Forests

Effects of soil microbes on forest recovery to climax community through the regulation of nitrogen cycling

Dandan Qi [1], Fujuan Feng [1,*], Yanmei Fu [1], Ximei Ji [1], Xianfa Liu[1]

[1] College of Life Science, Northeast Forestry University, Harbin 150040, China

*e-mail: ffj9018@nefu.edu.cn

Supplementary Material

1. Determination of soil physicochemical properties

Soil moisture content (Mc) was calculated after drying for at least 8 hours in an incubator at 105°C. Soil pH was determined in a 1:2.5 soil/water suspension. The soil bulk density (Bd) was measured using a soil cutting ring (100 cm^3). Total C and N were determined by combustion on a Vario EL III elemental analyzer (Elementar, Germany). Soil organic matter (SOM) was determined by the potassium dichromate titration method. NH_4^+ and NO_3^- concentrations were determined using an AA3 continuous flow analyzer (Seal Analytical, Germany) following 2 M KCl solution extraction. Soil total phosphorus (TP) and available phosphorus (AP) were determined using an AA3 continuous flow analyzer (Seal Analytical, Germany). Dissolved organic carbon (DOC) was extracted with 0.5 M potassium sulfate (K_2SO_4) solution by shaking for 30 min and was analyzed using a Multi 3100 N/C TOC analyzer (Analytik Jena, Germany).

2. Sequence quality control, assembly, gene prediction, and annotation

The 3' and 5' end adapter sequence were removed by SeqPrep (https://github.com/jstjohn/SeqPrep). The low-quality reads (a quality value <20 or containing N bases) and the sequences with lengths of less than 50 bp after splicing were removed by Sickle (https://github.com/najoshi/sickle). These clean reads were used for subsequent analysis.

The sequencing data were assembled using Megahit (https://github.com/voutcn/megahit, Multiple_Megahit) [1] and Newbler (https://ngs.csr.uky.edu/Newbler). Contigs with lengths of longer than 300 bp were selected as the final assembling results for subsequent gene prediction.

Open reading frames (ORFs) from each assembled contig were predicted using MetaGene [2] (http://metagene.cb.k.u-tokyo.ac.jp/). The predicted ORFs with lengths of longer than 100 bp were translated into amino acid sequences. The predicted gene sequences of all samples were clustered using CD-HIT (http://www.bioinformatics.org/cd-hit/, default parameters: 95% identity, 90% coverage) [3]. The longest sequence from each cluster was selected as the representative sequence to construct a non-redundant gene catalogue. Clean reads were mapped to the non-redundant gene catalogue with 95% identity using SOAPaligner (http://soap.genomics.org.cn/) to evaluate gene abundance in each sample [4]. In this study, we used the method of reads per kilobase per million mapped reads (RPKM) to calculate gene abundance, which reduced the impact of different sequencing depths and gene lengths between samples based on gene abundance [5].

The non-redundant gene catalogue was aligned to NCBI NR database with an e-value cutoff of $1e^{-5}$ using BLASTP (BLAST Version 2.2.28+, http://blast.ncbi.nlm.nih.gov/Blast.cgi) for taxonomic annotations [6] and calculating the microbial relative abundance in each sample at the corresponding taxonomy level. The functional annotation was performed using BLASTP (BLAST Version 2.2.28+, http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) database, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and the VFDB database with an e-value cutoff of $1e^{-5}$.

References

1. Li, D.;Liu, C.-M.;Luo, R.;Sadakane, K.; Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. **2015**, *31(10)*, 1674-1676.
2. Noguchi, H.;Park, J.; Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res*. **2006**, *34(19)*, 5623-5630.
3. Fu, L.;Niu, B.;Zhu, Z.;Wu, S.; Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **2012**, *28(23)*, 3150-3152.
4. Li, R.;Li, Y.;Kristiansen, K.; Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics*. **2008**, *24(5)*, 713-714.
5. Lawson, C. E.;Wu, S.;Bhattacharjee, A. S.;Hamilton, J. J.;McMahon, K. D.;Goel, R.; Noguera, D. R. Metabolic network analysis reveals microbial community interactions in anammox granules. *Nature Communications*. **2017**, *8(1)*, 15416.
6. Altschul, S. F.;Madden, T. L.;Schäffer, A. A.;Zhang, J.;Zhang, Z.;Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **1997**, *25(17)*, 3389-3402.

