found that season-specific differences are expressed in the *organizational* arrangement of networks. In the winter network, a small set of genera, including *Veillonella*, *Victivallis*, *Prevotella*, *Acidaminococcus*, and *Thermus*, are the primary keystones, possessing many more interactions (mostly through outgoing links) than the average. These genera may serve as leaders that are central to the networks. About twenty genera are the secondary keystones each with a much reduced number of links. A majority of genera tend to receive passive links, behaving like subordinates. It appears that over a half of all genera are solitary because they are passively linked to only a few other genera. The summer network contains a different set of primary keystones, i.e., *Lactonifactor, Odoribacter, Peptococcus, Paenibacillus*, and *Acetivibrio*. There are some overlaps in secondary keystones between winter and summer networks.

As observed above, microbial networks display season-specific differences in terms of their organization rather than structure (Fig. 1). This statement is supported by *keystoneness* as a

quantitative descriptor of the organization of microbial networks. Keystoneness contributes considerably to season-dependent difference (Fig. 1**B**). In the winter network, only a small portion of genera display a measureable capacity to link with other microbes through outgoing keystoneness (primary keystones) and incoming keystoneness (secondary keystones). Yet, keystoneness increases dramatically from winter to summer for a majority of genera, including primary keystones, secondary keystones, and non-keystones. Only a few genera, especially *Veillonella*, display a decrease from winter to summer

High variability was observed in keystoneness between different hosts in both winter and summer, although, for most genera, inter-host variability is much more pronounced in summer than in winter (Fig. 2), suggesting that keystoneness can serve as a major determinant of interpersonal variability in gut-microbial interaction networks. Taken together, our findings suggest that, on average, gut-microbial community assembly responds to seasonal change not through altering the overall architecture of its gut-microbial interactions, rather through quantitatively changing the role of individual microbes, especially those highly social microbes, in microbial network organization.

We also compared sex-, age-, and treatment-specific microbial networks. We did not found qualitative differences of network architecture between different sexes, ages (young, ages = 19–30 years, vs old, ages > 60 years), and treatments (use vs. or no use of antibiotics), but these context-specific network differences are displayed in network organization and can be quantitatively explained especially by keystoneness (**Fig. S1 – S3**).

We decomposed the overall abundance of each genus into its underlying endogenous component (determined by the intrinsic capacity of this genus) and exogenous component (affected by ecological interactions between other genera and this genus) and visualized how these two components contribute differently to microbial abundance across samples (Fig. 4). In summer, genus *Actinomyces* receives incoming promotion from *Dorea* and *Barnesiella* and incoming inhibition from *Enhydrobacter*, *Solobacterium*, *Subdoligranulum* (Fig. 4**A**). The strength of promotion and inhibition increases exponentially with HI, despite genus-dependent difference in increase slope. Together, the influences of these genera make *Actinomyces*'s overall abundance higher than its endogenous component. However, in summer, the endogenous component of *Actinomyces* is higher than its overall level because of a considerably high promotion extent by genus *Paenibacillus*. Also, the types of genera

that promote or inhibit *Actinomyces* change from winter to summer. In winter, genus *Bacteroides* increases its abundance with HI, a process influenced by numerous genera, of which two act through promotion and six through inhibition. Yet, in summer, its endogenous abundance decreases with HI, but with a positive promotion by *Lactonifactor* which makes *Bacteroides* outperform its endogenous component. These decomposition analysis provides a global and detailed picture of genus-genus interactions, their signs, magnitudes, and context-dependent changes, within the gut microbiota.

Networking the impact of microbiomes on animal evolution

More recently, a new concept of hologenome (host genome plus microbiome) has been proposed to postulate that the holobiont (host plus symbionts) evolves through natural selection because microbial symbionts can be transmitted from parent to offspring via cytoplasmic inheritance, coprophagy, direct contact during and after birth, and the environment [33,34]. This is contrast to the traditional Darwinian theory, stating that hosts are the units acted upon by natural selection [35]. Increasing studies have demonstrated that microbial symbionts contribute to the anatomy, physiology, development, innate and adaptive immunity, and behavior and finally also to genetic variation and the origin and evolution of species [36,37]. Given these findings, it is of great interest and importance to study how microbial interactions co-evolve with hosts.

We analyzed gut bacterial 16S rRNA gene sequence data collected on 265 specimens from 64 species [38]. The bacterial lineages collected were placed into phylogenetic taxa, approximately including 65 families that are common to all animal species. By modeling the total number of bacterial lineages per gut sample as a function of animal mass, a positive association was observed with a slope of 0.053 ($P < 0.05$), paralleling the known species-area relationship theory derived from island biogeography, consistent with the second law of thermodynamics [39,40]. We modified the system of qdODEs in equation (1) by replacing the habitat index derivative with the animal mass derivative and used these equations to reconstruct specimen-specific microbial networks which are further converted into species-specific networks. Figure 4 provides the examples of three animal species, bedbug (small size), hermitcrab (moderate size), and whale (large size). We found that the architecture of species-specific microbial network is predominated by commensalism and antagonism, together accounting for almost all links of microbial networks. The connectivity of

microbiome networks is proportional to animal size; network architecture is more complex on whale than on hermitcrab as well as on hermitcrab than on bedbug. Family *Oxalobacteraceae* is a microbe found to serve as a primary keystone in microbial networks on all animals, regardless of their size. Yet, families *Verrucomicrobiaceae* and *Enterobacteriaceae* add to a list of keystones in small-size bedbug mostly by receiving incoming links, whereas moderate-size hermitcrab and large-size whale have family *Helicobacteraceae* as a secondary keystone through outgoing links although, for the latter, many tertiary keystones are identified. Network properties, including the number of links for each bacterial family, changes dramatically among animal species, which may also be a determinant of animal adaptation to ecological environments.

Computer simulation

Results from simulation studies suggest that the model can capture a general trend of how each microbe interacts with others to form a network when the data contain a modest sample size ($n = 50$) and modest measurement error variance. The estimated networks approach the true networks when sample size increases and/or the error variance decreases. It appears that a sample size of 250 can reasonably well estimate a 100-node network from our model for error variance of intermediate size. **Figure S4** illustrates the comparison of estimated and true abundance curves as a function of habitat index for two microbes chosen from a pool of 100 microbes. A sample size of 50 and modest error variance can estimate and decompose the overall abundance curve of a microbe into its endogenous and exogenous components, but with a fairly large bias. To obtain an acceptable curve estimate, a sample size of 250 is needed with modest error variance. The bias of estimation due to unpredictable large error variance can be compensated with a larger sample size.

Discussion

Although microbial community assembly exists as a highly complex, heterogeneous, and dynamic system, its underlying inner workings may still obey some certain fundamental principles that guide how microbes interact with each other to achieve maximum inclusive fitness [20]. No single theory can effectively illustrate the overall picture of microbial communications and interactions and reveal the detailed mechanisms underlying these interactions. We propose a quasi-dynamic game model (qdGM) by combining elements of

biogeographic ecology theory and evolutionary game theory through statistical reasoning, aimed to unveil a basic rule for gut-microbial diversity, structure, and function. One main merit of qdGM lies in its capacity to pack steady-state abundance data into dynamic and mechanistic networks that encode and quantify all possible patterns of microbe-microbe interactions at play through bidirectional, signed, and weighted links. In the past, such fully informative networks only could be reconstructed from dense temporal data, although this type of data may be unavailable and less informative in gut microbiota studies.

In a population, the maximum payoff achieved by an organism may be quickly counteracted by the strategies of other organisms which also strive to maximize their payoffs [37]. This process will ultimately make the whole population achieve an optimal (equilibrium) payoff [41], which leads us to establish a statistical foundation for reconstructing a stable microbial network through maximizing the likelihood of the abundance data of all microbes [42]. The qdGM provides a way for quantitatively inferring such microbial networks, but not relying on temporal data. If the two microbes are mutualistic, this implies that they produce certain factors that are used by each other. In contrast, an antagonistic relationship implies that one microbe produces chemicals or signals that harm the other. Our model can also classify all microbes into social types, leaders vs. subordinates or social vs. solitary. The ecological and social classification of microbes can not only help researchers excavate and extract the mass, energetic, or signal basis for microbial interactions, but also make Kong et al.'s [43] designer microbial consortia study more efficient and predictive.

