Ecological Monographs, 89(2), 2019, e01358
© 2019 by the Ecological Society of America

Dominance network analysis provides a new framework for studying the diversity–stability relationship

Zhanshan (Sam) Ma[1,2,4] and Aaron M. Ellison[3]

[1]*Computational Biology and Medical Ecology Lab, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223 China*
[2]*Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223 China*
[3]*Harvard University, Harvard Forest, 324 North Main Street, Petersham, Massachusetts 01366 USA*

Citation: Ma, Z. S., and A. M. Ellison. 2019. Dominance network analysis provides a new framework for studying the diversity–stability relationship. Ecological Monographs 89(2): e01358. 10.1002/ecm.1358

Abstract. The diversity–stability relationship is a long-standing, central focus of community ecology. Two major challenges have impeded studies of the diversity–stability relationship (DSR): the difficulty in obtaining high-quality longitudinal data sets; and the lack of a general theoretical framework that can encompass the enormous complexity inherent in "diversity," "stability," and their many interactions. Metagenomic "Big Data" now provide high quality longitudinal data sets, and the human microbiome project (HMP) offers an unprecedented opportunity to reinvigorate investigations of DSRs. We introduce a new framework for exploring DSRs that has three parts: (1) a cross-scale measure of dominance with a simple mathematical form that can be applied simultaneously to individual species and entire communities and can be used to construct species dominance networks (SDNs); (2) analysis of SDNs based on special trio motifs, core-periphery, rich-club, and nested structures, and high salience skeletons; and (3) a synthesis of coarse-scale core/periphery/community-level stability modeling with fine-scale analysis of SDNs that further reveals the stability properties of the community structures. We apply this new approach to data from the human vaginal microbiome of the HMP, simultaneously illustrating its utility in developing and testing theories of diversity and stability while providing new insights into the underlying ecology and etiology of a human microbiome-associated disease.

Key words: community dominance; core-periphery network; diversity–stability relationship; human microbiome; mean crowding; nestedness; skeleton network; species dominance network; trio motifs.

Introduction

The relationship between diversity and stability (henceforth, the "diversity–stability relationship" or DSR; see Box 1 for acronyms and variables) has been a central focus of community ecology for more than one-half century (see, e.g., reviews by Pimm 1984, McCann 2000, Green et al. 2006, Donohue et al. 2016). Although the development of DSR theory and most data collected to test it have relied on the relatively large (>1 mm long) macroorganisms that we can see easily, the DSR is now receiving increasing attention from microbial community ecologists as DNA sequencing and metagenomic technology have revolutionized our understanding of microbial diversity (e.g., Gibbons and Gilbert 2015). As metagenomic technology has presented unprecedented opportunities for studying microbial communities, it also has raised new challenges. For example, millions of sequences of bacterial 16s rRNA can be collected rapidly and may be useful for resolving long-standing issues in theoretical ecology, but visualizing, let alone understanding, these high-dimensional "Big Data," requires new analytical approaches. In this article, we introduce and develop a new framework, dominance network analysis, to explore the DSR, and illustrate its application with a large microbial data set from the human vaginal microbiome (henceforth, the human vaginal microbial community or HVMC). Dominance network analysis includes three parts (Fig. 1): (1) a new cross-scale measure (or metric) of dominance (Ma and Ellison 2017) derived from Lloyd's mean-crowding index (Lloyd 1967, 1986) that is applicable not only to populations and individual species but also to entire communities; (2) the use of this dominance metric to construct species dominance networks (SDNs) and conduct dominance network analysis; and (3) topological and stability analyses of SDNs, including the synthesis of parts 1 and 2.

The motivation for using one part of the human microbiome as our illustrative application rather than a classical "macrobial" system is two-fold. First, the Human Microbiome Project (HMP) has revealed that

Manuscript received 10 January 2018; revised 14 August 2018; accepted 10 September 2018. Corresponding Editor: Steven D. Allison.
[4] E-mail: ma@vandals.uidaho.edu

Box 1. Acronyms and symbols

BV: bacterial vaginosis
CNS: core-nested-skeleton structures
CST: community state types
CNS: core, nested, skeleton structure
DDS: dominance-dependent stability
DIDS: dominance-inversely-dependent stability
DIS: dominance-independent stability
DSR: diversity–stability relationship
HMP: Human Microbiome Project
HSS: high salience skeleton
HVM: human vaginal microbiome
HVMC: human vaginal microbial community
MAO: most abundant OTU
MDO: most dominant OTU
OUT: operational taxonomic unit
SDN: species dominance network
SCN: species correlation network
SIN: species interaction network
P:N ratio: ratio of positive to negative links
PM: partial mutualism
TM: total mutualism
L-Q: linear-quadratic models
Q-Q: quadratic-quadratic models)
D_c: community dominance metric (index)
D_s: species dominance metric (index)
D_{sd}: species dominance distance metric (index)
m_c: mean population abundance (size) per species
m_c^*: community mean crowding
$S_c(t)$: community stability
$S_s(t)$: population(species) stability
ρ: strength of core
C:P: core : periphery ratio)\
$\overline{\rho(k)}$: rich club coefficient
r_{HSS}: assortativity

the number of microbes distributed within (gut, reproductive tract, oral, airway) or on (skin) our body is vast, at least an order of magnitude larger than the number of our own somatic cells, and these microbes express at least two orders of magnitude more genes than we do (e.g., Turnbaugh et al. 2007, Clemente et al. 2012, HMP Consortium 2012a, b, Lloyd-Price et al. 2016). Although the long coevolution between macrobes and microbes has forged pervasive, strong, and generally beneficial interconnections (e.g., digestive efficiencies of the host and a steady nutrient supply for microbes; for additional examples, see Bittleston et al. 2016), changes in these relationships often happen, leading to well-known pathological ("unstable") conditions (e.g., Hooper et al. 2012, Estrela et al. 2015). Indeed, ecological theory has been both a unifying driving force and tested for the metagenomic revolution (e.g., Costello et al. 2012, Lozupone et al. 2012, HMP Consortium 2012a, b, Faust et al. 2012, Barberán et al. 2014, Ding and Schloss 2014, Ma et al. 2012a, 2015, 2016a, b, Ma 2015, Coyte et al. 2016). Today, ecologists can test major ecological theories across not only taxa (plants, animals, and microbes) but also ecosystem types (e.g., forest, lakes, ocean, human, and animal microbiomes), and novel findings and insights are revealed more frequently than ever. For example, Sunagawa et al. (2015) found that the global ocean microbiome and human gut microbiome share a functional core of prokaryotic gene abundance with 73% similarity in the ocean and 63% in the gut, despite the huge physiochemical differences between the two ecosystems.

Second, clinical microbiologists (e.g., Sobel 1999) using culture-based technology have already applied basic ecological measures of species diversity and dominance to explore the etiology of a disease associated with the HVMC: bacterial vaginosis (henceforth, BV; Eschenbach et al. 1988). Nearly two decades on, however, metagenomics and additional assessment of the diversity of uncultivatable bacteria in the HVMC together have raised new and unanswered questions about bacterial diversity, stability, and the etiology of BV (Fredricks et al. 2005, Fredricks 2011, Ma et al. 2012a, Ravel et al. 2011, 2013, Srinivasan et al. 2010, White et al. 2011). If our framework for dominance network analysis can shed new light on BV and its etiology, it would suggest that it has predictive value in a field that long has been dominated by heuristic, phenomenological analysis (e.g., Donohue et al. 2016). Tackling the tough questions in BV etiology not only can illustrate the general utility of DSR in species-rich systems but also should be an ideal arena in which to assess whether our proposed framework and approach can be applied successfully in a system of real-world importance.

This article is the second in a three-part series in which we develop and test our novel framework for analyzing DSRs. In the first article (Ma and Ellison 2017), we presented and validated our new dominance concept (metrics), including phenomenological modeling of the community-level DSR. The focus of this article is to illustrate the analysis and synthesis of SDNs with the data sets from a longitudinal HMP metagenomic study of 32 healthy women (Gajer et al. 2012). The third paper (Ma and Ellison unpublished data) extends these analyses to additional HMP data sets from BV patients to hypothesize mechanisms underlying the ecological "causes" of BV associated with DSRs.

Dominance Network Analysis

General framework

Dominance network analysis includes phenomenological modeling of community dominance dynamics, species dominance network (SDN) analysis, and their joint synthesis (Fig. 1). Common to all three parts of this analysis is a set of three measures of dominance: D_c, D_s,

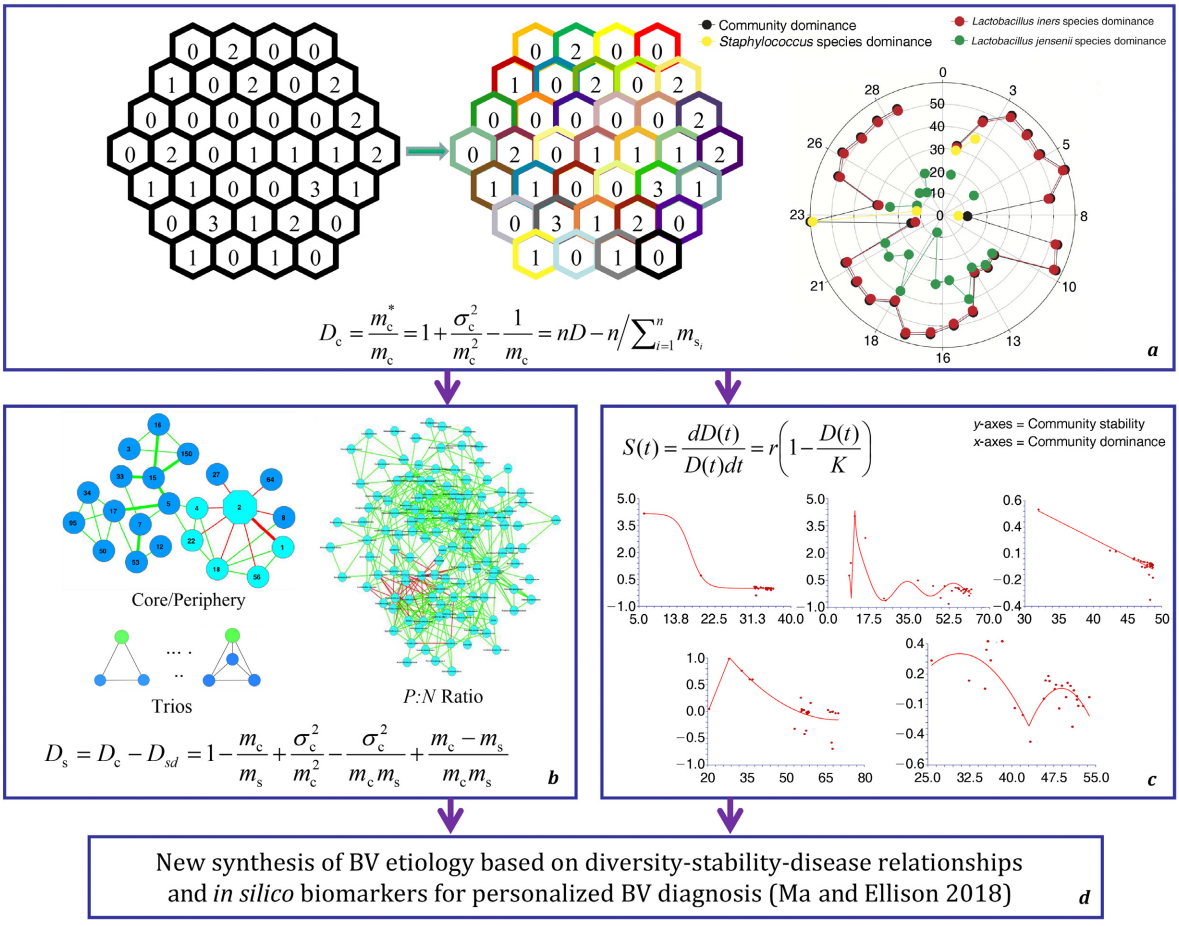

FIG. 1. A diagram showing the dominance-metric-based, species-dominance-network (SDN)-centered, framework for investigating the classic diversity–stability relationship (DSR): (a) The evolution from Lloyd's (1967) population mean crowding concept to a dominance concept applicable at both the community and species (population) levels, as indicated by the arrow. On the left are the individuals of the same species distributed over various sampling units, and on the right are the individuals of different species, shown in different colors. The formula refers to the *community dominance metric*, which is a linear function of the familiar Simpson's diversity index (D). The right-most polar coordinate graph shows the interactions of three species in a community, and their interactions seemed to control the dominance of the community, which inspired the methodology for detecting the special trios in SDN (see Ma and Ellison [2017] for additional details). (b) The main SDN methods for supporting our framework include special trio motifs, *P*:*N* ratio (positive to negative), core/periphery/nested networks, and skeleton networks. All of the networks were built with species dominance index (D_s). This block shows the focus of this article. (c) Phenomenological modeling of the DSR based on the dominance matrix, which can be performed for the population, community, or community guild (such as core and periphery identified with SDN analysis). (d) The third part of three-part series study focusing on diversity–stability–disease relationships and the extension of bacterial vaginosis (BV) etiology.

and D_{sd} (for community dominance, species dominance, and species dominance distance, respectively), which are based on Lloyd's mean crowding index for population aggregation (Lloyd 1967, 1986). We focus our attention on dominance as a proxy for diversity because it has intuitive meanings for both species and communities. For example, we often describe a community as high or low in *diversity* but rarely describe a species as high or low in *diversity* (except when referring to its genetic diversity). In contrast, we routinely describe one or more species as dominating a community, because individual species (e.g., keystone, dominant, or foundation species) may regulate community-wide diversity through bottom-up or top-down effects (e.g., Baiser et al. 2013). Since dominance (unevenness) and diversity (evenness) can be considered as both sides of the same coin, our new concept of dominance metric (Ma and Ellison 2017) provides a common currency applicable at both the species and community levels.