Supplementary tables (Table S1 - S6)

Table S1 Results of Student's t-test of the soil properties between the two forest types in July 2018

	PF	SF	*p*
SOM (%)	21.11	18.27	**
TN ($g.kg^{-1}$)	8.05	8.93	*
TP ($g.kg^{-1}$)	0.72	0.94	***
C/N	16.08	11.83	*
N/P	11.18	9.53	**
DOC ($mg.kg^{-1}$)	122.73	79.98	**
AP ($mg.kg^{-1}$)	15.19	30.53	**
NH_4^+ ($mg.kg^{-1}$)	11.93	6.39	***
NO_3^- ($mg.kg^{-1}$)	2.48	6.32	***
Mc (%)	47.57	46.84	*
Bd ($g.cm^{-3}$)	1.43	1.17	**
pH	5.80	5.86	**

The data in the table are average values. PF: primary Korean pine forests, SF: secondary broad-leaved forests. SOM: soil organic matter, TN: total nitrogen, TP: total phosphorus, DOC: dissolved organic carbon, AP: available phosphorus, NH_4^+: ammonium nitrogen, NO_3^-: nitrate nitrogen, Mc: moisture content, Bd: soil bulk density. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S2 The significance of differences in inorganic N content and N transformation rates between the two forest types assessed by Student's t-test

	May	Jun	Jul	Aug	Sep	Oct
NH_4^+ ($mg \cdot kg^{-1}$)	*	**	***	**	*	**
NO_3^- ($mg \cdot kg^{-1}$)	ns	ns	***	**	***	*
R_{amm} ($mg \cdot kg^{-1} \cdot d^{-1}$)	***	***	***	**	***	***
R_{nit} ($mg \cdot kg^{-1} \cdot d^{-1}$)	***	***	**	ns	*	***
R_{min} ($mg \cdot kg^{-1} \cdot d^{-1}$)	**	***	***	ns	***	***

R_{amm}: net ammonification rate. R_{nit}: net nitrification rate. R_{min}: net mineralization rate. ns: no significant difference. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S3 Two-way ANOVA analysis results of NH_4^+, NO_3^-, net ammonification rate, net nitrification rate and net mineralization rate

Difference source	NH_4^+		NO_3^-		Net ammonification rates（R_{amm}）		Net nitrification rates（R_{nit}）		Net mineralization rates（R_{min}）	
	F	*p*	F	*p*	F	*p*	F	*p*	F	*p*
Forest types (T)	136.67	***	310.18	***	43.15	***	209.67	***	284.74	***
Months (M)	194.31	***	27.86	***	331.98	***	47.59	***	61.89	***
T*M	56.53	***	45.24	***	84.55	***	35.87	***	76.95	***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S4 Comparative analysis of the relative abundance of Archaea, Bacteria, Eukarya, Viruses, and Unclassified in primary Korean pine forests and secondary broad-leaved forests

Domain	PF% (Mean ± SD)	SF% (Mean ± SD)	*p*-value
Bacteria	98.92 ± 0.01	98.48 ± 0.11	*
Archaea	0.70 ± 0.01	1.09 ± 0.05	**
Eukarya	0.19 ± 0.00	0.22 ± 0.06	ns
Viruses	0.03 ± 0.00	0.02 ± 0.00	**
Unclassified	0.16 ± 0.01	0.19 ± 0.00	*

PF: primary Korean pine forests, SF: secondary broad-leaved forests. ns: no significant difference. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S5 The comparison of orders with the contribution greater than 1% to N cycling between the two forest types

