

Functional Molecular Ecological Networks

Jizhong Zhou,^{a,b,c} Ye Deng,^{a,e} Feng Luo,^d Zhili He,^a Qichao Tu,^a and Xiaoyang Zhi^a

Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma, USA^a; Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA^b; Department of Environmental Science and Engineering, Tsinghua University, Beijing, People's Republic of China^c; School of Computing, Clemson University, Clemson, South Carolina, USA^d; and Glomics Inc., Norman, Oklahoma, USA^e

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

ABSTRACT Biodiversity and its responses to environmental changes are central issues in ecology and for society. Almost all microbial biodiversity research focuses on “species” richness and abundance but not on their interactions. Although a network approach is powerful in describing ecological interactions among species, defining the network structure in a microbial community is a great challenge. Also, although the stimulating effects of elevated CO₂ (eCO₂) on plant growth and primary productivity are well established, its influences on belowground microbial communities, especially microbial interactions, are poorly understood. Here, a random matrix theory (RMT)-based conceptual framework for identifying functional molecular ecological networks was developed with the high-throughput functional gene array hybridization data of soil microbial communities in a long-term grassland FACE (free air, CO₂ enrichment) experiment. Our results indicate that RMT is powerful in identifying functional molecular ecological networks in microbial communities. Both functional molecular ecological networks under eCO₂ and ambient CO₂ (aCO₂) possessed the general characteristics of complex systems such as scale free, small world, modular, and hierarchical. However, the topological structures of the functional molecular ecological networks are distinctly different between eCO₂ and aCO₂, at the levels of the entire communities, individual functional gene categories/groups, and functional genes/sequences, suggesting that eCO₂ dramatically altered the network interactions among different microbial functional genes/populations. Such a shift in network structure is also significantly correlated with soil geochemical variables. In short, elucidating network interactions in microbial communities and their responses to environmental changes is fundamentally important for research in microbial ecology, systems microbiology, and global change.

IMPORTANCE Microorganisms are the foundation of the Earth's biosphere and play integral and unique roles in various ecosystem processes and functions. In an ecosystem, various microorganisms interact with each other to form complicated networks. Elucidating network interactions and their responses to environmental changes is difficult due to the lack of appropriate experimental data and an appropriate theoretical framework. This study provides a conceptual framework to construct interaction networks in microbial communities based on high-throughput functional gene array hybridization data. It also first documents that elevated carbon dioxide in the atmosphere dramatically alters the network interactions in soil microbial communities, which could have important implications in assessing the responses of ecosystems to climate change. The conceptual framework developed allows microbiologists to address research questions unapproachable previously by focusing on network interactions beyond the listing of, e.g., the number and abundance of species. Thus, this study could represent transformative research and a paradigm shift in microbial ecology.

Received 18 June 2010 Accepted 15 September 2010 Published 5 October 2010

Citation Zhou, J., Y. Deng, F. Luo, Z. He, Q. Tu, et al. 2010. Functional molecular ecological networks. *mBio* 1(4):e00169-10. doi:10.1128/mBio.00169-10.

Editor David Relman, VA Palo Alto Health Care System

Copyright © 2010 Zhou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

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

An ecosystem is a complex system in which various species interact with each other to form complicated networks (1). Through such network interactions, an ecosystem is capable of accomplishing systems level functions (e.g., nutrient cycling, ecosystem stability) which could not be achieved by individual populations. Explaining and predicting network structures, dynamics, and the underlying mechanisms are essential parts of ecology. Although ecological networks of biological communities have been intensively studied in macrobial ecology (1–5), very limited studies have been carried out in microbial communities due to their vast diversity and as-yet-uncultivated status (6–8).

Massive amounts of data on microbial community diversity

and dynamics across various spatial and temporal scales can be generated with metagenomics and associated technologies, such as high-throughput sequencing and microarrays (9, 10), which offer an unprecedented opportunity to examine network interactions among different microbial species/populations (6). Recently, microarray-based high-throughput technologies, such as GeoChip (10), have been developed and are enabling microbial ecologists to address a variety of ecological questions on a community-wide scale (11, 12). However, identification of network structure based on such high-throughput microarray data is challenging.

Various network methods have been developed for inferring

cellular networks based on gene expression data (13), such as differential equation-based network methods, Bayesian network methods, and relevance/coexpression network methods (14). The correlation-based relevance network methods are most commonly used for identifying cellular networks (14) based on gene expression data because of their computational simplicity and the nature of microarray data (typically noisy, highly dimensional, and significantly undersampled) (13). However, most methods for relevance network analysis use arbitrary thresholds, which are often determined based on known biological information (15). Thus, the network structure largely depends on the thresholds chosen. It is difficult to select appropriate thresholds, especially for poorly studied organisms. We have previously developed a novel random matrix theory (RMT)-based approach to automatically identify cellular networks from microarray data (16, 17). Our results indicated that this approach is a reliable, sensitive, and robust tool for modular network identification and gene function prediction through high-throughput genomic data (16, 18, 19).

Here, we describe the novel RMT-based network approach (16, 17) to delineate and characterize functional molecular ecological networks (MENs) in microbial communities based on GeoChip hybridization data by addressing the following questions. (i) How can the functional molecular ecological networks in microbial communities be identified based on high-throughput GeoChip hybridization data? (ii) Does elevated CO₂ (eCO₂) have an impact on the functional network structure of soil microbial communities? To answer these questions, a novel conceptual framework of MENs has been developed for identifying and characterizing interaction networks in microbial communities based on high-throughput GeoChip hybridization data derived from the microbial communities under both eCO₂ and ambient CO₂ (aCO₂) in a multifactor grassland FACE (free air, carbon dioxide enrichment) experiment, BioCON (biodiversity, CO₂, and nitrogen deposition), at the Cedar Creek Ecosystem Science Reserve in Minnesota (20). Our results indicated that the functional ecological networks in microbial communities can be discerned using the RMT-based network approach and that eCO₂ has a significant impact on network interactions of microbial communities.

RESULTS

Molecular ecological networks. The detection and quantitation of microorganisms often rely on individual genes or gene-like DNA fragments such as 16S rRNA genes, functional genes, and intergenic regions. Based on gene abundance data, a network graph can be developed to represent the ecological interactions (links) of different gene markers (nodes) in a microbial community (21). Strictly speaking, the ecological networks determined in this way should reflect the interactions among different microbial populations carrying the OTUs (operational taxonomic units) or functional genes of interest rather than individual “species” in a microbial community. Thus, to avoid confusion, we especially refer to such molecule-based networks in microbial communities as MENs, in which OTUs or functional genes (nodes) are connected by pairwise interactions (links). In addition, MENs derived from phylogenetic gene markers (i.e., 16S rRNA gene sequencing data) are referred to as phylogenetic MENs (pMENs), whereas MENs derived from functional gene markers (e.g., GeoChip hybridization data) are called functional MENs (fMENs).

