Supplementary Materials and Methods

Experimental site and sampling

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, USA (Lat. 45 ° N, Long. 93 ° W). The main BioCON field experiment has 296 plots established in 1997 on a secondary successional grassland after removal of the previous vegetation (53). There are three treatments: (i) CO_2 concentration (ambient CO_2 at 368 μmol/mol, and elevated CO_2 at 560 μmol/mol), (ii) plant diversity (1, 4, 9 and 16 species grasses), and (iii) nitrogen addition (4 g NH_4NO_3 m^{-2} $year^{-1}$) (53). For the 24 plots used, 12 were from ambient CO_2, and 12 from eCO_2, all contained 16-species with no additional N supply. The soil samples were collected in July, 2007, and each sample was composited from five soil cores at a depth of 0-15 cm. All samples were immediately transported to the laboratory where they were frozen and stored at -80^{o}C for 454 pyrosequencing analysis of 16S rRNA genes.

Plant and soil data

Plant and soil data sets were used for Mantel tests and other statistical analyses to correlate network characteristics with plant and soil properties. Plant data, including plant species, aboveground and belowground biomass, plant C and N, and the C/N ratio were the same as previously described (54). Similarly, soil data including soil physical properties (e.g., volumetric moisture, pH), soil chemical properties (e.g., soil C, soil N, C/N), and biological processes (e.g., net N mineralization rate, nitrification rate) were the same as previously described (54, 55).

DNA extraction, purification and quantitation

Soil DNA was extracted by freeze-grinding mechanical lysis as described previously (56), and was purified using a low melting agarose gel followed by phenol extraction. DNA quality was assessed by the ratios of 260 nm/280 nm, and 260/230 nm using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE), and final soil DNA concentrations were quantified with PicoGreen (57) using a FLUOstar Optima (BMG Labtech, Jena, Germany).

454 pyrosequencing analysis

a. Sample tagging and PCR amplicon preparations

Based on the V4-V5 hypervariable regions of bacterial 16S rRNAs (*Escherichia coli* positions 515-907), the PCR primers, F515: GTGCCAGCMGCCGCGG, and R907: CCGTCAATTCMTTTRAGTTT were selected. Both primers were then checked with the ribosomal database (58), and covered > 98% of the 16S gene sequences in the database (July 2007). To pool multiple samples for one run of 454 sequencing, a sample tagging approach was used (59, 60). In this study, a unique 6-mer tag for each of 24 DNA samples was added to the 5'-end of both primers, and those tag-primers were synthesized by Invitrogen (Carlsbad, CA) and used for the generation of PCR amplicons. The amplification mix contained 10 units of Pfu polymerase (BioVision, Mountain View, CA), 5 µl Pfu reaction buffer, 200 µM dNTPs (Amersham, Piscataway, NJ), and a 0.2 µM concentration of each primer in a volume of 50 µl. Genomic DNA (10 ng) was added to each amplification mix. Cycling conditions were an initial denaturation at 94 ℃ for 3 min, 30 cycles of 95 ℃ 30 s, 58 ℃ for 60 s, and 72 ℃ for 60 s, a final 2-min extension at 72 ℃. Normally, multiple (5-10) 50-µl reactions were needed for each sample, and the products were pooled together after PCR amplification and purified by agarose gel electrophoresis. The amplified PCR products were recovered and then quantified with PicoGreen (57) using a FLUOstar Optima (BMG Labtech, Jena, Germany). Finally, amplicons of all samples were pooled in an equimolar concentration for 454 pyro-sequencing.

b. 454 pyrosequencing

The fragments in the amplicon libraries were repaired and ligated to the 454 sequencing adapters, and resulting products were bound to beads under conditions that favor one fragment per bead. The beads were emulsified in a PCR mixture in oil, and PCR amplification occurred in each droplet, generating millions of copies of a unique DNA template. After breaking the emulsion, the DNA strands were denatured, and beads carrying single-stranded DNA clones were deposited into wells on a PicoTiter-Plate (454 Life Sciences) for pyrosequencing (61) on a FLX 454 system (454 Life Sciences, Branford, CT). For this study, we recovered 85,399 sequence reads that represented both forward and reverse reads of 24 samples with an average length

around 240 bp. All pyrosequencing reads were initially processed using the RDP Pyrosequencing Pipeline (http://pyro.cme.msu.edu/pyro/index.jsp) (58).

c. Assignment of sequence reads to samples

The raw sequences were sorted and distinguished by unique sample tags. Since each sample had a unique tag, all sequence reads with the same tag were assigned to the same sample. The tag and primers were then trimmed for each sample. For all 24 samples, the number of sequence reads ranged from 4613 to 2341.

d. Removal of low-quality sequences

To minimize effects of random sequencing errors, we eliminated (*i*) sequences that did not perfectly match the PCR primer at the beginning of a read, (*ii*) sequences with non-assigned tags, (*iii*) sequence reads with < 150 bp after the proximal PCR primer if they terminated before reaching the distal primer, and (*iv*) sequences that contained any anonymous nucleotide (N). Similar approaches were also used in previous studies (62-65) to remove low quality sequences. As a result, a total of 77,653 sequences left with average length of 251 bp. There were 35,298 reads from forward and 42,355 reads from reverse strands. All these sequences are available at http://ieg.ou.edu/4download/. Chimeric sequences were checked using Mallard 1.02 (66), and removed from subsequent analysis.

e. Classification of 454 sequences and OTUs assignment

Because approximately 150 bp overlapped regions existed between the sequences from forward and reverse primers, we were able to identify the sequences from the same DNA molecule and combine them together into a single OTU based on the overlapped regions. The details were described in Supplemental methods. First, all forward sequences from 24 samples were pooled together, so did reverse sequences. These two datasets were independently aligned by RDP Infernal Aligner, a fast secondary-structure aware aligner (67) and then complete linkage clustering method was used to define OTUs within a 0.03 difference (68). Two sets of OTUs were obtained, one from forward primer and the other from reverse primer. Second, if the sequences from different strands had 100% match over a region of more than 100 bp based on BLASTn report, these sequences were considered from the same DNA molecule and hence

combined as a single sequence. Then, if the OTUs from different strand shared one or multiple pairs of sequences, all sequence reads of these OTUs were pooled together and composed to a new OTU. The other OTUs without any pairs of sequencing reads were still kept same. The singleton OTUs (with only one read) were removed and all other remained sequences from forward, reverse and combined OTUs were sorted into each sample based on OTUs. Eventually, the distributions of all OTUs were tabulated as an $n \times m$ matrix where n was the number of OTUs and m was the number of samples.