Name	PF-Mean (%)	PF-Sd (%)	SF-Mean (%)	SF-Sd (%)	*p*-value
o_Rhizobiales	24.24	0.76	19.42	0.41	**
o_Solirubrobacterales	5.32	0.03	5.34	0.14	ns
o_Corynebacteriales	5.96	0.06	4.01	0.03	***
o_Burkholderiales	4.50	0.08	5.22	0.09	**
o_Rhodospirillales	4.16	0.08	2.65	0.02	***
o_Pseudonocardiales	2.61	0.25	2.75	0.09	ns
o_unclassified_d_Bacteria	2.25	0.05	2.88	0.09	**
o_unclassified_c_Betaproteobacteria	2.14	0.22	2.85	0.11	*
o_Myxococcales	2.30	0.07	2.61	0.04	**
o_Streptosporangiales	2.72	0.18	2.06	0.10	*
o_Nitrospirales	1.78	0.17	2.79	0.14	**
o_unclassified_p_Acidobacteria	1.82	0.14	2.52	0.10	**
o_Streptomycetales	2.19	0.20	2.07	0.05	ns
o_Acidobacteriales	2.03	0.10	2.03	0.11	ns
o_unclassified_p_Candidatus_Rokubacteria	1.30	0.05	1.87	0.02	**
o_unclassified_p_candidate_division_NC10	1.27	0.08	1.74	0.07	**
o_Propionibacteriales	1.58	0.03	1.38	0.05	**
o_Planctomycetales	1.26	0.03	1.50	0.02	**
o_unclassified_c_Actinobacteria	1.26	0.12	1.33	0.06	ns
o_Micrococcales	1.25	0.08	1.34	0.04	ns
o_Gemmatimonadales	1.04	0.09	1.50	0.05	**

PF: primary Korean pine forests, SF: secondary broad-leaved forests. ns: no significant difference. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S6 Permutational multivariance analysis of variance (PERMANOVA) on the explanation of soil physicochemical properties to the differences in soil microbial community structure and functional composition (including microbes involved in N cycling)

	Module	species	N-Module	N-species
AP	0.90 **	0.89 *	0.83	0.57 *
TP	0.91 *	0.91 *	0.91	0.59 *
TN	0.66	0.68 *	0.72 *	0.48 *
SOM	0.89 *	0.87	0.88	0.57
NH_4^+	0.94 **	0.94 **	0.92	0.60 **
NO_3^-	0.93 *	0.92 *	0.94 *	0.59
DOC	0.77	0.78	0.87 *	0.53
Mc	0.57	0.56	0.61	0.41
Bd	0.79	0.80	0.90 **	0.53
PH	0.89 *	0.88	0.91 *	0.58 *

N-Module: Module related to N cycling. N-species: species involved in N cycling. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Supplementary figures (Figure S1 - S7)