A general framework of MEN analysis is illustrated in Fig. 1. First, high-throughput metagenomic data (e.g., large-scale se-

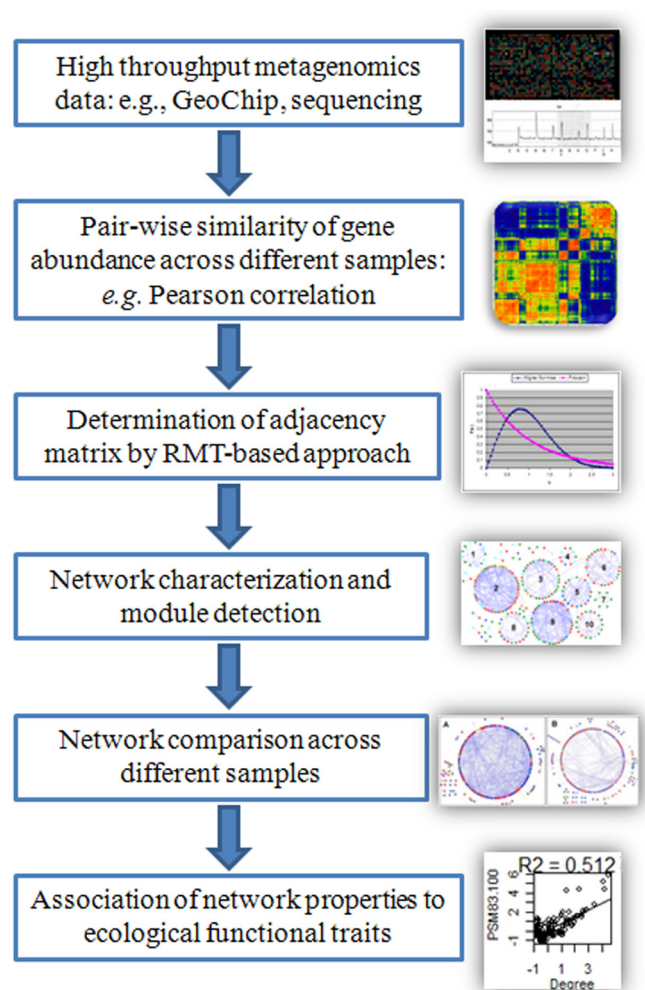


FIG 1 Overview of RMT-based MEN analysis. Six key steps are outlined here for molecular ecological analysis. A typical figure is placed in each step to highlight the main characteristics of such types of analysis.

quencing and functional gene array hybridization) are collected and appropriately transformed. Then, a pairwise Pearson correlation between any two genes is estimated based on the gene abundance data and the absolute value of the pairwise correlation coefficient is used to measure their similarity. As a result, a similarity matrix is obtained, which is subsequently transformed into an adjacency matrix by applying a threshold to the correlation values based on an RMT approach. Once the adjacency matrix, which measures the strengths of the connections between nodes, is defined, module analysis and network characterization are performed. Summing the strengths of the connections of each gene with all of the other connected genes yields a single network parameter, connectivity, which represents how strongly that gene is connected to all of the other genes in the network. In addition, the relationships of topological network characteristics (e.g., connectivity) to the sample traits of interest are evaluated to understand the importance of network properties in determining community functions.

Identification of functional MENs. Based on RMT, two universal extreme distributions of the nearest-neighbor spacing distribution of eigenvalues are predicted. One is Gaussian orthogonal

TABLE 1 Major topological properties of the empirical MENs of microbial communities under eCO₂ and aCO₂ and their associated random MENs

Community	Empirical networks				Random networks ^d						
	No. of original genes ^a	Similarity threshold (<i>s_i</i>)	Network size (<i>n</i>) ^b	<i>r</i> of scale free ^c (significance)	Avg connectivity (<i>avgK</i>)	Avg path length (GD) ^f	Avg clustering coefficient (<i>avgCC</i>)	Modularity (no. of modules)	Avg path length (GD)	Avg clustering coefficient (<i>avgCC</i>)	Modularity (<i>M</i>)
eCO ₂	371	0.80	245	-0.89 (<0.001)	7.13	3.09 ^e	0.22 ^e	0.44 ^e (18)	3.00 ± 0.03	0.099 ± 0.009	0.31 ± 0.01
aCO ₂	323	0.80	184	-0.94 (<0.001)	3.58	4.21 ^e	0.10 ^e	0.65 ^e (16)	3.84 ± 0.06	0.028 ± 0.007	0.52 ± 0.01

^a The number of genes that were originally used for network construction using the RMT-based approach.

^b The number of genes (i.e., nodes) in a network.

^c The correlation coefficient (*r*) of the linear relationship in $\log[P(k)] \sim -\gamma \log(k)$, where $P(k)$ is the fraction of connectivity k and γ is a constant.

^d The random networks were generated by rewiring all of the links of a MEN with the identical numbers of nodes and links to the corresponding empirical MEN.

^e Significant difference ($P < 0.001$) between aCO₂ and eCO₂.

^f GD, geodesic distance.

ensemble (GOE) statistics, which reflects the random properties of complex systems. The other is Poisson distribution, which is related to the system-specific, nonrandom properties of complex systems (22). The two predictions should be applicable to ecological communities if they are universal to complex systems based on RMT. Thus, we assume that there is a transition of the nearest-neighbor spacing distribution of eigenvalues from GOE to Poisson distributions, and this transition can serve as a reference point to distinguish random noise from system-specific, nonrandom properties embedded in high-throughput metagenomic data. This reference point is mathematically defined and can be used as a threshold to identify MENs in an automatic and objective fashion (16).

To discern the ecological network structure in microbial communities, GeoChip-based microbial community functional diversity data from aCO₂ and eCO₂ were analyzed using the RMT-based network approach (16, 17). Clear transitions of the nearest-neighbor spacing distribution of eigenvalues from GOE to Poisson distribution were observed for grassland microbial communities under aCO₂ and eCO₂, as indicated by the existence of a similarity threshold in Table 1. These results suggested that the two universal predictions based on RMT are applicable to the microbial communities examined. The transition points were used as the similarity thresholds (*s_i*) for network construction. Based on the RMT approach, the thresholds were determined to be 0.80 for both microbial communities (Table 1). If two genes have a correlation larger than this threshold, 0.80, this correlation is highly significant ($P < 0.0018$) based on the Fisher transformation test. This is also consistent with our previous analyses of gene coexpression networks which showed that the RMT-based approach is able to effectively remove random noise inherent in high-throughput microarray hybridization data (16, 23). Thus, these results suggested that the RMT-based network approach can be used to identify the network structure of microbial communities based on array hybridization data.

General characteristics of functional MENs. To understand whether the identified fMENs have a network topology similar to that of other complex systems, several important general network features, such as scale free, small world, modular, and hierarchical (15, 24, 25), were examined. A scale-free network is a network whose connectivity follows a power law, at least asymptotically; that is, a few nodes in the network have many connections with other nodes while most of the nodes have a few connections. Also, a small-world network is a network in which most of the nodes are

not neighbors of one another but most of the nodes can be reached from every other node by a small number of steps.

The degree distributions (or connectivity) in all of the constructed fMENs fit the power law model well, with typical correlation values of -0.89 to -0.94, respectively (Table 1; see Fig. S1 in the supplemental material), indicating that the fMENs in these microbial communities exhibited scale-free behavior, at least approximately. Also, the path length and clustering coefficients were significantly different from those of the corresponding random networks with the same network size and average number of links (Table 1) and were comparable to those in other networks displaying small-world behavior, indicating that the MENs in these microbial communities show typical small-world characteristics.

Many networks in biological and engineering systems are modular (24). In the fMENs examined here, a module is a group of functional genes that are highly connected among themselves but have few connections with the functional genes belonging to other modules. The two fMENs examined were modular, with a significantly higher modularity (*M*) than those from the corresponding random networks (Table 1).

