Phylogenetic Molecular Ecological Network of Soil Microbial Communities in Response to Elevated CO_2

Jizhong Zhou,[a,b,c] Ye Deng,[b,d] Feng Luo,[e] Zhili He,[b] and Yunfeng Yang[a]

State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China[a]; Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma, USA[b]; Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA[c]; Glomics, Inc., Norman, Oklahoma, USA[d]; and School of Computing, Clemson University, Clemson, South Carolina, USA[e]

ABSTRACT **Understanding the interactions among different species and their responses to environmental changes, such as elevated atmospheric concentrations of CO_2, is a central goal in ecology but is poorly understood in microbial ecology. Here we describe a novel random matrix theory (RMT)-based conceptual framework to discern phylogenetic molecular ecological networks using metagenomic sequencing data of 16S rRNA genes from grassland soil microbial communities, which were sampled from a long-term free-air CO_2 enrichment experimental facility at the Cedar Creek Ecosystem Science Reserve in Minnesota. Our experimental results demonstrated that an RMT-based network approach is very useful in delineating phylogenetic molecular ecological networks of microbial communities based on high-throughput metagenomic sequencing data. The structure of the identified networks under ambient and elevated CO_2 levels was substantially different in terms of overall network topology, network composition, node overlap, module preservation, module-based higher-order organization, topological roles of individual nodes, and network hubs, suggesting that the network interactions among different phylogenetic groups/populations were markedly changed. Also, the changes in network structure were significantly correlated with soil carbon and nitrogen contents, indicating the potential importance of network interactions in ecosystem functioning. In addition, based on network topology, microbial populations potentially most important to community structure and ecosystem functioning can be discerned. The novel approach described in this study is important not only for research on biodiversity, microbial ecology, and systems microbiology but also for microbial community studies in human health, global change, and environmental management.**

IMPORTANCE **The interactions among different microbial populations in a community play critical roles in determining ecosystem functioning, but very little is known about the network interactions in a microbial community, owing to the lack of appropriate experimental data and computational analytic tools. High-throughput metagenomic technologies can rapidly produce a massive amount of data, but one of the greatest difficulties is deciding how to extract, analyze, synthesize, and transform such a vast amount of information into biological knowledge. This study provides a novel conceptual framework to identify microbial interactions and key populations based on high-throughput metagenomic sequencing data. This study is among the first to document that the network interactions among different phylogenetic populations in soil microbial communities were substantially changed by a global change such as an elevated CO_2 level. The framework developed will allow microbiologists to address research questions which could not be approached previously, and hence, it could represent a new direction in microbial ecology research.**

Received 10 June 2011 **Accepted** 29 June 2011 **Published** 26 July 2011

Citation Zhou, J, Deng Y, Luo F, He Z, Yang Y. 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO_2. mBio 2(4):e00122-11. doi:10.1128/mBio.00122-11.

Editor David Relman, VA Palo Alto Health Care System

Address correspondence to Jizhong Zhou, jzhou@ou.edu.

J.Z. and Y.D. contributed equally to this article.

The global atmospheric concentration of CO_2 has increased by more than 30% since the industrial revolution due to fossil fuel combustion and land use changes (1). Many previous studies demonstrated that elevated CO_2 (eCO_2) stimulates plant growth and primary productivity (2–4), but the influences of eCO_2 on belowground microbial communities are poorly understood and controversial (5–8). It is expected that eCO_2 will alter microbial community composition and structure by increasing soil carbon input from plants and modifying soil chemical compositions (9). Recently, using metagenomic technologies, including high-throughput sequencing, functional gene microarrays, and 16S rRNA gene-based phylogenetic arrays, we found that the phylogenetic and functional structure of soil microbial communities was substantially altered by eCO_2 (10). However, it is less clear whether eCO_2 also alters interactions among different microbial phylogenetic groups/populations.

Biodiversity includes not only the number of species and their abundance but also the complex interactions among different species (11). Within habitats, various biological species/populations interact with each other through the flow of energy, matter, and

information to form large, complex ecological networks (12). Explaining and predicting such interactive network structures, dynamics, and the underlying mechanisms are essential parts of any study of biodiversity, and hence, ecological networks of biological communities have received great attention in plant and animal ecology (12–15) but only very recently in microbial ecology (16, 17). However, determining network structures and their relationships to environmental changes in microbial communities is a significant challenge (18). The availability of massive, community-wide, replicated metagenomic data under different environmental conditions provides an unprecedented opportunity to analyze the network interactions in a microbial community (17). Recently, we developed the random matrix theory (RMT)-based approach to delineate the network interactions among different microbial functional groups/populations based on GeoChip hybridization data (18). Our results indicated that the RMT-based network approach is very useful in defining the network interactions among different functional groups/populations with GeoChip hybridization data. Our results also revealed that eCO_2 has significant impacts on network interactions among different microbial functional groups/populations (18). However, its suitability for dealing with next-generation sequencing data remains unclear.

The objective of this study is to develop a novel RMT-based bioinformatic approach to define and characterize ecological networks in microbial communities based on high-throughput metagenomic sequencing data. We will focus on the following two questions. (i) How are phylogenetic ecological networks in microbial communities determined based on metagenomic sequencing data, and (ii) do human-induced factors, such as eCO_2, have an impact on the network structure of different phylogenetic groups in microbial communities? We hypothesize that the RMT-based network approach will be powerful in delineating phylogenetic molecular ecological networks (pMENs) in microbial communities using high-throughput sequencing data and that eCO_2 will dramatically alter the network interactions among different phylogenetic groups/populations. We tested these hypotheses with the experimental data from microbial communities under both ambient CO_2 (aCO_2) and eCO_2 in a long-term free-air carbon dioxide enrichment experimental facility at the Cedar Creek Ecosystem Science Reserve in Minnesota (4, 19). This facility was designed to examine the interactive effects of multiple climate factors, e.g., biodiversity, CO_2, and nitrogen deposition (Biocon), on grassland ecosystems.