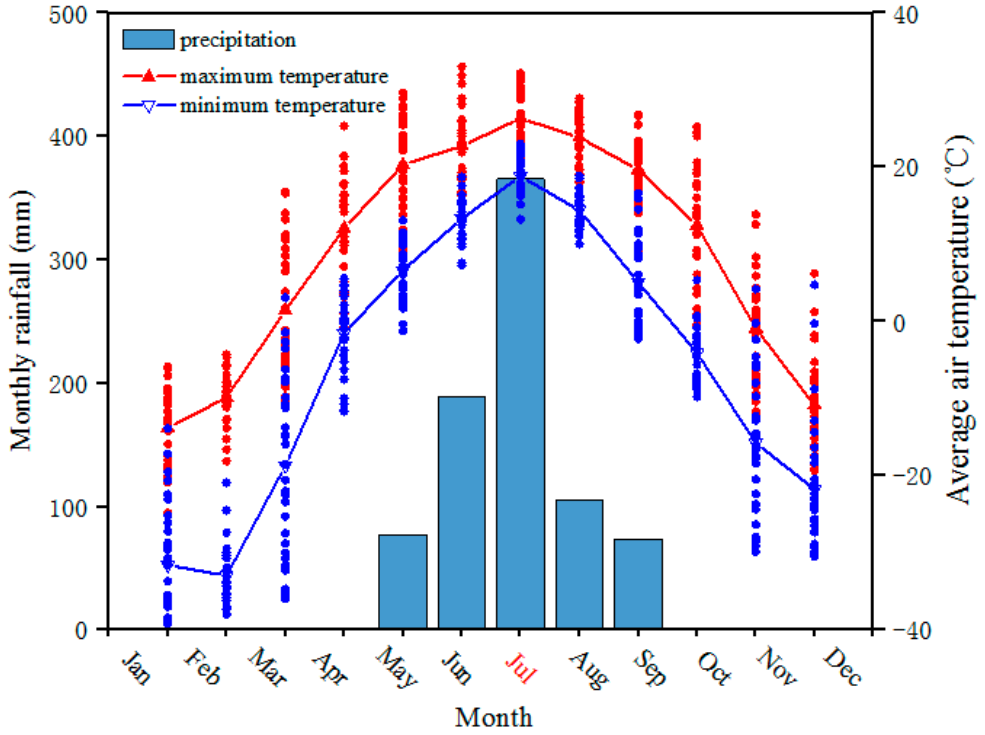

Figure S1. Climatic conditions of the study sites in 2018. Line chart represents air temperature. The monthly highest and lowest mean temperatures are coloured red and blue, respectively. The histogram represents rainfall from May to September

Figure S2. *In situ* N mineralization experiment in study sites

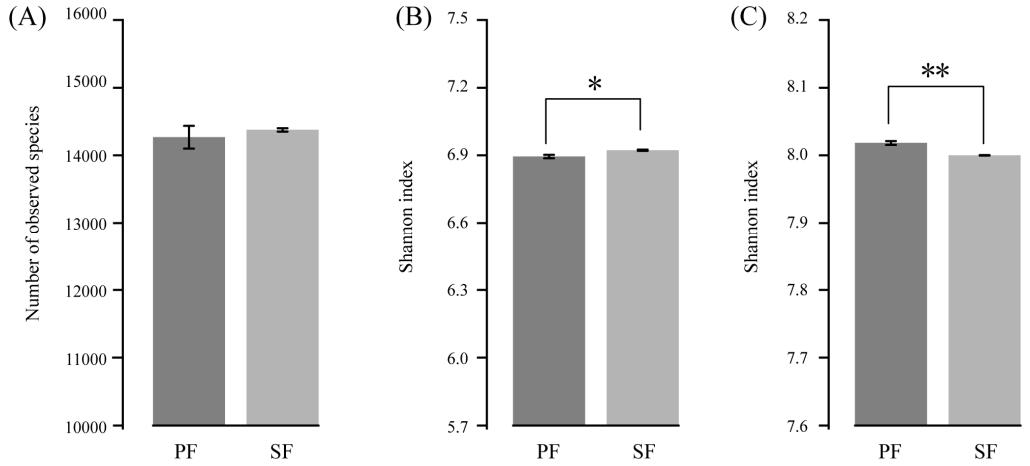

Figure S3. The histogram characterizes the differences of soil microbial richness (Species) (A), community diversity (Species) (B), and functional diversity (eggNOG) (C) between the two forest types. Dark grey and light grey represent the PF and SF, respectively. Asterisks indicate the significant differences of variable means between the PF and SF. PF: primary Korean pine forests, SF: secondary broad-leaved forests

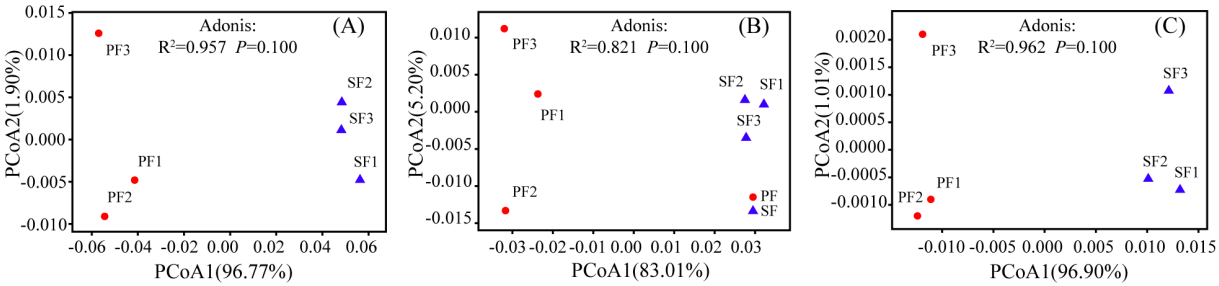

Figure S4. The PCoA of soil microbial community structure (genus) (A), functions (genes, eggNOG) (B), and metabolism pathways (level 3, KEGG) (C) between the two forest types based on Bray-Curtis distances. PF: primary Korean pine forests, SF: secondary broad-leaved forests