Specifically, our measure of community dominance, D_c, can be interpreted directly as community-level diversity along the lines of, for example, Simpson's D (Ma and Ellison 2017). At the same time, our measure of species dominance, D_s, identifies which species dominates the community. A unique ecological property of this set of measurements is that they have the same

mathematical form when applied to either individual species or multispecies communities (Ma and Ellison 2017). In brief, D_c is a linear function of the well-known Simpson's (1949) diversity index, whereas for individual species, D_s measures the difference between the community dominance and the dominance of a virtual community whose mean population size (per species) is equal to the population size of the focal species whose contribution D_s is being measured. Each measure is defined in more detail, below; technical details are given in Ma and Ellison (2017).

Roadmap

We first lay out key mathematical details and notation used throughout the paper. Less detail is provided for the measures of dominance and methods of phenomenological modeling, which are described fully by Ma and Ellison (2017). The characterization of species dominance networks is more comprehensive because it is presented here for the first time. Additional technical details on all elements of the mathematical methods are provided in Ma and Ellison (2017). Second, we use our three dominance measures to model phenomenologically the dominance (or diversity)–stability relationship. Specifically, we apply six models (linear, quadratic, reciprocal, logistic, liner-quadratic, and quadratic-quadratic) to examine dominance-dependent stability (DDS), dominance-inversely-dependent stability (DIDS), and dominance-independent stability (DIS) in the HVMC. These models help identify and determine the (in)stability of potential equilibria in the community. An additional useful feature that emerges from this modeling approach is that resilience is defined quantitatively as the derivative of stability (rather than qualitatively, as is done more commonly: see, e.g., Fig. 4a of Donohue et al. 2016). Third, we move beyond relatively simple definitions of diversity and stability based on species numbers to explore the DSR in a community defined as a network of interacting species. This approach is especially important for microbial communities. Data sets from longitudinal studies of microbial community dynamics typically are "high-dimensional." Microbial metagenomic data sets, for example, usually have hundreds, if not thousands, of operational taxonomic units (OTUs) sampled repeatedly through time or space. Network analysis is one of the most powerful approaches for dealing with complex, high-dimensional data because the basic network structure still holds and can be computed when the dimensions of the data far exceed the number of data points (Lau et al. 2017).

Our approach to analyzing SDNs differs from most previous analyses of ecological networks because it uses species dominance, not abundance, as a characteristic of individual species (network nodes). We quantify core-nested-skeleton (CNS) structures in a SDN because these structures play critical roles in stabilizing ecological networks. By integrating species-scale CNS network analysis with community-scale dominance dynamics, we reveal how underlying topological structures shape the DSR. We search for special node-connected trios (trio motifs) that include the most dominant OTU (MDO), which has the highest value of D_s in the network; the most abundant OTU (MAO) in the network; hubs (those nodes with the largest number of connections); and the sign (positive or negative) of interactions between OTUs. Through this analysis of trio motifs, we uncover network properties that existing standard network analysis cannot reveal, such as the existence of specific functional groups, including BV-associated bacteria in the HVMC. Last, we synthesize results from the phenomenological DSR models, the SDN analysis, and characteristics of the microbial host (human). This synthesis provides a "big picture" of community dynamics, especially the potential influences of environmental, habitat, and host-specific factors on diversity and stability (Ma and Ellison unpublished data).

Community dominance and species dominance derived from mean crowding

We start by defining community dominance, D_c, as:

$$D_c = \frac{m_c^*}{m_c} = 1 + \frac{\sigma_c^2}{m_c^2} - \frac{1}{m_c} \tag{1}$$

where m_C^* is community mean crowding:

$$m_c^* = m_c + \sigma_c^2/m_c - 1 \tag{2}$$

m_c is the mean of the population abundances (size) across all species (i.e., per species) in the community, and σ_c^2 is corresponding variance. Note that m_c^* also can be interpreted as a measurement of community unevenness, and there is a direct linear relationship between D_c and Simpson's diversity index (D; Ma and Ellison 2017). As a result, D_c is well-correlated with other measures of diversity, such as Shannon-Weiner's H', the Hill-number equivalents of H' and D (Chao et al. 2014), and the Berger-Parker dominance index (Berger and Parker 1970).

One advantage of D_c is that it can be used to define a dominance index D_s for each species in the community. To do this, we first define the species dominance distance D_{sd} as

$$D_{sd} = \frac{m_c^*}{m_s} = \frac{m_c}{m_s} + \frac{\sigma_c^2}{m_c m_s} - \frac{1}{m_s} \tag{3}$$

where m_s is the population abundance (size) of the focal species of interest (s) in the community. We then define species dominance (D_s) as the difference between D_c and D_{sd}:

$$D_s = D_c - D_{sd} = \frac{m_c^*}{m_c} - \frac{m_c^*}{m_s} = \frac{\sigma_c^2}{m_c^2} - \frac{\sigma_c^2}{m_c m_s} - \frac{1}{m_c} + \frac{1}{m_s}. \tag{4}$$

Like D_{sd}, D_s is species specific. The latter measures the dominance of a specific (focal) species in the community; more dominant species have larger values of D_s. In subsequent sections, we use D_c to model community-level DSRs, and use D_s to perform dominance network analyses such as searching for CNS and special trio motifs with significant DSR implications.

Phenomenological modeling of community stability

Despite its widespread use and perceived importance, there is no commonly accepted definition for community stability in the literature (Grimm and Wissel 1997, Ma 2012, Donohue et al. 2016). Here, we define community dominance stability (henceforth, community stability) as

$$S_c(t) = \frac{D_c(t+1) - D_c(t)}{D_c(t)}. \quad (5)$$

This definition of community stability measures the temporal change in community dominance. Similarly, we can define population dominance stability (henceforth, population stability) as:

$$S_s(t) = \frac{D_s(t+1) - D_s(t)}{D_s(t)}, \quad (6)$$

noting that there is a separate value of population stability for each species.

Next, we make the reasonable assumption that that temporal dynamics of dominance can be described by a differential equation:

$$S(t) = \frac{dD(t)}{D(t)dt} = f[D(t), Z] \quad (7)$$

where $D(t)$ is species- or community-level dominance at time t, Z is an optional vector of covariates, and $S(t)$ is the stability of the community (S_c) or the species (S_s) of interest. Since we do not know the form of the stability function $f[\cdot]$, our modeling strategy is phenomenological and data driven; for the latter, we use discrete difference equations. In a previous paper (Ma and Ellison 2017), we identified five models for $f[\cdot]$, linear (L), linear-quadratic (L-Q), quadratic-quadratic (Q-Q), logistic, and sine-logistic models, which could capture a broad spectrum of complex DSRs.

As does density-dependent theory for population regulation (*e.g.*, Cushing et al. 2003, Pastor 2008), our generalized stability model (Eqs. 5–7) may exhibit three types of dynamic behavior: dominance-dependent stability (DDS), in which (local) stability increases with dominance; dominance-inversely-dependent stability (DIDS), in which stability decreases with dominance; and dominance-independent stability (DIS), in which stability does not change with dominance. That is, if

$$S(t) \propto kD(t)dt \quad (8)$$

then $k < 0$, $k > 0$, and $k = 0$ correspond to DDS, DIDS, and DIS, respectively.

In practice, except for the simple linear model for which the slope $(b) \equiv k$ in Eq. 8, we may not be able to determine "the" value of k since it may be infeasible to describe a nonlinear relationship with a single parameter. However, the nonlinear models we use are simple enough that we can estimate piecewise (i.e., local) dominance–stability dependence relationships based on multiple parameter combinations. Ma and Ellison (2017) illustrated this estimation for basic DSRs.

Also in practice, necessary caution should be taken in adopting our phenomenological modeling approach. First, like many other approaches using time-series data, sufficient length of data points is a must for obtaining reliable analysis. Second, both the choice and interpretations of the stability models are both science and art. Furthermore, the expectation from the modeling analysis should be realistic, only identifying the three DSR mechanisms (DDS, DISS, DIS). It is generally not possible to use the approach for predictive purpose. As to the model choice, we recommend fitting multiple models to the same data sets, and choosing a most appropriate one based on model-fitting performance (r^2, standard error of coefficients), ecological realism, and parsimony, as demonstrated in Ma and Ellison (2017). That is, conditional upon statistically satisfactory model fitting judged by sufficiently high r^2 and simultaneously sufficiently low standard errors of model parameters (guarding against over-fitting), one should choose a model that can reliably and parsimoniously identify the above-mentioned three DSR mechanisms.

Species dominance networks (SDNs)

A variety of different methods grouped under the rubric of network analysis that emphasize species interactions and food webs can be used to analyze ecological systems (e.g., Junker and Schreiber 2008, Ings et al. 2009, Lau et al. 2017). We constructed SDNs for the HVMC using Spearman's rank correlation coefficient (ρ) computed between pairs of species-specific species dominance indices D_s (Eq. 4). Our resulting SDNs are similar to species interaction networks in macroecology (e.g., Ings et al. 2009) or cooccurrence networks in microbial ecology (e.g., Barberán et al. 2012), but we use D_s rather than population abundance as species weights. Other than the adoption of dominance metric, we can use the standard correlation network analysis to build SDNs.

We used the iGraph package (Csárdi 2006) in the R statistical software environment (R Core Team 2015) and *Cytoscape* software (Shannon et al. 2003) to do the standard correlation network analysis (iGraph *available online*).[5] The former was used to compute the Spearman rank correlation coefficients (r) between species

[5] http://igraph.org/r

dominance indexes (D_s). We choose the significance level of $P = 0.05$, and the correlation cutoff thresholds of $r \leq -0.5$ or $r \geq 0.6$ to build the species dominance networks (SDN). After filtering at the specified significance and correlation coefficient thresholds, the resulting correlation coefficient values were input into Cytoscape to produce SDN graphs.

We computed these SDNs for each of 32 individuals from a NIH-HMP longitudinal study of healthy women at reproductive ages (Gajer et al. 2012; henceforth, the "32-healthy cohort data set"). This study applied metagenomic sequencing technology to 16s-rRNA marker genes and generated a time series of vaginal bacterial OTU abundance data for each of the 32 subjects over a 16-week period. Data are publicly available as described by Gajer et al. (2012). Figs. 2 and 3 illustrate two such graphs corresponding to two selected subjects, and the graphs for all 32 subjects are displayed as Appendix S1: Fig. S1.