Existing approaches can only identify a single overall microbial network from a large number of samples, failing to characterize network variability. The most remarkable advantage of qdGM is that it can reconstruct sample-specific networks (i.e., mobile networks) and convert them into context-specific networks through which to test how microbial interactions alter from treatment to treatment and vary along spatiotemporal gradients. Reanalyzing Davenport et al.'s [32] Hutterites data, we found that the overall structure of gut-microbiome networks is predominated by commensalistic or amensalistic interactions. This phenomenon was also observed in animal species-specific microbial networks. The overwhelming existence of commensalism or amensalism is consistent with widely identified cyclic synergism or antagonism (e.g., cyclic dominance) in nature, guided by the rock–paper–scissors game [44]. In winter's gut microbiota, we identified a few leaders, such as genera *Prevotella*,

Veillonella, and *Victivallis*, which manipulate many other microbes and play a pivotal role in mediating the organization and function of microbial networks. Over a half of microbes in the gut are solitary because they only receive one or two incoming links from other microbes. Although gut microbial networks do not make qualitative structural changes in response to seasonal change, sex difference, aging, and antibiotic treatment in their structure, they respond to these environmental or developmental changes in a quantitative way through network organization. For example, keystoneness, as a parameter defined to quantify the capacity of individual microbes to establish social links, displays a pronounced season-dependent difference. Part of the reason why gut-microbial networks change quantitatively in organization rather than qualitatively in structure may result from the strategy of sampling a healthy population. We postulate that the transition from organizational to structural change in microbial networks may form a hallmark for the change of host health.

The qdGM can quantify the popularity of a given microbe in microbial community assembly and identify which microbes are more social and which are more solitary. The classification of these two types of microbes has great implications for designing precision medicine. For example, if both social microbes and solitary microbes impact health risks, their clinical implications should be strikingly different. For the latter, simply activating or repressing their abundance with little need to consider other microbes could be a sufficient strategy for improving human health, but this strategy may not work for the former because any alteration of their abundance may perturb the microbial equilibrium in the gut. Our result further shows that the gross abundance of each microbe is not merely determined by its own intrinsic capacity but also by the strategy of how it interacts with other microbes. Quantitatively, the latter contributes to about 40% of the microbial diversity in the gut. This result implies the importance of both direct and indirect effects on health risks by microbial networks. The use of qdGM in the Hutterites data analysis seems to gain new insights into the gut microbiota, although we emphasize that this analysis is an illustrative starting point and future extensions are required.

The application of the qdGM to analyzing microbial data from different taxa of animals show its promising utility to reveal the evolution and ecology of holobionts, a new concept that has received increasing attention in modern biology [36,37]. The new model can not only identify which microbes drive animal evolution and ecology, but also reveal how different microbes

work together to shape evolutionary processes. Based on data reanalysis of microbial abundance from 64 animals, we found that the role of most microbes as a mediator of networks change from animal to animal. As a primary keystone, family *Oxalobacteraceae* affects microbial networks across all animals of different sizes, but its role is played differently, depending animal size; by both incoming and outgoing links in small- and moderate-size animals but only by outgoing in large-size animals. We found a clear trend, i.e., microbial networks have greater connectivity for large than small animals. Taken together, the qdGM is armed with a capacity to identify key microbial indicators that may have played an important role in mediating animal energy expenditure and gain and, further, animals' adaptation and evolution over environmental change.

Results from simulation studies show that the qdGM has favorable statistical properties, and is expected to produce reasonably accurate microbial network with a modest sample size. As with all network inference methods, however, we cannot expect the networks inferred from the qdGM to cover all issues typical of microbial communities. The key step of the network inference is to solve a system of ODEs. Currently, we implemented the widely used fourth-order Runge-Kutta algorithm to estimate equation parameters. More robust algorithms, especially those with a capacity to handle high- or even ultrahigh-dimensional systems of differential equations and tackle errors due to experimental uncertainty or mis-assignment of sequencing reads into operational taxonomic units through stochastic differential equation modeling [11], are needed. The qdGM focuses on a set of homogeneous samples, but it can be readily extended to assemble microbial data from different sources with various perturbations [45]. With these and other modifications, the qdGM could potentially provide a powerful approach for inferring biologically meaningful microbial networks that contribute substantially to translational medicine and evolutionary research.

Conclusions

Networks are central to the functionality of complex systems. Existing approaches can only reconstruct a single aggregate microbial network from a number of samples, failing to characterize population heterogeneity. Furthermore, identifying the full information of networks that encapsulate bidirectional, signed, and weighted microbe-microbe interactiomes

critically relies on the availability of temporal or perturbed data, which are difficult to collect or not informative for biological modelling. We develop a statistical model that can recover sample-specific informative networks from steady-state data. We integrate allometric scaling theory and evolutionary game theory to quantify sample-specific changes of microbial abundance and interactions by deriving a system of quasi-dynamic ordinary differential equations (qdODEs). We integrate functional clustering and variable selection to identify distinct network communities from high- or ultrahigh-dimensional microbiome data and the most significant links a given microbe receive to form sparse networks. We incorporate a maximum likelihood approach for estimating qdODE parameters that constitute the stability of microbial networks. We formulate a statistical platform for comparing and testing how microbial networks evolve over a phylogenetic clade of hosts, how they change across spatiotemporal gradients, and how they establish causal interrelationships with biogeochemical processes. We derive the mathematical properties of qdODE solving and perform extensive computer simulation studies to investigate the statistical behavior of microbial networks. The networks inferred from model can reveal the full information of microbial interactions, interrogate their ecological properties, and serve as an approach for visualizing how microbial communities assemble, stabilize, and evolve in a range of biological and biotechnological contexts.

Statistical Methods

Regression model

Given microbial community assembly containing m microbes measured on n sample, let $\mathbf{y}_j = (y_j(W_1), \ldots, y_j(W_n))$ denote a vector of observed abundance values for microbe j ($j = 1, \ldots, m$). Based on the structure of qdODEs in equation (1), this microbe's habitat index-varying abundance level can be described by a multiple regression model, expressed as

$$y_j(W_i) = G_j\big(y_j(W_i)\!:\!\Theta_j\big) + \sum_{j'=1, j' \neq j}^{m} G_{jj'}\big(y_{j'i}(W_i)\!:\!\Theta_{jj'}\big) + e_j(W_i), \qquad (4)$$

$$= a_j(W_i) + X_j^T \mathbf{b}_j(W_i) + e_j(W_i). \qquad (5)$$

In equation (4), $G(\cdot)$ and $G_{jj'}(\cdot)$ are the habitat index-varying independent and dependent

abundance of microbe j, whose derivatives are $g_j(\cdot)$ and $g_{jj'}(\cdot)$ of equation (1), respectively, and $e_j(W_i)$ is the residual error of microbe j at sample i, obeying a multivariate normal distribution with mean vector $\mathbf{0}$ and sample-dependent covariance matrix $\mathbf{\Sigma}_j$ for microbe j. We assume that the residual errors of microbial abundance are independent among samples so that $\mathbf{\Sigma}_j$ is structured as $\mathbf{\Sigma}_j = \sigma_j^2 \mathbf{I}_n$ where σ_j^2 is the residual variance of microbe j at the same sample and $\mathbf{I}_n$ is the identity matrix. In equation (4B), we have $a_j(W_i) = G(\cdot)$ and $X_j^T \mathbf{b}_j(W_i) = \sum_{j'=1,j'\neq j}^{m} G_{jj'}(\cdot)$, where X_j^T is the vector containing $m-1$ ones and $\mathbf{b}_j(W_i)$ = $(b_{j1}(W_i), \ldots, b_{jm}(W_i))$ is a vector of the dependent value of microbe j determined by all microbes, except for j.