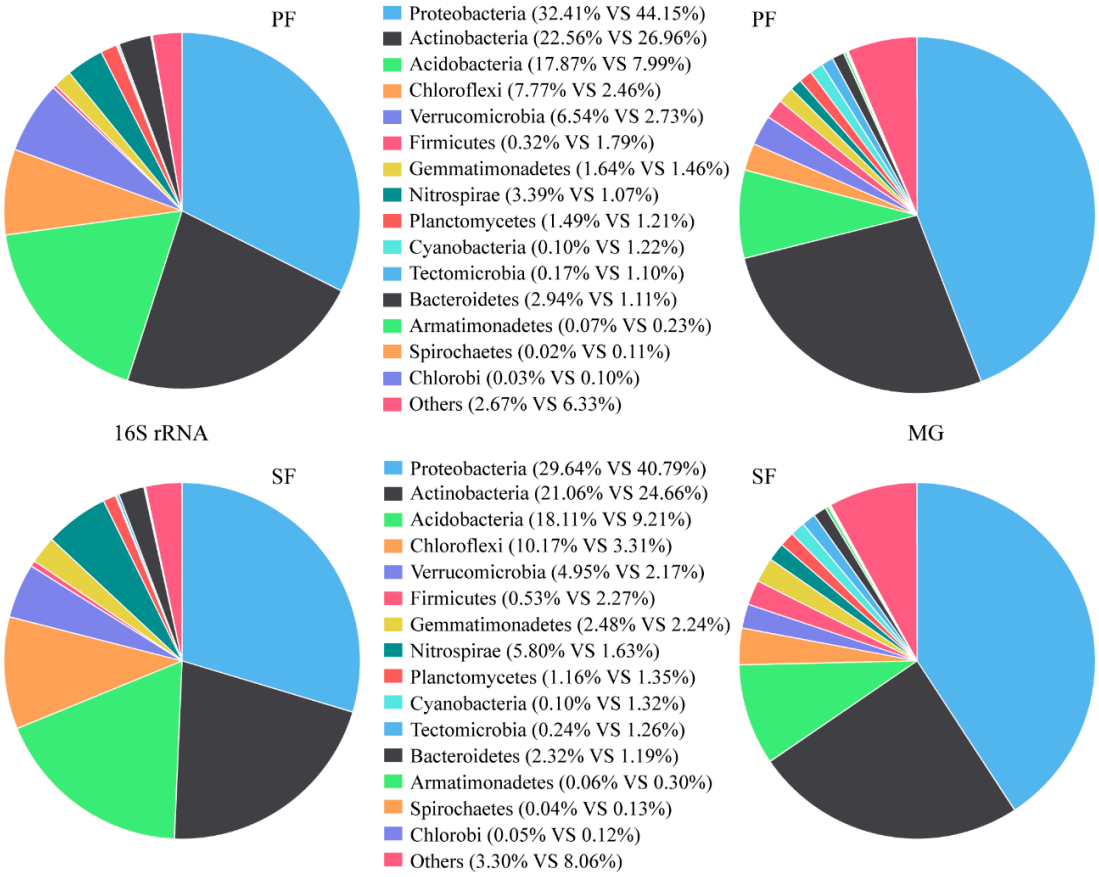

Figure S5. Pie plot of the average composition resulting from 16S rRNA sequencing data and metagenome sequencing data (MG). The percentage values in the legend are the relative abundance of the primary Korean pine forests (PF, up) and the secondary broad-leaved forests (SF, down), respectively