Characterizing species dominance networks (SDNs) with new approaches

Identification of special trio motifs.—By identifying what we call trio motifs in SDNs, we aim to resolve several deficiencies associated with standard correlation network analysis as applied to analysis of ecological systems. First, standard network analysis does not account for identities or characteristics of nodes other than their topological properties (e.g., Erdös and Renyi 1960, Watts and Strogatz 1998, Barabasi and Albert 1999; also see Bollobás 2001, Durrett 2006, Newman and Clauset 2016). For example, in a microbial network, aerobes and anaerobes have different roles that are not distinguishable from network topology alone. Second, although many ecological networks include both positive and negative relationships, network topology does not distinguish between these types of connections (edges) between nodes. Computation of the degree of a node requires knowing only the numbers of edges into or out of a node, not their sign. In real ecological networks, however, the number of negative and positive connections, and their magnitude, can make a large difference in network dynamics (e.g., Newman and Clauset 2016). Third, although many types of modular techniques can be applied to topological networks (e.g., Lecca and Re 2015), most of them are computationally expensive because the general module detection problem is "NP-hard:" the search time for finding all instances grows exponentially or even faster with the size of network (Fortunato 2010). This is especially problematic for large networks of OTUs derived from metagenomic data. Our method for identifying trio motifs overcomes these two ecologically important issues and also is computationally simple and efficient.

Our trio motifs are similar to triads identified in social network analysis (O'Malley and Marsden 2008, Kitts and Huang 2010). Whereas social-network triads are sub-graphs consisting of three nodes and possible edges between them, Ma and Ye (2017) identified 19 trio motifs (15 of which are of particular biomedical importance) and classified them hierarchically based on the special role of node, the interaction type (+ or –), and their combination.

Core-periphery, rich club, and nested structures in SDNs.—Three related concepts of network structures, core-periphery, rich-club, and nestedness, were applied to enhance our understanding of SDNs. The identification of structuring within networks was inspired by May's (1973) observation that increasingly complex systems should show decreasing stability, but that network stability could be achieved through modularity or other substructuring (e.g., Allesina and Tang 2012, 2015). Such substructuring is thought to increase network resilience in fluctuating or stochastic environments (e.g., Grilli et al. 2016). It has been found that networks with strong mutualistic interactions, such as the human microbiome network and pollination networks, are more robust if they have nested subnetworks (Scheffer et al. 2012). Hence, the three network structures we explore here should be particularly useful for exploring the stability (resilience) of SDNs.

We used the equations described below and coded in Python (version 2.7, www.python.org) to detect core-periphery, rich-club, and nested structures in the 32-healthy cohort data set. OTUs unconnected to any other OTUs (*i.e.*, nodes of degree 0) were deleted prior to analysis because they are irrelevant for these connectedness-related topological structures.

A core-periphery network consists of two groups of nodes. Nodes in the first group are connected tightly to one another and form a cohesive sub-graph as the core of the network. Nodes in the second group, i.e., the periphery, are connected more loosely to the core or its nodes and lack cohesion with the core (Rombach et al. 2014). A simple measure of how well the structure of a real network approximates the ideal or perfect core is characterized by the following definitions:

$$\phi = \sum_{i,j} a_{ij}\,\delta_{ij} \tag{9}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } c_i = \text{core or } c_j = \text{core} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

In the equations, a_{ij} indicates the presence or absence of a tie in two species (OTUs), and it is an element of the *adjacency matrix* (A) of the species (OTU) interaction (correlation) network; c_i refers to the class (core or periphery) to which node i is assigned, and δ_{ij} (subsequently called the pattern matrix) indicates the presence or absence of a tie in the idealized pattern (Borgatti and Everett 1999, Csermely et al. 2013). With a fixed distribution of values, the measure (ϕ) attains its maximum value if and only if **A** (the matrix of a_{ij}) and **D** (the

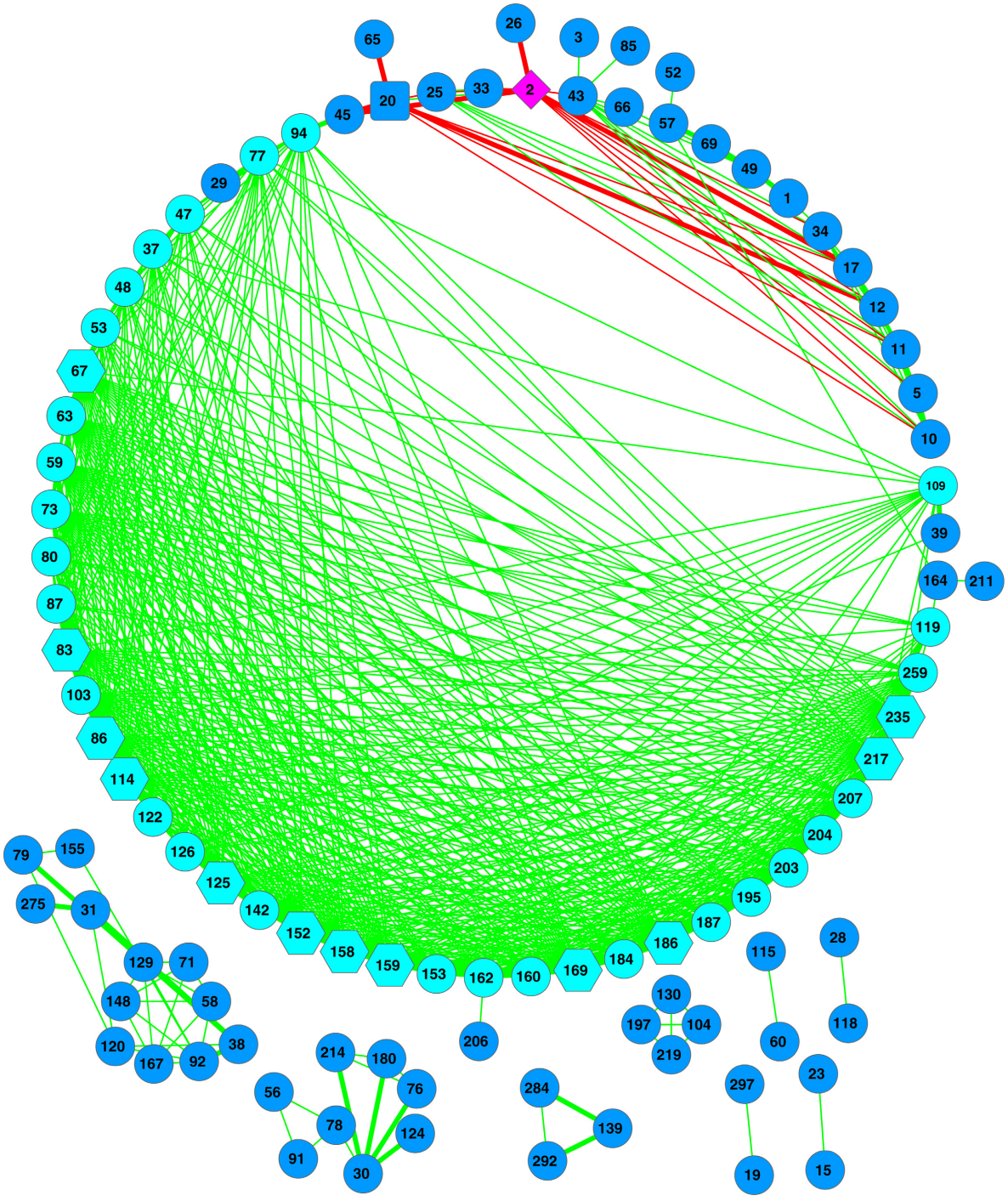

FIG. 2. The SDN (species dominance network) for subject number 424 (multiple hubs and separate hub, MDO [most dominant operational taxonomic unit (OTU)] and MAO [most abundant OTU], core nodes [dark blue], periphery nodes [light blue], and high salience skeletons [heavy line for links]), see Appendix S1: Table S1 for detailed legends of the network symbols. Note that the edges in this network are somewhat too dense to see the skeletons, and readers are referred to Fig. 3 or Appendix S1: Fig. S1 for better visualization of the skeletons.

matrix of δ_{ij}) are identical, which indicates that the observed and ideal interaction matrices are identical, and the network represented by matrix **A** has a perfect core/periphery structure.

We can also compute the core ratio CR of core nodes (OTUs) to periphery nodes in the network, and the linkage density $2L/n(n-1)$, where L is the number of links and n is the number of nodes (OTUs), for each of the

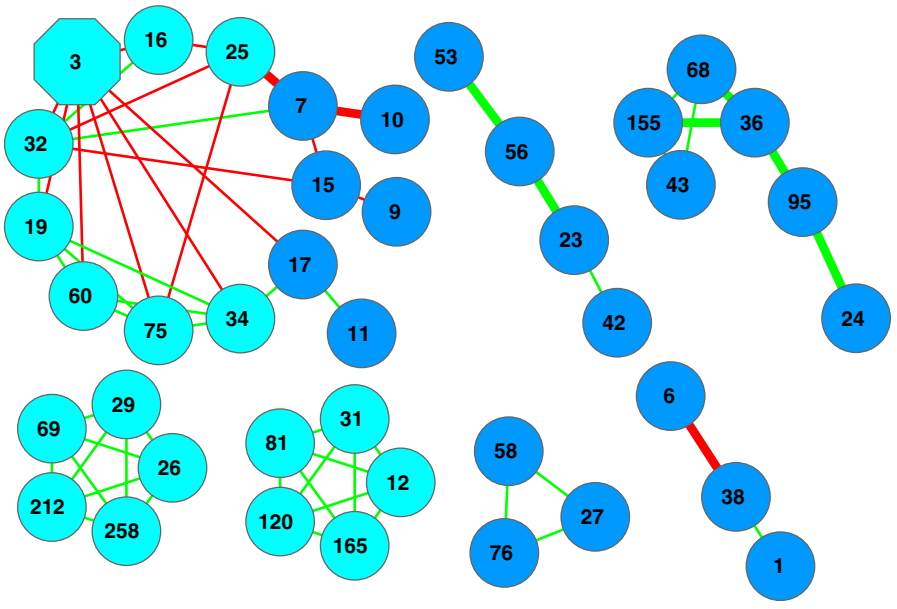

FIG. 3. The SDN (species dominance network) for subject number 408 (hub, MDO, and MAO are the same), core nodes (dark blue), periphery nodes (light blue), and high salience skeletons (heavy lines for links), see Appendix S1: Table S1 for detailed descriptions of the network symbols.

four quadrants (or blocks) in the core-periphery matrix, i.e., block 11 (core-core), block 12 (core-periphery), block 21 (periphery-core), and block 22 (periphery-periphery). Block density measures how densely connected are any two nodes in each block.

First characterized for the topology of the Internet, rich clubs in complex networks are a particular type of core-periphery network. Rich clubs are observed when hubs (i.e., high-degree nodes, which have many links to other nodes) are more densely connected with one another than are low-degree nodes (Zhou and Mondragon 2004, Julian et al. 2007). In general, a rich club is defined when nodes of degree $> k$ are more densely connected to one another than are nodes of degree $\leq k$ (Colizza et al. 2006). That is, richer nodes (with degree of a threshold of $>k$) tend to connect with each other in the rich club.

The topological rich-club coefficient $\phi(k)$, *i.e.* the proportion of edges connecting the rich nodes, with respect to all possible number of edges between them is

$$\phi(k) = \frac{E_{fk}}{\binom{N_{fk}}{2}} = \frac{2E_{fk}}{N_{fk}(N_{fk}-1)} \tag{11}$$

where N_{fk} refers to the nodes having a degree higher than k, and E_{fk} denotes the number of edges among the N_{fk} nodes in the rich-club. If $\phi(k) = 0$, the nodes do not share any links; if $\phi(k) = 1$ the rich nodes form a fully connected sub-network, a clique. Plotting $\phi(k)$ as a function of k can provide insights into network topology and functioning (Colizza et al. 2006). Colizza et al. (2006) proposed the first null model to detect rich clubs using the randomization procedure of Maslov and Sneppen (2002), and the null model preserved the degree sequence of the original network. Formally, the rich-club coefficient is defined as

$$\rho(k) = \frac{\phi(k)}{\phi_{\text{rand}}(k)}, \tag{12}$$

where $\phi_{\text{rand}}(k)$ is the topological rich-club coefficient of the null model.

In nested networks, nodes (OTUs) of low degree are linked primarily (or exclusively) to nodes (OTUs) of higher degree (Atmar and Patterson 1993). For a network matrix $\mathbf{A} = \{a_{ij}\}$ with elements indexed by OTUs i and j, Lee et al. (2012) defined nestedness S as

$$S = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\sum_l a_{il} a_{jl}}{\min(k_i, k_j)}, \tag{13}$$

where $k_i = \sum a_{il}$ is the degree of node i, and $k_j = \sum a_{jl}$ is the degree of node j.