Hierarchy is a central organizing principle of complex networks, but there is no formal definition of hierarchical topology (26). One of the most important signatures of hierarchical modular organizations is that the scaling of clustering coefficients follows $C(k) \sim k^{-\gamma}$, in which k is connectivity and γ is a constant. By log transformation, we will have $\log[C(k)] \sim -\gamma \log(k)$; that is, the logarithms of clustering coefficients have a linear relationship with the logarithms of connectivity. The clustering coefficients for the MENs examined followed $\log[C(k)] \sim -\gamma \log(k)$ ($r = -0.31 \sim -0.85$, $P < 0.001$), suggesting that all of the MENs examined here appeared to be hierarchical. However, it should be cautioned that global network properties such as the average shortest pathway, degree distribution, and clustering coefficient may fail to capture potentially important network structure features (27).

Impact of eCO₂ on the architecture of whole MENs. Although identical thresholds were used to define the network, the network size was considerably bigger under eCO₂ than under aCO₂ (Table 1). Also, the network composition was substantially different. Only 129 (43%) nodes of the fMENs were shared by eCO₂ and aCO₂. However, the connectivity values for the genes shared by these two networks were significantly correlated ($r = 0.379$, $P < 0.001$).

Various network indexes were calculated separately for both fMENs under eCO₂ and aCO₂. To test their statistical significance,

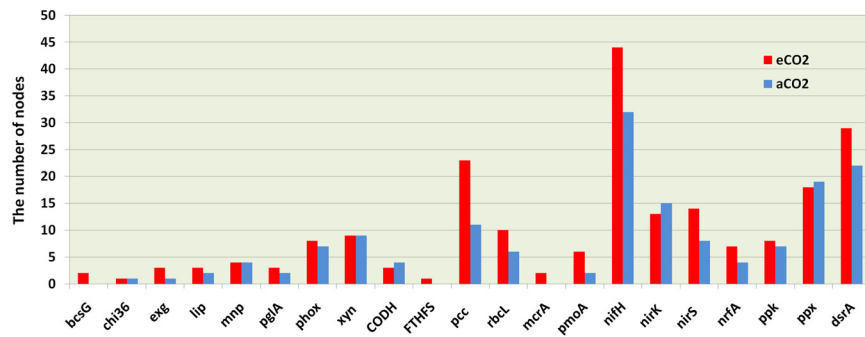


FIG 2 Distributions of major functional genes in the network under aCO₂ (blue) and eCO₂ (red). The distribution of genes varies substantially among different functional groups. The gene designations are explained in the legend to Fig. 3.

100 random networks corresponding to both fMENs were also generated, respectively. The standard deviations of individual indexes were estimated based on their corresponding random networks and used for the Student *t* test of their significance between eCO₂ and aCO₂. All of these indexes were significantly different ($P < 0.001$) between aCO₂ and eCO₂ (see Table S2 in the supplemental material), indicating that the overall network structures of these two microbial communities were distinctly different. Interestingly, compared to aCO₂, the fMENs at eCO₂ generally had significantly higher connectivity, shorter path lengths, higher clustering efficiencies, and more modules (Table 1), which are key network properties in terms of system efficiency and robustness (15, 28). All of the above results suggested that eCO₂ could have a significant impact on the overall architecture of the fMENs in these grassland microbial communities and that the overall network composition and structure are not well conserved between eCO₂ and aCO₂.

Totals of seven and eight modules with more than five nodes were obtained for the networks under aCO₂ and eCO₂, respectively (see Fig. S2 in the supplemental material). The sizes of these modules vary substantially, ranging from 6 to 59 nodes (see Fig. S2B and D in the supplemental material). These networks also differed significantly from each other in modularity (*M*) (Table 1). Fisher's exact test showed that no modules can be statistically paired between eCO₂ and aCO₂, suggesting that these two networks are even less conserved at the modular level.

Effects of eCO₂ on the network interactions of functional gene categories/groups. To examine the effects of eCO₂ on the network interactions at the levels of functional gene categories/groups, the network members were classified based on their associated functional gene groups. Consistent with the overall network topology, the network structures, as measured by average connectivity, average clustering coefficients, average path lengths, and modularity, were significantly different between eCO₂ and aCO₂ for the genes involved in C, N, P, and S cycling (see Fig. S3 in the supplemental material). They were also distinctly different from the corresponding random networks. Similar to the entire networks, all of these subnetworks had significantly higher connectivity, shorter path lengths, and higher clustering efficiencies under eCO₂ than under aCO₂ (see Fig. S3 in the supplemental material).

The impact of eCO₂ on network structure could also be re-

flected at the functional gene group level. Because most (57%) of the nodes were not shared by these two networks (see Table S3 in the supplemental material), for making a meaningful direct comparison, only the shared nodes from individual functional gene groups in these two networks were considered. As expected, the network complexities of individual functional gene groups, as measured by the node number, the average connectivity of the shared nodes, and the Shannon diversity of the connectivity among various functional groups, were also considerably different between aCO₂ and eCO₂ (Fig. 2; see Table S3 in the supplemental material). The network complexity for most functional gene groups was generally higher under eCO₂ than under aCO₂. For instance, the numbers of nodes for C fixation (*pmL*, *rbcL*), N fixation (*nifH*), and sulfate reduction (*dsrA*) were substantially higher under eCO₂ than under aCO₂ (Fig. 2). These results suggested that eCO₂ substantially altered the network structure among various microbial functional gene groups in this grassland ecosystem.

Differential influences of eCO₂ on network interactions of individual functional genes/sequences. To determine how eCO₂ affects the network structure of individual functional genes/sequences, the connectivity and clustering coefficients for the individual nodes shared by these two networks were estimated, followed by paired *t* tests to examine their statistical significance. Significant differences in connectivity ($P = 8.7 \times 10^{-11}$) or clustering coefficients ($P = 1.63 \times 10^{-5}$) were observed between these two networks under eCO₂ and aCO₂. These results suggested that eCO₂ significantly shifted the network structure of individual functional genes.

Since the structure of the entire networks is too complicated to display, only very limited numbers of key nodes with higher connectivity were considered. The top six functional genes with the highest connectivities under eCO₂ were examined (Fig. 3). The network interactions among the top six functional genes of the entire networks under eCO₂ (Fig. 3A) were distinctly different from the network interactions of the corresponding genes under aCO₂ (Fig. 3B). These six genes under eCO₂ had far more complicated network interactions than the corresponding genes under aCO₂ in terms of network size, connectivities, and clustering coefficients (Fig. 3C). In contrast, the identities of the highest-ranked genes under aCO₂ (see Fig. S4A in the supplemental material), based on connectivities, were substantially different from those under eCO₂ (Fig. 3A). The network interactions of these top key genes under aCO₂ are also quite different from the corresponding genes under eCO₂ (see Fig. S4A and B in the supplemental material).

eCO₂ had differential influences on the network interactions of various individual functional genes/sequences (see Table S3 in the supplemental material). Due to limited space, it is not possible to describe them in detail here. Thus, we picked *nifH* genes as an example because our previous results showed that the total abundance of *nifH* genes increased much more than that of other functional genes under eCO₂ (29) and the connectivities for all of the shared *nifH* genes in N fixation were significantly different be-