RESULTS

pMENs. An ecological network is a representation of various biological interactions (e.g., predation, competition, and mutualisms) in an ecosystem in which species (nodes) are connected by pairwise interactions (links) (12, 20–23). Since microorganisms are invisible to the naked eye and the majority of them are not yet cultivated, their detection often relies on molecular markers, including rRNA genes, various functional genes (e.g., *nifH*, *amoA*, *nirS*, and *dsrA*), and intergenic regions. The abundance of each gene marker in a sample is often determined based on the number of sequences, hybridization signal intensity, or PCR amplification band intensity. The gene richness and abundance determined are then used to describe the compositions and structures of microbial communities of interest. Based on such experimental data, a network graph can be constructed to illustrate the ecological interactions (edges) of different gene markers/populations (nodes) in a microbial community (16). We refer to such a network graph as an ecological network in general. As a result, an ecological network generated from a microbial community is actually based on individual genes rather than individual species. As previously described, we refer to such molecule-based ecological networks in microbial communities as molecular ecological networks (MENs) (18) in which different nodes are linked by lines (i.e., interactions) and to the MENs derived from phylogenetic gene markers as pMENs (18).

RMT-based approach for pMEN construction. Based on a modification of the previous procedure for constructing functional MENs (fMENs) (18), the framework for constructing pMENs can be divided into nine key steps, which include metagenomic sequence collection, data standardization, Pearson correlation estimation, adjacency matrix determination by an RMT-based approach, network characterization, module detection, eigengene network analysis, network comparison, and association of network properties with ecological functional traits (18). Eigengene analysis is important for revealing higher-order organization and identifying key populations based on network topology (24–26).

Based on the adjacency matrix, a network graph is constructed to represent positive or negative interactions among different operational taxonomic units (OTUs). Since the pairwise correlations of OTU abundance across different samples are used to define the adjacency matrix, the pMENs constructed here, in fact, describe the co-occurrence of OTUs across different samples. Thus, positive interactions (positive correlations) indicate that the abundance of OTUs changes along the same trend, whereas negative interactions (negative correlations) signify that the abundance of OTUs changes in the opposite direction.

pMENs in grassland soil microbial communities. We analyzed the phylogenetic diversity data of grassland soil microbial communities under aCO_2 and eCO_2 (10) using the RMT-based network method. The phylogenetic diversity of these communities (12 replicate samples each from aCO_2 and eCO_2, respectively) was determined by barcode-based amplicon pyrosequencing of 16S rRNA genes (10). After data preprocessing, 343 and 358 OTUs remained in the eCO_2 and aCO_2 data sets, respectively (Table 1). Following the RMT-based network analysis outlined above, very close similarity thresholds (s_t) were obtained for eCO_2 (0.78) and aCO_2 (0.77) (Table 1). Applying these thresholds, two networks of similar sizes for eCO_2 (263 nodes) and aCO_2 (292 nodes) were constructed (Table 1).

General features of many complex networks are scale free, small world, modular, and hierarchical (27–29). Four commonly used complementary network indexes can be used to describe network difference: (i) connectivity or degree of distribution, which is the number of links of a node to other nodes; (ii) geodesic distance or path length, which is the shortest path between two nodes; (iii) the clustering coefficient, which describes how well a node is connected with its neighbors; and (iv) modularity, which measures the degree to which the network was organized into clearly delimited modules. For the first three indexes, the means over a number of nodes can be calculated to describe the overall features of the entire network.

The degrees of distribution (i.e., connectivity) in all of the constructed pMENs were fitted with three models: (i) power law (scale free), (ii) truncated power law (broad scale), and (iii) expo-

TABLE 1 Topological properties of the empirical pMENs of microbial communities under eCO_2 and aCO_2 and their associated random pMENs

	Empirical networks							Random networks[c]		
Condition	No. of original OTUs[a]	s_t	Network size[b]	Avg connectivity	Avg geodesic distance	Avg clustering coefficient	Modularity (no. of modules)	Avg geodesic distance ± SD	Avg clustering coefficient ± SD	Avg modularity ± SD
eCO_2	343	0.78	263	3.10	3.95[d]	0.25[d]	0.81[d] (34)	3.98 ± 0.22	0.015 ± 0.006	0.61 ± 0.02
aCO_2	358	0.77	292	3.06	4.26[d]	0.27[d]	0.85[d] (36)	4.10 ± 0.20	0.017 ± 0.005	0.59 ± 0.01

[a] The number of OTUs was originally used for network construction by the RMT-based network approach.
[b] The number of OTUs (i.e., nodes) in a network.
[c] The random networks were generated by rewiring all of the links of a pMEN with the identical numbers of nodes and links to the corresponding empirical pMEN.
[d] Significant difference ($P < 0.001$) between aCO_2 and eCO_2 values.

nential distributions (i.e., faster-decaying functions) (30). Although the degrees of distribution fit well with all three models, truncated power law had the best fit, with determinant coefficients of 0.97 and 0.98 for eCO_2 and aCO_2, respectively (see Fig. S1 in the supplemental material), which are consistent with many mutualistic ecological networks (30). Also, significant differences between these two pMENs and their corresponding random networks with identical network sizes and average numbers of links were observed in terms of the average geodesic distance, average clustering coefficient, and modularity (Table 1). These network properties were comparable to those of other networks displaying small-world behavior (31), indicating that the pMENs obtained possessed typical small-world characteristics. All of the above results suggest that, similar to other complex systems, the constructed pMENs appeared to be, at least approximately, scale free, small world, and modular.