f. OTUs filtering and phylotype assignment

Based on the OTU dataset, OTU that appeared in only one sample among the total of 12 samples for each CO_2 condition was removed, resulting in 3500 OTUs for further analysis. This stringent OTU filtering criterion was to denoise the potential sequence errors (*e.g.* chimeras). Then the longest sequences of each OTU were selected to assign a taxonomy by the RDP classifier (69). The confidence cutoff was set to 0.5. The lineages of the longest sequences were summarized as the phylotypes of OTUs.

g. Relative abundance calculation

The sequence numbers for individual samples were different. In order to compare the difference between aCO_2 and eCO_2, we standardized the OTU distribution matrix into the relative abundance (RA) based on the following equation:

$$RA_{ij} = \frac{S_{ij}}{\sum_{j=1}^{n} S_{ij}} * 100 \qquad (1)$$

where i was the i^{th} sample (1 to m), j is the j^{th} OTU (1 to n) and S_{ij} was the number of sequences in i^{th} sample and j^{th} OTU. $\sum_{j=1}^{n} S_{ij}$ was the sum of sequences in i^{th} sample. Here m equals 24 samples and n equals 3500 OTUs.

h. Initial dataset preparation for network construction

After all data were transformed to RA, only the OTUs detected in equal or more than 9 of the 12 biological replicates were kept for network construction. The missing values were filled through the nearest neighbors method (70), which chose ten most correlated data from the remained values and calculated the mean values for the missing positions. Finally, two tables with OTU distributions of RA across all replicates for eCO_2 and aCO_2 respectively were obtained for network construction.

i. Data standardization and correlation analysis

Since the sequence numbers of individual OTUs obtained varied significantly among different samples, the relative proportions of sequence numbers were used for subsequent analysis. Let $y_{i,j}$ represent the relative proportion of the i-th OTU in the j-th sample ($i \in \{1,...,n\}$, $j \in \{1,...,m\}$).$Y^{nxm} = [y_{i,j}]$ is the relative sequence abundance matrix. The mean and standard deviation of $y_{i,j}$ across these samples are

$$\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{i,j} \quad (2)$$

$$\sigma_{i,y}^2 = \frac{1}{m-1}\sum_{j=1}^{m}(y_{i,j} - \bar{y}_i)^2 \quad (3)$$

Let $x_{i,j}$ represent the standardized relative proportion of the i-th OTU in the j-th sample ($i \in \{1,...,n\}$, $j \in \{1,...,m\}$), then

$$x_i = \frac{(y_i - \bar{y})}{\sigma_y} \quad (4)$$

x_i is the standardized relative proportion of the i-th OTU with mean 0 and variance 1.

Pairwise Pearson correlations of the a-th and b-th OTU (r_{ab}) are calculated based on the standardized relative abundance data ($x_{i,j}$) across all samples,

$$r_{ij} = cor\,(x_i, x_j) = \frac{\sum_q^m (x_{i,q} - \bar{x}_i)(x_{j,q} - \bar{x}_j)}{\sqrt{\sum_q^m (x_{i,q} - \bar{x}_i)^2 \sum_q^m (x_{j,q} - \bar{x}_j)^2}} \quad (5)$$

A correlation matrix, $R^{nxn} = [r_{ij}]$, is obtained, and it is then used for subsequent network analysis.

Network characterization

Cytoscape 2.6.0 (71) software was used to visualize the network graphs. Other OTU information, e.g. taxonomy, relative abundances, and edge information, e.g. weights and positive and negative correlations, were also imported into the software and visualized in the network figures. Since we were interested in the impacts of eCO_2 on network interactions, the pMENs were constructed separately under aCO_2 and eCO_2.

Various indexes, including average degree (connectivity) (72), betweenness (73), stress and eigenvector centrality (73), average clustering coefficient (74, 75), vulnerability (76), average geodesic distance (72), geodesic efficiency and harmonic geodesic distance (77), density and transitivity (78) and connectedness (79), were used to describe the individual nodes properties in the network and the overall topology or structure of the different networks. In general, the network index, connectivity (k_i), is calculated by summing the connection strengths (i.e. links) of each OTU (i.e. node) with all other connected OTUs in the network. Connectivity provides information on how strongly an OTU is connected to other OTUs and is one of the most commonly used network indexes. Most calculations were accomplished through *sna* and *igraph* packages in R project (80, 81).

To characterize the modularity property of pMENs, each network was separated into modules which were usually considered as functional units in biological systems (82, 83). Modularity (M) measures the extent to which nodes have more links within their own modules than expected if linkage is random. The modules were detected by the fast greedy modularity optimization (84). Then the modularity value of each network (M) was calculated as previously described (84), which is used to describe how well the modules are separated. Compared to other module separation algorithms (85, 86), the fast greedy modularity optimization had much higher M values and clearer separations in graphs.

Each node can be assigned a role based on its topological properties (87). The role of each node can be determined based on its position compared to other nodes in its own module and how well it connects to nodes in other modules. Therefore, the role of node i is characterized by its within-module connectivity (z_i) and among-module connectivity (P_i) (87). In this study, we

followed the simplified classification as follows: (i) Peripheral nodes ($z_i \leq 2.5$, $P_i \leq 0.62$), which have only a few links and almost always to nodes within their modules; (ii) Connectors ($z_i \leqslant 2.5$, $P_i > 0.62$), which are highly linked to several modules; (iii) Module hubs ($z_i > 2.5$, $P_i \leq 0.62$), which are highly connected to many species in their own modules; and (iv) Network hubs ($z_i > 2.5$, $P_i > 0.62$), which act as both module hubs and connectors. All of above calculations were carried out by a Perl scripted program.

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. One way to characterize a network is by comparing it with a random network. Two major approaches can be used to generate random networks (88, 89) although several other approaches were also available (90, 91). One is to completely randomly reshuffle all links in an original network by keeping the total nodes and total links constant, which is referred to Erdos–Renyi "random" network (88). The other is to randomly reshuffle all links of the original network by holding the total nodes, total links and the degree of connectivity for each node constant (89). While the former generates the random network which could contain some nodes without any links and thus the network size decreases, the latter is able to keep the size of random network consistent with empirical network. For direct comparison, we used the later approach to generate random networks.