The custom Python program we wrote to implement above described core-periphery and nestedness network analyses is provided in Appendix S1 of the OSI.

High salience skeleton (HSS) structure.—Trio motifs, core-periphery structure, rich clubs, and nestedness all focus on the OTUs themselves (network nodes). In contrast, the high salience skeleton (HSS) of a network focuses on the links between OTUs (network edges). Edges with high salience can be thought of as the

highways of a network. Following Grady et al. (2012), we computed the weights of the edges in our SDN between nodes (OTUs) as $w_{ij} = 1/|\rho_{ij}|$, where ρ_{ij} is the Spearman's correlation coefficient between the species dominance index of OTU i and OTU j.

Grady et al. (2012) then defined a salience matrix (**S**) for a network, whose elements (s_{ij}) are the salience values for each edge. The computation of the salience (s_{ij}) is based on the notion of shortest paths in weighted networks. Given a weighted network defined by the matrix of weights w_{ij} and a shortest path that originates at node x and terminates at node y, the indicator function is defined as

$$\sigma_{ij}(y,x) = \begin{cases} 1 & \text{if link } i \to j \text{ is on the shortest} \\ & \text{path from } x \text{ to } y \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

The shortest path tree (SPT) $T(x)$ rooted at node x can be represented as a matrix $\mathbf{T} = (\mathbf{x})$ whose elements are

$$T_{ij} = \begin{cases} 1 & \text{if } \sum_y \sigma_{ij}(y,x) > 0 \\ 0 & \text{otherwise} \end{cases}. \quad (15)$$

Then, the salience matrix **S** is a linear superposition of all SPTs, i.e.,

$$S = (s_{ij}) = \langle T \rangle = \frac{1}{N}\sum_x T(x) \quad (16)$$

Intuitively, the process of finding the HSS is to convert a weighted network matrix (which here is the SDN weighted by the inverse of Spearman's correlation coefficients of species dominance) into a new network matrix of edges. In the conversion process, those links whose weights were reduced to zero are removed, and the remaining links and nodes form a new network, the HSS. The custom Python program we wrote to implement Grady (2012) HSS detection approach is provided in Data S3.

Statistical distributions of properties of SDNs

We also used fits of three statistical distributions to the aforementioned network properties, topological structures, nestedness, and stability, of the SDNs to shed light on inter-subject heterogeneity. We tested the fit of the data to Poisson, normal, and power distributions. The two distributions for continuous variables (normal, power) differ in their symmetry (the normal is symmetric around its mean, the power is asymmetric and long tailed) and representativeness of their expected value (the mean of a normal distribution is a good estimator of the expected value, but the power law has a "no-average" property). The power-law distribution is an asymmetric, long-tail probability distribution that has some unique properties not possessed by the normal distribution. For example, the power-law distribution usually suggests heavy heterogeneity or skewed data points. It has the so-termed "no-average" property, which means that the average of the power law distribution can hardly represent majority of the data points because of the highly skewed long tail. The probability density function of the power distribution is

$$p(x) = \frac{K-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-K} \quad (17)$$

where x is the random variable, $x_{\min}$ is the minimum value of x, and K is the exponent or the scaling parameter ($K > 1$), which can be considered as a measure of asymmetricity (skewness) of the heterogeneity in the power distribution. A comprehensive discussion on the power distribution can be found in Clauset et al. (2009). Details of the Poisson and normal distributions can be found in most standard statistical texts (e.g., Gotelli and Ellison 2012).

In our analyses, we constructed individual SDNs, one SDN for each of the subjects in the 32-healthy cohort, based on their individual, longitudinal data sets. If the power distribution provided the best fit for a network measure (property), then we inferred that its inter-subject heterogeneity was asymmetrical and heterogeneous, and that the variance or skewness would help define network structure. The "no average" property of the power distribution suggests that there is not an average Joe who can represent the cohort with respect to the network property. In contrast, if the Poisson distribution provided the best fit, we inferred that inter-subject heterogeneity was random and homogenous. Last, if the Normal distribution provided the best fit, we inferred that inter-subject heterogeneity was symmetrical with a meaningful expected value (i.e., that an average property can represent a majority of the individuals in a cohort or population).

Although the average may be a poor indicator for the network properties of a cohort or population, variance (or equivalently, the standard error) and skewness can be rather useful for assessing the heterogeneity of the network properties. We therefore compute standard error and skewness in the following section for each network property, together with the distributions parameters mentioned previously, for the 32-healthy cohort, to cross-verify each other.

Results and Discussion

Species dominance networks (SDN)

Reconstructing SDNs and visual inspection.—The individual networks (graphs) of the 32-healthy cohort were heterogeneous, and the networks of most individuals differed from one another. We illustrate our results with two of these that have extreme properties (overloaded

triple roles of MAO, MDO, and hub vs. separate roles by three separate nodes; Figs. 2 and 3) to illustrate the network topology of the SDNs; the remaining 30 networks and their associated properties are provided in supplemental online material (Appendix S1). The symbols used in all 32 SDN network graphs are described in Appendix S1: Table S1, where the types of nodes and edges for all individuals in the 32-healthy cohort are illustrated in detail. In addition, in those networks, the OTU numbers, rather than their Latin (scientific) names are used to avoid overly crowded graphs. The look-up lists for the OTUs and their Latin names are available in Gajer et al. (2012), and in our Appendix S1: Table S16.

In Fig. 2 (subject 424), 12 nodes have the same highest degree, or are hubs (green hexagons). Note that we use the term hub slightly differently from its common usage in the literature. We designate a node as a hub only if it has the highest degree, whereas in the broader literature, a hub must have "high" degree, but not necessarily the highest degree. Our designation of node(s) with the highest degree (including ties) as hub(s) facilitates the quantitative analysis of the network properties. Besides the hub, two other types of nodes that deserve special attention are the most dominant OTU (MDO; pale blue square in Fig. 2) and the most abundant OTU (MAO; pink diamond in Fig. 2). The former is a unique feature in the SDN that is identified by our dominance metric at the species level. MDO is the OTU node with the highest D_{sd} in a SDN; for subject 424, it is OTU 20 (*L. reuteri*). MAO is the OTU node with the highest species abundance in a SDN; for subject 424 it is OTU 2 (*L. crispatus*).

In the SDN of subject 424, the hub, MDO, and MAO are separate nodes. However, two or all three of these may be identical. Among all the SDNs for the individuals in the 32-healthy cohort, there are only two subjects (412, 424) whose hub, MDO, and MAO are all distinct. For two other subjects (403, 408), their hub, MDO, and MAO are identical.

When three roles, hub, MDO and MAO, have the same node, we represented the overloaded node with a hexagon in red color (taking the shape of hub, but with a different color from the hub). When the hub and either the MDO or MAO overlap, the overloaded symbol shape still follows hub, but the color follows either MDO (pale blue) or MAO (pink). When MDO and MAO are the same OTU, the overloaded symbol takes its shape from MDO and color from MAO, i.e., a square colored in pink.

Another visually apparent property of the SDN for subject 424 is the prevalence of cooperative relationships (the ratio of positive links to negative links, $P{:}N$ in Appendix S1: Table S8 $\gg 1$, also see Ma 2017). This is consistent with the biological reality that HVMC is primarily a symbiosis-dominated community. Compared with the other networks in the 32-healthy cohort data set (Appendix S1: Fig. S1), that of subject 424 has the most edges (628, 3× the mean of the 32 individuals), the highest average degrees (12.4, 2× the cohort mean), and the second-most number of nodes (101, 1.7× the cohort mean) (Appendix S1: Table S2). Despite these visually conspicuous differences, the functional implication of these differences is obscure. A fundamental difficulty arises from the extreme individual heterogeneity in network topology exhibited by the 32 SDN networks; that is, everyone is different, and so are her SDNs. We therefore use additional computational and statistical analyses to gain further insights (see subsequent subsections).

The extreme inter-subject heterogeneity is obvious even from a simple comparison between the two exemplary networks here. The SDN of subject 408 (Fig. 3) has many fewer nodes (40) and edges (*61*) than the SDN of subject 424. An interesting observation from SDN-408 is the two conspicuous cliques of five: one consisting of OTU# 26, 29, 69, 212, and 258, another of 12, 31, 81, 120, and 165. Also in SDN of subject 408, OTU 3 (*Atopobium*) assumes the triple role of hub, MDO, and MAO. Formally, the fewer nodes and edges in SDN-408 can be quantified by network density, which measures how a network is densely populated with edges, and it has a value between 0 and 1. In the case of SDN-408, network density is (0.078 ± 0.001) smaller than the others.

General network properties of SDNs.—Appendix S1: Table S2 lists the results of 10 standard (general) network properties of the SDNs for all individuals in the 32-healthy cohort. Appendix S1: Table S2 also includes the results from fitting Poisson, normal, and power law distributions to them that reveal the inter-subject heterogeneity of these network properties. Table 1 shows the results for subjects 408 and 424.

The average degree per node (average number of neighbors) measures the average connectivity of a node in the network. Network density measures how a network is densely populated with edges, and it has a value between 0 and 1. A network of totally isolated nodes has zero density and a clique (totally connected network) has the density of 1. Network density is related to average degree, but the relationship is not a simple positive correlation. Network modularity measures how a network can be naturally divided into communities or modules (also known as cluster or groups). It also takes a value between 0 and 1. In gene regulatory networks, a module or community is considered to be a functional group, and we hypothesize that in species dominance networks, a module or community may represent a functionally similar group, such as a group of anaerobes. A network with high modularity should have dense connections between the nodes within modules but sparse connections between nodes in different modules.

The local clustering coefficient of a node n is defined as $C_n = 2e_n/k_n(k_n - 1)$ where k_n is the number of neighbors of n and e_n is the number of connected pairs between all neighbors of n. The network-clustering coefficient is the average of the clustering coefficients for all nodes in the network. It measures the degree to which nodes in a network tend to cluster together. The above-

TABLE 1. The properties of the species-dominance networks (SDNs) of the 32-healthy cohort constructed with Spearman's correlation coefficient (r) [$r \leq -0.5$ or $r \geq 0.6$; $P < 0.05$], illustrative results excerpted from full results in Appendix S1: Table S2.

Subject ID and parameters	No. nodes	No. edges	Average degree	Average cluster coefficient	Diameter	Average path length	No. connected components	Network density	Modul-arity	No. communities
408	40	61	3.050	0.670	5	1.949	7	0.078	0.769	8
424	101	628	12.436	0.760	8	3.062	9	0.124	0.291	10
...	...	...	...	...	...	...	...	...	...	...
Mean ± SE	58.438 ± 4.424	191.406 ± 27.351	5.892 ± 0.565	0.667 ± 0.015	7.531 ± 0.564	2.913 ± 0.199	5.969 ± 0.534	0.110 ± 0.009	0.548 ± 0.032	9.938 ± 0.833
Skewness	0.386	1.182	0.965	−0.561	0.750	0.791	0.216	0.656	−0.103	1.332
Assessing the inter-subject heterogeneity by analyzing the statistical distributions of the network properties										
Power law (PL)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
K of PL	6.4	3.6	5.2	14	5.6	3.2	5.5	3	26	3.9

described cluster coefficient (in Table 1 and Appendix S1: Table S2) is the local clustering coefficient, which is an indication of the embeddedness of single nodes. We further computed the global clustering coefficient, which is defined as: 3× the number of type-2 triangle motifs divided by the number of type-1 triangle motifs (Appendix S1: Table S4).

Scale-free, small-world, and stability properties of SDNs.—Appendix S1: Table S3 shows three additional network properties: scale-free, small-world, and stability. Scale-free networks are networks whose degree distributions follow a power distribution. Twenty-six of the 32 (81%) SDNs are scale-free (Appendix S1: Table S3), providing additional insights into the HVMCs. The hubs play a critical role in maintaining the connectedness of these networks. On the one hand, scale-free SDNs are quite robust against random removal of a node, since the chance that hubs are removed at random should be small. But on the other hand, scale-free SDNs are vulnerable to the loss of hubs due to targeted removal of highly connected critical nodes (hubs). Given the critical importance of hubs, we present more detailed analysis of hubs, along with analysis of MDOs and MAOs, in the next subsection.

