

**Integrated metagenomics and network analysis of soil microbial community of
the forest timberline**

Junjun Ding^{1,2}, Yuguang Zhang^{1,*}, Ye Deng³, Jing Cong¹, Hui Lu¹, Xin Sun¹, Caiyun
Yang³, Tong Yuan³, Joy D Van Nostrand³, Diqiang Li¹, Jizhong Zhou^{2,3,4} and Yunfeng
Yang^{2,*}

¹Institute of Forestry Ecology, Environment and Protection, and the Key Laboratory
of Forest Ecology and Environment of State Forestry Administration, the Chinese
Academy of Forestry, Beijing 100091, China

²State Key Joint Laboratory of Environment Simulation and Pollution Control, School
of Environment, Tsinghua University, Beijing 100084, China

³Institute for Environmental Genomics and Department of Botany and Microbiology,
University of Oklahoma, Norman, OK 73019, USA

⁴Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA
94720, USA

⁵School of Minerals Processing and Bioengineering, Central South University,
Changsha 410083, China

*For GeoChp and network analysis, please contact yangyf@tsinghua.edu.cn; Phone:
+86-010-62784692; Fax: +86-010-(+86) 10-62785687; For field sampling,
environmental parameter measurements and MiSeq sequencing, please contact
zhangyg@caf.ac.cn; Phone: +86-010-62889240; Fax: +86-010-(+86) 10-62884972.

Figure S1. Rarefaction curves of high-throughput sequencing data of the samples from the shrubland and the coniferous forest.

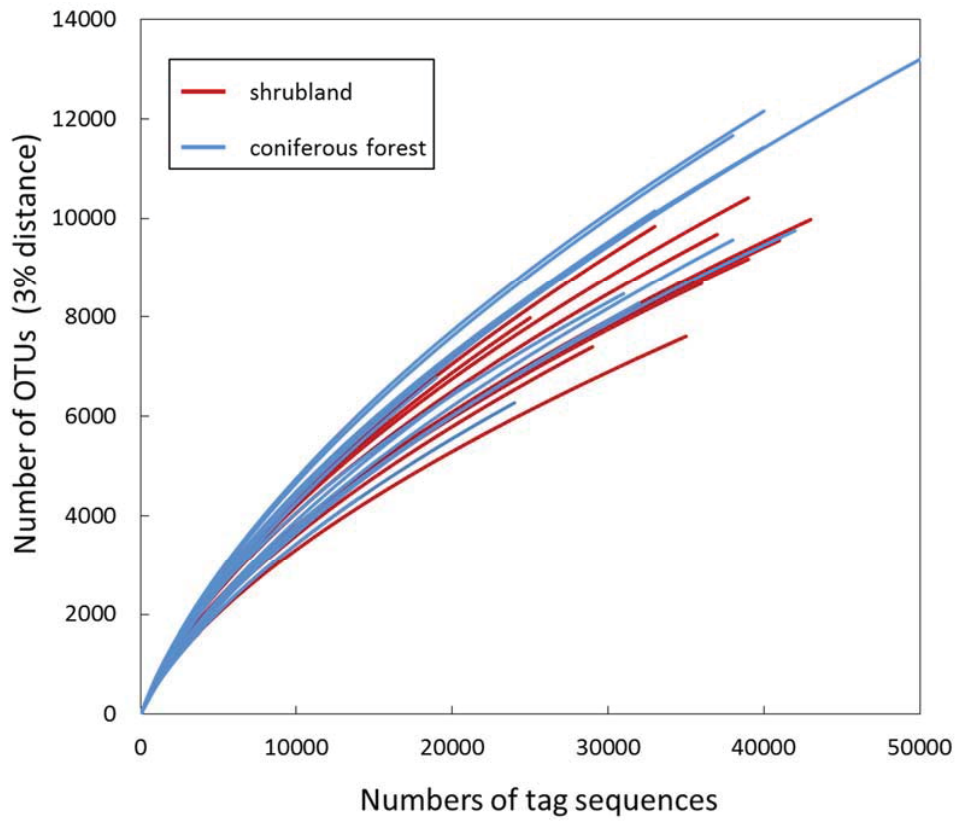


Figure S2. Correlations between (A) microbial biomass (indicated by soil microbial biomass carbon and soil microbial biomass nitrogen), microbial diversity (taxonomic and functional diversity) and NH_4^+ -N content; and (B) microbial α -diversity (taxonomic and functional diversity) and soil microbial biomass nitrogen. Microbial diversity was calculated by Simpson index. White and black dots represent samples from the shrubland and the coniferous forest, respectively.