Variable selection

In statistics, it is extremely difficult to model and estimate the effects of all other microbes on a focal microbe when the number of microbes is large. Indeed, this rarely happens in biology. For these reasons, we need to choose a small set of the most significant microbes that affect a focal microbe. To do so, we need to choose appropriate functions to fit $G_j(\cdot)$ and $G_{jj'}(\cdot)$ for equation (4) where the observed abundance level of microbe j is treated as a response and the observed abundance levels of all other microbes treated as predictors. Although their biological meanings can be hardly validated, non-parametric approaches, such as B-spline or Legendre Orthogonal Polynomials (LOP), can be statistically effective and efficient in fitting such functions through their linearization. As a first step toward network reconstruction, we incorporate the LOP into model (5) for the following variable selection step.

Group LASSO: Considering a focal microbe j as a response, we use nonparametric independent parameters $\mathbf{a}_j$ to fit its $G_j(\cdot)$. As predictors, $(m-1)$ microbes contribute to microbe j's dependent component through unknown nonparametric unknown nonparametric dependent parameters $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{j1},\ldots,\boldsymbol{\beta}_{j(j-1)},\boldsymbol{\beta}_{i(j+1)},\ldots,\boldsymbol{\beta}_{jm})$. Thus, we have $m-1$ groups of dependent parameters that reflects the influence of other microbes on the focal microbe. We implemented group LASSO [46] to select those nonzero groups. The group LASSO estimators of dependent parameters, denoted as $\dot{\boldsymbol{\beta}}_j = (\boldsymbol{\beta}_{j1}, \ldots, \boldsymbol{\beta}_{j(j-1)}, \boldsymbol{\beta}_{j(j+1)}, \ldots, \dot{\boldsymbol{\beta}}_{jd_i})$, where d_j $\ll m$ is the number of the most significant microbes that interact with microbe j. It can be

obtained by minimizing the following penalized weighted least-square criterion,

$$L_1(\dot{\boldsymbol{\beta}}_j, \lambda_j) = (\mathbf{y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j)^T \mathbf{Z}_j (\mathbf{y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j) + \lambda_{1j} \sum_{j'=1, j' \neq j}^{m} \left\| \beta_{j|j'} \right\|_2, \qquad (6)$$

where $\mathbf{y}_j = (y_j(W_1), \ldots, y_j(W_n))$, $\boldsymbol{\mu}_j = (\mu_j(W_1), \ldots, \mu_j(W_n))$, and $\mathbf{b}_j = (\mathbf{b}_j(W_1), \ldots, \mathbf{b}_j(W_n))$ determined by $\boldsymbol{\beta}_j$ through a nonparametric link; λ_{1i} is a penalty parameter determined by BIC or extended BIC; and $\mathbf{Z}_j = \text{diag}\{z_j(W_1), \ldots, z_j(W_n)\}$ where $z_j(W_i)$ is a prescribed nonnegative weight function on $[W_1, W_n]$ with boundary conditions $z_j(W_1) = z_j(W_N) = 0$. This weight function is used to speed up the rate of convergence.

Adaptive group LASSO: In group LASSO, penalty parameters of each group are treated equally, without considering the relative importance of different groups. It has been recognized from traditional linear regression analysis that the over-penalization of parameters for some predictors may reduce the efficiency of parameter estimation and the continuity of variable selection [46]. To overcome this limit of group LASSO, Wang and Leng [47] integrated it with adaptive LASSO to create adaptive group LASSO. This integrative approach selects significant groups by weighted penalty parameters. Weight $w_{j|j'}$ is obtained as $\left\| \beta_{j|j'} \right\|_2^{-1}$ if $\left\| \beta_{j|j'} \right\|_2 > 0$ and ∞, otherwise. The adaptive group LASSO estimators of dependent parameters are obtained by minimizing the penalized weighted least-square criterion as follow:

$$L_2(\dot{\boldsymbol{\beta}}_j, \lambda_j) = (\mathbf{Y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j)^T \mathbf{Z}_j (\mathbf{Y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j) + \lambda_{2j} \sum_{j'=1, j' \neq j}^{m} w_{j|j''} \left\| \beta_{j|j''} \right\|_2 \quad (7)$$

where λ_{2j} is a penalty parameter determined by BIC or extended BIC.

Likelihood formulation

Through variable selection, we change the number of incoming links for a microbe j from m to d_j. A number of approaches, including non-linear least-squares and maximum likelihood, can be implemented to solve the shrunk qdODEs from equation (1). Since we argue that microbial community assembly tends to reach its maximum overall payoff guided by evolutionary game theory, a maximum likelihood approach that is founded on the most probable existence of all microbes is chosen for our qdODE solving. Let $\boldsymbol{\phi} = (\boldsymbol{\mu}; \boldsymbol{\Sigma})$ denote the parameters that explain the regression model. The likelihood function of $\boldsymbol{\phi}$ given the abundance data is written as

$$\mathcal{L}(\boldsymbol{\mu}; \boldsymbol{\Sigma}) = f(\boldsymbol{y}_1, \dots, \boldsymbol{y}_m | \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m; \boldsymbol{\Sigma}) \tag{8}$$

where $f(\cdot)$ is the n-dimensional m-variate normal distribution for m microbes across n samples with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Specifically, we model the mean vector by equation (5) subject to variable selection, i.e.,

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1; \dots; \boldsymbol{\mu}_m) = \big(\mu_1(W_1), \dots, \mu_1(W_n); \dots; \mu_m(W_1), \dots, \mu_m(W_n)\big)$$

$$= \Bigg(G_1(y_1(W_1){:}\,\Theta_1) + \sum_{j'=2}^{d_1} G_{1j'}\big(y_{j'}(W_1){:}\,\Theta_{1j'}\big), \dots$$

$$G_1(y_1(W_n){:}\,\Theta_1) + \sum_{j'=2}^{d_1} G_{1j'}\big(y_{j'}(W_n){:}\,\Theta_{1j'}\big), \dots,$$

$$G_m(y_m(W_1){:}\,\Theta_m) + \sum_{j'=1}^{d_m} G_{mj'}\big(y_{j'}(W_1){:}\,\Theta_{mj'}\big), \dots,$$

$$G_m(y_m(W_n){:}\,\Theta_m) + \sum_{j'=1, j'\neq m}^{d_m} G_{mj'}\big(y_{j'}(W_n){:}\,\Theta_{mj'}\big) \Bigg), \tag{9}$$

where the number of microbes that regulate microbe j has been reduced to d_j. Now, we strive to find a biologically relevant approach for fitting $G_j(\cdot)$ and $G_{jj'}(\cdot)$ and a statistically robust approach for modeling $\boldsymbol{\Sigma}$.

Our qdODEs (1) establishes a part-whole relationship by modeling how NI scales with HI across samples. According to the metabolic theory of ecology [48], we argue that this relationship may comply with allometric scaling law. To test our argument, we reanalyze Davenport et al.'s [32] data by plotting the NI of individual genera against HI across all samples. We found that the power equation can reasonably well fit the HI-dependent change of microbial abundance. For example, genus *Actinomyces* decreases its abundance exponentially with niche index (β = -2.43; P < 0.001) (Fig. S5**A**). The abundance of genus *Bacteroides* follows a reverse pattern of change, i.e., increasing exponentially with niche index (β = 3.12; P < 0.001) (Fig. S5**B**). We found the random distribution of residuals independent from the power equation fitting of these two genera across niche index (**Fig. S6**). In reanalyzing Sherrill-Mix et al.'s [38] data, a similar phenomenon was also observed (**Fig. S7**), with an additional finding that the abundance level of individual bacterial families against animal mass tends to follow a power curve (P < 0.05). Taken together, real-data analysis supports the hypothesis on the role of the power law in governing the part-whole relationship of individual microbes with the total microbes.