One of the widely used procedures for the later approach is developed by Maslov and Sneppen (55, 89, 92-101). To obtain a meaningful comparison, the random network is restricted so that all OTUs have exactly the same connectivity as in the original network, but their interaction partners are totally randomly selected. For each network identified in this study, a total of 100 randomly rewired networks (89) were generated and all network indexes were calculated individually. Multiple sampling of the randomized networks enabled us to calculate both the average expectation and the standard deviation for each index of the random networks. The statistical *Z*-test was used to test the differences of the indexes between the pMEN and random networks. Meanwhile, for the comparisons between the network indexes under different conditions, the Student *t*-test was employed using the standard deviations derived from corresponding random networks.

Eigengene network analysis

One of the grand challenges in dealing with high throughput metagenomics data is the high dimensionality. Various statistical approaches are used for data reduction, including principal component analysis (PCA), detrended correspondence analysis (DCA), and singular value decomposition (SVD). Based on SVD analysis, Langfelder and Horvath (102) proposed eigengene network analysis to summarize the gene expression data from each module as a centroid. Eigengene network analysis is very useful to reveal higher order organization among gene co-expression modules (102-104). Therefore, we adapted this method for analyzing modules in pMENs as described elsewhere (55, 102-104).

OTU significance based on a sample trait

In gene expression network analyses, the gene significance (GS_{il}) is the correlation between the expression profile of the i-th gene and the l-th sample trait, T_l (103). The higher GS_{il}, the more biologically significant is Gene i related to the sample Trait l. In this study, the OTU significance is defined as:

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

where x_i is the i-th OTU abundance $i \in \{1,...,n\}$ and T_l is the l-th sample trait (*e.g.* soil pH, N content, total plant biomass) ($l \in \{1,...,q\}$) . Since the measurement units for different traits vary, all trait data were standardized prior to statistical analysis.

Massive soil and plant trait data are available for this long-term experimental site (53, 54, 105) as described above, and they were used for estimating OTU significance. The correlation coefficients between each OTU and each soil or plant variable was calculated across 12 replicate samples under both eCO_2 and aCO_2, respectively. Thus OTU significance matrix, GS^{nxl}, was obtained.

Relationships of microbial interaction networks with soil variables

To discern the relationships between the phylogenetic molecular ecological networks and soil properties, Mantel tests were performed. The following soil variables were selected: the percentages of C or N at the depths of 0-10, 10-20, 20-40, and 40-60 cm, the proportion of soil

moistures at the depths of 0-17, 42-59 and 83-100 cm, and soil pH. The relationships between the pMENs and soil variables were determined as follows: First, the OTU significances of all soil variables were calculated with the above equation to generate the OTU significance matrix GS^{nxl}. Then the Euclidean distance matrix $D_{GS}^{n\times n}$ was generated by calculating the Euclidean distance between every two OTUs. The distance matrix among all OTUs' connectivity ($D_k^{n\times n}$) was calculated as well. In addition, Mantel tests were performed between the distance matrices of the connectivity ($D_k^{n\times n}$) and OTU significance ($D_{GS}^{n\times n}$) to examine the relationships between network structure (i.e., connectivity) and soil variables. Similar processes were carried out based on individual phylogenetic groups. The Mantel tests were performed using the programs available in R vegan package (106).

Phylogenetic analysis

To illustrate the phylogenetic relationships among different nodes (OTUs), the hierarchical trees were constructed based on individual modules. First of all, a representative sequence was randomly picked from a group of sequences (> 1) belongs to that OTU. All OTU sequences in the networks were searched against the type strain sequences in the Ribosomal Database Project (RDP) and the most nearest high- quality sequences were selected as references for comparison. Then all of the OTU sequences and the selected reference sequences were aligned together by using ClustalW2 (107). The trees were constructed using Neibor-Joining approach with bootstrap of 1,000 times. This distance was calcuated based on Kimura 2-parameter mode using MEGA4 software (108).

Supplemental References

53. **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. Bengston.** 2001. Plant diversity enhances ecosystem responses to elevated CO2 and nitrogen deposition. Nature **410:**809-812.

54. **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.

55. **Zhou, J., Y. Deng, F. Luo, Z. He, Q. Tu, and X. Zhi.** 2010. Functional molecular ecological networks. mBio **1:**e00169-10.

56. **Zhou, J. Z., M. A. Bruns, and J. M. Tiedje.** 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. **62:**316-322.

57. **Ahn, S., J. Costa, and J. Emanuel.** 1996. PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR. Nucl. Acids Res. **24:**2623-2625.

58. **Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucl. Acids Res. **37:**D141-145.

59. **Binladen, J., M. T. P. Gilbert, J. P. Bollback, F. Panitz, C. Bendixen, R. Nielsen, and E. Willerslev.** 2007. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. PLoS ONE **2:**e197.

60. **Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods **5:**235-7.

61. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani,**

K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-380.

62. **Huber, J. A., D. Mark Welch, H. G. Morrison, S. M. Huse, P. R. Neal, D. A. Butterfield, and M. L. Sogin.** 2007. Microbial population structures in the deep marine biosphere. Science **318:**97-100.

63. **Huse, S. M., J. A. Huber, H. G. Morrison, M. L. Sogin, and D. Mark Welch.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. **8:**R143.

64. **Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Natl Acad. Sci. USA **103:**12115-12120.

65. **Sun, Y. J., Y. P. Cai, L. Liu, F. H. Yu, M. L. Farrell, W. McKendree, and W. Farmerie.** 2009. ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. Nucleic Acids Res. **37:**e79.

66. **Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman.** 2006. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. Appl. Environ. Microbiol. **72:**5734-5741.

67. **Nawrocki, E. P., and S. R. Eddy.** 2007. Query-dependent banding (QDB) for faster RNA similarity searches. PLoS Comput. Biol. **3:**e56.

68. **Stackebrandt, E., and B. M. Goebel.** 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. Int. J. Syst. Bacteriol. **44:**846-849.

69. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole.** 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. **73:**5261-7.

70. **Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman.** 2001. Missing value estimation methods for DNA microarrays. Bioinformatics **17:**520-5.

71. **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. Schmulevich, 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:**2366-82.

72. **West, D. B.** 1996. Introduction to Graph Theory. Prentice Hall, Upper Saddle River, N.J.

73. **Brandes, U., and T. Erlebach.** 2005. Network analysis: methodological foundations. Springer-Verlag, Berlin.

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

75. **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.

76. **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.

77. **Latora, V., and M. Marchiori.** 2001. Efficient behavior of small-world networks. Phys. Rev. Lett. **87:**198701.