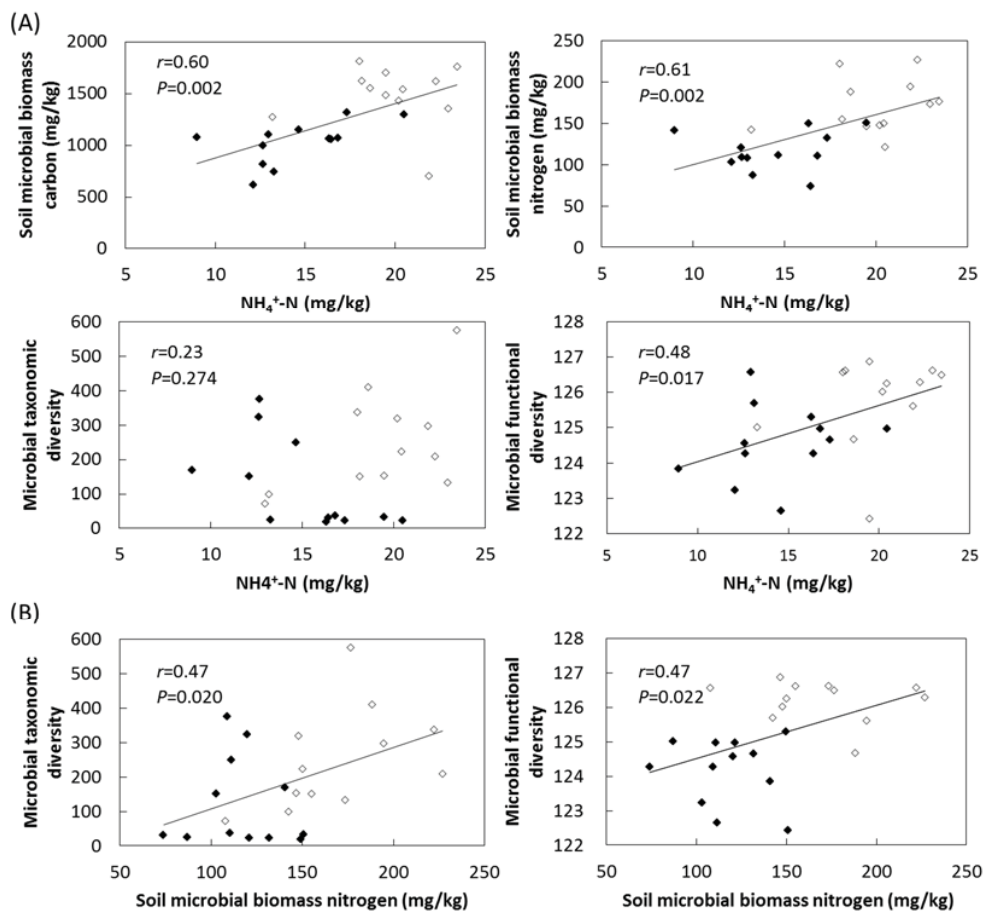


Figure S3. Comparison of the taxonomic distribution of the most important phyla between (A) the shrubland and (B) the coniferous forest.