The overall network structure of pMENs is not preserved under aCO_2 and eCO_2. Although the network sizes of the pMENs obtained were very similar under eCO_2 and aCO_2, substantial differences were observed in terms of network composition. These two pMENs shared 44.5% (171) of their nodes, but all of the shared OTUs showed no significant correlations of the connectivity values ($r = 0.123$, $P = 0.109$). In addition, the average geodesic distance, average clustering coefficient, and modularity (as well as many other network parameters) of these pMENs were significantly different between aCO_2 and eCO_2 (Table 1). These results indicated that the network composition and structure were not conserved between eCO_2 and aCO_2.

The majority of the nodes in these two networks belonged to 10 phyla, but their distribution varied substantially among different phylogenetic groups (see Fig. S2 in the supplemental material). Among them, *Actinobacteria*, *Alphaproteobacteria*, and *Acidobacteria* were more dominant (Table S1; Fig. S2). Also, for most phylogenetic groups, the relative proportions of OTUs were not obviously different under eCO_2 and aCO_2. However, under eCO_2 the relative proportions of *Acidobacteria* in the network decreased considerably, while the percentages of *Alphaproteobacteria*, *Betaproteobacteria*, and *Bacteriodetes* increased substantially (Fig. S2). In addition, eCO_2 had differential impacts on the network structures of different phylogenetic groups. Two typical examples are related to *Actinobacteria* and *Verrucomicrobia*, the former of which plays an important role in the decomposition of organic materials and the production of secondary metabolites with very diverse physiology, but the physiological and ecological roles of the latter are not well known. The top 10 *Actinobacteria* OTUs under eCO_2 (Fig. 1A) had more complex interactions than their corresponding OTUs under aCO_2 (Fig. 1B), while *Verrucomicrobia* had much less complex interactions under eCO_2 (see Fig. S3B in the supplemental material) than under aCO_2 (Fig. S3A). It appears that eCO_2 selected for *Actinobacteria* but against *Verrucomicrobia*. Altogether, the above results suggested that the interactions among different microbial taxa in the grassland microbial communities were substantially changed by eCO_2. However, such impacts varied considerably among different microbial groups.

Modules of the pMENs under aCO_2 and eCO_2 are even less preserved. In the pMENs examined here, a module is a group of OTUs which are well connected among themselves but are less linked with OTUs belonging to other modules. A module in a pMEN could indicate that the microbial populations within it may have similar ecological niches. In this study, modules were detected by fast greedy modularity optimization (32). While a total of 11 modules with ≥5 nodes were obtained for the networks under eCO_2, the pMEN under aCO_2 had 14 modules with >5 nodes (see Fig. S4A and C in the supplemental material). The module sizes varied considerably, ranging from 5 to 39 nodes (Fig. S4B and D).

Distinct individual modules were observed for the pMENs under eCO_2 (Fig. S4A) and aCO_2 (Fig. S4C). It appears that the relationships between phylogenetic relatedness and ecological relatedness are complicated, depending on individual microbial groups, as well as environmental conditions (16). For instance, most of the members of *Verrucomicrobia*, a recently described phylum of abundant soil bacteria with a few described species, were found to be in the same module (9 of 12 OTUs) under aCO_2 (module A13, orange, Fig. S4C) and 7 of 9 were found to be in the same module under eCO_2 (module E9, orange, Fig. S4A). Also, the interactions among them (26 and 13 links for aCO_2 and eCO_2) in this network were far more significant ($P < 0.01$) than those (14 and 5 links) predicted by chance based on their corresponding random networks. These results suggested that these *Verrucomicrobia* might have similar ecological niches. However, the *Actinobacteria* under both aCO_2 (81 nodes) and eCO_2 (76 nodes) were spread over all major modules (Fig. S4A and C, green), consistent with the fact that this group of bacteria has very diverse physiology and could occupy different ecological niches. Furthermore, many of the actinobacterial OTUs are very closely related phylogenetically, each with unique combinations of relationships to other microorganisms and network parameters (data not shown), indicating that they could represent different "ecological species" or ecotypes (16) at this site.

Since the node composition is substantially different among different modules under eCO_2 and aCO_2, Fisher's exact test was

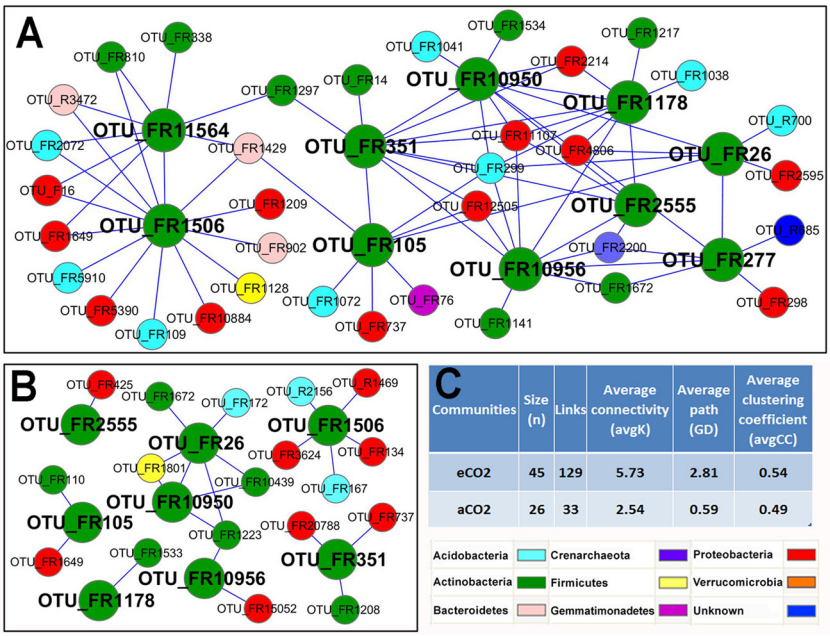

FIG 1 Effects of eCO_2 on the network interactions of *Actinobacteria*. (A) Network interactions of the top 10 OTUs of *Actinobacteria* with the highest connectivities under eCO_2. (B) Network interactions of the corresponding OTUs of *Actinobacteria* under aCO_2. (C) Summary of several key parameters of network topology. Since many *Actinobacteria* were observed, only the top 10 OTUs of this group under eCO_2 are presented here. Two of these OTUs (FR385 and FR4675) were not observed under aCO_2. GD, geodesic distance.

used to statistically pair various modules between eCO_2 and aCO_2. A total of 5 module pairs (40%) were obtained from these two networks (see Table S2 in the supplemental material), but the majority of the modules (60%) could not be paired together. These paired modules contained 37% of the total nodes in these two networks. Within the paired modules, only 12.2% (25/205) of the total nodes shared between these two networks were identical. These results indicated that these two networks are poorly conserved at the modular level.