78. **Wasserman, S., and K. Faust.** 1994. Social Network Analysis: Methods and applications. Cambridge Univerisity Press, Cambridge.

79. **Krackhardt, D.** 1994. Graph Theoretical Dimensions of Informal Organizations. Lawrence Erlbaum and Associates, Hillsdale, NJ.

80. **Butts, C. T.** 2008. Social Network Analysis with sna. J. Stat. Softw. **24:**1-51.

81. **Csardi, G.** 2006, posting date. The igraph library. [Online.]

82. **Newman, M. E. J.** 2006. Modularity and community structure in networks. Proc. Natl Acad. Sci. USA **103:**8577-8582.

83. **Olesen, J. M., J. Bascompte, Y. L. Dupont, and P. Jordano.** 2007. The modularity of pollination networks. Proc. Natl Acad. Sci. USA **104:**19891-19896.

84. **Clauset, A., M. E. Newman, and C. Moore.** 2004. Finding community structure in very large networks. Phys. Rev. E **70:**066111.

85. **Girvan, M., and M. E. Newman.** 2002. Community structure in social and biological networks. Proc. Natl Acad. Sci. USA **99:**7821-6.
86. **Pons, P., and M. Latapy.** 2005. Computing communities in large networks using random walks. Computer and Information Sciences - Iscis 2005, Proceedings **3733:**284-293.
87. **Guimera, R., and L. A. N. Amaral.** 2005. Functional cartography of complex metabolic networks. Nature **433:**895-900.
88. **Erdos, P., and A. Renyi.** 1959. On random graphs. I. . Publicationes Mathematicae **6:**290-297.
89. **Maslov, S., and K. Sneppen.** 2002. Specificity and stability in topology of protein networks. Science **296:**910-3.
90. **Beygelzimer, A., G. E. Grinstein, R. Linsker, and I. Rish.** 2005. Improving network robustness by edge modification. Physica A **357:**593-612.
91. **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. USA **99:**12917-12922.
92. **Bernhardt, B. C., Z. Chen, Y. He, A. C. Evans, and N. Bernasconi.** 2011. Graph-Theoretical Analysis Reveals Disrupted Small-World Organization of Cortical Thickness Correlation Networks in Temporal Lobe Epilepsy. Cereb Cortex.
93. **Wen, W., W. Zhu, Y. He, N. A. Kochan, S. Reppermund, M. J. Slavin, H. Brodaty, J. Crawford, A. Xia, and P. Sachdev.** 2011. Discrete neuroanatomical networks are associated with specific cognitive abilities in old age. J Neurosci **31:**1204-12.
94. **Sabri, M., R. Hauser, M. Ouellette, J. Liu, M. Dehbi, G. Moeck, E. Garcia, B. Titz, P. Uetz, and S. Moineau.** 2011. Genome annotation and intraviral interactome for the Streptococcus pneumoniae virulent phage Dp-1. J Bacteriol **193:**551-62.
95. **Spoormaker, V. I., M. S. Schroter, P. M. Gleiser, K. C. Andrade, M. Dresler, R. Wehrle, P. G. Samann, and M. Czisch.** 2010. Development of a large-scale functional brain network during human non-rapid eye movement sleep. J Neurosci **30:**11379-87.
96. **Centola, D.** 2010. The spread of behavior in an online social network experiment. Science **329:**1194-7.

97. **Nayak, R. R., M. Kearns, R. S. Spielman, and V. G. Cheung.** 2009. Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. Genome Res. **19:**1953-62.
98. **Seo, C. H., J. R. Kim, M. S. Kim, and K. H. Cho.** 2009. Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. Bioinformatics **25:**1898-904.
99. **Ma'ayan, A., G. A. Cecchi, J. Wagner, A. R. Rao, R. Iyengar, and G. Stolovitzky.** 2008. Ordered cyclic motifs contribute to dynamic stability in biological and engineered networks. Proc. Natl Acad. Sci. USA **105:**19235-40.
100. **Smart, A. G., L. A. Amaral, and J. M. Ottino.** 2008. Cascading failure and robustness in metabolic networks. Proc. Natl Acad. Sci. USA **105:**13223-8.
101. **Mondragon, R. J.** 2008. Topological modelling of large networks. Philos Transact A Math Phys Eng Sci **366:**1931-40.
102. **Langfelder, P., and S. Horvath.** 2007. Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. **1:**54.
103. **Horvath, S., and J. Dong.** 2008. Geometric interpretation of gene coexpression network analysis. PLoS Comput. Biol. **4:**e1000117.
104. **Oldham, M. C., G. Konopka, K. Iwamoto, P. Langfelder, T. Kato, S. Horvath, and D. H. Geschwind.** 2008. Functional organization of the transcriptome in human brain. Nat. Neurosci. **11:**1271-82.
105. **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:**922-5.
106. **Dixon, P.** 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. **14:**927-930.
107. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-8.

108. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596-9.

1

Supplemental Tables

Table S1. Summary of the network complexity of various phylogenetic groups. For direct comparison, the same parameters under aCO_2 and eCO_2 were highlighted with same colors

Domain	Class	Number of Shared nodes	aCO_2			eCO_2		
			Number of nodes	Average connectivity in shared nodes	Shannon index of connectivity	Number of nodes	Average connectivity in shared nodes	Shannon index of connectivity
pMEN all organisms		171	292	3.25	5.37	263	3.32	5.29
Archaea	*Crenarchaeota*	2	3	3.38	1.05	2	2.67	0.50
Bacteria	*Acidobacteria*	24	45	2.81	3.52	34	3.46	3.25
	Actinobacteria	48	81	2.60	4.15	76	3.00	4.01
	Bacteroidetes	10	23	2.50	2.86	24	4.50	2.82
	Chloroflexi	2	3	1.50	0.96	2	5.00	0.53
	Firmicutes	6	11	2.67	2.11	8	4.17	1.98
	Gemmatimonadetes	8	10	4.25	2.09	8	3.13	1.97
	Nitrospira	1	2	1.00	0.64	1	3.00	0.00
	Planctomycetes	2	4	2.50	1.28	5	1.50	1.52
	α-Proteobacteria	27	45	2.37	3.65	42	2.93	3.48
	β-Proteobacteria	15	22	2.73	2.94	24	4.53	2.93
	δ-Proteobacteria	3	6	2.00	1.64	8	2.00	2.01
	γ-Proteobacteria	9	11	2.89	2.07	9	3.44	1.96
	Verrucomicrobia	8	12	12.38	2.23	9	5.13	2.01
	unclassified	6	14	2.33	2.44	11	1.33	2.23