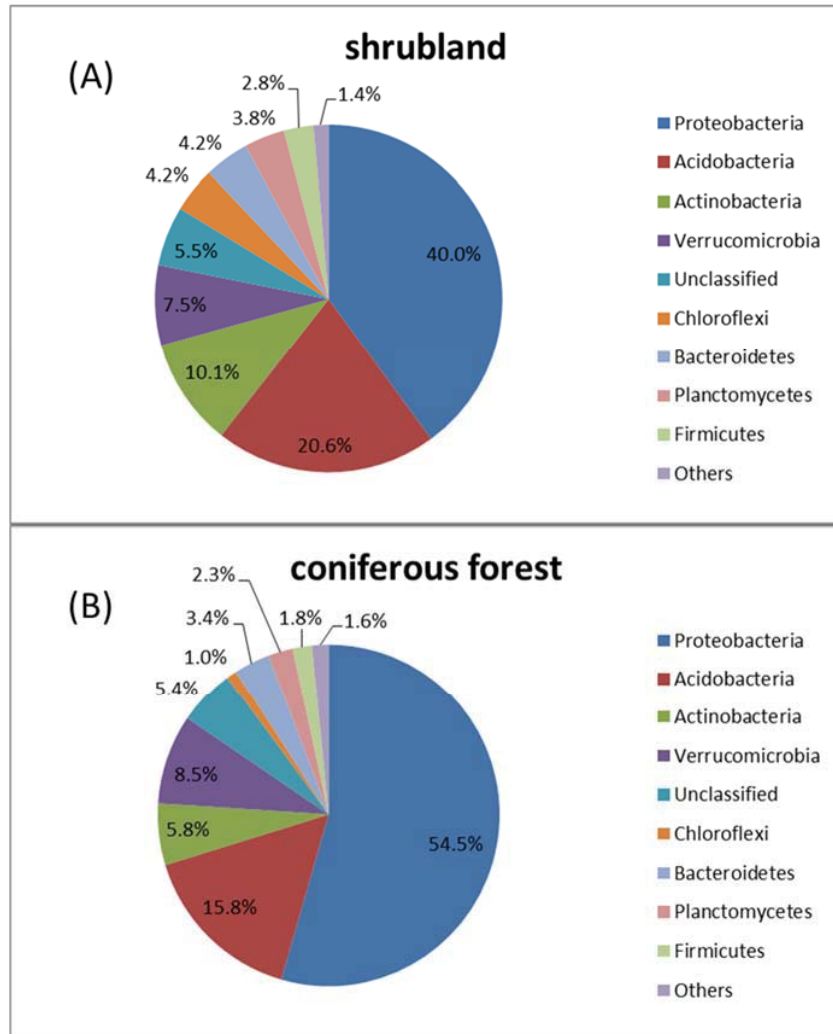


Figure S4. Response ratios to compare DNA abundance of gene categories of the coniferous forest to those associated with the shrubland at the 95% confidence level.

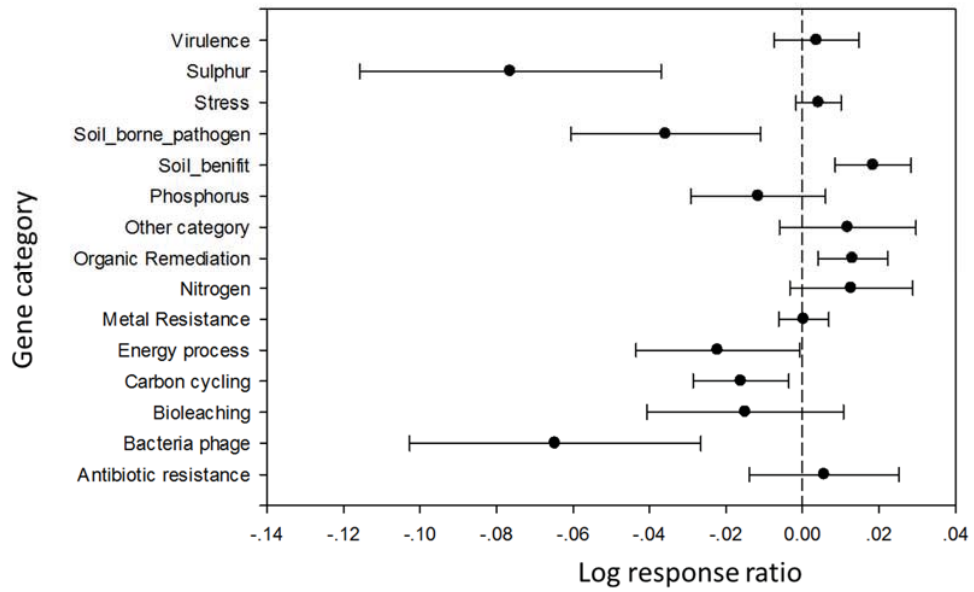


Figure S5. Correlations between (A) $\text{NH}_4^+\text{-N}$ and total abundance of *amoA* and *hao* genes; and (B) $\text{NO}_3^-\text{-N}$ and total abundance of *nirS* and *nirK* genes. White and black dots represent samples from the shrubland and the coniferous forest, respectively.