Twenty-five of the 32 (80%) SDNs of the HVMCs were small-world networks (Appendix S1: Table S3). A small-world network is characterized by the property that most nodes are not neighbors of one another, yet most nodes are reachable from every other by a small number of hops. In Appendix S1: Table S3, whether the average path length (p) exceeded the $\ln(N_v)$ in the network was used to test the small-world network property. The average path length of the 32-healthy cohort was approximately 3, whereas the average of $\ln(N_v) \approx 4$.

Appendix S1: Table S3 also shows that all 32 SDNs were unstable to linear perturbations. This suggests that SDN network may be "rewired" significantly under certain environmental perturbations (host factors), such as menses possibly.

The fact that a majority of SDNs were scale-free, small-world networks, and were unstable supports our hypothesis that most fundamental network properties are conserved across individuals. That is, the inter-subject heterogeneities are bounded.

General network motifs without reference to the interaction types and node identities.—We investigated the general network motifs without considering the types (positive vs. negative interactions) or the identities of special network nodes (MDO, MAO, or hubs). Such motif finding can be done with standard network analysis software such as iGraph (Appendix S1: Table S4, Table 2). As before, we fitted Poisson, normal, and power distributions to the numbers of various motifs listed in the top section of Appendix S1: Table S4; results are at the bottom of Appendix S1: Table S4. Once again, the power distribution fit the distributions of all eight

TABLE 2. Key parameters of the core-periphery, rich-club, and nested structures in the SDNs of the 32-healthy cohort, illustrative results excerpted from full results in Appendix S1: Table S8.

			Density matrix			P:N				Rich-club	Nested-ness	Mean C/P dominance	
SDN	ρ	C:P	B11	B12 (21)	B22	Whole	C	P	C–P	$\overline{\rho(k)}$	*S*	C	P
408	0.380	0.250	0.607	0.020	0.079	2.81	0.89	8.75	0.67	1.629	0.095	5.99	14.43
424	0.735	0.639	0.717	0.003	0.055	33.89	Inf.	4.61	Inf.	1.178	0.180	1.75	38.91
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Mean	0.515	0.593	0.639	0.029	0.078	14.67	*N* inf. = 20†	*N* inf. = 7†	*N* inf. = 14†	1.298	0.160	7.14	24.09
SE	0.031	0.069	0.051	0.004	0.007	2.30				0.060	0.013	1.07	2.02
Skewness	−0.055	1.625	−0.244	1.468	1.025	0.984				2.765	0.663	1.16	0.36

†Many SDN exhibited infinite P:N ratio (lack of negative links, N) in their core (C), periphery (P), and core-periphery modules, marked here are the number of infinite P:N ratios in these modules (e.g., "N.Inf. = 20" indicating that 20 subjects or SDNs exhibited infinite P:N ratio with their core structures). The core appears to be more likely free of negative links, indicating predominantly cooperative nature in core structure.

Subject ID and parameters	Three-motif type 1	Three-motif type 2	Four-motif type 1	Four-motif type 2	Four-motif type 3	Four-motif type 4	Four-motif type 5	Four-motif type 6	Global clustering coefficient $\frac{3\times\text{Column 3}}{\text{Column 2}}$
408	53	39	11	66	50	3	11	15	2.208
424	2884	4019	4463	5532	21568	342	14837	20653	4.181
Mean	635	576	961	1881	2889	94	1455	1824	3.183
SE	153.1	151.862	341.26	624.136	984.184	29.658	558.958	713.099	0.551
Skewness	2.027	2.572	2.837	3.200	2.631	2.312	3.050	3.605	2.082
Assessing the inter-subject heterogeneity by analyzing the *statistical distributions* of the motifs									
Power law (*k*)	yes (2.4)	yes (2.9)	yes (2)	yes (2.3)	yes (2)	yes (2)	yes (1.7)	yes (2.4)	yes (2.3)

FIG. 4. The mean number of various basic motifs found in the SDNs of the 32-healthy cohort, illustrative results excerpted from full results in Appendix S1: Table S4.

types of motifs, whereas the normal and Poisson distributions did not fit any; inter-subject heterogeneity was highly asymmetrical and followed a long-tail distribution. The bottom of Appendix S1: Table S4 and Fig. 4 also give the exponent (*K*) of the power distribution, which measures the degree of asymmetry of the inter-subject heterogeneity of the motifs.

The standard motif detection reveals inter-subject heterogeneity, but it does not consider the identities of nodes of the types of their interactions. It offers little help in identifying which node dominates a network, which may be of fundamental importance in applications such as investigating disease etiology. Our new mean-crowding-based dominance index makes this possible, but we still need to use the special trio-motif detection technique (Ma and Ye 2017) discussed below.

Characterizing species dominance networks (SDNs) with special trio-motifs

The previous motif detection with standard correlation network analysis, which considered neither the identities nor the interaction types of nodes, offered limited value for analyzing the species dominance networks or providing insights into DSRs. We therefore analyzed special trio motifs by (1) considering the types of interactions (positive vs. negative) and (2) the roles of the special network nodes in (MDO, MAO, and hubs). As in many other studies of biological networks, our analysis was based on the correlation of a species dominance index between OTUs. Although correlation is not equivalent to causation, correlative data are still the only available data type for the human microbiome. Further, because computational time increased exponentially with arbitrarily sized motif searching problem, we limited our search to 3- and 4-node trio motifs. The trios not only are the most parsimonious motifs but also are sufficient to reveal important functional insights (Ma and Ye 2017).

We introduce two major categories of triangle motifs or trios: three-node "trios without handle" and four-node "Trios with handle," where the handle can be any of the MDO, MAO, or hub. Here, we discuss the MDO-connected trio motifs (Appendix S1: Table S5, Fig. 5);

Subject ID and parameters	Trios without handle								Trios with MDO handle														
	Trios with MDO			Trios without MDO			Total & Per Node Num.		Single-link MDO			Double-link MDO				Triple-link MDO					Degree of MDO	Trios per MDO degree
	+ + +	+ – –	Σ	+ + +	+ – –	Σ	Σ	Trios per node	+	–	Σ	+ +	+ –	– –	Σ	+ + +	+ + –	+ – –	– – –	Σ		
408	0	9	9	27	3	30	39	1	0	2	2	0	0	1	1	0	0	0	4	4	7	1
424	0	16	16	3979	24	4003	4019	40	3	9	12	0	11	12	23	0	0	10	10	20	8	7
Mean	0	19	19	515	41	557	576	8	2	31	33	0	4	22	26	0	1	8	26	35	8	7
SE	0.1	4.68	4.72	149.9	16.5	150.6	151.9	1.713	1.	16.4	17.5	0.03	2.31	8.05	10.0	0.03	0.47	3	9	11	1.091	1.669
Skewness	4.1	1.51	1.49	2.744	2.95	2.644	2.572	1.941	4	3.92	3.87	5.39	3.22	2.33	2.54	5.39	2.83	2	2.	2	0.935	1.370
Revealing the inter-subject heterogeneity by analyzing the *statistical distributions* of the network triangle motifs																						
Power law (k)	no 0.25	yes 3.8	yes 4.8	yes 2.5	yes 2.4	yes 2.8	yes 2.9	yes 2.5	yes 1.8	yes 1.4	yes 1.4	yes Inf.	yes 1.7	yes 1.7	yes 1.6	yes Inf.	yes 4.9	yes 1.7	yes 1.8	yes 4.4	yes 4.5	yes 2.5

FIG. 5. The occurrence of the MDO-triangle motifs (trios) in 32-healthy cohort, illustrative results excerpted from full results in Appendix S1: Table S5.

the results of MAO- and hub-connected trio motifs are presented in Appendix SI. The primary objective was to reveal the role of MDO in shaping the interactions within the fundamental functional groups in the SDN.

We report two categories of trios in Fig. 5 and Appendix S1: Table S5: the left side of each table shows Trios without handle, which are three-node trios without links to the external handle (MDO), whereas the right side of each table shows trios with MDO handle, which consist of three-node trios with links to the external handle (MDO). In the first category, there are two subtypes: one with MDO and the other without MDO. The type of trios without MDO is the same as the three-motif type 2 in Appendix S1: Table S4 that was detected with standard motif detection methods and that should be compared with our trios with MDO. In the second category, there are three sub-types: single-link MDO, double-link MDO, and triple-link MDO, with one, two, and three links to the MDO handle from the trio, respectively.

We also report trios per node in the section of Trios without handle because the number of nodes or edges should influence the number of trios. In the section of trios with MDO handle, we report the trios per MDO degree because this parameter should reflect the influence of MDO connections on the number of trios. The *trios* per MDO degree refers to the average trios per MDO handle to which it is connected, excluding those trios without a MDO handle.

In summary (Fig. 5, Appendix S1: Table S5):

1) Trios without MDO handle. We only detected two of the four possible trios without handle: "+ + +", "+ – –". We name the first of these total mutualism, and the second partial mutualism. That the other two possible patterns "+ + –" (strongly partial mutualism), "– – –" (total competitive) are missing in the 32-healthy cohort is puzzling. However, this apparent puzzle is interpretable. A trio system consisting of three players who try to compete with one another (or dominate each other in term of our species dominance metric) (– – –) would break down if selection continued to exert ever-increasing pressure on each. Similarly, when there are two pairs of cooperative allies, a third player may be "coerced" to cooperate with them (+++), rather than to oppose them (+ + –). In other words, for a third player, its "life pressure" from natural selection should be lower if it simply acts as a collaborator in the system (+ + +), than to act as a "competitor" in the system (+ + –).

The ratio of the two observed trio patterns depends on whether the MDO is part of the trio (Fig. 5 and Appendix S1: Table S5, left). In the trios with MDO, the ratio of partial mutualism trios (+ – –) to total mutualism trios (+ + +) is 608:5. However, in trios without MDO, the ratio is 1312:16480. We hypothesize that, in trios with MDO, the partial mutualism trio (+ – –) represents a trio relationship with a cooperative ally (+) collectively opposing a third non-cooperator (– –). In this trio pattern, the existence of negative feedback (relationship) should limit resource overconsumption, and hence, the partial mutualism system can be more stable because of its moderate resource consumption. On the other hand, in the full mutualism trio (+ + +), resource overconsumption may occur because of all positive feedbacks. The potential of resource overconsumption should be particularly higher when the full mutualism trio contains MDO. Therefore, in the trios with MDO, the full mutualism trios can be harder to sustain than partial-mutualism trios. In contrast, in the trios without MDO, the potential of resource overconsumption should be low when the trio species are moderate or low in resource consumption in the absence of MDO. This low or moderate resource consumption can allow more full-mutualism trios to occur when MDO is absent. The presence or absence of MDO in the trios therefore determines the "carrying capacity" to accommodate different levels of the full mutualism with and without MDOs.

2) Trios with MDO handle. Each type (single-link, double-link, or triple-link) of *Trios with MDO handle* was further classified based on the type of the interaction between MDO and the trio base, either positive (+) or negative (–) (Fig. 5 and Appendix S1: Table S5, right). The MDO (the handle) almost always inhibits the trios connected to it, as exhibited by the "–" (inhibitive) column vs. "+" (facilitative)

column in the single-link MDO pattern, and similar prevalence of "– –" in the double-link and "– – –" in the triple-link patterns (Fig. 5 and Appendix S1: Table S5). The difference is ≈15-fold. Similar inhibitive roles of MDO are also prevalent in the "Double-link MDO" and "Triple-link MDO". In the former trios, the ratio of "– –" vs. sum of ("++" and "+ –") is ≈5×. In the latter, the ratio of "– – –" vs. sum of ('+++', "++ –", "+ – –") is ≈3×, although the inhibitive force seems to decrease in the latter two types.

We hypothesize that the inhibitory role of MDO may be a signature of healthy vaginal microbial communities, and this hypothesis also is consistent with the conventional wisdom that dominant species are beneficial for protecting women from BV disease. In this study, we quantitatively measured such dominance (inhibition) in a vaginal community; such quantitative measures are absent in the previously published literature. Dominance is only one aspect of BV etiology; which OTU dominates and to what extent it dominates are important too. We hypothesize that the inhibitive role of MDO should play a critical role in preventing potentially harmful anaerobes from changing the vaginal environment to BV-prone states. In the third paper in this series (Ma and Ellison unpublished data), we will test this hypothesis with a new data set including ABV and BV patients.