Eigengene network analysis reveals dramatic impacts of eCO_2 on the higher-order network structure. To reveal the higher-order organization of the constructed pMENs, eigengene network analysis (24–26) was performed. In this analysis, each module is summarized through singular value decomposition analysis with a single representative abundance profile, which is referred to as the module eigengene. A conceptual example of eigengene network analysis for a module is illustrated in Fig. 2. Eigengene network analysis comprises several key components: (i) a heat map showing the relative abundances of individual OTUs within a module; (ii) eigengene, showing a representation of the abundance profile; (iii) module membership, showing key OTUs within a module; (iv) module visualization, showing the interactions among different OTUs; and (v) a phylogenetic tree showing relationships among the different OTUs within a module.

In this study, the module eigengenes explained 35 to 79% of the variations in relative OTU abundance across different samples under eCO_2 and 43 to 79% of that under aCO_2 (Fig. 2, module E9; see the supplemental material for the others). Most of the eigengenes (18/25) explained more than 50% of the variations observed, similar to observations from human eigengene network analysis (33). These results suggest that these eigengenes relatively well represent the changes in OTUs across different samples in individual modules.

Eigengenes from different modules often showed considerable correlations, and such correlations (Pearson) were used to define the eigengene network (25). The relationships among eigengenes are usually visualized by average-linkage hierarchical analysis as a clustering dendrogram (25). Groups of eigengenes in this dendrogram are referred to as metamodules in an eigengene network, which describes a higher-order structure of the constructed network. In this study, the eigengenes from many modules showed significant correlations. A total of 4 metamodules were clustered for both aCO_2 and eCO_2 networks (see Fig. S5 in the supplemental material). However, the eigengenes from the four paired modules were clustered differently with other eigengenes (Fig. S5), indicating that the higher-order organization of the paired modules was not preserved between eCO_2 and aCO_2 either.

To determine the extent to which an OTU is associated with a module, module membership was evaluated, which is the square of the Pearson correlation between the abundance profile of a given OTU and a given module eigengene (25). Most of the OTUs had significant module memberships with their respective modules (Fig. 2, module E9; see the supplemental material for the others). However, for the conserved nodes, divergent patterns of module memberships were observed among these four paired modules (see Fig. S6 in the supplemental material). While module pairs 1, 4, and 5 had significantly positive relationships (rho = 0.18 to 0.39, $P < 0.001$), two other module pairs (pairs 2 and 3) had no correlations on module memberships (Fig. S6). These results indicated that eCO_2 also significantly altered the topological positions of individual OTUs.

Visualization of topological roles of individual nodes reveals differential impacts of eCO_2 on key microbial populations. Topologically, different OTUs (nodes) play distinct roles in the network (34). The topological roles of different OTUs can be described by two parameters. One is within-module connectivity (Z_i), which describes how well a node is connected to other nodes within its own module. The other is connectivity among modules (P_i), which reflects how well a node connects to different modules. According to the simplified classification used in pollination networks (11), the nodes in a network are divided into the following four subcategories: (i) peripheral nodes, which have low Z and P values (i.e., they have only a few links and almost always to the species within their modules); (ii) connectors, which have a low Z but a high P value (i.e., these nodes are highly linked to several modules); (iii) module hubs, which have a high Z but a low P value (i.e., they are highly connected to many species in their own modules); and (iv) network hubs, which have high Z and P values (i.e., they act as both module hubs and connectors). From an ecological