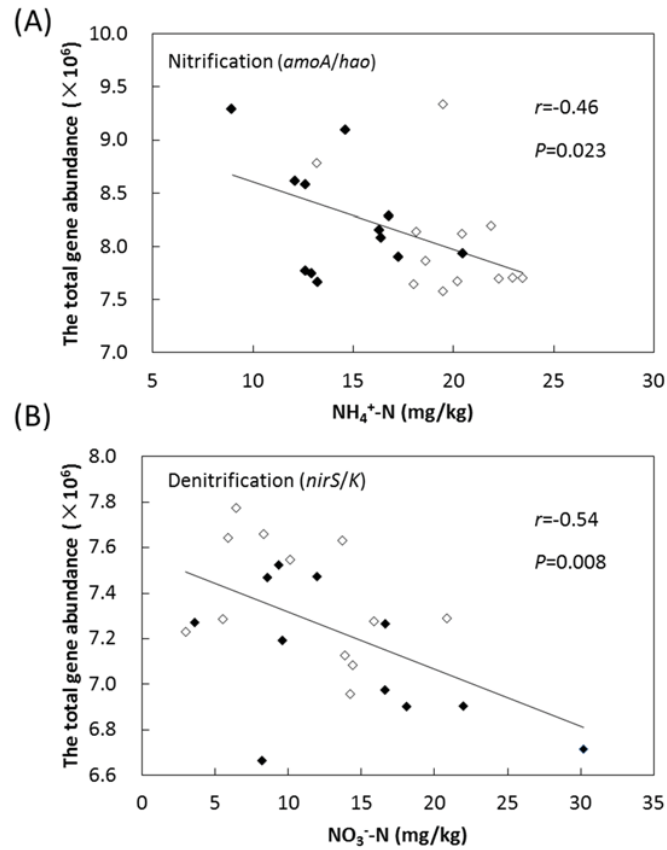


Figure S6. Response ratios to compare DNA abundance of the coniferous forest to those associated with the shrubland at the 95% confidence level for the gene category of (A) carbon cycle and (B) phosphorus cycle.

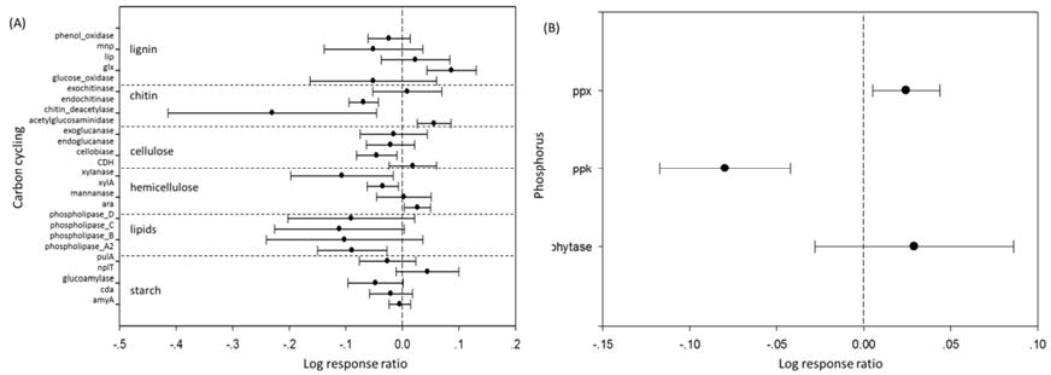


Figure S7. Z-P plot showing the distribution of nodes based on their topological roles.

Each symbol represents (A) an OTU or (B) a functional gene in the coniferous forest (blue) and the shrubland (red) samples. The topological role of each node was determined according to within-module connectivity (Z_i) and among-module connectivity (P_i).