If we define the ratio of inhibition to facilitation links as *dominance power*, more precisely, the number of trios with totally inhibitory MDO-trios to the total number of facilitative (*total mutualism* in the terminology in [*i*]) and partially facilitative (*partial mutualism* in [*i*]), the *dominance power* of MDO-trios is ≈15, 5, and 3, for the single-link MDO, double-link MDO and triple-link MDO, respectively. The decreasing dominance power series of 15, 5, and 3 is prima faciae reasonable because it measures the power for MDO to inhibit one, two, or three species, respectively, *i.e.,* the more species, the more difficult to inhibit. The dominance power of MDO can be very useful for investigating the critical role of dominant species such as *Lactobacillus inners* in inhibiting some potentially harmful functional groups of anaerobes, and their potential implications for BV etiology.

To further investigate the inter-subject heterogeneity in the MDO-trios motifs, we also fitted the numbers of trios to Poisson, normal, and power distributions (Fig. 5, Appendix S1: Table S5). Only the power distribution fit all the trio patterns, whereas both Poisson distribution and normal distribution failed to fit in most cases. The ubiquitous success of the power distribution suggests that the inter-subject heterogeneity in these trio patterns is not symmetrical (as would be implied by the normal distribution), but rather is highly asymmetrical and follows a long-tail distribution. The bottom of Fig. 5 and Appendix S1: Table S5 also gives the exponent (in parentheses) of the power distribution, which measures the degree of asymmetry of the inter-subject heterogeneities of those trio patterns.

3) MAO- and hub-connected trios. We also searched for the trio motif patterns connected with MAO (most abundant OTU), and the trios connected with hubs, respectively (Appendix S1: Tables S6 and S7). In general, the results of trio analysis with MAO were similar to those with MDO as, in many cases, MAO and MDO were the same OTU in some subjects. This appears to suggest that either MDO or MAO can be utilized for motif analysis. However, there are cases where MDO and MAO are different. One advantage of using MDO is that it is based on the mean-crowding-based dominance index that is applicable at both species and community levels.

In addition, we expect that the primary role of a hub is connection, linking nodes together, rather than dominating other nodes or being dominated by its neighbors. In other words, a hub should be "easy" rather than "hard" with which to interact and to connect more nodes (i.e., to become a hub). Natural selection should favor this strategy. We conjecture that the "ally (clique) of hubs" in the SDN may have important functional implications. From general principles of complex networks, we may conclude that the network with larger hub cliques (i.e., cliques of the hubs with more nodes) should be more robust than those with smaller hub cliques. For example, SDN-424 has 12 hubs that form a clique of 12 hub nodes, which are displayed as 12 mutually connected hexagons in the network (Fig. 2).

Core-periphery, rich-club, and nestedness analyses

Although standard correlation network analysis provided limited insights into the SDN, the analysis of trio motifs was a powerful tool to detect important local topological structures. In this study, trio motif analysis showed extremely high inter-subject heterogeneity: it also can be used to identify BV-associated bacteria or trios of anaerobes and sheds important light on the investigation of BV etiology (Ma and Ellison unpublished data). In this and next section, we focus on global topological structures, including core-periphery, rich-club, nestedness, and skeleton, respectively, all of which have significant effects on network stability. The first three analyses (core-periphery, rich-club, and nestedness) are measured for network nodes, whereas the last analysis (the skeleton) is measured for network links. Together, the CNS (core, nestedness, and skeleton) analyses not only complement the standard network analysis and our trio-motif technique, but also provide an effective approach to revealing the global topological structures that shape the stability of underlying network (community). Finally, we investigate the dynamics (stability) of these global topological structures by integrating SDN analysis with phenomenological modeling.

Table 2 (illustrative results) and Appendix S1: Table S8 (full results) show key parameters from core-periphery, rich-club, and nestedness analyses for each of

the SDNs in the 32-healthy cohort. The parameter ρ (2nd column) shows that the core-periphery structure is rather strong, except for subject 444 (ρ = 0.097), with an average ρ = 0.52 across the 32 subjects. The average ratio of cores to periphery nodes (C:P; third column) is 0.592, indicating that, on average, the number of nodes in the core is ≈ 40% less than in the periphery. In the density matrix of core-periphery (columns 4–6), blocks 12 and 21 are equal, as they represent the interaction density between core and periphery nodes, or vice versa. The average density of core-core is 0.639 vs. 0.029 and 0.078 of core-periphery and periphery-periphery, respectively, indicating far stronger interconnections within the core.

The next four columns (Column 7–10) report the P/N ratio—the number of positive links (correlations) to negative links (correlations) within the whole network, core, periphery, and core-periphery, respectively. A finding from the P/N ratios is that the core has fewer negative interactions than the periphery, suggesting that the core is a more cooperative structure. In 20 out of the 32 SDNs in the healthy-cohort study, negative links were missing from the core. The numbers of missing links in the periphery and core-periphery are 7 and 14, respectively. This difference among the three network structures is consistent with the nature of human microbiomes, which are predominantly mutualistic systems. Compared with periphery and core-periphery structures, the level of mutualism within core structures seems to be stronger.

For most of the 32 SDNs, $\rho(k) > 1$ (column 11), meaning that the rich-club phenomenon is ubiquitous. The SDN of subject 435 has the highest rich-club index (2.749), whereas that of subject 429 has the lowest (0.956), which is still close to 1, suggesting that rich clubs predominate. The last column, nestedness (S), ranges from 0.044 to 0.321 with an average of 0.160 (little nesting observed). The index S measures the relative level of nestedness with a range between 0 to 1; the larger the S, and the higher the nestedness ($S = 1$ for perfectly nested).

Appendix S1: Table S9 also gives the result of fitting normal and power distributions to these network properties. The near ubiquitously successful fitting of the power distribution to all parameters tested suggests that these parameters are highly heterogeneous and asymmetrical among individuals.

Among individuals, members of the core were more variable than those of the periphery among individuals, as illustrated by the OTU frequency distribution in the core and periphery, respectively (see Data S1: OTU-Frequency-A.csv & OTU-Frequency-B.csv). For example, the most widely distributed periphery OTU occurred in 24 subjects, but its counterpart in the core occurred in only 11 subjects. The four indicator species of Ravel et al.'s (2011) community state types (CSTs) of the HVMC, *Lactobacillus inners, Lactobacillus jensenii, Lactobacillus gasseri,* and *Lactobacillus crispatus*, occurred more frequently in the periphery. In Figs. 2 and 3, as well as Appendix S1: Figs. S1–S32, the core and periphery nodes are colored in dark blue and light blue, respectively.

High salience skeleton (HSS) network analysis

Because it is based on edges, the HSS is often utilized to detect "highways" in a network, rather than the nodes identified as hubs. We used the inverse of the correlation coefficient of species dominance to construct a weighted network in which the HSS represents species interactions (correlations) that play critical roles (both direct and indirect interactions) in the overall SDN. A link is assigned high salience not only because itself has sufficiently high weight, but also because it is an important route of the global highway network.

Table 3 (illustrative results) and Appendix S1: Table S10 (full results) lists the HSS statistics, including percentage of links with salience >0 (links%, i.e., the percentage of links from the full network [weighted network] that are preserved in the HSS), and the maximum, mean, standard error, skewness, and kurtosis of the salience in each SDN. On average, approximately one-third of the links are salient skeletons (Table 3, Appendix S1: Table S10). The assortativity (r_{HSS}) is the Pearson's correlation coefficient of degree between pairs of linked nodes, and its mostly negative but rather small absolute values suggest that the HSS network is mostly neutral or slightly disassortative. Assortativity also is related to network resilience; Newman (2002) found through both analytic and simulation studies that disassocitative networks tend to be less resilient because network is less easily percolated in such networks. The last two columns in Fig. 5 list the P value from fitting power distributions to salience values for each SDN, and the parameter k of the power distribution. The salience in most (26 out of 32) SDNs followed a highly skewed power distribution, which was also supported by the rather large skewness and kurtosis. Both skewness (the third standardized moment) and kurtosis (the fourth standardized moment) are descriptors for the shape of probability distribution, measuring "tailedness" of the distribution, while a higher kurtosis is the result of outliers, as opposed to frequent modestly sized deviations.

We also analyzed the inter-subject heterogeneity of HSS parameters by fitting both normal and power distributions (Appendix S1: Table S11). All HSS parameters were successfully fitted to the power distribution, whereas the normal distribution failed to fit any of the parameters other than skewness and kurtosis. Thus, as were the parameters of core networks, the parameters of HSS network were rather heterogeneous and asymmetrical among individuals.

Appendix S1: Table S12 gives the salience values corresponding to certain percentiles in each SDN, and the table also shows that in most SDNs, >50% links have zero salience. Table S13 displays the distribution of

TABLE 3. Key statistical properties of the high-salience skeletons (HSS) in the SDNs of 32-healthy cohort, illustrative results excerpted from full results in Appendix S1: Table S10.

Subject ID and parameters	Statistics of HSS						Assortativity	Power law	
	Links (%)	Maximum	Mean	SE	Skewness	Kurtosis	r_{HSS}	P	K
408	32.1	0.782	0.036	0.096	3.989	18.271	−0.019	0.435	4.763
424	23.7	0.967	0.016	0.061	7.153	68.196	−0.008	0.923	5.082
...	...	...	...	...	...	...	...	...	...
Mean	28.178	0.921	0.031	0.086	5.235	37.358	−0.016	0.441	3.743
SE	2.193	0.015	0.002	0.005	0.171	2.694	0.001	0.058	0.229
Skewness	0.910	−1.305	0.886	0.413	0.248	0.344	−0.889	0.169	1.319

salience over various intervals of the salience in the SDNs of the 32-healthy cohort. Supplementary Data S2 (Top 50 skeleton links.csv and Full skeleton links.csv) gives skeleton salience values of each SDN of the 32 SDNs, in descending order or salience (top 50 skeleton links.csv and full skeleton links.csv list in Data S2). We failed to detect any skeleton with salience >0 and shared by the SDNs of all 32 individuals, which again demonstrated the extreme inter-subject heterogeneity in SDNs (Appendix S1: Tables S14, S15). That is, the highway structure of SDNs differed among individuals. In Figs. 2 and 3, and Appendix S1: Figs. S1–S32, the HSSs were drawn with heavily thickened edges. Strictly speaking, skeleton links cannot be identified fully in the graphs of SDNs we drew in this article, because they can only be fully identified in weighted SDNs or the skeleton networks, both of which were indeed constructed in this study (to compute those HSS parameters in Table 3 and Appendix S1: S11–S15) but not drawn to save page space and also improve clarity. That is, in Figs. 2 and 3, and Appendix S1: Figs. S1–S32, the network links we marked (with thickened lines) are but a subset of the HSSs. The full list of HSSs is given in Full skeleton links.csv in Data S2.

Synthesis of SDN and phenomenological modeling of DSR

The primary objective of the synthesis here is to ask which network structure, the core or periphery, is more stable, and which one may exert more control or regulation over the stability of the whole SDN. Previously, we used data-driven phenomenological modeling (Ma and Ellison 2017) to investigate the DSR at the whole community level in each of the 32 SDNs. Here, we integrated that modeling approach with the SDN analysis presented herein to investigate the DSR of core and periphery structure in each SDN (community). The fundamental patterns we expected to obtain from the phenomenological modeling, whether at the coarser scale of the entire community or the finer scale of the core-periphery, were the three dominance–stability dependence relationships (diversity–stability dependence relationships in Ma and Ellison [2017]): dominance-dependent stability (DDS), dominance-inversely-dependent stability (DIDS), and dominance-independent stability (DID). All three types of relationships may occur in a single community, and indeed they often alternate with one another. Furthermore, the six phenomenological models are not necessarily exclusive. Rather, some communities may be fit equally well by multiple models because they can describe the same or similar trends over the range of the observed data. We thus pursued a compromise between realism and simplicity, and assigned a "best" model for each community based on a set of rules that considered the statistical appropriateness of model fitting (R^2, standard error of parameter, model parsimony) and biological interpretations.