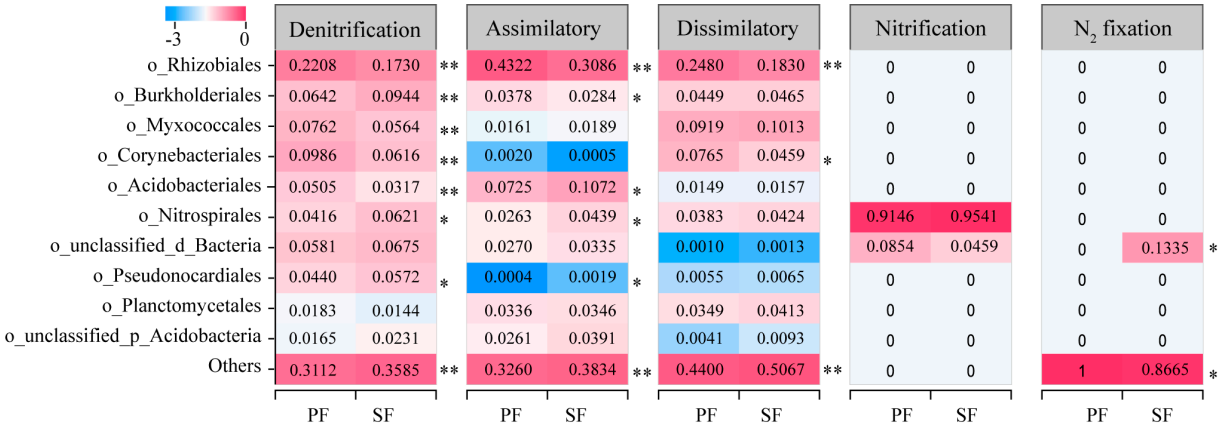

Figure S6. The heatmap of species (order) and functional contribution analysis. The comparison of taxonomic groups involved in the major N cycling processes between PF and SF, following by Welch's t-test with Benjamini-Hochberg FDR correction. Only the top 10 most abundant microbial orders are shown in the figure. Values indicate the relative abundance of order species in PF and/or SF. PF: primary Korean pine forests, SF: secondary broad-leaved forests. Assimilatory: assimilation nitrate reduction, Dissimilatory: dissimilatory nitrate reduction. $p < 0.05$ *， $p < 0.01$ **， $p < 0.001$ ***

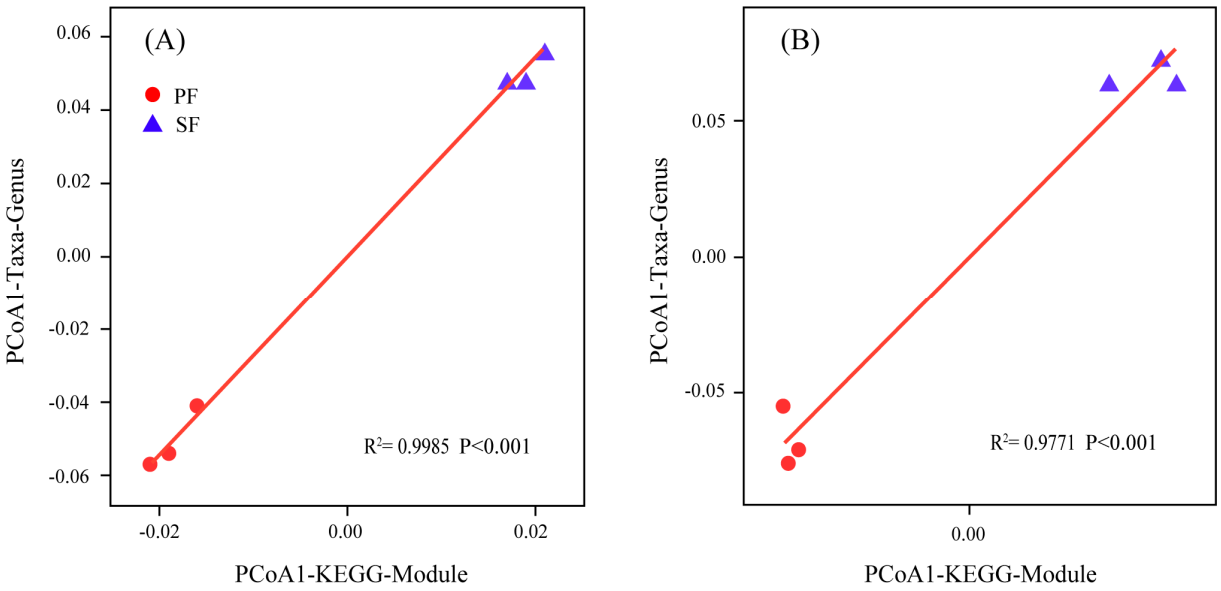

Figure S7. Linear regression analysis of the correlation of soil microbial functional (KEGG Module) similarity and community composition (genus) similarity, based on β diversity. (A) soil microbes (B) soil microbes involved in N cycling. The higher determination coefficient R^2, the higher consistency of β diversity between community and function