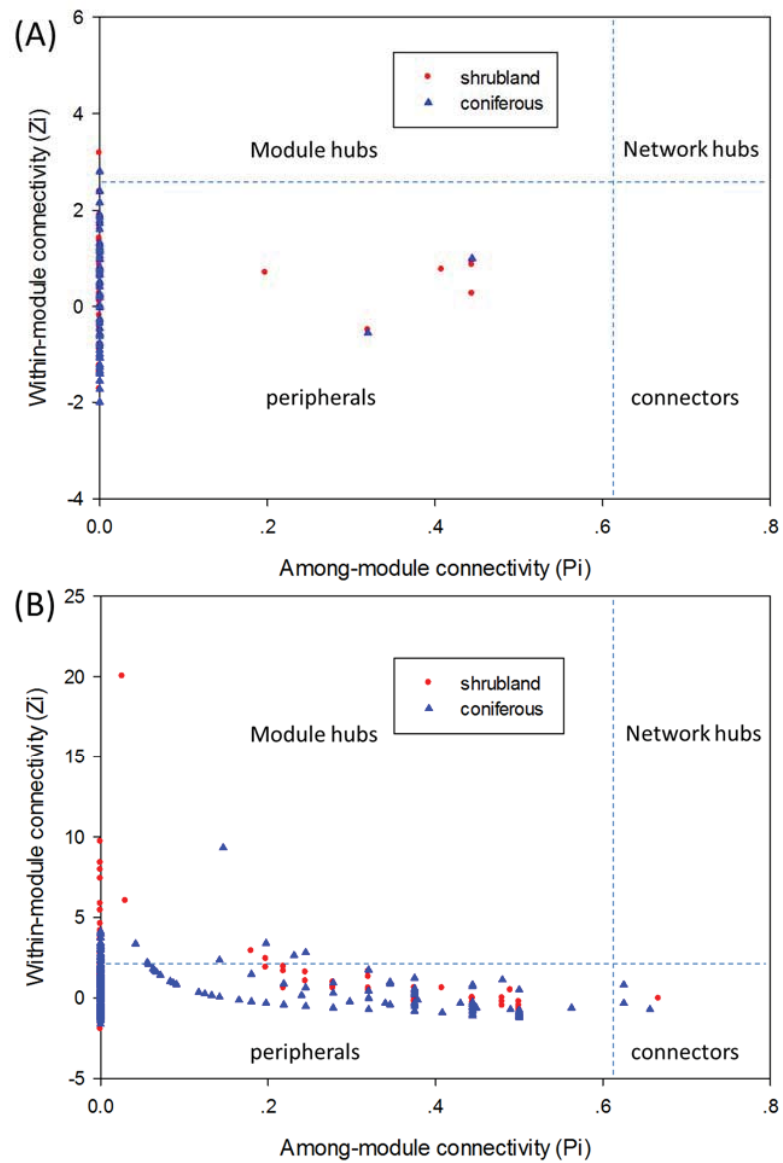


Table S1. Topological properties of networks.

Networks	Empirical networks										Random networks ^d		
	Community	Number of original genes ^a	Similarity threshold (S)	network size(n) ^b	R square of scale free	Modularity	Avg connectivity (avgK)	Avg clustering coefficient (avgCC)	Avg path length (GD) ^c	Transitivity (Trans)	Avg clustering coefficient (avgCC)	Avg path length (GD)	Transitivity (Trans)
OTUs	Shrubland	1287	0.86	732	0.95	0.85	3.31	0.20	2.55	0.52	0.012±0.003	3.67±0.14	0.026±0.003
	Coniferous forest	1036	0.86	641	0.89	0.79	5.51	0.28	5.24	0.58	0.032±0.004	3.45±0.08	0.056±0.003
Functional genes	Shrubland	7530	0.95	2431	0.75	0.93	2.25	0.13	4.06	0.02	0.014±0.002	2.81±0.09	0.004±0.001
	Coniferous forest	6176	0.95	1724	0.93	0.78	3.41	0.19	1.74	0.35	0.011±0.002	3.68±0.10	0.024±0.002

^aThe number of genes originally used for network construction using the RMT-based approach.

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

^cGD, geodesic distance

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