Tables 4 and 5 and Appendix S1: Table S16 (full results) summarize the description of the DSR for the core-periphery of each community (SDN), key model parameters for judging the DSR types, and the R^2, based on the detailed modeling results presented in Appendix S1: Tables S15-1–S15-6. The linear, quadratic, and quadratic-quadratic (Q-Q) models were the most widely applicable for the core structure, with 6, 7, and 6 SDNs, respectively, fitted successfully. The linear, quadratic, and linear-quadratic (L-Q) models were the best for the periphery structure, with 8, 12, and 7 SDNs, respectively, fitted successfully. The DSRs of the core and periphery of the same SDN may be the same type, but were more likely to be different types. For example, both the core and periphery of subject 400 had a DSR described by a parabola (quadratic model). Since the parabola opens upward ($c < 0$), the DSR is DDS with a possible stable equilibrium at the vertex of the parabola. The core of subject 446 exhibited logistic stability, but its periphery was linearly stable. In addition, the DSRs of core-periphery structure of an SDN may differ from that of the whole SDN (Ma and Ellison 2017). For example, the stability model of entire SDN of subject 400 is logistic (Ma and Ellison 2017), rather than the quadratic of the core-periphery. In eight cases, we failed to find appropriate models for their core-periphery structures (Appendix S1: Table S14).

Results from the integrated SDN analysis and phenomenological modeling complemented the insights from SDN analysis by further revealing the dynamics (i.e., dominance-stability dependence types) of core-periphery structures (i.e., network components identified by

TABLE 4. An exemplary summary of the stability-model selection from six candidate models (linear, quadratic, reciprocal, logistic, liner-quadratic, quadratic-quadratic models) fitted to each subject in the 32-healthy cohort, illustrative results excerpted from full results in Appendix S1: Table S14.

Networks	Model	Dominance–stability relationship	R (linear), R^2 (other models)
408			
Core	Linear ($b < 0$)	globally DDS, but the mechanism may be complex locally	0.645
Periphery	Q ($c > 0$)	DDS followed by a possible stable equilibrium and DIDS	0.320
424			
Core	Q ($c > 0$]	DDS followed by a possible stable equilibrium and DIDS	0.421
Periphery	Linear ($b < 0$)	globally DDS, but the mechanism may be complex locally	0.709
...	...	...	...
Mean			
Core			0.597
Periphery			0.657

TABLE 5. A summary of the stability-model selection from six candidate models (linear, quadratic, reciprocal, logistic, liner-quadratic, quadratic-quadratic models) fitted to each subject in the 32-healthy cohort, illustrative results excerpted from full results in Appendix S1: Table S14.

Model	Whole†	Core	Periphery	Stability implications
Linear	14	6	8	Globally DDS, but the mechanism may be complex locally.
Quadratic	0	7	12	DDS followed by a possible stable equilibrium and DIDS.
Reciprocal	0	2	1	DDS with an asymptotic equilibrium line when $D_c \to \infty$
Logistic	3	2	0	DDS with an asymptotic equilibrium line when $D_c \to \infty$
Logistic-Sine	2	0	0	DDS and DIS alternate periodically
Linear-Quadratic	10	1	7	DIDS followed by DDS, a possible stable equilibrium and another piece of DIDS
Quadratic-Quadratic	3	6	4	DIDS and DDS alternate, two parabolas connected at $D_c = d \approx 2$, the equilibriums are uncertain (stable or unstable)
NA (failed to fit)	0	8	0	

Note: D_c and d are defined in Box 1.

†The counting of fitted models for the whole community was from Ma and Ellison (unpublished data), where the whole community level modeling was presented to demonstrate the utility of our dominance metric at the community level.

the SDN analysis). Taken together, these results help to identify which of the network structures (core or periphery) is more stable, and which one exerts more control (regulation) of the stability of whole SDN. For example, both core and periphery of the SDN of subject 400 have the same dominance-stability relationship, but the overall SDN exhibits a different type. This suggests that network stability of this SDN is likely determined by the interaction between the core and periphery, rather than by either of them alone.

SUMMARY AND PERSPECTIVES

SDN analysis with standard correlation network analysis

The SDN for each HVMC captures the community's species dominance dynamics in a multidimensional space, and the network properties encapsulate the community's DSR in a vector of network parameters. Summarizing the information from basic network properties, we reiterate three exemplary insights we obtained previously. First, SDNs are highly heterogeneous among individuals, but the inter-subject heterogeneities are bounded: some network properties are highly conserved (constant) but others are highly variable among individuals or among various groups (Ma and Ellison unpublished data). We use an analogy with inheritance vs. variation in genetics to describe the phenomenon of "bounded heterogeneity," i.e., the balance between homogeneity (conservation) and heterogeneity (variation). Second, most of the inter-subject heterogeneities among individual subjects follow a power distribution, which is asymmetrical and long tailed, and for which the mean is a poor representation of most individuals in a cohort. Third, the adoption of species dominance index, rather than species abundance, in the SDNs is advantageous in distinguishing between healthy and diseased microbiomes (Ma and Ellison unpublished data).

The lack of difference in some network properties suggests that they are highly robust and conserved across individuals. We hypothesized that, from an evolutionary perspective, inter-subject heterogeneity is similar to variation, and the lack of difference among groups or cohorts is similar to inheritance in genetics. In other words, evolution of the human metagenome is like evolution of the human genome; both variation and

inheritance are necessary for the evolution and adaptation to occur. Furthermore, the evolutions of both genome and metagenome are likely intermingled with complex interactions.

We also tested the applicability of two of the most extensively studied complex network models, scale-free (Barabasi and Albert 1999) and small-world networks (Watts and Strogatz 1998, Gray et al. 2009), to the SDNs of HVMCs. The good fit of a scale-free model is consistent with the biological reality of the HVMC; the community is robust to random losses of OTUs in the community, but vulnerable to a targeted loss of, e.g., an MDO or hub. The small-world network property indicates that the SDNs are tightly connected and the average path distances between nodes are rather small (<3 nodes or species). A third property we tested is the linear stability, which suggests that SDN may be rewired (unstable) under certain environmental perturbations, such as woman's menses.

Special trio motifs

Methodologically, our SDN analysis introduces two innovations. The first is the use of species dominance index, rather than population abundance, and the second is the special trio-motif detection techniques. The first enables us to capture the dynamical co-dominance relationships between species (OTUs) in a multi-dimensional setting. The second allows us to answer the question of which OTU dominates, its identity (MDO, MAO, hub), and interaction types (mutualism or inhibitive). The significance of the dominance-based SDN network analysis, especially with special trio-motif detection, lies in the fact that what matters most in practice (such as the clinical diagnosis and treatment of BV) may be more about who dominates and less about the degree of dominance at the community level because there may be multiple trajectories that can lead to the observed dominance (diversity) or stability level. The functional redundancy can be achieved by different guilds of species, such as the trios of anaerobes. Furthermore, it may be necessary to capture the behaviors of species that are not in dominant groups. For example, it has been conjectured that BV may be associated with the resurgence or invasion by normally less abundant species.

Network motifs are analogous to guilds or functional groups in ecological communities. In the trio motif without MDO handle, we only detected two kinds of trios (+ + +: total mutualism or TM) and (+ – –: partial mutualism or PM); the other two combinations (– – –) and (+ + –) were missing. The ratio of TM : PM was opposite in trios with and without MDOs, and the existence of MDO in the trios led to a 1000-fold difference TM/PM. We propose a resource-limited mutualism hypothesis to explain this interesting phenomenon. That is, when the MDO (possibly a resource sink) exists in a trio, the community can support few total mutualism trios because of limited resources. In contrast, when there is no MDO member in the trio, more mutualism trios can thrive because of overall moderate or small resource consumptions by non-MDO species. As discussed by Ma and Ellison (unpublished data), the extremely rare total mutualism trios with MDO members are often associated with BV. They have the potential to become biomarkers for personalized diagnosis of BV.

In the trio with MDO handle, we found that the MDO handle, whether it connected to a trio with one (single-link MDO), two (double-link MDO), or three links (triple-link MDO), predominantly played an inhibitory role to its connected trio. The decreasing dominance power (inhibitory links/facilitative links) of MDO, double-link MDO, and triple-link MDO trios, reflects the reality that inhibiting more species is more difficult. The dominance power of an MDO can be used to assess the critical role of dominant species such as *L. inners* in inhibiting some potentially harmful functional groups of anaerobes, and to study their implications to BV etiology (Ma and Ellison unpublished data).

Core-periphery, rich-club, nested network structures, and high skeleton networks

The focus of special trio motif analysis is local, but the core-periphery (including nestedness, and rich-club) analysis reveals global characteristics of network structure and stability. In our core-periphery network analysis of the HVMC, approximately one-third of nodes constitute the core, and the remaining constitutes the periphery. Nevertheless, the parameters of core-periphery structures are highly heterogeneous and skewed among individuals. The members of the core also are more variable than those of the periphery among individuals.

Interestingly, the four indicator species of community state types (CSTs; Ravel et al. 2011, Ma and Li 2017), *L. inners, L. jensenii, L. gasseri,* and *L. crispatus,* occurred more frequently in the periphery than in the core. In addition, nodes in the core are more cooperative than those in periphery; more negative interactions (small *P*: *N* ratios) were identified in the periphery of the SDNs. We hypothesize that interactions in the periphery are more dynamic than those involving the dominant species (e.g., the indicator species) in the community. However, whether the core or periphery is more stable depends on the type of dominance-stability dependence relationship.

Although the core-periphery, rich-club, and nestedness reveal the global topological and stability characteristics of SDNs, the analysis of high salience skeleton network accomplishes a similar mission but from a contrastingly different perspective, the link perspective, or detecting "highways" in the SDN. It was shown that on average approximately one-third of the edges in the SDNs are high salience skeletons (HSSs), and they constitute the highways in the SDNs. However, the HSS is highly heterogeneous among individual SDNs and follows highly skewed, long-tail power distribution.

Inter-subject heterogeneities in SDN properties

The nearly ubiquitous success of the power-law distribution in fitting to the various SDN properties indicates that the inter-subject heterogeneity in the healthy cohort is asymmetrical and heterogeneous, and there is not an average Joe that can represent a majority of individuals in the cohort. A biomedical implication for the microbiome-associated diseases such as BV is that personalized medicine would be necessary because of the frequently extreme heterogeneities among individuals (Ma et al. 2011). The extreme heterogeneities or individualities appear to be a characteristic of the medical ecology of human microbiome-associated diseases such as BV. Obviously, this characteristic also allowed us to more thoroughly demonstrate and test our proposed SDN analysis framework.

Synthesis of dominance dynamics modeling and SDN analysis

Our framework consists of SDN analysis, phenomenological modeling of dominance–stability dynamics, and their integrations (Fig. 1). We demonstrated a synthesis at the community level DSR in part one of this series (Ma and Ellison 2017). Here we demonstrated the synthesis for investigating the dynamics of core-periphery structures. This synthesis aimed to address a fundamental question, i.e., which structure, core or periphery, is more stable, and which structure is likely to control the community stability to a higher extent. The answer is that it depends on the types of dominance–stability dependence types (i.e., DDS, DIDS, and DIS, and more likely their alternations even in the same community).

The synthesis may be conducted in other junctions, such as analyzing the dynamics of high salience skeletons, as well as incorporating the effects of meta-factors of important biomedical significance, as demonstrated in part three of this series. Finally, it should be pointed out that, although the HVM time-series data sets of the 32-healthy cohort we utilized are sufficiently long for demonstrating the SDN analysis (the primary objective of this study), the data sets are far from ideal for conducting the synthesis. Furthermore, rigorous tests against null models or control treatments are also missing in our synthesis.

Acknowledgments

We thank Larry Forney, University of Idaho, and Jacques Ravel, University of Maryland for discussion of BV at various occasions. We are also indebted to D. D. Ye, L. W. Li, and J. Li, from the Chinese Academy of Sciences, for programming and verifying the computational results. Z. S. Ma work received funding from the following sources: National Science Foundation of China (Grant No. 71473243), Yun-Ridge Industry Technology Leader Grant, A China-US International Cooperation Project on Genomics/Metagenomics Big Data. A. M. Ellison's work on this project was supported by the Chinese Academy of Sciences (CAS) President International Fellowship Initiative for Visiting Scientists, Grant no. 2016VBA074. The authors declare no conflict of interests.

Literature Cited

Allesina, S., and S. Tang. 2012. Stability criteria for complex ecosystems. Nature 483:205–208.