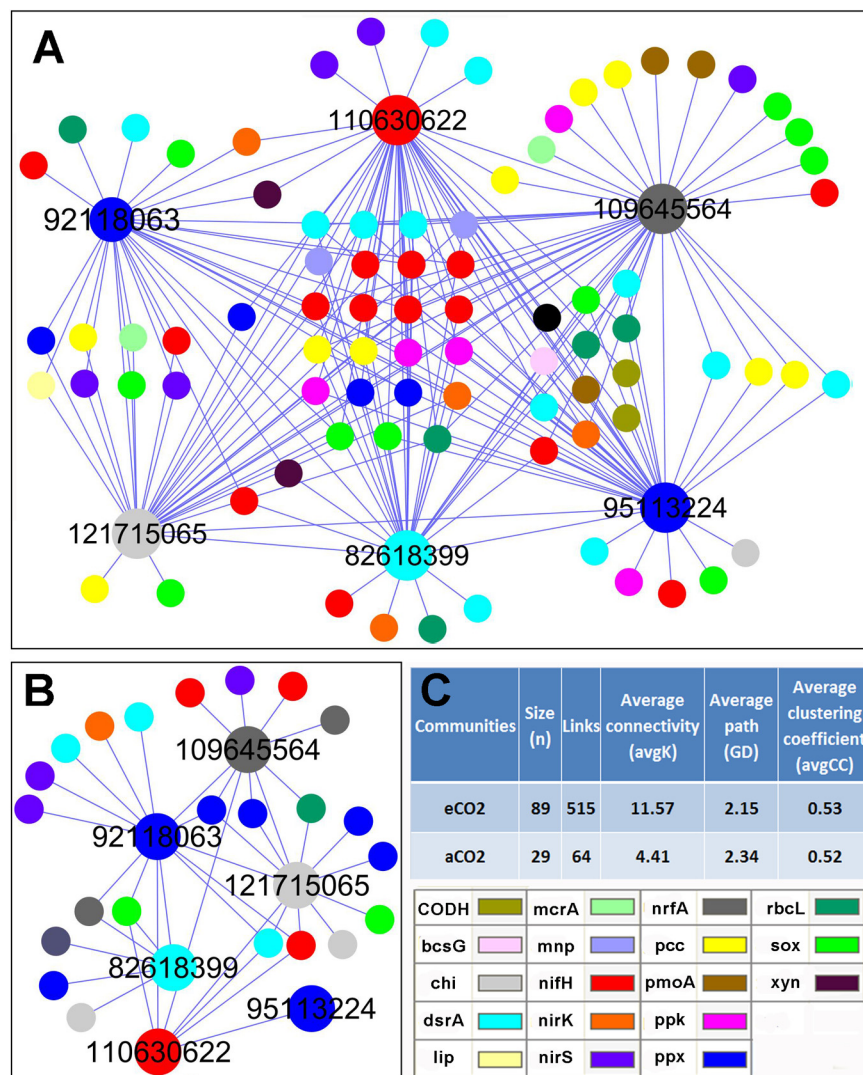


FIG 3 Impact of eCO₂ on the network interactions of key functional genes. (A) Network interactions of the top six functional genes with the highest connectivities under eCO₂. (B) Network interactions of the corresponding functional genes under aCO₂. Each node signifies a functional gene. Colors of the nodes indicate different functional genes. A blue line indicates a positive interaction between two individual nodes, while a red line indicates a negative interaction. The networks were constructed by the RMT-based approach with the GeoChIP data. The network interactions for these microbial functional genes were complex under eCO₂, but simple under aCO₂, suggesting that eCO₂ has a significant impact on the network interactions among key functional genes/populations in the grassland soil microbial communities. The genes are *chi* (endochitinase), *bcsG* (endoglucanase), *chi36* (exochitinase), *exg* (exoglucanase), *lip* (lignin peroxidase), *mnp* (manganese peroxidase), *pglA* (pectinase), *phox* (phenol oxidase), *xyn* (xylanase), *CODH* (carbon monoxide dehydrogenase), *FTHFS* (tetrahydrofolate formylase), *pcc* (propionyl coenzyme A carboxylase), *rbcl* (ribulose-1,5-bisphosphate carboxylase oxygenase), *mcrA* (methyl coenzyme M reductase), *pmoA* (methane monooxygenase), *nifH* (nitrogenase reductase), *nirK* (nitrite reductase), *nirS* (nitrite reductase), *nrfA* (*c*-type cytochrome nitrite reductase), *ppk* (polyphosphate kinase), *ppx* (exopolyphosphatase), *dsrA* (dissimilatory sulfite reductase), and *sox* (sulfite oxidase).

tween eCO₂ and aCO₂ based on a pairwise *t* test ($P < 0.001$) (see Table S4 in the supplemental material). The N-fixing populations monitored by *nifH* genes had far more complex network interactions with other functional groups of diverse phylogenetic compositions under eCO₂ (Fig. 4A) than under aCO₂ (Fig. 4B). Also, under eCO₂, several *nifH* hubs (e.g., 110630622, 89512768, and 76667345) were observed and each formed a separate module in-

teracting with various other functional groups (Fig. 4). In addition, the top *nifH* network hub from an uncultivated bacterium (GenBank ID: 110630622) (see Table S4 in the supplemental material) had positive interactions with many functional gene groups of diverse phylogenetic origins (indicated by the colors of the nodes), such as those involved in N fixation, denitrification, C fixation, C degradation, sulfate reduction, sulfur oxidation, and P utilization (Fig. 5A). Positive interactions may reflect commonly preferred conditions or cooperative behaviors such as cross feeding, syntrophic interactions, and mutualistic interactions (6). However, the same N fixer had very few interactions with other functional gene groups under aCO₂ (Fig. 5B; see Table S4 in the supplemental material). These results also suggested that eCO₂ dramatically changed the network interactions of various microbial functional genes/populations in the grassland ecosystem.

Association of network structure with ecological functional traits. Since microorganisms mediate important biogeochemical cycles of C, N, and P in soils, one intriguing question is whether the network interactions altered under eCO₂ are relevant to soil geochemistry and plant productivity. To discern the relationships among microbial network interactions, soil properties, and plant variables, Mantel tests were performed. Because using many unrelated individual variables may mask the signature of any significant variables, the trait-based gene significance measure (30), defined as the square of the correlation between the signal intensity of a gene and each soil or plant variable, was used to identify common subsets of soil and plant variables important to network interactions. Partial Mantel tests revealed very strong correlations between gene connectivity and the gene significance of the selected soil variables based on all of the genes detected or on subsets of the genes involved in recalcitrant C degradation ($P < 0.05$) or N fixation ($P < 0.01$) under eCO₂. Also, a strong correlation between gene

connectivity and the gene significance of the selected plant variables was obtained based on all of the genes detected that are involved in N cycling ($P < 0.001$). However, none of them were significant ($P < 0.05$) under aCO₂. These results suggested that the microbial community network interactions were, to some extent, related to soil and plant variables and that eCO₂ could have a significant impact on such relationships.