Table S2. The paired modules ($p<0.05$) between eCO_2 and aCO_2 pMENs

	Modules under eCO_2	Module size under eCO_2	Modules under aCO_2	Module size under aCO_2	No of Shared nodes	p value from Fisher exact test
Pair 1	E9	20	A13	27	10	<0.001
Pair 2	E4	5	A5	36	3	0.010
Pair 3	E11	26	A6	11	3	0.046
Pair 4	E7	6	A12	19	2	0.041
Pair 5	E3	32	A2	23	7	0.003

Table S3. Mantel test on connectivity vs. the OTU significances of soil geochemical variables[a]

Phylogeny	aCO_2			eCO_2		
	Network size	r_M[b]	**P**[c]	**Network size**	r_M	**P**
All detected OTUs	292	0.039	0.169	263	0.368	**0.001**
Acidobacteria	45	0.054	0.262	34	0.137	0.124
Actinobacteria	81	0.381	**0.002**	76	0.562	**0.001**
Bacteroidetes	23	-0.084	0.622	24	0.487	**0.012**
Firmicutes	11	0.023	0.338	8	0.310	0.114
Gemmatimonadetes	10	-0.220	0.958	8	0.299	0.168
Planctomycetes	4	0.604	0.221	5	-0.651	1.000
α-Proteobacteria	45	-0.034	0.638	42	0.472	**0.001**
β-Proteobacteria	22	0.029	0.319	24	0.297	**0.044**
δ-Proteobacteria	6	0.498	0.138	8	0.043	0.416
γ-Proteobacteria	11	-0.159	0.723	9	0.020	0.399
Verrucomicrobia	12	0.184	0.057	9	0.086	0.253

[a]Soil variables used for OTU significance calculations: %C20-40 (soil carbon content at 20-40cm depth), %N20-40 (soil nitrogen content at 20-40cm depth), C/N ratio20-40 (Carbon/Nitrogen ratio at 20-40cm depth), PSM0-17 (soil moisture at 0-17cm), PSM42-59 (soil moisture at 42-49cm) and PSM83-100 (soil moisture 83-100cm).

[b]Correlation based on Mantel test.

[c]The significance (probability) of Mantel test

Supplemental Figures

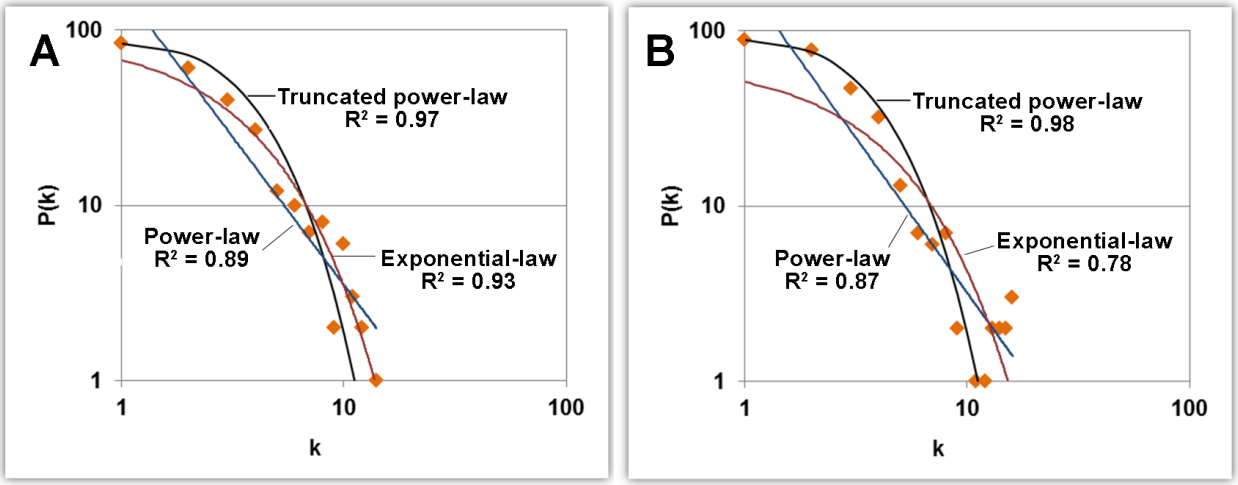


Fig. 1. The scatter plots showing the fittings of the OTU connectivity distributions of the pMENs under both aCO_2 and eCO_2. The x-axis is the node connectivity (*k*). The y-axis is the number of nodes under a given connectivity. The values in both axes were log-transformed. Lines and R^2 values show the best fit of the data to the model. (A) pMEN under eCO_2; and (B) pMEN under aCO_2.