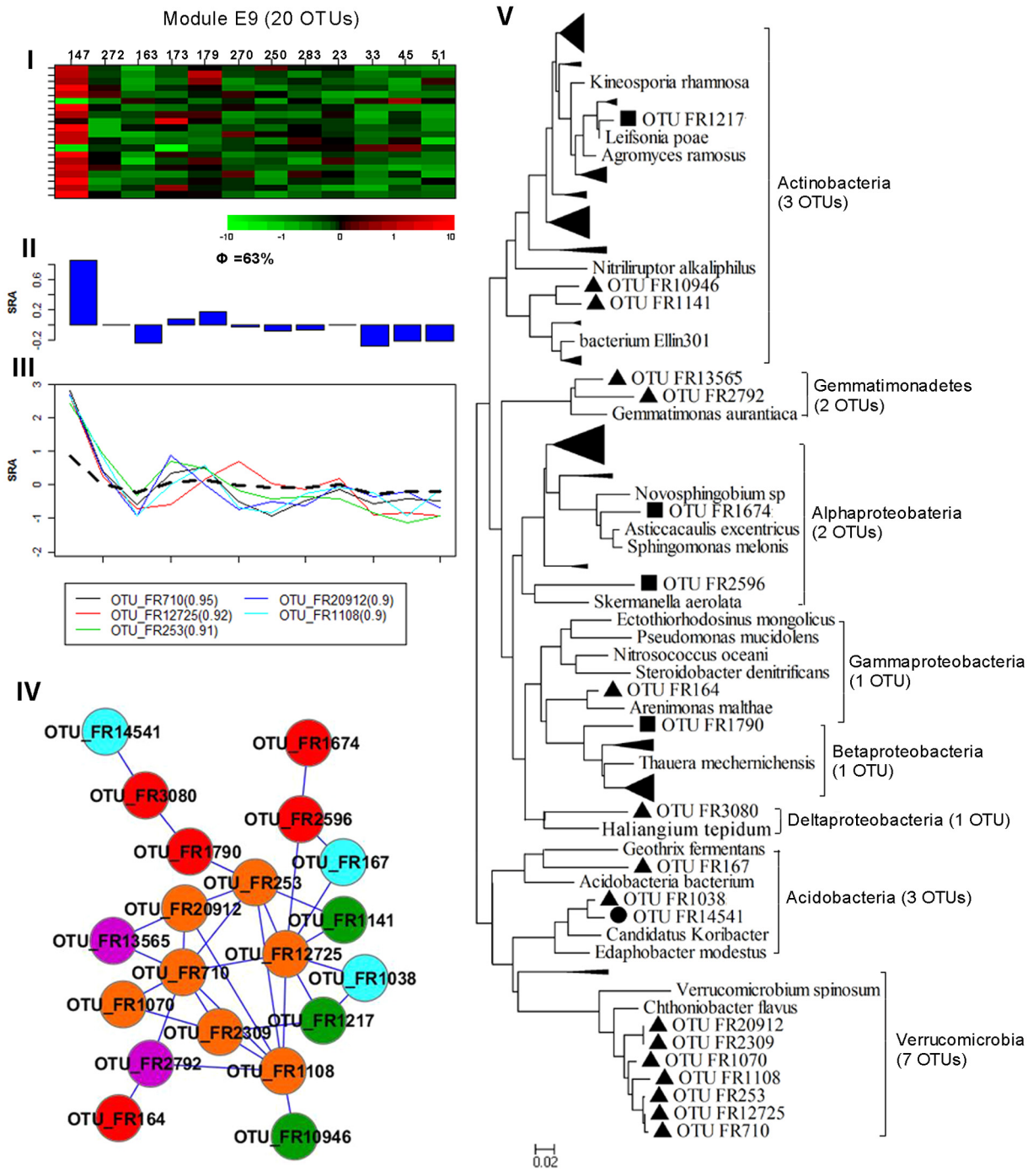

FIG 2 Conceptual example of eigengene network analysis with module E9 under eCO_2. (I) Heat map of the standardized relative abundance (SRA) of OTUs across different samples. Rows correspond to individual OTUs in the module, whereas columns are the samples. The number above each column is the experimental plot number in the Biocon experiment. Red corresponds to the OTUs whose SRAs are >0, and green signifies those whose SRAs are <0. (II) SRA of the corresponding eigengene (*y* axis) across the samples (*x* axis). The parameter Φ indicates the percentage of the total variance explained by the eigengene. (III) Module memberships identify groups of OTUs that consistently coexist in these microbial communities. Only 5 OTUs with signifcant module memberships are shown here, where the *y* axis is SRAs and the *x* axis is individual samples. The values in parentheses are module memberships. (IV) Module visualization showing the interactions among different OTUs within the module. Blue line, positive interactive (positive correlation); red line, negative interaction (negative correlation). The different colors of the shading of nodes represent different phylogenetic groups. (V) Phylogenetic tree showing the relationships of the OTUs observed in the corresponding modules. The tree was constructed by the neighbor-joining approach with 1,000 bootstrap values. Due to space limitation, bootstrap percentages are not shown on the tree. The symbols before individual OTUs signify different features of nodes in the module. The symbol ▲ indicates that the OTU exists in both aCO_2 and eCO_2 networks with significant module memberships, ■ indicates that the OTU has significant module membership but is not shared by the corresponding network under aCO_2, ● indicates that the OTU is shared but without significant module membership, while ▼ indicates that the OTU is not shared and has no significant module membership. The eigengene analysis figures of all other modules under aCO_2 and eCO_2 and a detailed description of each module can be downloaded through http://ieg.ou.edu/4download/.

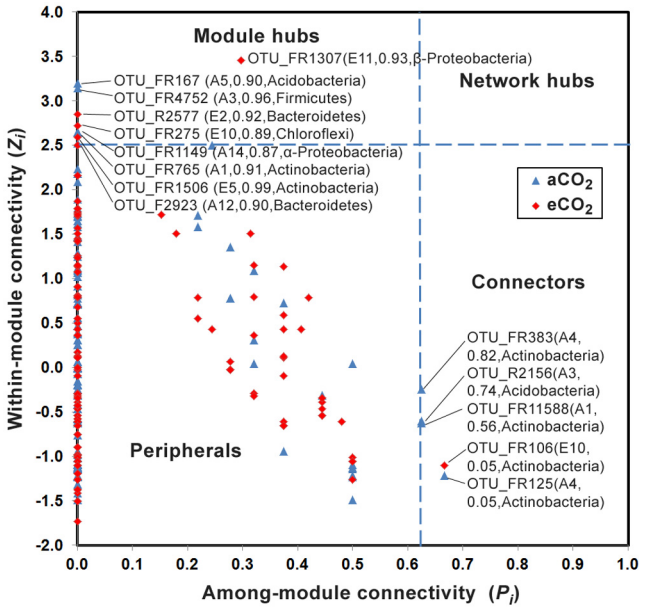

FIG 3 *Z-P* plot showing the distribution of OTUs based on their topological roles. Each symbol represents an OTU under eCO_2 (red) or aCO_2 (blue). The topological role of each OTU was determined according to the scatter plot of within-module connectivity (Z_i) and among-module connectivity (P_i) (11, 34). The module hubs and connectors are labeled with OTU numbers. In parentheses are the module number, module membership, and phylogenetic associations.

perspective, peripheral nodes represent specialists whereas module hubs and connectors are similar to generalists. Network hubs are supergeneralists (11).