To the end, we implement power equation to fit $G_j(\cdot)$ by exponent parameters $\Theta_j = (\alpha_j, \beta_j)$ and an LOP-based nonparametric approach to fit $G_{jj'}(\cdot)$. We will model the covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_1 & \cdots & \mathbf{\Sigma}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{m1} & \cdots & \mathbf{\Sigma}_m \end{pmatrix}, \tag{10}$$

where $\mathbf{\Sigma}_j$ is the sample-dependent covariance matrix of microbe j, and $\mathbf{\Sigma}_{j_1 j_2}$ is the sample-dependent covariance matrix between microbes j_1 and j_2. $\mathbf{\Sigma}_{j_1 j_2}$ is structured as $\mathbf{\Sigma}_{j_1 j_2} = \sigma_{j_1 j_2} \mathbf{I}_n$, where $\sigma_{j_1 j_2}$ is the residual covariance of microbes j_1 and j_2 at the same sample. If there is a mix of static and temporal data involved, we may implement the AR(1) model to model autocorrelative structure of the covariance.

Under mean-covariance structure modeling by equations (9) and (10), model parameters $\boldsymbol{\phi} = (\boldsymbol{\mu}; \boldsymbol{\Sigma})$ become model parameters $\boldsymbol{\phi}' = [\Theta_j, \Theta_{jj'}(j = 1, \dots, m, j' = 1, \dots, j-1, j+1, \dots, d_j); \sigma_j^2, \sigma_{j_1 j_2}(j_1 \neq j_2 = 1, \dots, m)]$, whose optimal solution can be obtained, by maximizing the likelihood (8), as

$$\widehat{\boldsymbol{\phi}'} \in \left\{ \arg \max_{\boldsymbol{\phi}' \in \boldsymbol{\Phi}'} \mathcal{L}(\boldsymbol{\phi}') \right\}. \tag{11}$$

Intuitively, this maximization implies an optimal topological structure and organization by which microbes interact with each other to maximize the overall abundance of microbial community assemble as a whole. The convex optimization formulation under equation (11) ensures the stability and sparsity of the network reconstructed from qdODEs. Since no constraints are given on the number of outgoing links, the resulting network can be high-dimensional and omnidirectional.

If n samples include C contexts, we need to test whether if the data can be better modeled according to different contexts. In this case, we can still formulate a likelihood like equation (8), but the mean vector is now modelled as

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1; \dots; \boldsymbol{\mu}_m) = \big(\mu_1(W_1), \dots, \mu_1(W_n); \dots; \mu_m(W_1), \dots, \mu_m(W_n)\big)$$

$$= \left(\sum_{c=1}^{C} \left[G_1(y_1(W_1)\!:\!\Theta_1^c) + \sum_{j'=2}^{d_1} G_{1j'}\big(y_{j'}(W_1)\!:\!\Theta_{1j'}^c\big) \right], \dots \right.$$

$$\sum_{c=1}^{C} \left[G_1(y_1(W_n)\!:\!\Theta_1^c) + \sum_{j'=2}^{d_1} G_{1j'}\big(y_{j'}(W_n)\!:\!\Theta_{1j'}^c\big) \right], \dots,$$

$$\sum_{c=1}^{C} \left[G_m(y_m(W_1)\!:\!\Theta_m^c) + \sum_{j'=1}^{d_m} G_{mj'}\big(y_{j'}(W_1)\!:\!\Theta_{mj'}^c\big) \right], \dots,$$

$$\left. \sum_{c=1}^{C} \left[G_m(y_m(W_n)\!:\!\Theta_m^c) + \sum_{j'=1}^{d_m} G_{mj'}\big(y_{j'}(W_n)\!:\!\Theta_{mj'}^c\big) \right] \right), \tag{12}$$

and the covariance matrix modelled similarly as above.

By plugging in the MLEs of mean vectors (9) and (12) into likelihood (9), we obtain the likelihood values L_1 (assuming that there are no differences among C contexts) and L_2 (assuming that there are difference among C contexts), respectively. We further estimate the log-likelihood ratio,

$$\text{LR} = -2\log(L_0/L_1) \tag{13}$$

as a statistic used to test if n samples should be sorted into C contexts. By reshuffling n samples randomly into C groups, we calculate the LR value. If this permutation procedure is repeated 1000 times, we obtain the 95th percentile from 1000 LR values and use it as a critical threshold.

Quantifying the keystoneness of microbes

Since its proposal, the keystone species concept has been widely used to define entire ecosystems [31]. The qdODE model allows us to expand the definition of this concept. Indeed, every species, as long as it links to any another, may play a bridging role in network organization. Based on the maximum likelihood estimates of $G(\cdot)$ and $G_{jj'}(\cdot)$ under equation (11), we derive the parameter of keystoneness to assess the degree with which a microbe potentially behaves as the keystone. A similar strategy for keystoneness assessment was used by Berry and Widder [49], but our definition is quantitative and extendible to any cases. We define the keystoneness of a microbe as the sum of ratios of its dependent abundance to independent abundance over all microbes linked with it. Because of different behaviors of active and passive influences, outgoing links may contribute to keystoneness differently from incoming links. For a given microbe j, we define its outgoing keystoneness and incoming keystoneness, respectively, as

$$K_{ji}^{\text{out}} = \sum_{j'=1}^{D_j} \left| \frac{G_{j'j}\left(y_j(W_i : \Theta_{j'|j})\right)}{G_{j'}\left(y_{j'i}(W_i : \Theta_{j'})\right)} \right| \tag{14A}$$

$$K_{ji}^{\text{in}} = \sum_{j'=1}^{d_j} \left| \frac{G_{j|j'}\left(y_{j'}(W_i : \Theta_{j|j'})\right)}{G_j\left(y_j(W_i : \Theta_j)\right)} \right| \tag{14B}$$

where D_j and d_j are the numbers of microbe j's outgoing links and incoming links, respectively. If the outgoing keystoneness of a microbe is larger than its incoming keystoneness, this microbe is regarded as a socially aggressive microbe. If a microbe has a smaller value of keystoneness, including outgoing and incoming, than the average of all microbes, it is regarded as a solitary microbe. As can be seen, keystoneness provide a quantitative classification of all microbes co-existing within the gut.

Computer simulation

The statistical properties of our qdODE network inference are empirically investigated through computer simulation studies. We simulated the abundance data of 100 microbes, $y_j(W_i)$ ($j = 1, \ldots, 100$), for n samples, which vary with HI, W_i ($i = 1, \ldots, n$), based on equation (5). The residual error of microbe j at sample i is assumed to obey a multivariate normal distribution with mean vector $\mathbf{0}$ and structured covariance matrix. We design different scenarios by changing sample size and variance, (1) $n = 50$, $\sigma_j^2 = 1.5$, (2) $n = 50$, $\sigma_j^2 = 0.7$, (3) $n = 250$, $\sigma_j^2 = 1.5$, and (4) $n = 250$, $\sigma_j^2 = 0.7$. We analyze the data simulated under each of these scenarios. By comparing the estimated and true HI-varying abundance curves, we assess the estimation precision of our new model.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: Fecal microbial abundance data
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0140301#ack
The computer code is available at https://github.com/JiangLibo/microbialnetworks or may be requested from the corresponding author.

Authors' contributions: LJ performed data analysis and wrote the computer code. GC developed evolutionary game theory. RW formulated the model, supervised the project, and wrote the manuscript.

References

1.   Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320: 1034–1039.

2.   Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A et al. (2015) Structure and function of the global ocean microbiome. Science 348: 1261359.

3.   Edwards JA, Santos-Medellín CM, Liechty ZS, Nguyen B, Lurie E, Eason S, et al. (2018) Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. PLoS Biol 16(2): e2003862.

4.   Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486: 207–214.

5.   Goldford JE, Lu NX, Bajic D, Estrela S., Tikhonov M, Sanchez-Gorostiaga, Segre D, Mehta P, Sanchez A (2018) Emergent simplicity in microbial community assembly. Science 361: 469-474.

6.   Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA (2012) The application of ecological theory toward an understanding of the human microbiome. Science 336: 1255–1262.