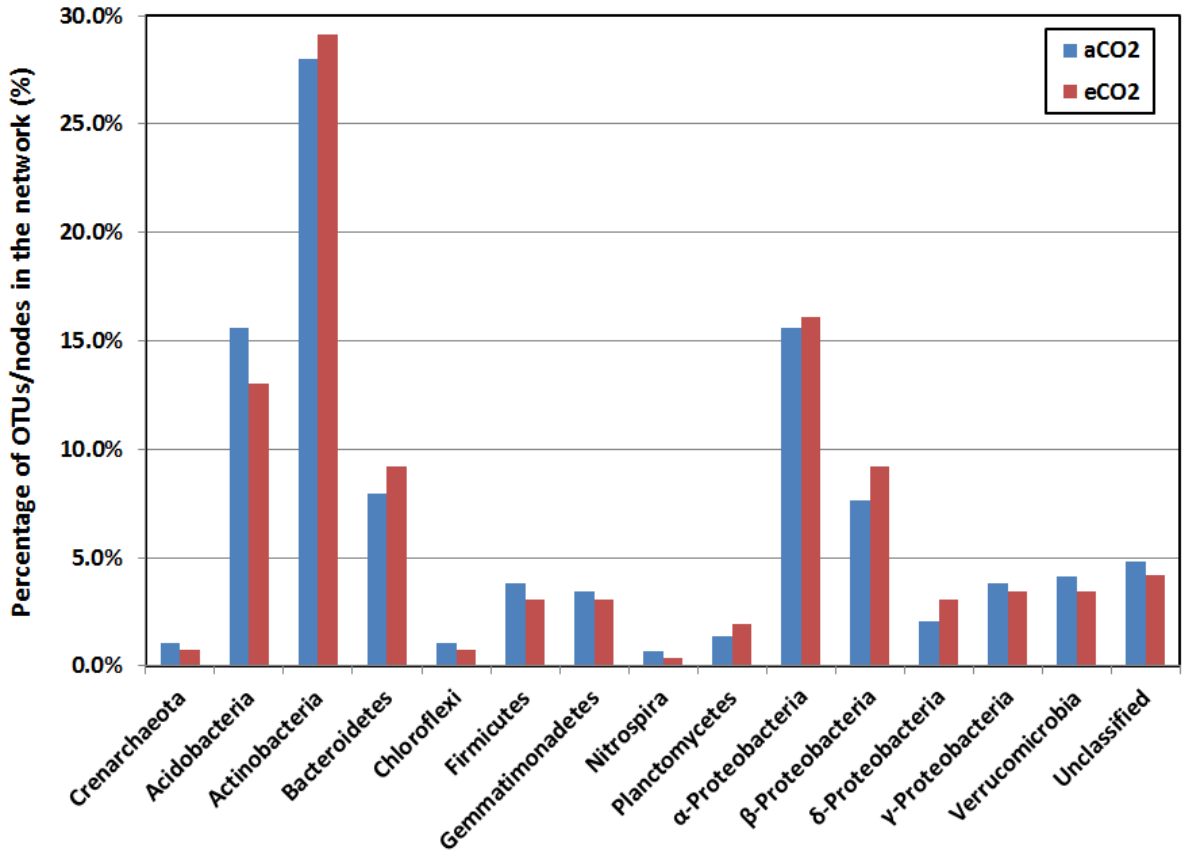
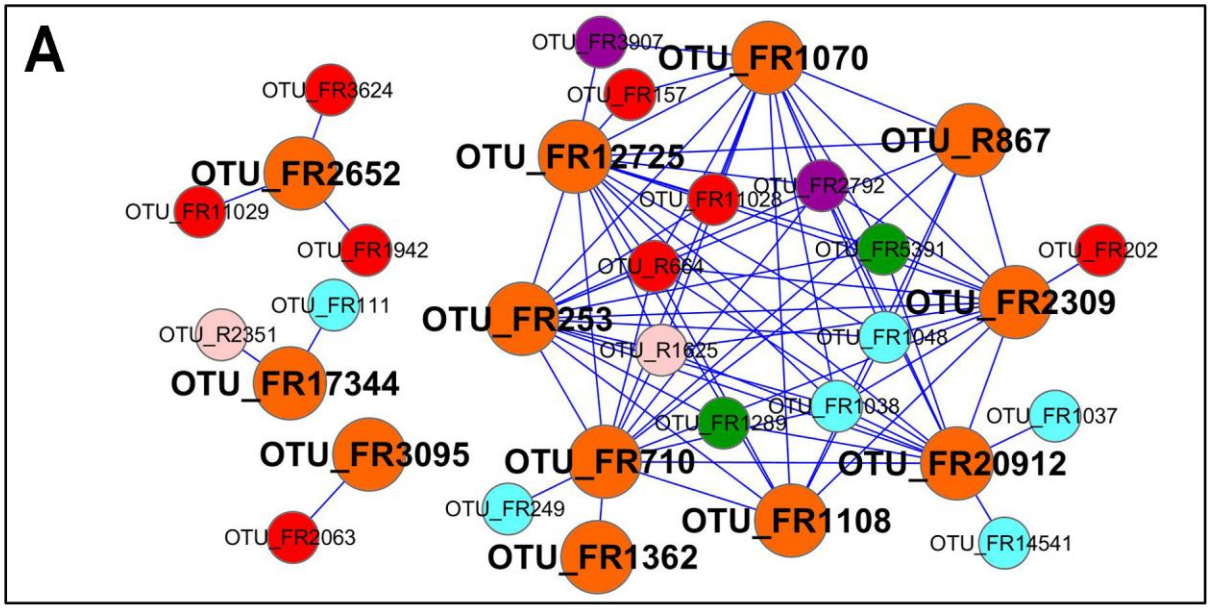
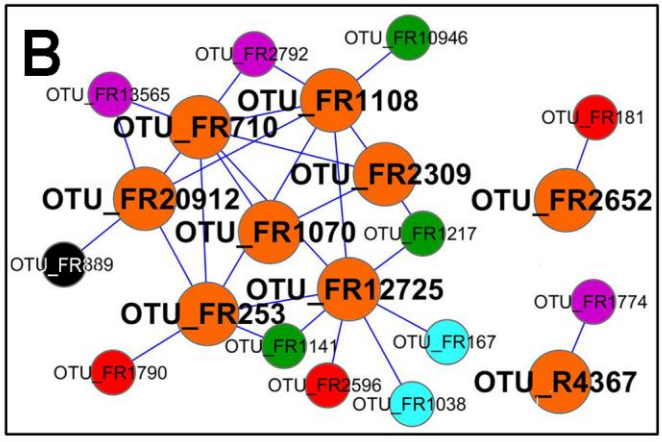


Fig. S2. Phylogenetic distributions of nodes (OTUs) in the network under aCO_2 (blue) and eCO_2 (red). The distribution of OTUs varies substantially among different phylogenetic groups.