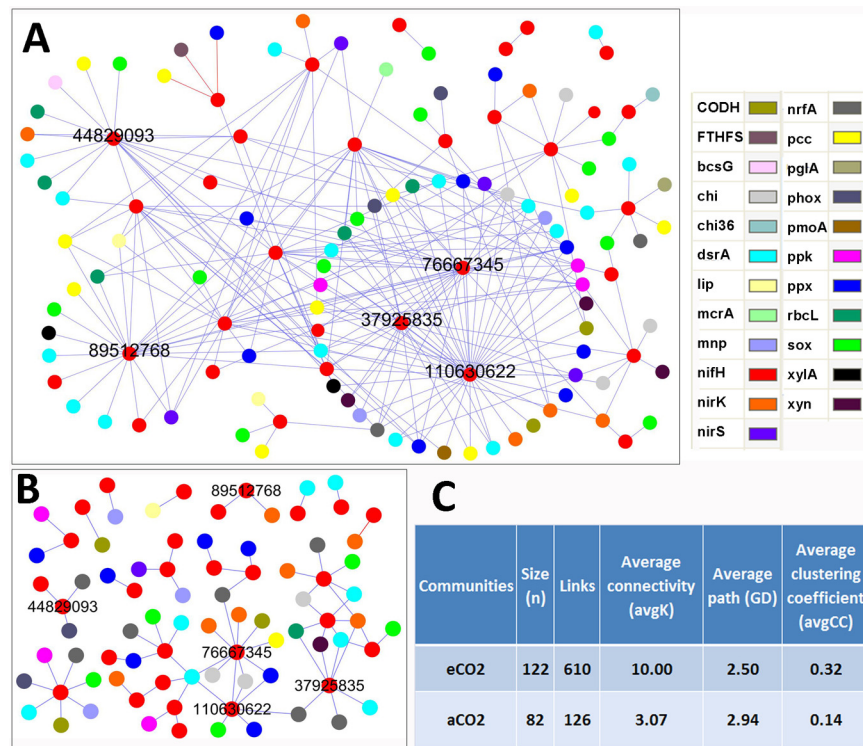


FIG 4 Network interactions of microorganisms containing *nifH* genes under eCO₂ (A) and aCO₂ (B). Microorganisms containing *nifH* genes formed complex network interactions with other functional groups, and some *nifH*-containing populations serve as central hubs in this community. The networks were constructed by the RMT-based approach with the GeoChip data from eCO₂ and aCO₂ and only shared *nifH* nodes, and their nearest neighbors in the network are shown here. The *nifH* genes detected in both fMENS of aCO₂ and eCO₂ are displayed with a bigger node size. The gene designations are explained in the legend to Fig. 3. The numbers represent the GenBank protein IDs to differentiate different *nifH* genes because most of them represent uncultivated microorganisms.

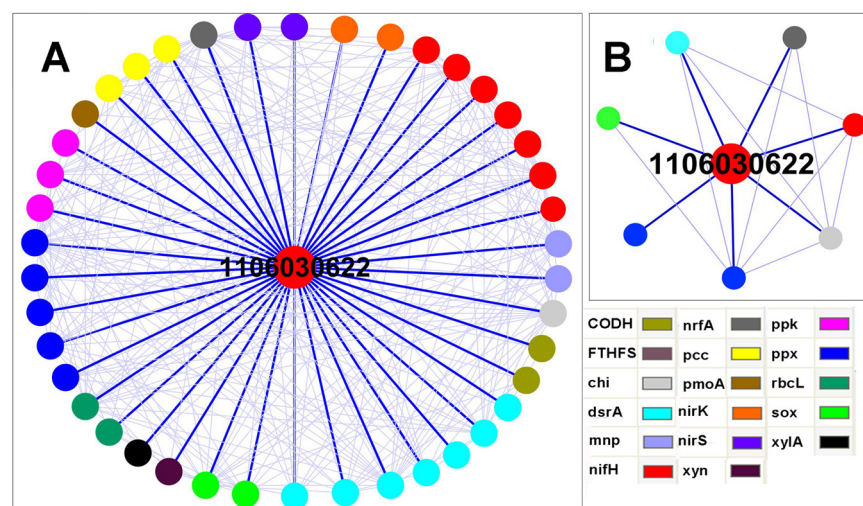


FIG 5 Network interactions of a *nifH* hub under both eCO₂ and aCO₂. The *nifH*-containing uncultivated microorganism had intensive positive interactions with many functional groups of diverse phylogenetic origins under eCO₂ (A) but very simple interactions with other functional groups under aCO₂ (B). Only this *nifH* gene node (1106030622) and its nearest neighbors are shown. The direct interactions with this *nifH* gene are labeled with thick lines, whereas the indirect interactions are marked with thin lines.

DISCUSSION

Microorganisms play critical roles in biogeochemical cycling of C, N, S, P, and various metals, but the precise roles of many microorganisms in these cycles are unknown (31). Elucidating their network interactions and linking them to ecosystem processes and functions is difficult. Although the availability of community-wide metagenomic data across many replicated samples offers an unprecedented opportunity to examine the network interactions in microbial communities, identifying network structure with such high-throughput metagenomic data is very challenging. Based on high-throughput metagenomic data such as array hybridization, a conceptual framework for studying functional network interactions in microbial communities has been developed in this study. The applications of network approaches to microbial communities could also provide a general framework for assessing the consequences of environmental disturbances at the whole-community level, which can serve as the first step toward a predictive microbial ecology within the context of global environmental change (6).

The RMT-based approach presented here provides a reliable, sensitive, and robust tool for identifying MENs with several main advantages. First, this approach was developed based on the two universal laws of RMT, and hence it is based on a sound theoretical foundation. Thus, it should be applicable to a variety of complex systems such as cells, communities, and ecosystems. Second, the threshold for defining a MEN is automatically defined based on the data structure itself rather than artificially chosen, and thus, no ambiguity occurs in identifying MENs. Third, since RMT is powerful for removing noise from nonrandom, system-specific features, the identified network is reliable, as clearly demonstrated in identifying transcriptional networks (16). In contrast to other approaches, such as those based on permutation testing (7, 32), the thresholds of correlation defined by RMT for defining networks are generally substantially higher. Consequently, the networks identified should be highly robust. Fourth, this RMT approach has potential for analyzing heterogeneous ecological data sets (e.g., hybridizations, sequencing, geochemistry) or combinations thereof. This could be particularly

important for linking network structure to ecosystem functioning. In addition, unlike local similarity analysis (7), which deals with time series data, this approach can be used to analyze both spatial and temporal data sets. However, the networks in this study are constructed based on Pearson correlation, which assumes a linear relationship between correlated variables. Some other, nonlinear, correlation methods, such as local similarity analysis (7), should be considered for further improvement of the RMT-based network approach.

Knowledge of network topology (e.g., scale free and small world) is important because the same or similar architecture among different types of networks (e.g., biological, physical, and social networks) may reveal common organizing principles of complex systems, and the shape of the degree distribution greatly affects the stability of the complex systems (15). Several recent studies indicated that most food webs did not display small-world patterns and scale-free structure, except for a few (33, 34). Food webs could be fitted with different functional forms, including power law, truncated power law, exponential, and uniform distributions (34). It appears that most food webs follow exponential distribution (35). Also, most of the mutualistic networks examined follow the truncated power law distribution. In contrast to the above two types of ecological networks, the fMENs presented here displayed small-world effects and the power law distribution, which are more similar to many cellular networks (15) such as protein-protein interaction networks, gene expression networks, and some metabolic networks. Further studies are needed to understand the unique characteristics, origins, evolutionary mechanisms, and dynamics of MENs.

Modularity is an inherent characteristic of many large complex systems, including many technological, biological (e.g., protein-protein interaction networks, gene expression), and social networks (15, 24, 26, 28, 36). In cellular networks, modularity is an evolved property that could enhance the flexibility of generation of various phenotypes during development (28). Modularity of gene regulation is essential to handle diverse and complex stimuli and responses. In ecology, a module is a group of species that interact strongly among themselves but little with species in other modules (known as compartmentalization) (37). Modularity in an ecological community may reflect habitat heterogeneity, physical contact, functional association, divergent selection, and/or phylogenetic clustering of closely related species (36). Modules with their component species may even be the key units of coevolution. Food webs are traditionally considered representative examples of ecological modularity. Recently, it was shown that all larger pollination networks are modular (36). Similar to food webs and pollination networks, our results showed that functional MENs are also modular. The presence of genes in the same modules could signify that the microorganisms carrying these genes have similar ecological niches. In addition, many different functional genes, such as *nifH*, belong to different modules and serve as hubs in different modules. This could be a unique predominant feature in ecological networks due to the existence of many redundant populations.