7.   Coyte KZ, Schluter J, Foster KR (2015) The ecology of the microbiome: networks, competition, and stability. Science 350: 663-666.

8.   Röttjers L, Faust K (2018) From hairballs to hypotheses–biological insights from microbial networks. FEMS Microb Rev 42: 761-780.

9.   Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. Nat Rev Microbiol 10: 538-550.

10. Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG et al. (2013) Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol 9: e1003388.

11. Fisher CK, Mehta P (2014) Identifying keystone species in the human gut microbiome from metagenomics time series using sparse linear regression. PLoS ONE 9(7): e102451.

12. Oliverio AM, Bradford MA, Fierer N (2017) Identifying the microbial taxa that consistently respond to soil warming across time and space. Glob Chang Biol 23: 2117-2129.

13. Rath KM, Maheshwari A, Rousk J (2019) Linking microbial community structure to trait distributions and functions using salinity as an environmental filter. mBio 10(4): e01607-19.

14. Duran-Pinedo AE, Frias-Lopez J (2015) Beyond microbial community composition: functional activities of the oral microbiome in health and disease. Microbes Infect 17: 505-516.

15. McGeachie MJ, Sordillo JE, Gibson T, Weinstock GM, Liu YY, Gold DR, Weiss ST, Litonjua A (2016) Longitudinal prediction of the infant gut microbiome with dynamic Bayesian networks. Sci Rep 6: 20359.

16. Xiao Y, Angulo MT, Friedman J, Waldor MK, Weiss ST, Liu YY (2017) Mapping the ecological networks of microbial communities. Nat Commun 8(1): 2042.

17. Chen C, Jiang L, Fu G et al. An omnidirectional visualization model of personalized gene regulatory networks. NPJ Syst Biol 2019; 5: 38.

18. Hutchinson GE (1957) Concluding remarks. Population studies: animal ecology and demography. Cold Spr Harb Sym Quant Biol 22: 415–427.

19. Smith JM, Price GR (1973) The logic of animal conflict. Nature 246: 15-18.

20. Godbold JA, Bulling MT, Solan M (2011) Habitat structure mediates biodiversity effects on ecosystem properties. Proc Roy Soc B Biol Sci 278: 2510-2518.

21. Polechová J, Storch D (2019) Ecological Niche. Encyclopedia of Ecology (Second Edition), edited by Fath BD. Elsevier, Amsterdam, Netherlands.

22. Peterson AT, Soberôn J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MP (2011) Species-environment relationships. In: Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press.

23. Spigler RB, Woodard AJ (2019) Context-dependency of resource allocation trade-offs highlights constraints to the evolution of floral longevity in a monocarpic herb. New Phytol 221: 2298-2307.

24. Loison A, Strand O (2005) Allometry and variability of resource allocation to reproduction in a wild reindeer population. Behav Ecol 16: 624–633.

25. Pereira FC, Berry D (2017) Microbial nutrient niches in the gut. Environ Microbiol 19: 1366-1378.

26. Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M et al. (2018) Enterotypes in the landscape of gut microbial community composition. Nat Microbiol 3(1): 8-16.

27. Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant breeding program. Aust J Agr Res 14: 742-754.

28. Lobell DB, Roberts MJ, Schlenker W, Braum N, Little BB, Rejesus RM, Hammer GL (2014) Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. Science 344: 516-519.

29. von Neumann J, Morgenstern S (1944) Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ.

30. Hofbauer, J. and Sigmund, K. (1998) Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge, UK.

31. Banerjee S, Schlaeppi K, van der Heijden MGA (2018) Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol 16: 567-576.

32. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y (2015) Genome-wide association studies of the human gut microbiota. PLoS ONE 10: e0140301.

33. Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev 32: 723–735.

34. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S (2017) The evolution of the host microbiome as an ecosystem on a leash. Nature 548: 43-51.

35. Rosenberg E, Zilber-Rosenberg I (2018) The hologenome concept of evolution after 10 years. Microbiome 25: 78.

36. Rosenberg E, Zilber-Rosenberg I (2016) Microbes drive evolution of animals and plants: the hologenome concept. mBio 7: e01395.

37. Hassani MA, Durán P, Hacquard S (2018) Microbial interactions within the plant holobiont. Microbiome 6: 58.

38. Sherrill-Mix S, McCormick K, Lauder A, Bailey A, Zimmerman L, Li Y, Django JN, Bertolani P, Colin C, Hart JA et al. (2018) Allometry and Ecology of the Bilaterian Gut Microbiome. mBio 9: e00319-18.

39. Rosenzweig ML (1995) Species Diversity in Space and Time. Cambridge University Press, Cambridge.

40. Würtz P, Annila A (2008) Roots of diversity relations. J Biophys 2008: 654672.

41   Parker GA, Smith MJ (1990) Optimality theory in evolutionary biology. Nature 348: 27-33.

42. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL et al. (2013) The long-term stability of the human gut microbiota. Science 341: 1237439.

43. Kong W, Meldgin DR, Collins JJ, Lu T (2018) Designing microbial consortia with defined social interactions. Nat Chem Biol 14(8): 821-829.

44. Sinervo B, Lively CM (1996) The rock–paper–scissors game and the evolution of alternative male strategies. Nature 380: 240-243.

45. Bucci V, Tzen B, Li N, Simmons M, Tanoue T, Bogart E, Deng L, Yeliseyev V, Delaney ML, Liu Q et al. (2016). MDSINE: microbial dynamical systems inference engine for microbiome time-series analyses. Genome Biol 17: 121.

46. Zou H. The adaptive Lasso and its oracle properties. J Am Stat Assoc 2006;101:1418-1429.

47. Wang H, Leng C (2008) A note on adaptive group LASSO. Comput Stat Data Anal 52: 5277-5286.

48. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Toward a metabolic theory of ecology. Ecology 85:1771–1789.

49. Berry D, Widder S (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol 5: 219.

Figure Legends

Figure 1. Network analysis of 101 genera within the gut microbiota of Hutterites samples measured in winter (W) and summer (S). (**A**) Voronoi treemaps that visualize microbial networks at the genus level. Each polygon area (node) is represented by a genus (with its name and ID shown), with the color metric being proportional to the relative endogenous abundance level of this genus. Promotion and inhibition are denoted by arrowed red and blue lines, respectively, with the thickness of lines being proportional to the strength of microbial interactions. (**B**) The frequency distributions of microbial interactions and keystoneness across 101 genera. The numbers of all interactions, commensalism, and amensalism are each composed of outgoing links (red) and incoming links (blue). Keystoneness, including outgoing keystoneness (red) and incoming keystoneness (blue), is shown for the winter network. To show its season-dependent change, keystoneness is displayed as the difference of its summer value from its winter value (W- S), positive in purple and negative in green.

Figure 2. Inter-host variation in the keystoneness of each genus within gut-microbial networks of Hutterites samples measured in winter (W) and summer (S).

Figure 3. Phylogenetic change of microbial networks at the family level among three representative species. Networks, visualized as Voronoi treemaps, involve 65 bacterial families, in which each polygon area (node) is represented by a family (with its ID shown), with the color metric being proportional to the relative independent abundance level of this family. Promotion and inhibition are denoted by arrowed red and blue lines, respectively, with the thickness of lines being proportional to the strength of microbial interactions. The number of links (including outgoing and incoming) for each family is plotted below the networks. Correlations in network edges among animal species are given at the lower right side.