Allesina, S., and S. Tang. 2015. The stability-complexity relationship at age 40: a random matrix perspective. Population Ecology 57:63–75.

Atmar, W., and B. D. Patterson. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. Oecologia 96:373–382.

Baiser, B., N. Whitaker, and A. M. Ellison. 2013. Modeling foundation species in food webs. Ecosphere 4:146.

Barabasi, A.-L., and R. Albert. 1999. Emergence of scaling in random networks. Science 286:509–512.

Barberán, A., S. T. Bates, E. O. Casamayor, and N. Fierer. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME Journal 6:343–351.

Barberán, A., E. O. Casamayor, and N. Fierer. 2014. The microbial contribution to macroecology. Frontiers in Microbiology 5:article 203.

Berger, W. H., and F. L. Parker. 1970. Diversity of planktonic *Foraminifera* in deep-sea sediments. Science 168: 1345–1347.

Bittleston, L. S., N. E. Pierce, A. M. Ellison, and A. Pringle. 2016. Convergence in multispecies interactions. Trends in Ecology and Evolution 31:269–280.

Bollobás, B. 2001. Random graphs. Cambridge University Press, Cambridge, UK.

Borgatti, S. P., and M. G. Everett. 1999. Models of core-periphery structures. Social Networks 21:375–395.

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Snader, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs 84:45–67.

Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2009. Power-law distributions in empirical data. SIAM Review, 51:661–703.

Clemente, J. C., L. K. Ursell, L. W. Parfrey, and R. Knight. 2012. The impact of the gut microbiota on human health: an integrative view. Cell 148:1258–1270.

Colizza, V., A. Flammini, M. A. Serrano, and A. Vespignani. 2006. Detecting rich-club ordering in complex networks. Nature Physics 2:110–115.

Costello, E. K., K. Stagaman, L. Dethlefsen, B. J. Bohannan, and D. A. Relman. 2012. The application of ecological theory toward an understanding of the human microbiome. Science 336:1255–1262.

Coyte, K. Z., J. Schluter, and K. R. Foster. 2016. The ecology of the microbiome: networks, competition, and stability. Science 350:663–666.

Csardi, G, and Nepusz T. 2006. The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org.

Csermely, P., A. London, L. Y. Wu, and B. Uzzi. 2013. Structure and dynamics of core-periphery networks. Journal of Complex Networks 1:93–123.

Cushing, J. M., R. F. Constantino, B. Dennis, R. A. Desharnais, and S. M. Henson. 2003. Chaos in ecology: experimental nonlinear dynamics. Academic Press, Amsterdam, The Netherlands.

Ding, T., and P. D. Schloss. 2014. Dynamics and associations of microbial community types across the human body. Nature 509:357–360.

Donohue, I., et al. 2016. Navigating the complexity of ecological stability. Ecology Letters 19:1172–1185.

Durrett, R. 2006. Random graph dynamics. Cambridge University Press, Cambridge, UK.

Erdös, P., and A. Renyi. 1960. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5:17–61.

Eschenbach, D. A., S. Hiller, C. Critchlow, C. Stevens, T. DeRouen, and K. K. Holmes. 1988. Diagnosis and clinical manifestations of bacterial vaginosis. American Journal of Obstetrics and Gynecology 158:819–828.

Estrela, S., M. Whiteley, and S. P. Brown. 2015. The demographic determinants of human microbiome health. Trends in Microbiology 23:134–141.

Faust, K., J. F. Sathirapongsasuti, J. Izard, N. Segata, D. Gevers, J. Raes, and C. Huttenhover. 2012. Microbial co-occurrence relationships in the human microbiome. PLoS Computational Biology 8:e1002606.

Fortunato, S. 2010. Community detection in graphs. Physics Reports 486:75–176.

Fredricks, D. N. 2011. Molecular methods to describe the spectrum and dynamics of the vaginal microbiota. Anaerobe 17:191–195.

Fredricks, D. N., T. L. Fiedler, and J. M. Marrazzo. 2005. Molecular identification of bacteria associated with bacterial vaginosis. New England Journal of Medicine 353:1899–1911.

Gajer, P., et al. 2012. Temporal dynamics of the human vaginal microbiota. Science Translational Medicine 4:132ra52.

Gibbons, S. M., and J. A. Gilbert. 2015. Microbial diversity – exploration of natural ecosystems and microbiomes. Current Opinion in Genetics & Development 35:66–72.

Gotelli, N. J., and A. M. Ellison. 2012. A primer of ecological statistics. Second edition. Sinauer Associates Inc., Sunderland, MA, USA.

Grady, D., C. Thiemann, and D. Brockmann. 2012. Robust classification of salient links in complex networks. Nature Communications 3:864.

Gray, R. T., C. K. C. Fung, and P. A. Robinson. 2009. Stability of small-world networks of neural populations. Neurocomputing 72:1565–1574.

Green, D. G., N. Klomp, G. Rimmington, and S. Sadedin. 2006. Complexity in landscape ecology. Springer, Berlin, Germany.

Grilli, J., T. Rogers, and S. Allesina. 2016. Modularity and stability in ecological communities. Nature Communications 7:12031.

Grimm, V., and C. Wissel. 1997. Babel, or the ecological stability discussions: an inventory and analysis of terminology and a guide for avoiding confusion. Oecologia 109:323–334.

HMP Consortium (Human Microbiome Project Consortium). 2012*a*. A framework for human microbiome research. Nature 486:215–221.

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

Hooper, L. V., D. R. Littman, and A. J. Macpherson. 2012. Interactions between the microbiota and the immune system. Science 336:1268–1273.

Ings, T. C., et al. 2009. Ecological network—beyond food webs. Journal of Animal Ecology 78:253–269.

Julian, J., L. F. C. McAuley, and T. S. Caetano. 2007. Rich-club phenomenon across complex network hierarchies. Applied Physics Letters 91:084103.

Junker, B. H., and F. Schreiber. 2008. Analysis of biological networks. John Wiley and Sons, Hoboken, New Jersey, USA.

Kitts, J. A., and J. Huang. 2010. Triads. Pages 873–874 *in* G. A. Barnett, editor. Encyclopedia of social networks. Volume 2. Sage Publications, New York, New York, USA.

Lau, M. K., S. R. Borrett, B. Baiser, N. J. Gotelli, and A. M. Ellison. 2017. Ecological network metrics: opportunities for synthesis. Ecosphere 8:e01900.

Lecca, P., and A. Re. 2015. Detecting modules in biological networks by edge weight clustering and entropy significance. Frontiers in Genetics 6:265.

Lee, D. S., S. E. Maeng, and J. W. Lee. 2012. Scaling of nestedness in complex networks. Journal of the Korean Physical Society 60:648–656.

Lloyd, M. 1967. Mean crowding. Journal of Animal Ecology 36:1–30.

Lloyd, M. 1986. This week's citation classic. ISI current contents.

Lloyd-Price, J., G. Abu-Ali, and C. Huttenhower. 2016. The healthy human microbiome. Genome Medicine 8:51.

Lozupone, C. A., J. I. Stombaugh, J. I. Gordon, J. K. Jansson, and R. Knight. 2012. Diversity, stability and resilience of the human gut microbiota. Nature 2012:220–230.

Ma, Z. S. 2012. A unified definition for reliability, survivability and resilience inspired by the handicap principle and ecological stability. International Journal of Critical Infrastructures 8:242–272.

Ma, Z. S. 2015. Power law analysis of the human microbiome. Molecular Ecology 24:5428–5445.

Ma, Z. S. 2017. The P/N (positive-to-negative links) ratio in complex networks—a promising *in silico* biomarker for detecting changes occurring in the human microbiome. Microbial Ecology 74:1063–1073.

Ma, Z. S., and A. M. Ellison. 2017. A new dominance metric and its application to diversity–stability analysis. Ecosphere (Accepted). https://arxiv.org/abs/1703.08835

Ma, Z. S., and L. W. Li. 2017. Quantifying the human vaginal community state types (CSTs) with the species specificity index. PeerJ 5:e3366.

Ma, Z. S., and D. D. Ye. 2017. Trios—promising *in silico* biomarkers for differentiating the effect of disease on the human microbiome network. Scientific Reports, 7:Article No. 13259.

Ma, Z. S., Z. Abdo, and L. J. Forney. 2011. Caring about trees in the forest: incorporating frailty in risk analysis for personalized medicine. Personalized Medicine 8:681–688.

Ma, B., L. J. Forney, and J. Ravel. 2012*a*. The vaginal microbiome: rethinking health and disease. Annual Review of Microbiology 66:371–389.

Ma, Z. S., J. W. Geng, Z. Abdo, and L. J. Forney. 2012*b*. A bird's eye view of microbial community dynamics. Pages 57–70 *in* L. A. Ogilvie and P. R. Hirsch, editors. Microbial ecological theory: current perspectives. Caister Academic Press, Norfolk, UK.

Ma, Z. S., Q. Guan, C. Ye, C. Zhang, J. Forster, and L. J. Forney. 2015. Network analysis suggests a potentially 'evil' alliance of opportunistic pathogens inhibited by a cooperative network in human milk bacterial communities. Scientific Reports 5:8275.

Ma, Z. S., L. W. Li, W. Li, J. Li and H. J. Chen. 2016*a*. Integrated network-diversity analyses suggest suppressive effect of Hodgkin's lymphoma and slightly relieving effect of chemotherapy on human milk microbiome. Scientific Reports 6:Article Number 28048.

Ma, Z. S., C. C. Zhang, Q. P. Zhang, J. Li, L. Li, L. Qi, and X. Yang. 2016*b*. A brief review on the ecological network analysis with applications in the emerging medical ecology. *In* Hydrocarbon and lipid microbiology protocols. Springer. https://link.springer.com/protocol/10.1007/8623_2016_204

Maslov, S., and K. Sneppen. 2002. Specificity and stability in topology of protein networks. Science 296:910–913.

May, R. M. 1973. Stability and complexity in model ecosystems. Princeton University Press, Princeton, New Jersey, USA.

McCann, K. S. 2000. The diversity-stability debate. Nature 405:228–233.

Newman, M. E. J. 2002. Assortative mixing in networks. Physical Review Letters 89:208701.

Newman, M. E. J., and A. Clauset. 2016. Structure and inference in annotated networks. Nature Communications 7:11863.

O'Malley, A. J., and P. V. Marsden. 2008. The analysis of social networks. Health Services and Outcomes Research Methodology 8:222–269.

Pastor, J. 2008. Mathematical ecology of populations and ecosystems. Wiley-Blackwell, Hoboken, New Jersey, USA.

Pimm, S. L. 1984. The complexity and stability of ecosystems. Nature 307:321–326.

R Core Team. 2015. R: a language and environment for statistical computing. The R Project for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Ravel, J., et al. 2011. Vaginal microbiome of reproductive-age women. Proceedings of the National Academy of Sciences USA 108(Suppl. 1):4680–4687.

Ravel, J., et al. 2013. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. Microbiome 2013:29.

Rombach, M. P., M. A. Porter, J. H. Fowler, and P. J. Mucha. 2014. Core-periphery structure in networks. SIAM Journal of Applied Mathematics 74:167–190.

Scheffer, M., et al. 2012. Anticipating critical transitions. Science 338:344.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, n. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13:2498–2504.

Simpson, E. H. 1949. Measurement of diversity. Nature 163:688.

Sobel, J. D. 1999. Is there a protective role for vaginal flora? Current Infectious Disease Reports 1:379–383.

Srinivasan, S., C. Liu, C. M. Mitchell, T. L. Fiedler, K. K. Thomas, K. J. Agnew, J. M. Marrzaao, and D. N. Fredricks. 2010. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. PLoS ONE 5:e10197.

Sunagawa, S., L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, and G. Salazar. 2015. Structure and function of the global ocean microbiome. Science 348:1261359-1–1261359-9. https://doi.org/10.1126/science.1261359

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. 2007. The human microbiome project. Nature 449:804–810.

Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. Nature 393:440–442.

White, B. A., D. J. Creedon, K. E. Nelson, and B. A. Wilson. 2011. The vaginal microbiome in health and disease. Trends in Endocrinology and Metabolism 22:389–393.

Zhou, S., and R. J. Mondragon. 2004. The rich-club phenomenon in the Internet topology. IEEE Communication Letters 8:180–182.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecm.1358/full