The fMENs identified possess the general features of many cellular networks with the hierarchical, modular, small-world, or scale-free network architecture, which could have important implications for the robustness and functional stability of ecosystems (15, 28). A small-world pattern facilitates efficient, rapid communication among different members within a system so the system

can make quick responses to environmental changes such as elevated atmospheric CO₂. On the other hand, the short path length will allow the local perturbations to reach the whole network quickly so that that network structure, as well as the functions, could be altered. However, the characteristics of modularity will help to minimize the effects of local perturbations on the system as a whole by containing perturbations and damage at a local level (28), whereas the hierarchical organization of various modules ensures that communication between modules and network hubs is relatively quick. Also, while a scale-free network is unperturbed by the random loss of nodes, it is vulnerable to attacks to network hubs (15). However, such vulnerability can be reduced or improved by the nonrandom organization of similar functional genes as hubs in multiple modules. As a result, a loss of one module hub will not have too much of an impact on the functional stability of the system as a whole. Therefore, as a whole, the overall microbial community could rapidly respond to environmental changes and remain robust in the face of random and specific perturbations via the balance of the advantages and disadvantages of various network topological characteristics for system functional stability.

Understanding the responses of biological communities to environmental change, especially anthropogenic change, is a central issue in ecology and evolution and for society. Due to the increased input of C into soil and associated chemistry under eCO₂, it is also expected that the composition and structure of microbial communities will be markedly altered under eCO₂, as demonstrated by our previous study (29). Although it is known that environmental changes such as acidification, habitat modification, and hydrological disturbance have profound effects on ecological networks (38), this has not yet been explored in microbial communities due to the lack of appropriate technologies. The MENs developed in this study provide an appropriate framework in which to explore the possible effects of environmental changes on microbial community structure. Also, our results indicated that under eCO₂, the network interactions for most of the functional gene groups become more complex than those under aCO₂. This is consistent with our previous results showing that both the functional and phylogenetic structures of microbial communities were dramatically altered by eCO₂ (29). Besides, such a shift in network structure is significantly correlated with soil geochemical variables. These results, along with those of our previous studies (29), suggested that global climate change factors such as eCO₂ have a significant impact on not only the functional structure of grassland microbial communities but also their network interactions.

In summary, the analytical approaches and results presented here are important for studies on ecology, global change biology, and microbiology. First, traditionally, most biodiversity studies of microbial communities have just used information on the number of species and the abundance of each species, but no sufficient experimental data with many genes/populations across different samples were available for characterizing the interactions among different microorganisms. This study provides a novel conceptual framework for studying network interactions among different microbial populations, which is an essential component of biodiversity studies. Also, understanding the responses and mechanisms of biological communities to global change is a central goal for ecologists (39, 40). As demonstrated by this study, the network interactions for most of the functional gene groups become more com-

plex under eCO₂ than under aCO₂. Thus, it is apparent that global climate change factors such as eCO₂ could have a fundamental impact on network interactions at the levels of entire communities, individual functional gene categories/groups, and functional genes/sequences. To our knowledge, this is the first study to demonstrate the changes in the network interactions of microbial communities in response to eCO₂. In addition, metagenomics has emerged as a cutting-edge 21st-century science but one of the greatest challenges is how to use such information to understand community-level functional processes. This study provides unique tools for discerning network interactions based on metagenomic data. Thus, in short, the MEN framework developed should have a profound impact on the study of biodiversity, global change biology, ecosystem ecology, and systems microbiology.

MATERIALS AND METHODS

Sampling site and GeoChip data collection. Network analysis was conducted with the microbial communities of the soil samples from the Biocon experiment located at the Cedar Creek Ecosystem Science Reserve in Minnesota (45°N, 93°W). Plots were established in 1997 on a secondary successional grassland on a sandy outwash soil after removing the previous vegetation (41). The main Biocon field experiment has 296 (of a total of 371) evenly distributed plots (2 by 2 m) in six 20-meter-diameter rings, three for an aCO₂ concentration of 368 μmol/mol and three for an eCO₂ concentration of 560 μmol/mol using a FACE system (42). In this study, 24 plots (12 for aCO₂ and 12 for eCO₂, all with 16 species and no additional N supply) were used. The experimental analyses of these plots (e.g., soil chemistry, plants, GeoChip hybridization, and data preprocessing) are described elsewhere (29). Since this study focused on the impact of eCO₂ on ecosystem functional processes, only the genes involved in the cycling of nutrient such as carbon (C), nitrogen (N), phosphorus (P), and sulfur (S) were used for network analysis. Because we are more interested in the network interactions among different microbial functional groups, only the representative commonly used signature genes for various functional gene groups were selected for the network analysis. In most cases, only those genes detected in half or more than half of the total samples (majority rule) were kept for subsequent network constructions.

Network construction. Two MENs were constructed. The experimental data used for constructing fMENs were generated by GeoChip-based microarray analysis (29). The GeoChip hybridization intensity data were log transformed before the construction of a Pearson correlation matrix, which is commonly used for constructing gene expression networks (30). Logarithmic transformation improves data statistical properties by stabilizing the variations in signal intensity. The correlation matrix was then converted to a similarity matrix, which measures the degree of concordance between the abundance profiles of genes across different samples by taking the absolute values of the correlation matrix (30, 43). Subsequently, an adjacency matrix, which encodes the connection strength between each pair of nodes, was derived from the similarity matrix by applying an appropriate threshold, s_p , which was defined using the RMT-based network approach as previously described (16, 17).

Network characterization. The Cytoscape 2.6.0 (44) software was used to visualize the network graphs. Other information about genes, e.g., taxonomy, relative abundance, and edge information, e.g., weights and positive and negative correlations, was also imported into the software and visualized in the network figures. Since we are interested in the impact of eCO₂ on network interactions, the fMENs were constructed separately under aCO₂ and eCO₂.

Various indexes, including average degree (connectivity) (45), betweenness (46), stress and eigenvector centrality (46), average clustering coefficient (47, 48), vulnerability (49), average geodesic distance (45), geodesic efficiency and harmonic geodesic distance (50), density and transitivity (51), and connectedness (52), were used to describe the properties of individual nodes in the network and the overall topologies or structures

of the different networks. In general, the network index, connectivity (k_i), is calculated by summing the strengths of the connections (i.e., links) of each gene (i.e., node) with all of the other connected genes in the network. Connectivity represents how strongly a gene is connected to other genes and is one of the most commonly used network indexes. The definitions and calculations of other indexes are provided in Table S1 in the supplemental material. Most calculations were accomplished through the *sna* and *igraph* packages in the R project (53, 54). Those overall topological indexes in part II (see Table S1 in the supplemental material) describe the overall network topology from different angles, and thus they are useful in characterizing various fMENs identified under different conditions.