Figure 4. Decomposition of the overall abundance level (thin green line) into its endogenous component (thick blue line) and exogenous component (red line) for genera *Actinomyces* (7) (**A**) and *Bacteroides* (20) (**B**) expressed within the microbiota gut of the Hutterites samples, including 124 hosts for winter and 103 hosts for summer. IDs of the genera that produce outgoing links are 36 for *Dorea*, 21 for *Barnesiella*, 38 for *Enhydrobacter*, 86 for *Solobacterium*, 91 for *Subdoligranulum*, 68 for *Paenibacillus*, 55 for *Lactonifactor*, 3 for

Acetivibrio, 73 for *Peptococcus*, 50 for *Holdemania*, 66 for *Oribacterium*, 39 for *Enterococcus*, 93 for *Thermus*, 89 for *Staphylococcus*, and 55 for *Lactonifactor*. Dots denote the abundance values of genera *Actinomyces* (**A**) and *Bacteroides* (**B**) across individual hosts in winter (blue) and summer (red).
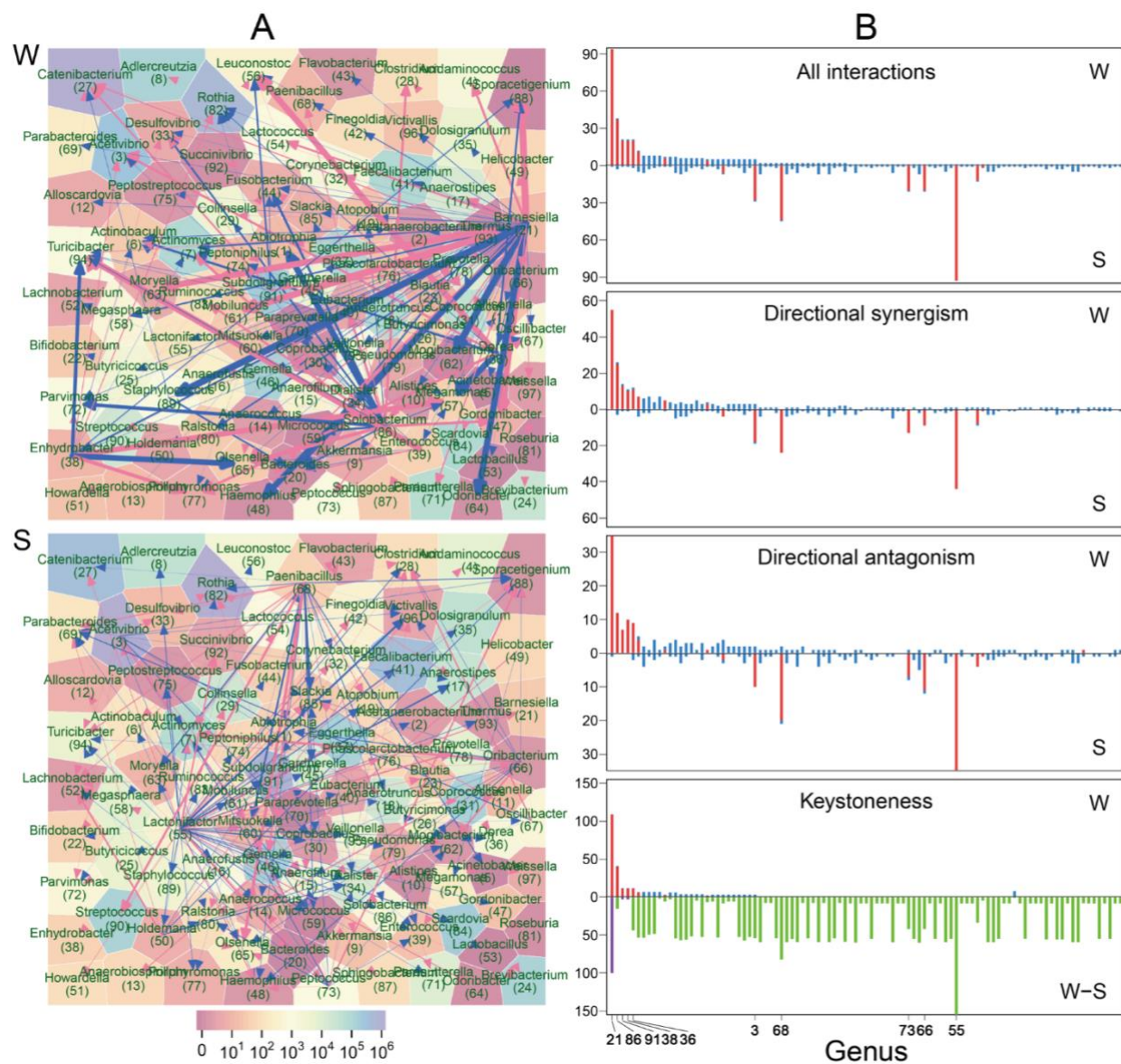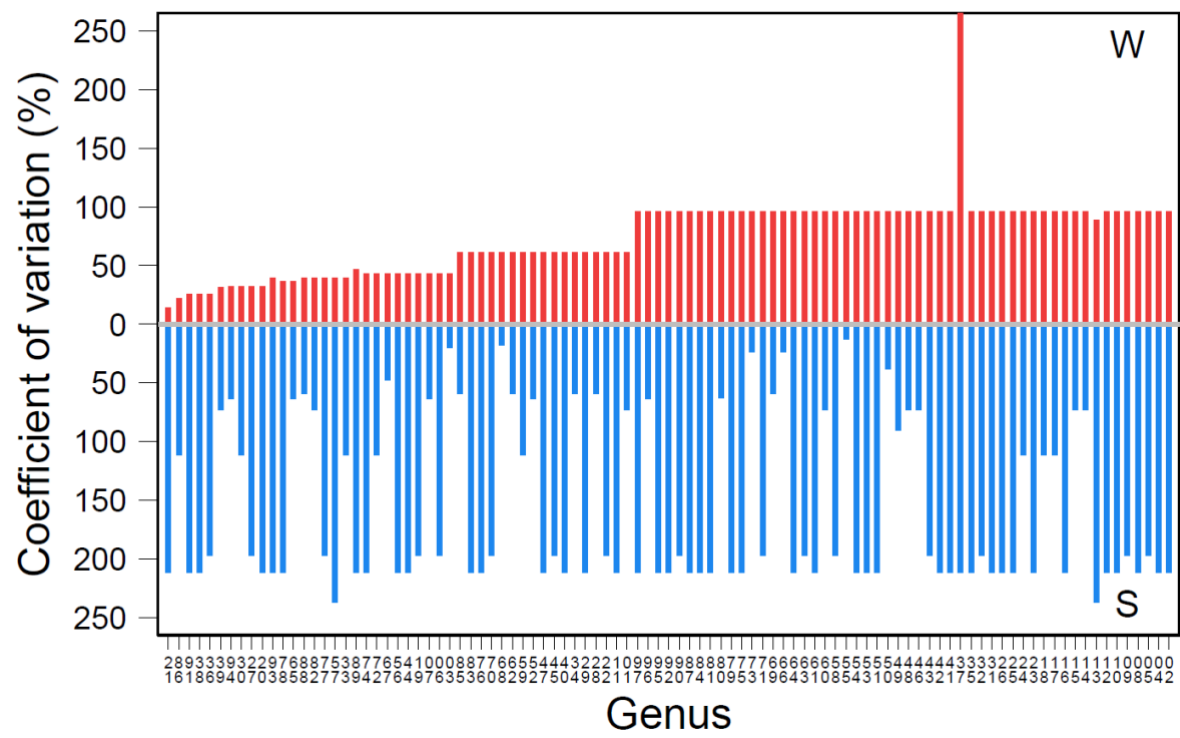
Figure 1