The topological roles of the OTUs identified in these two networks are shown as a *Z-P* plot in Fig. 3. The majority (97.5%) of the OTUs were peripherals with most of their links inside their modules. Among these peripherals, 89% even had no links at all outside their own modules (i.e., $P_i = 0$). About 2.5% of the OTUs were generalists, including 1.6% that were module hubs and 0.9% that were connectors. However, no network hubs (supergeneralists) were observed in these two networks. Two (OTUs FR765 and FR1506) of the nine module hubs under eCO_2 belonged to *Actinobacteria* (Fig. 3) that are closely related to *Ferrithrix thermotolerans* and *Ilumatobacter fluminis*, respectively. Two module hubs (OTUs R2577 and F2923) could be assigned to *Bacteroidetes*, while the others belonged to different major taxa (i.e., *Alpha-*,and *Betaproteobacteria*, *Firmicutes*, *Chloroflexi*, and *Acidobacteria*). All module hubs were from different modules and had significant module memberships with their respective module eigengenes. Interestingly, four of the five connector OTUs (FR383, FR11588, FR106, and FR125) were *Actinobacteria* that are closely related to *Streptosporangium roseum*, *Ferrithrix thermotolerans*, *Friedmanniella lacustris*, and *Rubrobacter* sp., respectively. The other connector (OTU R2156) was derived from *Acidobacteria* close to *Edaphobacter modestus*. In addition, for the shared OTUs, no significant relationships were observed for *Z* values ($r = 0.06$, $P = 0.44$) under eCO_2 and aCO_2. These results also suggested that eCO_2 greatly altered the network structure and topological roles of individual OTUs and key microbial populations.

Association of network structure with environmental characteristics. Similar to our previous network study based on GeoChip hybridization data (18), the relationships between microbial network interactions and soil properties were established with Mantel tests. We used the trait-based OTU significance measure (24), which is referred to as the square of the correlation between the signal intensity of an OTU and each soil variable, to determine a common subgroup of soil properties important to network interactions. Under eCO_2, very strong correlations were observed between the connectivity and the OTU significance of the selected soil variables based on all detected OTUs ($P = 0.001$) or several phylogenetic groups such as *Actinobacteria* ($P = 0.001$), *Bacteroidetes* ($P = 0.012$), *Alphaproteobacteria* ($P = 0.001$), and *Betaproteobacteria* ($P = 0.044$) under eCO_2 (see Table S3 in the supplemental material). Under aCO_2, the connectivities of *Actinobacteria* ($P < 0.05$) were also significantly correlated with the OTU significance of the selected soil variables. All of the results together suggested that the network interactions among different phylogenetic groups/populations were dramatically shifted by eCO_2 and that such changes in network interactions are significantly related to soil properties.

DISCUSSION

Metagenomics is a rapidly developing emerging scientific field which generates tremendous amounts of experimental data via high-throughput sequencing technologies. However, it comes with the two-part challenge of how to handle these vast quantities of data and how to use such information to further address biological questions, aiming to understand community level functional processes. In this study, we describe a novel framework and approach for discerning network interactions using high-throughput sequencing-based metagenomic data. The approach developed would allow microbiologists to address research questions (network interactions) which could not be approached previously and thus should represent a research paradigm shift in metagenomic analysis.

In this study, the pairwise correlations of relative OTU abundance across different samples were used to delineate an adjacency matrix for network construction. Based on this adjacency matrix, a network graph was constructed to represent positive or negative interactions among different OTUs. Thus, a network connection between two OTUs in fact describes the co-occurrence of these two OTUs across different samples but not necessarily their physical interactions. In other words, both OTUs might be responding to a common environmental parameter rather than interacting directly.

Compared to other Pearson correlation-based relevance network approaches (35–37), the network approach described here has several advantages (18, 38). First, this approach was developed based on the two universal laws of RMT, and thus, it should be suitable for various biological systems (e.g., cells, populations, communities, and ecosystems). Theoretically, the results obtained with an RMT approach should be more robust and consistent and should more accurately reflect the nature of the complex systems under study. Second, the majority of relevance network analysis methods define the adjacency matrix for network construction using arbitrary thresholds based on known biological information (28, 35–37, 39). As a result, the networks obtained vary with the thresholds selected. However, it is a great challenge in selecting an appropriate threshold for network construction, especially for poorly studied organisms and/or microbial communities. In contrast, the novel RMT-based approach developed here automati-

cally defines thresholds for network construction and hence no ambiguity exists for the networks constructed. Moreover, RMT is useful in removing noise from nonrandom, system-specific features, and hence the networks identified should be more accurate and reliable (18, 38). This is particularly important for dealing with high-throughput metagenomic data because such data generally have an inherently high noise level.

The identification and characterization of OTU co-occurrence modules represent a new approach for detecting the interactions of microbial populations in a community. Based on the oft-invoked principle of guilt by association (26), the abundance changes in the microbial populations with strong module memberships are probably driven by the same underlying factors. Thus, it is reasonable to hypothesize that the microbial populations with strong module memberships are physically and/or functionally associated in a microbial community. This hypothesis has important implications not only for our understanding of the interactions and ecological functions of the known cultivated microorganisms but also for predicting the potential ecological roles of as-yet-uncultivated microorganisms. As shown in this study, the modularity, module memberships, topological roles, interaction patterns (positive, zero, or negative), and phylogenetic relationships of individual OTUs are rich sources of new hypotheses for identifying key microbial populations and for understanding their interactions and ecological roles in grassland microbial communities.