To characterize the modularity property of fMENs, each network was separated into modules, which were usually considered functional units in biological systems (36, 55). Modularity (M) measures the extent to which nodes have more links within their own modules than expected if linkage were random. The modules were detected by the modularity detection method as described previously (56). After scanning all of the branches of the hierarchical tree of a graph, the level with the maximum modularity score was used to separate the graph into multiple dense subgraphs. The modularity of each network (M) was calculated as previously described (56).

Random network construction and network comparison. Since only a single data point of each overall network index was available for each network parameter, standard statistical analysis could not be performed to assess their statistical significance. Thus, random networks were generated using the Maslov-Sneppen procedure (57). This method keeps the numbers of nodes and links unchanged but rewires all of the links' positions in the fMENs so the sizes of networks are the same and the random rewired networks are comparable to the original ones. For each network identified in this study, a total of 100 randomly rewired networks were generated and all of the network indexes were calculated individually. Then, the average and standard deviation for each index of all of the random networks were obtained. The statistical Z test was used to test the differences between the indexes of the fMEN and random networks. Meanwhile, for comparisons of the network indexes under different conditions, the Student t test was employed using the standard deviations derived from corresponding random networks.

Gene significance based on a sample trait. In gene expression network analyses, gene significance (GS_{il}) is the correlation between the expression profile of the i th gene and the l th sample trait, T_l (30). The higher the GS_{il} value, the more biologically significant is gene i with respect to sample trait l . In this study, gene significance was defined as follows:

$$GS_{il} = [\text{cor}(x_i, T_l)]^2$$

where x_i is the i th gene signal intensities $i \in \{1, \dots, n\}$ and T_l is the l th sample trait (e.g., soil pH, N content, total plant biomass) [$l \in \{1, \dots, q\}$]. Since the measurement units for different traits vary, all of the trait data were standardized before statistical analysis.

Massive soil and plant trait data from this long-term experimental site are available (20, 29, 41) as described above, and they were used for estimating gene significance. The coefficient of correlation between each gene and each soil or plant variable was calculated across 12 replicate samples under both eCO₂ and aCO₂, respectively. Thus, the gene significance matrix, $GS^{n \times q}$, was obtained.

Relationships of microbial interaction networks with soil and plant variables. To discern the relationships between the fMENs and soil properties and plant variables, Mantel tests were performed. Since many unrelated individual variables could mask the signature of any significant variables, a common subset of soil or plant variables were selected for Mantel tests as follows. First, the correlations between the gene significance of the l th soil or plant variable (GS_{il}) and the connectivity of individual genes (k_i) were calculated across all of the genes detected. The statistical significances of these correlations were estimated based on P values. Second, all of the soil or plant variables with significant r values ($P < 0.1$) under either eCO₂ or aCO₂ were selected and combined as a common set of soil or plant variables for Mantel tests. The following soil

variables were selected: the percentages of C or N at depths of 0 to 10, 10 to 20, 20 to 40, and 40 to 60 cm; the proportions of soil moisture at depths of 0 to 17, 42 to 59, and 83 to 100 cm; and soil pH. The following plant variables were also selected: total root biomass, percent belowground C, aboveground total biomass, aboveground N, aboveground C/N ratio, coarse roots 0 to 20 (g/m²), annual root ingrowth fine roots (g/m²), forb biomass, and biomass for four plant species, *Amorpha canescens* (legume), *Andropogon gerardii* (C4), *Lespedeza capitata* (legume), and *Sorghastrum nutans* (C4). In addition, simple or partial Mantel tests were performed for the connectivity of all of the genes detected and all of the selected soil or plant variables to examine the relationships between network structure (i.e., connectivity) and soil or plant variables. Mantel tests were performed using the programs available in the R vegan package (58).

ACKNOWLEDGMENTS

This work has been partially supported through contract DE-SC0004601 and contract DE-AC02-05CH11231 (as part of ENIGMA, a Scientific Focus Area) by the US 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, and the Oklahoma Bioenergy Center (OBC).

We all contributed intellectual input and assistance to this study and manuscript preparation. The original concept, framework, and strategies were developed by J.Z. All of the data analyses and network construction were performed by Y.D. with help from F.L., Q.T., X.Z., and Z.H. J.Z. synthesized all of the data and wrote the paper with help from Y.D. and Z.H.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00169-10/-/DCSupplemental>.

Table S1, DOC file, 0.093 MB.
Table S2, DOC file, 0.036 MB.
Table S3, DOC file, 0.067 MB.
Table S4, DOC file, 0.045 MB.
Figure S1, TIF file, 0.331 MB.
Figure S2, TIF file, 4.817 MB.
Figure S3, TIF file, 0.213 MB.
Figure S4, TIF file, 2.710 MB.