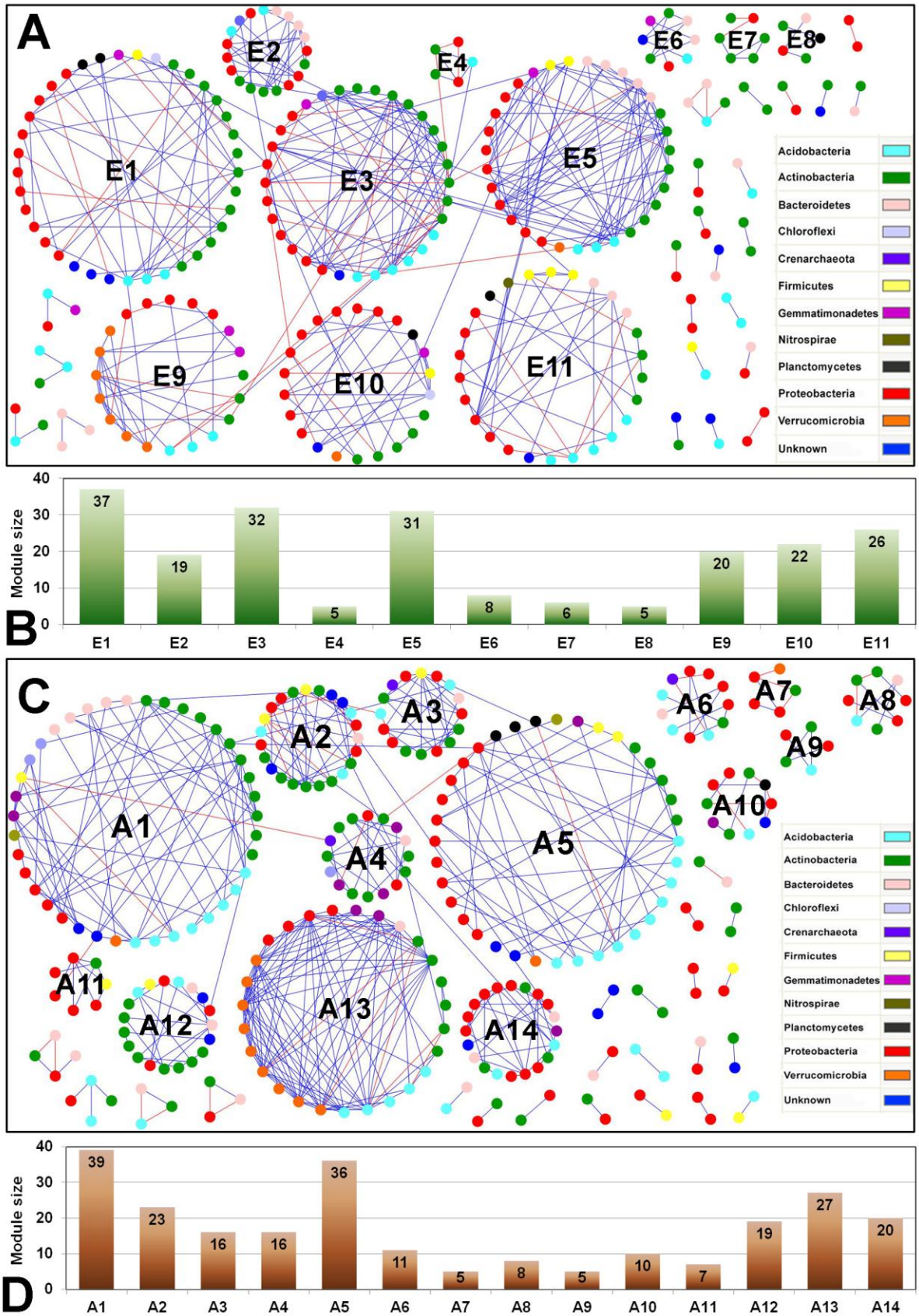


C

Communities	Size (n)	Links	Average connectivity (avgK)	Average path (GD)	Average clustering coefficient (avgCC)
eCO2	21	31	2.95	1.38	0.44
aCO2	32	110	6.88	0.88	0.73

Proteobacteria, Acidobacteria, Verrucomicrobia, Actinobacteria, Gemmatimonadetes, Bacteroidetes, Planctomycetes

Fig. S3. Effects of eCO_2 on the network interactions of *Verrucomicrobia*. All *Verrucomicrobia* nodes and their nearest neighbors were shown under aCO_2 (A) and eCO_2 (B). (C) Summary of several key parameters for these sub-networks topology. *Verrucomicrobia* is a recently described phylum of abundant bacteria with a few described species, including members of the microbial communities of soil and fresh and marine waters. Their physiological roles are not well known. The network interactions under eCO_2 were much simpler than those under aCO_2, indicating that an increased C influx to soil did not favor their interactions with other members in the community. Thus it appears that eCO_2 had negative impacts on the network interaction of this group of bacteria.


A
E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
E11
Acidobacteria
Actinobacteria
Bacteroidetes
Chloroflexi
Crenarchaeota
Firmicutes
Gemmatimonadetes
Nitrospirae
Planctomycetes
Proteobacteria
Verrucomicrobia
Unknown
B
Module size
0
10
20
30
40
37
19
32
5
31
8
6
5
20
22
26
E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
E11
C
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
Acidobacteria
Actinobacteria
Bacteroidetes
Chloroflexi
Crenarchaeota
Firmicutes
Gemmatimonadetes
Nitrospirae
Planctomycetes
Proteobacteria
Verrucomicrobia
Unknown
D
Module size
0
10
20
30
40
39
23
16
16
36
11
5
8
5
10
7
19
27
20
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14

1

Fig. S4. Modular organization of the pMENs with 16S rRNA gene-based metagenomics data. The networks were constructed with the RMT-based approach with the pyrosequencing data from (A) eCO_2 (12 samples) and (C) aCO_2 (12 samples). Colors of the nodes indicate different major phyla. Clear modular architecture was observed in this pMEN. Each node signifies an OTU, which could correspond to a microbial population. Colors of the nodes indicate different major phyla. A blue line indicates a positive interaction between two individual nodes, while a red line indicates a negative interaction. The numbers indicate different modules or submodules determined by the fast greedy modularity optimization method. All data showed that the phylogenetic MENs have a modular architecture. Besides, the sizes for individual modules were plotted in B (eCO_2) and D (aCO_2).