Identification of keystone populations is a critical issue in ecology, but it is very difficult to achieve, especially in microbial communities given their extreme complexity, high diversity, and uncultivated status. As demonstrated in this study, key populations could be identified based on network topology, module memberships, and/or their relationships to ecosystem functional traits. The conceptual framework developed in this study could provide important information on candidate genes/populations most important to certain ecosystem processes and functioning. This could be particularly important in ecosystem modeling studies in which microbial community structure must be appropriately simplified prior to their incorporation into ecosystem models.

Knowledge of the responses of biological communities to eCO_2 and their mechanisms is critical for projecting future climate change (6, 8). In this study, we demonstrated the impacts of eCO_2 on the network interactions among different phylogenetic groups/populations based high-throughput metagenomic sequencing data and the relationships between network structure and soil properties. It is obvious that the network interactions among different microbial phylogenetic groups/populations are greatly affected by eCO_2 in this grassland ecosystem. These results are consistent with our previous study of fMENs (18) and other studies of macroecology (40). To the best of our knowledge, this is the first study to document the changes in network interactions among different phylogenetic groups/populations of microbial communities in response to eCO_2.

The relationship between biodiversity and ecosystem functioning has emerged as a central issue in ecological and environmental sciences (41–47) and is one of the great challenges of the 21st century's sciences (48). Traditionally, almost all biodiversity studies in microbial ecology consider just species richness and abundance and ignore the interactions among different microorganisms. However, network interactions could be more important to ecosystem processes and functions than species diversity (parts list). In this study, we developed a novel conceptual framework for determining network interactions among different phylogenetic groups/populations in microbial communities based on high-throughput metagenomic sequencing data. This novel framework will allow microbial ecologists to examine research issues beyond microbial species richness and abundance. The developed pMEN framework and information on the responses of network structure to eCO_2 should have a profound impact on the study of biodiversity, ecosystem ecology, systems microbiology, and climate change.

MATERIALS AND METHODS

In this study, 24 soil samples used for network analysis of microbial communities were collected from the Biocon (biodiversity, CO_2, and N) experimental site located at the Cedar Creek Ecosystem Science Reserve in Minnesota (45°N, 93°W). Of these 24 samples, 12 were from aCO_2 replicate plots and 12 were from eCO_2 replicate plots. All of the plots contained 16 species without additional N supply. The soil samples were collected in July 2007, and each sample was a composite of five soil cores from depths of 0 to 15 cm (10).

Two MENs were constructed with the following steps. First, the experimental data used for constructing pMENs were generated by pyrosequencing of 16S rRNA genes (10). Since the sequence numbers of individual OTUs obtained varied significantly among different samples, the relative proportions of sequence numbers were used for subsequent Pearson correlation analysis. Second, a similarity matrix was obtained by taking the absolute values of the correlation matrix. This similarity matrix measures the degree of concordance between the abundance profiles of individual OTUs across different samples. Third, an appropriate threshold for defining network structure, s_t, is defined using the RMT-based network approach (38, 49) to obtain an adjacency matrix, which encodes the strength of the connection between each pair of nodes. Fourth, the submodules within a large module were detected by fast greedy modularity optimization (32). In addition, for network comparison, random networks corresponding to all pMENs were generated using the Maslov-Sneppen procedure (50) and keeping the numbers of nodes and links constant but rewiring all of the links' positions in the pMENs. A standard Z or *t* test was employed to determine the significance of network indexes between the pMENs and random networks and across different experimental conditions. Finally, sample trait-based significance (24) was defined and a Mantel test was used to examine the relationships between the trait-based gene significance and soil variables for understanding the importance of network interactions in ecosystem functioning. More detailed information about the Materials and Methods used in this study is provided in the supplemental material.

ACKNOWLEDGMENTS

We thank Joy D. Van Nostrand and James W. Voordeckers for editing this paper. We also thank Meiying Xu, Peter Reich, and Sarah Hobbie, who made the experimental data for this study available.

This work has been supported, through contracts DE-SC0004601, DE-AC02-05CH11231 (as part of ENIGMA, a Scientific Focus Area), and DE-SC0004601, by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics: GTL Foundational Science, the United States Department of Agriculture (project 2007-35319-18305) through the NSF-USDA Microbial Observatories Program, the Oklahoma Bioenergy Center of the State of Oklahoma, and

the State Key Joint Laboratory of Environment Simulation and Pollution Control (grant 11Z03ESPCT) at Tsinghua University.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00122-11/-/DCSupplemental.

Text S1, PDF file, 0.362 MB.
Figure S1, TIF file, 0.293 MB.
Figure S2, TIF file, 0.363 MB.
Figure S3, TIF file, 2.672 MB.
Figure S4, JPG file, 2.315 MB.
Figure S5, TIF file, 0.132 MB.
Figure S6, TIF file, 0.087 MB.
Table S1, DOC file, 0.066 MB.
Table S2, DOC file, 0.029 MB.
Table S3, DOC file, 0.038 MB.