REFERENCES

- Montoya, J. M., S. L. Pimm, and R. V. Sole. 2006. Ecological networks and their fragility. *Nature* 442:259–264.
- Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci. U. S. A.* 100:9383–9387.
- Bastolla, U., M. A. Fortuna, A. Pascual-Garcia, A. Ferrera, B. Luque, and J. Bascompte. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* 458:1018–1020.
- Cattin, M. F., L. F. Bersier, C. Banasek-Richter, R. Baltensperger, and J. P. Gabriel. 2004. Phylogenetic constraints and adaptation explain food-web structure. *Nature* 427:835–839.
- Holland, M. D., and A. Hastings. 2008. Strong effect of dispersal network structure on ecological dynamics. *Nature* 456:792–794.
- Raes, J., and P. Bork. 2008. Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* 6:693–699.
- Ruan, Q. S., D. Dutta, M. S. Schwalbach, J. A. Steele, J. A. Fuhrman, and F. Z. Sun. 2006. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* 22:2532–2538.
- Fuhrman, J. A. 2009. Microbial community structure and its functional implications. *Nature* 459:193–199.
- Handelsman, J., J. M. Tiedje, L. Alvarez-Cohen, M. Ashburner, I. K. O. Cann, E. F. DeLong, W. F. Doolittle, C. M. Fraser-Liggett, A. Godzik, J. I. Gordon, M. Riley, M. B. Schmid, and A. H. Reid. 2007. Committee on metagenomics: challenges and functional applications. The National Academies Press, Washington, DC.
- He, Z., T. J. Gentry, C. W. Schadt, L. Wu, J. Liebich, S. C. Chong, Z. Huang, W. Wu, B. Gu, P. Jardine, C. Criddle, and J. Zhou. 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* 1:67–77.
- Wang, F., H. Zhou, J. Meng, X. Peng, L. Jiang, P. Sun, C. Zhang, J. D. Van Nostrand, Y. Deng, Z. He, L. Wu, J. Zhou, and X. Xiao. 2009. GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *Proc. Natl. Acad. Sci. U. S. A.* 106:4840–4845.
- Zhou, J., S. Kang, C. W. Schadt, and C. T. Garten, Jr. 2008. Spatial scaling of functional gene diversity across various microbial taxa. *Proc. Natl. Acad. Sci. U. S. A.* 105:7768–7773.
- Gardner, T. S., and J. J. Faith. 2005. Reverse-engineering transcription control networks. *Phys. Life Rev.* 2:65–88.
- Zhang, B., and S. Horvath. 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:Article17.
- Barabasi, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5:101–115.
- Luo, F., Y. Yang, J. Zhong, H. Gao, L. Khan, D. K. Thompson, and J. Zhou. 2007. Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. *BMC Bioinform.* 8:299.
- Luo, F., J. X. Zhong, Y. F. Yang, R. H. Scheuermann, and J. Z. Zhou. 2006. Application of random matrix theory to biological networks. *Phys. Lett. A* 357:420–423.
- Yang, Y., D. P. Harris, F. Luo, W. Xiong, M. Joachimiak, L. Wu, P. Dehal, J. Jacobsen, Z. Yang, A. V. Palumbo, A. P. Arkin, and J. Zhou. 2009. Snapshot of iron response in *Shewanella oneidensis* by gene network reconstruction. *BMC Genomics* 10:131.
- Zhou, A., Z. He, A. M. Redding-Johanson, A. Mukhopadhyay, C. L. Hemme, M. Joachimiak, F. Luo, Y. Deng, K. S. Bender, Q. He, J. D. Keasling, D. A. Stahl, M. W. Fields, T. C. Hazen, A. P. Arkin, J. D. Wall, and J. Zhou. May 2010, posting date. Hydrogen peroxide-induced oxidative stress responses in *Desulfovibrio vulgaris* Hildenborough. *Environ. Microbiol.* [Epub ahead of print.] doi: 10.1111/j.1462-2920.2010.02234.x.
- Reich, P. B., S. E. Hobbie, T. Lee, D. S. Ellsworth, J. B. West, D. Tilman, J. M. Knops, S. Naeem, and J. Trost. 2006. Nitrogen limitation constrains sustainability of ecosystem response to CO₂. *Nature* 440:922–925.
- Fuhrman, J. A., and J. A. Steele. 2008. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquat. Microb. Ecol.* 53:69–81.
- Mehta, M. L. 1990. Random matrices, 2nd ed. Academic Press, Inc., New York, NY.
- Luo, F., J. Zhong, Y. Yang, and J. Zhou. 2006. Application of random matrix theory to microarray data for discovering functional gene modules. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 73:031924.
- Alon, U. 2003. Biological networks: the tinkerer as an engineer. *Science* 301:1866–1867.
- Barabasi, A. L. 2009. Scale-free networks: a decade and beyond. *Science* 325:412–413.
- Muller-Linow, M., C. C. Hilgetag, and M. T. Hutt. 2008. Organization of excitable dynamics in hierarchical biological networks. *PLoS Comput. Biol.* 4:e1000190.
- Guimera, R., M. Sales-Pardo, and L. A. Amaral. 2007. Classes of complex networks defined by role-to-role connectivity profiles. *Nat. Phys.* 3:63–69.
- Kitano, H. 2004. Biological robustness. *Nat. Rev. Genet.* 5:826–837.
- He, Z., M. Xu, Y. Deng, S. Kang, L. Kellogg, L. Wu, J. D. van Nostrand, S. E. Hobbie, P. Reich, and J. Zhou. 2010. Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO₂. *Ecol. Lett.* 13:564–575.
- Horvath, S., and J. Dong. 2008. Geometric interpretation of gene co-expression network analysis. *PLoS Comput. Biol.* 4:e1000117.
- Fitter, A. H., C. A. Gilligan, K. Hollingworth, A. Kleczkowski, R. M. Twyman, J. W. Pitchford, and N. S. B. Programme. 2005. Biodiversity and ecosystem function in soil. *Funct. Ecol.* 19:369–377.
- Curtis, D., B. V. North, and P. C. Sham. 2001. Use of an artificial neural network to detect association between a disease and multiple marker genotypes. *Ann. Hum. Genet.* 65:95–107.
- Berlow, E. L., J. A. Dunne, N. D. Martinez, P. B. Stark, R. J. Williams, and U. Brose. 2009. Simple prediction of interaction strengths in complex food webs. *Proc. Natl. Acad. Sci. U. S. A.* 106:187–191.
- Dunne, J. A., R. J. Williams, and N. D. Martinez. 2002. Food-web

- structure and network theory: the role of connectance and size. *Proc. Natl. Acad. Sci. U. S. A.* **99**:12917–12922.
35. Fortuna, M. A., and J. Bascompte. 2006. Habitat loss and the structure of plant-animal mutualistic networks. *Ecol. Lett.* **9**:278–283.
 36. Olesen, J. M., J. Bascompte, Y. L. Dupont, and P. Jordano. 2007. The modularity of pollination networks. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 19891–19896.
 37. Bascompte, J., and D. B. Stouffer. 2009. The assembly and disassembly of ecological networks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**: 1781–1787.
 38. Tylianakis, J. M., T. Tcharntke, and O. T. Lewis. 2007. Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature* **445**:202–205.
 39. Heimann, M., and M. Reichstein. 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451**:289–292.
 40. Gruber, N., and J. N. Galloway. 2008. An Earth-system perspective of the global nitrogen cycle. *Nature* **451**:293–296.
 41. Reich, P. B., J. Knops, D. Tilman, J. Craine, D. Ellsworth, M. Tjoelker, T. Lee, D. Wedin, S. Naeem, D. Bahauddin, G. Hendrey, S. Jose, K. Wrage, J. Goth, and W. Bengton. 2001. Plant diversity enhances ecosystem responses to elevated CO₂ and nitrogen deposition. *Nature* **410**: 809–812.
 42. Lewin, G. R., E. Mckintosh, and S. B. McMahon. 1994. Nmda receptors and activity-dependent tuning of the receptive-fields of spinal-cord neurons. *Nature* **369**:482–485.
 43. Horvath, S., B. Zhang, M. Carlson, K. V. Lu, S. Zhu, R. M. Felciano, M. F. Laurance, W. Zhao, S. Qi, Z. Chen, Y. Lee, A. C. Scheck, L. M. Liau, H. Wu, D. H. Geschwind, P. G. Febbo, H. I. Kornblum, T. F. Cloughesy, S. F. Nelson, and P. S. Mischel. 2006. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl. Acad. Sci. U. S. A.* **103**:17402–17407.
 44. Cline, M. S., M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P. L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schumovich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader. 2007. Integration of biological networks and gene expression data using cytoscape. *Nat. Protoc.* **2**:2366–2382.
 45. West, D. B. 1996. *Introduction to graph theory*. Prentice Hall, Upper Saddle River, NJ.
 46. Brandes, U., and T. Erlebach. 2005. *Network analysis: methodological foundations*. Springer-Verlag, Berlin, Germany.
 47. Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of “small-world” networks. *Nature* **393**:440–442.
 48. Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297**:1551–1555.
 49. Costa, L. D., F. A. Rodrigues, G. Travieso, and P. R. V. Boas. 2007. Characterization of complex networks: a survey of measurements. *Adv. Phys.* **56**:167–242.
 50. Latora, V., and M. Marchiori. 2001. Efficient behavior of small-world networks. *Phys. Rev. Lett.* **87**:198701.
 51. Wasserman, S., and K. Faust. 1994. *Social network analysis: methods and applications*. Cambridge University Press, Cambridge, United Kingdom.
 52. Krackhardt, D. 1994. Graph theoretical dimensions of informal organizations. Lawrence Erlbaum, Hillsdale, NJ.
 53. Butts, C. T. 2008. Social network analysis with sna. *J. Stat. Softw.* **24**:1–51.
 54. Csardi, G. 2006. The Igraph library. <http://cneurocv.s.rnki.kfki.hu/igraph/>.
 55. Newman, M. E. J. 2006. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U. S. A.* **103**:8577–8582.
 56. Clauset, A., M. E. Newman, and C. Moore. 2004. Finding community structure in very large networks. *Phys. Rev. E* **70**:066111.
 57. Maslov, S., and K. Sneppen. 2002. Specificity and stability in topology of protein networks. *Science* **296**:910–913.
 58. Dixon, P. 2003. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**:927–930.