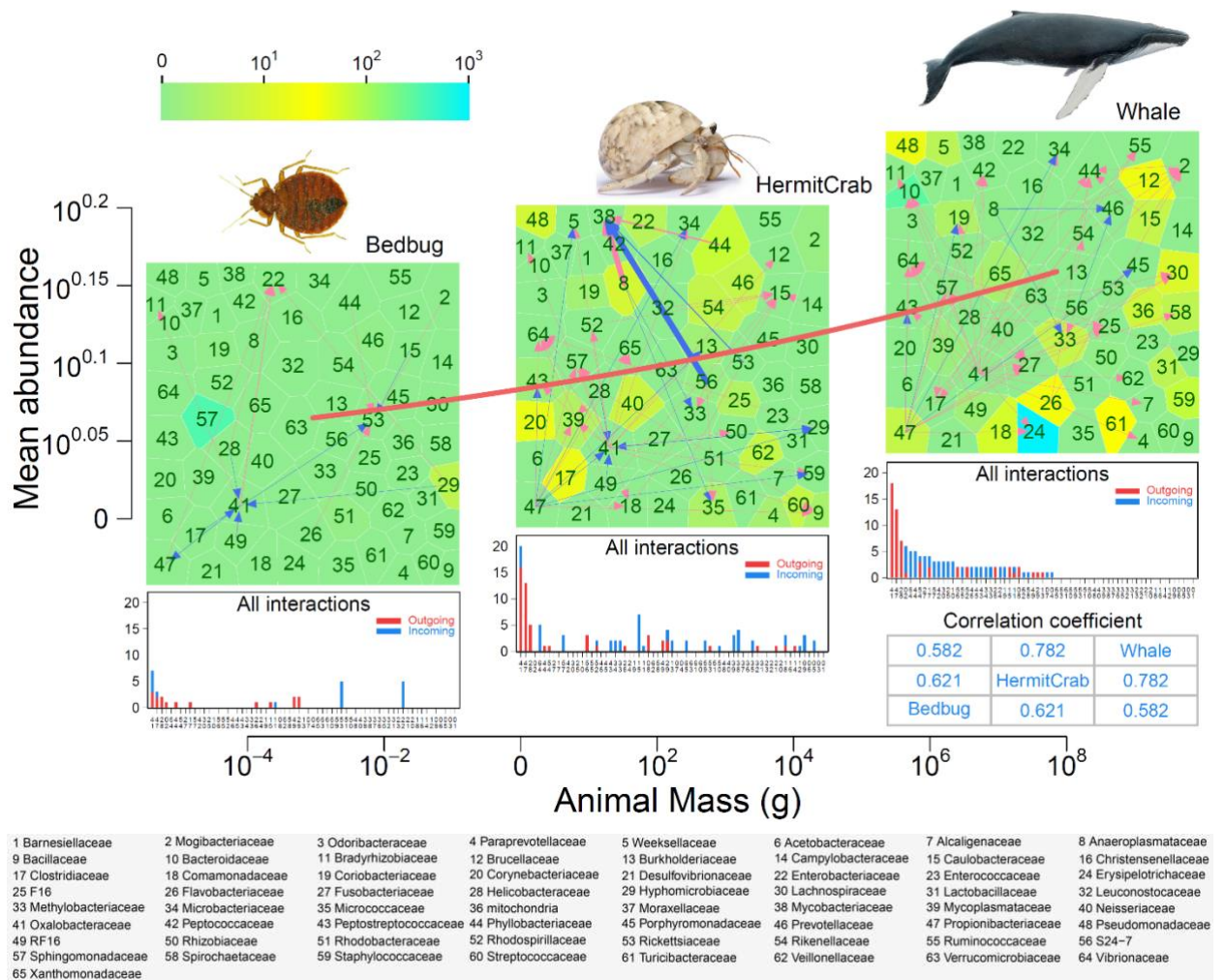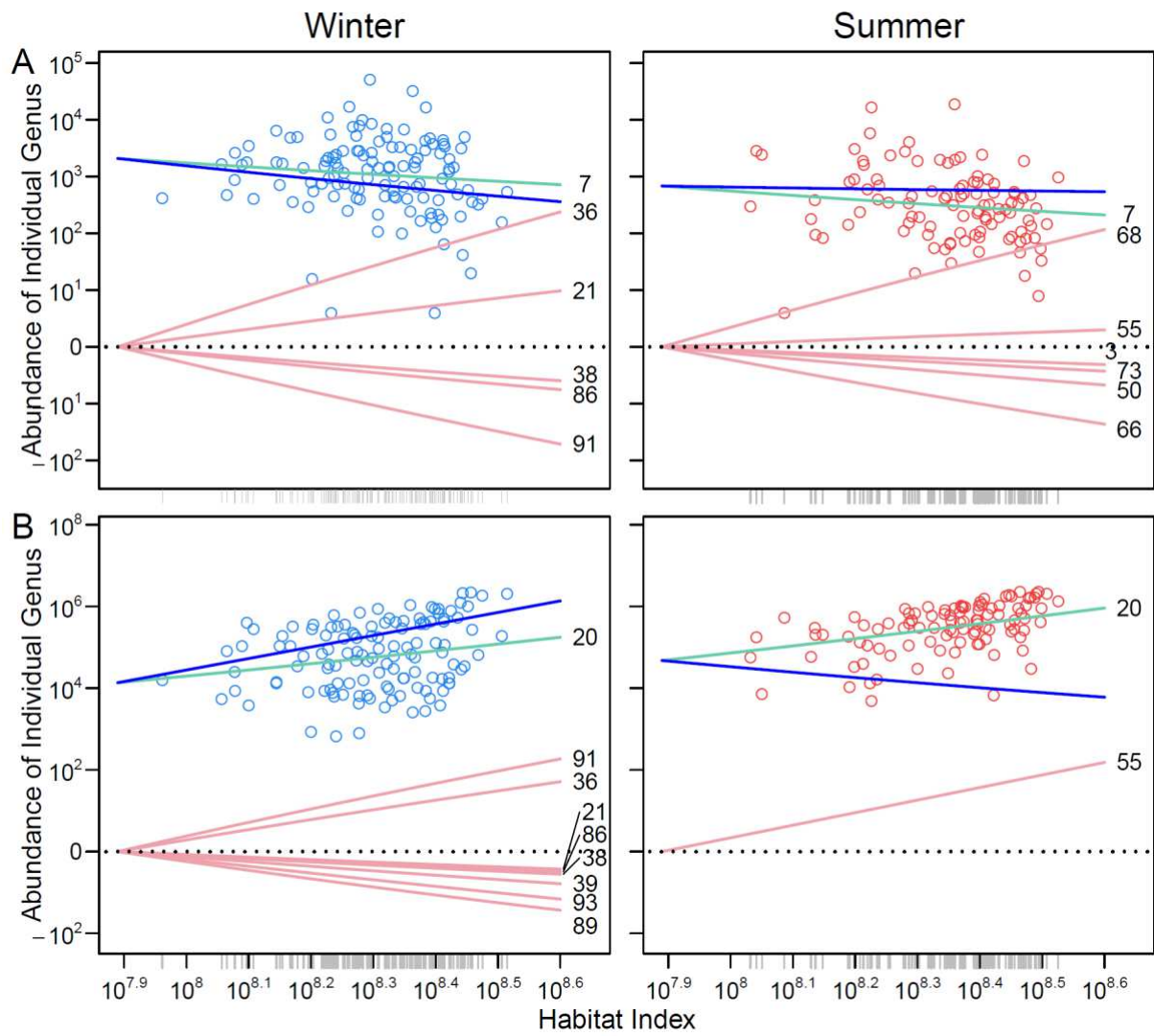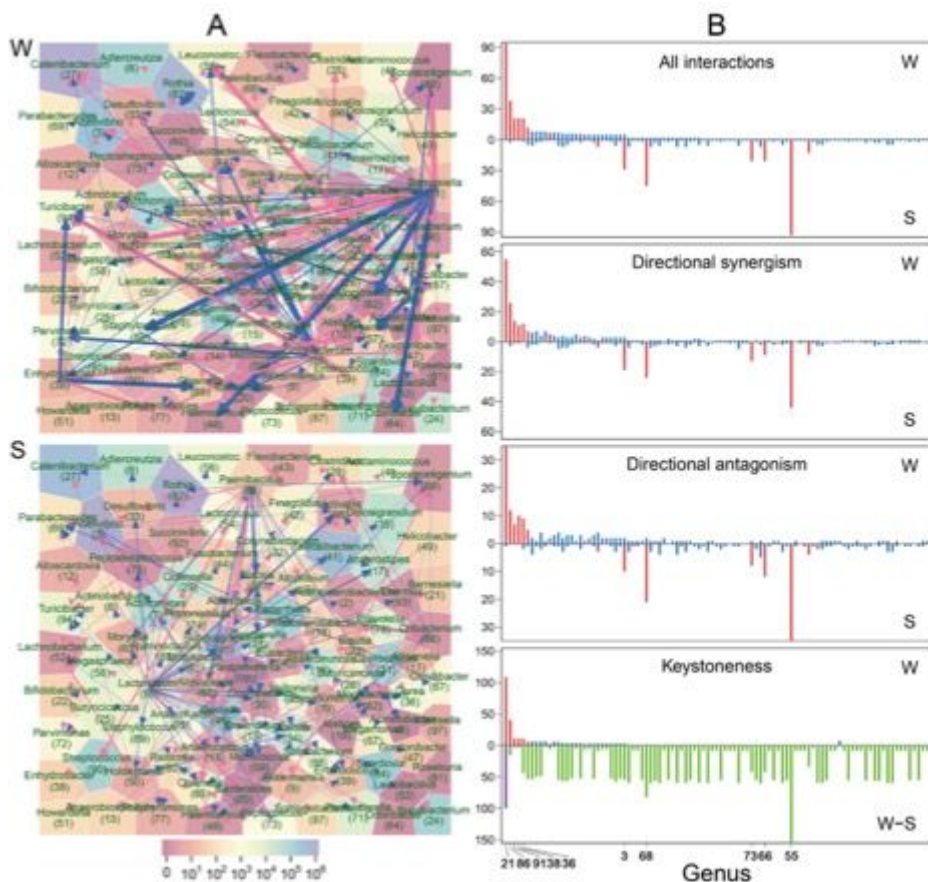**Figure 2**