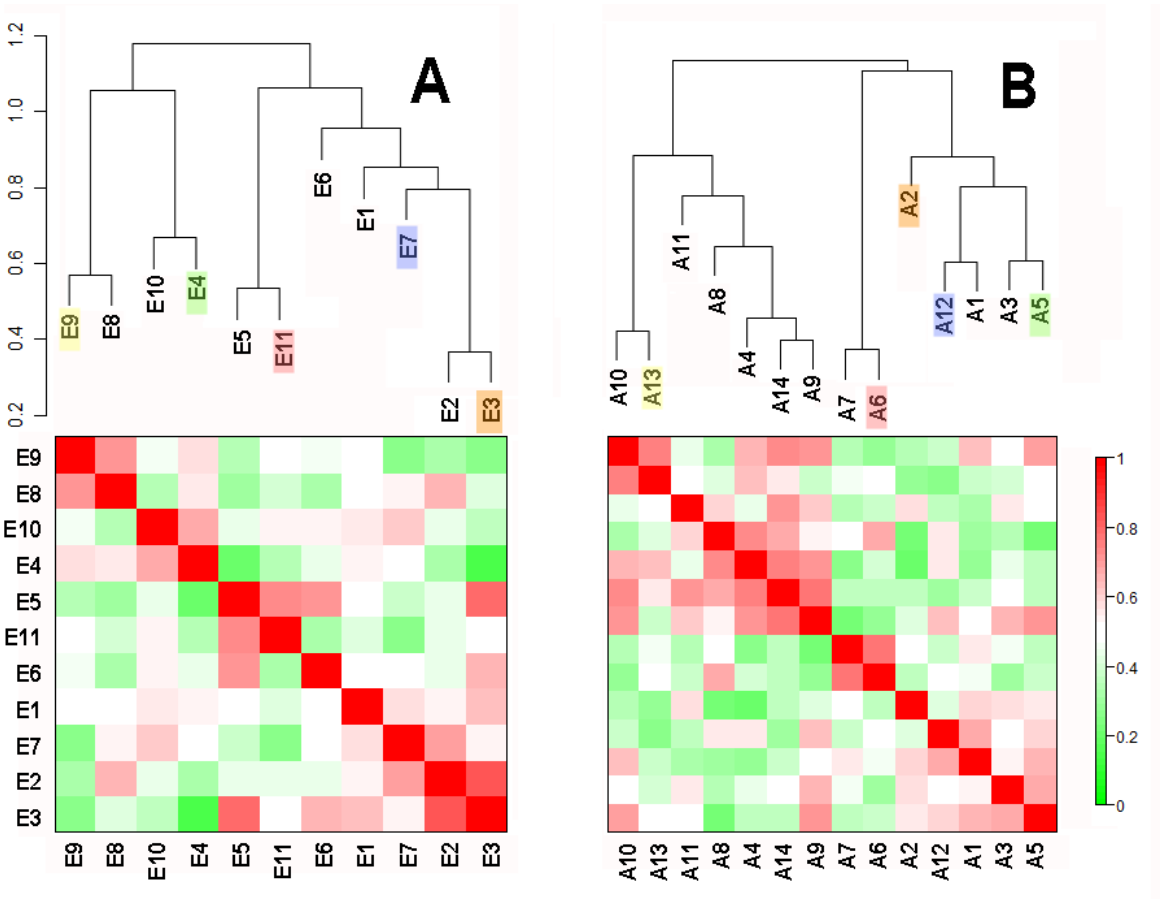


Fig. S5. Module eigengene networks for eCO_2 (A) and aCO_2 (B) pMENs. The upper plots are the hierarchical clustering dendrograms to illustrate the relationships among module eigengenes. The paired modules between aCO_2 and eCO_2 networks were marked in same colors (details in Table S2). The below plots show the correlations among modules. Each color grid represents the signed correlation between the two corresponding eigengenes (Equation 24 in Methods online). Red color means higher correlation whereas green color signified lower correlation.

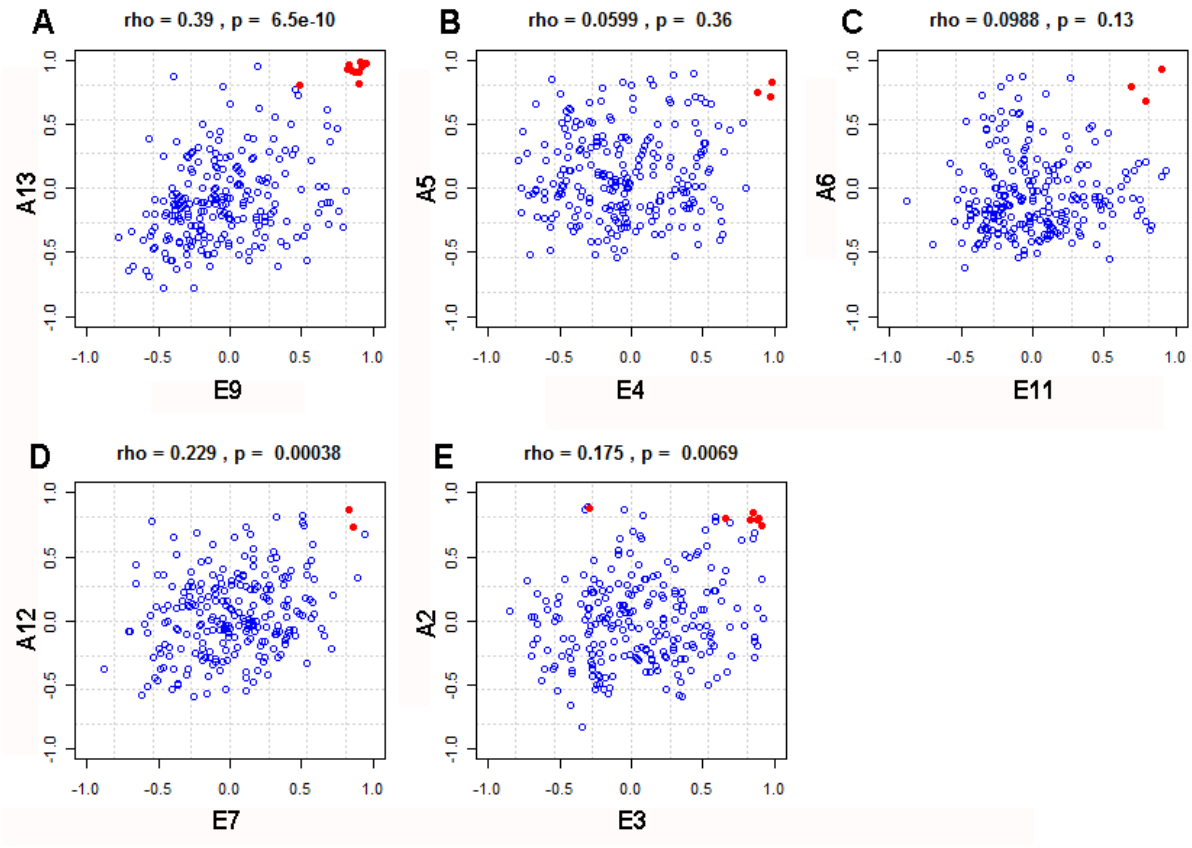


Fig. S6. Module memberships of the shared nodes for the paired modules. Comparison of module memberships between these two networks for Pair 1 (E9-A13) (A), Pair 2 (E4-A5) (B), Pair 3 (E11-A6) (C), Pair 4 (E7-A12) (D) and Pair 5 (E3-A2) (E). Module memberships were determined as the Pearson correlations between the eigengene and the shared nodes in these two networks (See Methods online). Red dots represent the nodes shared between the two modules examined; Blue dots represent the nodes from other modules which were shared between these two networks. Divergent patterns of the module memberships (negative, positive or no relationship) were obtained for the paired modules.