REFERENCES

1. **Houghton JT, et al. (ed).** 2001. Climate change 2001: the scientific basis: contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom.
2. **Ainsworth EA, Long SP.** 2005. What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. New Phytol. **165**:351–371.
3. **Luo YQ, Hui DF, Zhang DQ.** 2006. Elevated CO_2 stimulates net accumulations of carbon and nitrogen in land ecosystems: a meta-analysis. Ecology **87**:53–63.
4. **Reich PB, et al.** 2001. Plant diversity enhances ecosystem responses to elevated CO_2 and nitrogen deposition. Nature **410**:809–812.
5. **Carney KM, Hungate BA, Drake BG, Megonigal JP.** 2007. Altered soil microbial community at elevated CO_2 leads to loss of soil carbon. Proc. Natl. Acad. Sci. U. S. A. **104**:4990–4995.
6. **Gruber N, Galloway JN.** 2008. An earth-system perspective of the global nitrogen cycle. Nature **451**:293–296.
7. **Heath J, et al.** 2005. Rising atmospheric CO_2 reduces sequestration of root-derived soil carbon. Science **309**:1711–1713.
8. **Heimann M, Reichstein M.** 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. Nature **451**:289–292.
9. **Adair EC, Reich PB, Hobbie SE, Knops JMH.** 2009. Interactive effects of time, CO_2, N, and diversity on total belowground carbon allocation and ecosystem carbon storage in a grassland community. Ecosystems **12**: 1037–1052.
10. **He Z, et al.** 2010. Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO_2. Ecol. Lett. **13**:564–575.
11. **Olesen JM, Bascompte J, Dupont YL, Jordano P.** 2007. The modularity of pollination networks. Proc. Natl. Acad. Sci. U. S. A. **104**:19891–19896.
12. **Montoya JM, Pimm SL, Solé RV.** 2006. Ecological networks and their fragility. Nature **442**:259–264.
13. **Dunne JA, Williams RJ, Martinez ND.** 2002. Food-web structure and network theory: the role of connectance and size. Proc. Natl. Acad. Sci. U. S. A. **99**:12917–12922.
14. **Allesina S, Alonso D, Pascual M.** 2008. A general model for food web structure. Science **320**:658–661.
15. **Bastolla U, et al.** 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. Nature **458**:1018–1020.
16. **Fuhrman JA, Steele JA.** 2008. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. Aquat. Microb. Ecol. **53**:69–81.
17. **Raes J, Bork P.** 2008. Molecular eco-systems biology: towards an understanding of community function. Nat. Rev. Microbiol. **6**:693–699.
18. **Zhou J, et al.** 2010. Functional molecular ecological networks. MBio **1**:e00169-10.
19. **Reich PB, et al.** 2006. Nitrogen limitation constrains sustainability of ecosystem response to CO_2. Nature **440**:922–925.
20. **Bascompte J.** 2007. Networks in ecology. Basic Appl. Ecol. **8**:485–490.
21. **Dunne JA, Williams RJ, Martinez ND, Wood RA, Erwin DH.** 2008. Compilation and network analyses of Cambrian food webs. PLoS Biol. **6**:e102.
22. **Dunne JA.** 2006. The network structure of food webs, p. 27–86. *In* Pascual M, Dunne JA (ed), Ecological networks: linking structure to dynamics in food webs. Oxford University Press, Oxford, United Kingdom.
23. **Chaffron S, Rehrauer H, Pernthaler J, von Mering C.** 2010. A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res. **20**:947–959.
24. **Horvath S, Dong J.** 2008. Geometric interpretation of gene coexpression network analysis. PLoS Comput. Biol. **4**:e1000117.
25. **Langfelder P, Horvath S.** 2007. Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. **1**:54.
26. **Oldham MC, et al.** 2008. Functional organization of the transcriptome in human brain. Nat. Neurosci. **11**:1271–1282.
27. **Alon U.** 2003. Biological networks: the tinkerer as an engineer. Science **301**:1866–1867.
28. **Barabási AL, Oltvai ZN.** 2004. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. **5**:101–113.
29. **Clauset A, Moore C, Newman ME.** 2008. Hierarchical structure and the prediction of missing links in networks. Nature **453**:98–101.
30. **Jordano P, Bascompte J, Olesen JM.** 2003. Invariant properties in coevolutionary networks of plant-animal interactions. Ecol. Lett. **6**:69–81.
31. **Watts DJ, Strogatz SH.** 1998. Collective dynamics of "small-world" networks. Nature **393**:440–442.
32. **Clauset A, Newman ME, Moore C.** 2004. Finding community structure in very large networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **70**: 066111.
33. **Dong J, Horvath S.** 2007. Understanding network concepts in modules. BMC Syst. Biol. **1**:24.
34. **Guimerà R, Sales-Pardo M, Amaral LA.** 2007. Classes of complex networks defined by role-to-role connectivity profiles. Nat. Phys. **3**:63–69.
35. **Arumugam M, et al.** 2011. Enterotypes of the human gut microbiome. Nature **473**:174–180.
36. **Fuhrman JA.** 2009. Microbial community structure and its functional implications. Nature **459**:193–199.
37. **Muegge BD, et al.** 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science **332**: 970–974.
38. **Luo F, et al.** 2007. Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. BMC Bioinform. **8**:299.
39. **Qin J, et al.** 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464**:59–65.
40. **Tylianakis JM, Tscharntke T, Lewis OT.** 2007. Habitat modification alters the structure of tropical host-parasitoid food webs. Nature **445**: 202–205.
41. **Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK.** 2005. The contribution of species richness and composition to bacterial services. Nature **436**:1157–1160.
42. **France KE, Duffy JE.** 2006. Diversity and dispersal interactively affect predictability of ecosystem function. Nature **441**:1139–1143.
43. **Hooper DU, et al.** 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecol. Monogr. **75**:3–35.
44. **Loreau M, et al.** 2001. Ecology—biodiversity and ecosystem functioning: current knowledge and future challenges. Science **294**:804–808.
45. **May RM.** 1973. Stability and complexity in model ecosystems. Princeton University Press, Princeton, NJ.
46. **Symstad AJ, et al.** 2003. Long-term and large-scale perspectives on the relationship between biodiversity and ecosystem functioning. BioScience **53**:89–98.
47. **Tilman D, Reich PB, Knops JM.** 2006. Biodiversity and ecosystem stability in a decade-long grassland experiment. Nature **441**:629–632.
48. **Omenn GS.** 2006. Grand challenges and great opportunities in science, technology, and public policy. Science **314**:1696–1704.
49. **Luo F, Zhong JX, Yang YF, Scheuermann RH, Zhou JZ.** 2006. Application of random matrix theory to biological networks. Phys. Lett. A **357**: 420–423.
50. **Maslov S, Sneppen K.** 2002. Specificity and stability in topology of protein networks. Science **296**:910–913.