Figure 3

Figure 4

Figures

Figure 1

Network analysis of 101 genera within the gut microbiota of Hutterites samples measured in winter (W) and summer (S). (A) Voronoi treemaps that visualize microbial networks at the genus level. Each polygon area (node) is represented by a genus (with its name and ID shown), with the color metric being proportional to the relative endogenous abundance level of this genus. Promotion and inhibition are denoted by arrowed red and blue lines, respectively, with the thickness of lines being proportional to the strength of microbial interactions. (B) The frequency distributions of microbial interactions and keystoneness across 101 genera. The numbers of all interactions, commensalism, and amensalism are each composed of outgoing links (red) and incoming links (blue). Keystoneness, including outgoing keystoneness (red) and incoming keystoneness (blue), is shown for the winter network. To show its season-dependent change, keystoneness is displayed as the difference of its summer value from its winter value (W- S), positive in purple and negative in green.
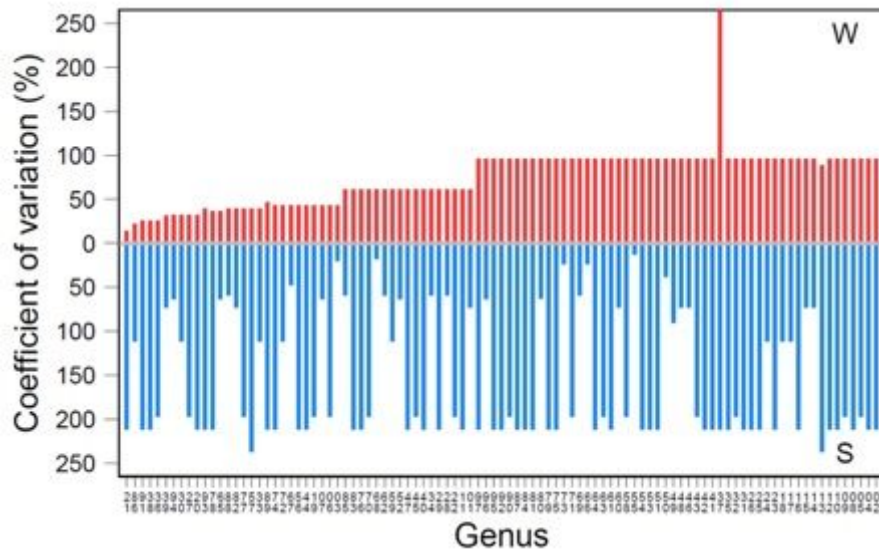
Figure 2

Inter-host variation in the keystoneness of each genus within gut-microbial networks of Hutterites samples measured in winter (W) and summer (S).

Figure 3

Phylogenetic change of microbial networks at the family level among three repres¬¬entative species. Networks, visualized as Voronoi treemaps, involve 65 bacterial families, in which each polygon area (node) is represented by a family (with its ID shown), with the color metric being proportional to the relative independent abundance level of this family. Promotion and inhibition are denoted by arrowed red and blue lines, respectively, with the thickness of lines being proportional to the strength of microbial interactions. The number of links (including outgoing and incoming) for each family is plotted below the networks. Correlations in network edges among animal species are given at the lower right side.
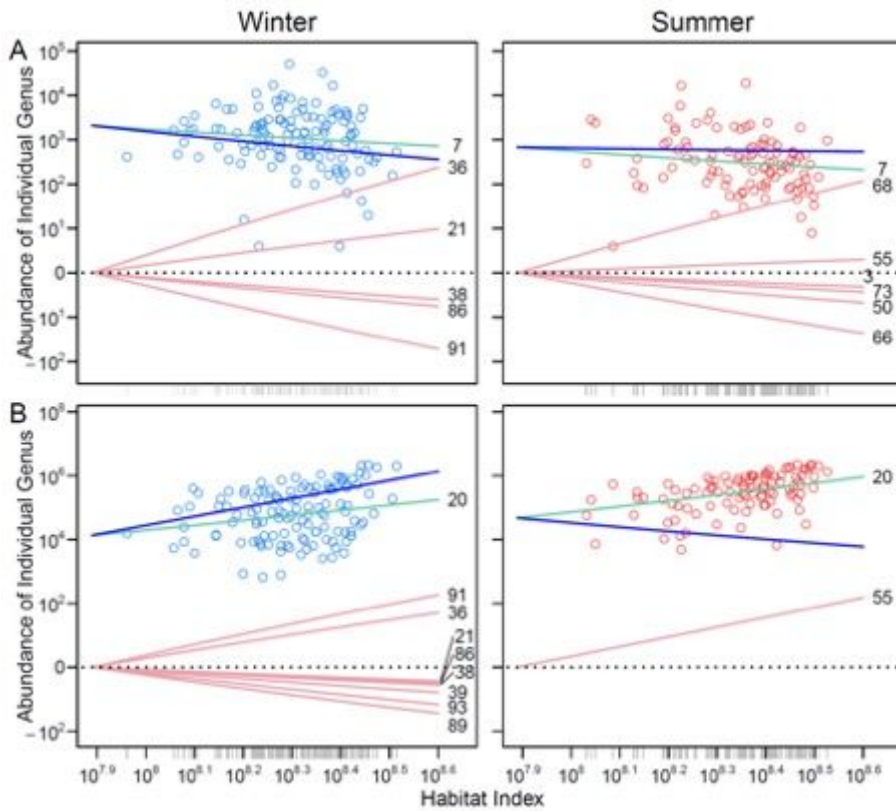
Figure 4

Decomposition of the overall abundance level (thin green line) into its endogenous component (thick blue line) and exogenous component (red line) for genera Actinomyces (7) (A) and Bacteroides (20) (B) expressed within the microbiota gut of the Hutterites samples, including 124 hosts for winter and 103 hosts for summer. IDs of the genera that produce outgoing links are 36 for Dorea, 21 for Barnesiella, 38 for Enhydrobacter, 86 for Solobacterium, 91 for Subdoligranulum, 68 for Paenibacillus, 55 for Lactonifactor, 3 for Acetivibrio, 73 for Peptococcus, 50 for Holdemania, 66 for Oribacterium, 39 for Enterococcus, 93 for Thermus, 89 for Staphylococcus, and 55 for Lactonifactor. Dots denote the abundance values of genera Actinomyces (A) and Bacteroides (B) across individual hosts in winter (blue) and summer (red).