Supporting information

Experiment design

The long-term experiment was initiated in 1982 at Madian Agro-Ecological Station in Mengcheng county, Anhui province, China (33°13′N, 116°35′E). The soil type is Calcic Kastanozems. Five treatments were involved in this study, including (1) control (no fertilization); (2) NPK (applied with mineral NPK fertilizer only); (3) NPK+HS (NPK with wheat straw); (4) NPK+PM (NPK with pig manure); (5) NPK+CM (NPK with 30000 kg fresh cow manure). Each treatment contained four replicate plots with the size of 70 m^2 (14.9 m × 64.7 m). Mineral N, P and K fertilizers were applied as urea, calcium superphosphate and potassium chloride with the amount of 180 kg N, 90 kg P_2O_5, and 135 kg K_2O ha^{-1} y^{-1}, respectively. The amount of wheat straw, pig and cow manure were 7500, 15000, and 30000 kg ha^{-1} y^{-1}, respectively. All the treatments were subjected to wheat-soybean crop rotation with same agricultural management practices.

Bioinformatic analysis

Quality control and assembly were performed using MetaWRAP (Uritskiy et al., 2018). MetaWRAP-Read-qc module was run to trim the raw reads and remove human contamination. Then the clean reads were assembled using metaWRAP-Assembly module with the method of MegaHIT (Li et al., 2015). Gene identification of the metagenomic sequences were carried out using

MetaGeneMark (Zhu et al., 2010). Functional annotation of the predicted genes were performed using eggNOG-mapper v2 (Cantalapiedra et al., 2021). Then the genes involved in nitrogen cycle were extracted. The taxonomic annotation of these genes was performed using the Basic Local Alignment Search Tool (BLAST) with the nucleotide sequence (nt) database. The relative abundance of each gene was assessed by RPKM (Reads Per Kilobase Million) with some modification which calculated by the formula:

$$\mathrm{RPKM} = \frac{\text{Number of reads mapped to gene} \times 10^3 \times 10^2}{\text{Total number of sequence reads} \times \text{gene length in bp}}$$

Hear, the number of reads mapped to gene were determined by Bowtie2 (Langmead and Salzberg, 2012).

Statistical analysis

Statistical analysis was performed using R program (version 4.0.5). Kruskal-Wallis rank sum test was performed to check the significance of the difference of variate between treatments using the “dplyr” library. Response ratio analysis was performed using the “ARPobservation” library to compare the effect size of treatment on relative abundance of the functional genes with a 95% confidence interval. PCoA (principal co-ordinates analysis) based on Bray–Curtis distance was performed using the “ape” library to show the community clustering between treatments. Beta-null deviation was calculated to assess the roles of deterministic and stochastic processes in microbial community assembly following the reference (Tucker et al., 2016). Correlations

between microbial communities and soil properties were assessed by Mantel test using “vegan” library. Partial redundancy analysis (RDA) was carried out using “vegan” library to partition the contribution of each soil property in explaining the variation in microbial community. The significance of partial RDA models was assessed by ANOVA based on 999 permutations. SPIEC-EASI (Sparse InversE Covariance Estimation for Ecological Association Inference) was used to infer microbial ecological network at genus level in this soil (Kurtz et al., 2015). Only the genera of each gene detected in no less than 80% of the samples were involved in eco-network construction. In order to distinguish the ecological clusters based on microbial habitat preferences, the genus with positive connections were extracted to re-construction the eco-network. The characteristics of the network were analyzed using the “igraph” library.

Reference

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Molecular Biology and Evolution, 38, 5852-5829.

Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A., 2015. Sparse and compositionally robust inference of microbial ecological networks. PLOS Computational Biology 11, e1004226.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie

2. Nature Methods 9, 357-359.

Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674-1676.

Tucker, C.M., Shoemaker, L.G., Davies, K.F., Nemergut, D.R., Melbourne, B.A., 2016. Differentiating between niche and neutral assembly in metacommunities using null models of beta-diversity. Oikos 125, 778-789.

Uritskiy, G.V., DiRuggiero, J., Taylor, J., 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6, 158.

Zhu, W., Lomsadze, A., Borodovsky, M., 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Research 38, e132.

Table S1 Soil properties in different fertilization regimes

	Control	NPK	NPK+HS	NPK+PM	NPK+CM
pH	6.51(0.14) d	5.02(0.09) c	5.02(0.03) c	6.24(0.02) b	7.23(0.04) a
TC	5.94(0.14) d	8.08(0.60) c	11.35(0.32) b	12.5(0.77) b	21.83(1.83) a
TN	0.72(0.09) d	0.85(0.03) c	1.46(0.18) b	1.6(0.06) b	2.27(0.13) a
NH_4^-–N	4.57(0.67) c	6.95(1.67) ab	7.80(0.86) b	6.16(0.75) a	4.79(1.23) ac
NO_3^-–N	1.16(1.02) c	11.26(1.02) a	6.28(1.49) b	4.99(0.89) b	10.97(1.81) a
DOC	48.07(11.95) d	137.81(33.08) c	148.79(49) bc	286.49(34.59) a	94.26(14.62) b
DON	4.11(1.46) c	5.97(0.54) c	8.54(1.10) b	6.21(0.72) ac	9.16(2.49) ab
AP	2.23(0.43) c	24.21(1.16) b	26.97(1.82) b	138.19(15.24) a	109.55(13.11) a
AK	111.59(3.93) c	109.03(3.33) c	171.45(2.78) b	170.84(3.00) b	582.74(47.44) a

The average of four replicate with standard deviation in bracket.

Same letters indicate no significant difference between treatments checked by Kruskal–Walis rank sum test (P> 0.05).

Control, non-fertilization; NPK, chemical fertilization; NPK+HS, NPK with wheat straw; NPK+PM, NPK with pig manure; NPK+CM, NPK with cow manure.

Table S2Parameters of the ecological network

Parameter	Value
Number of nodes	156
Clustering coefficient	0.408
Connected components	8
Network density	0.043
Network diameter	11
Network heterogeneity	0.869
Network radius	1
Network centralization	0.133
Characteristic path length	3.507
Average number of neighbors	6.615

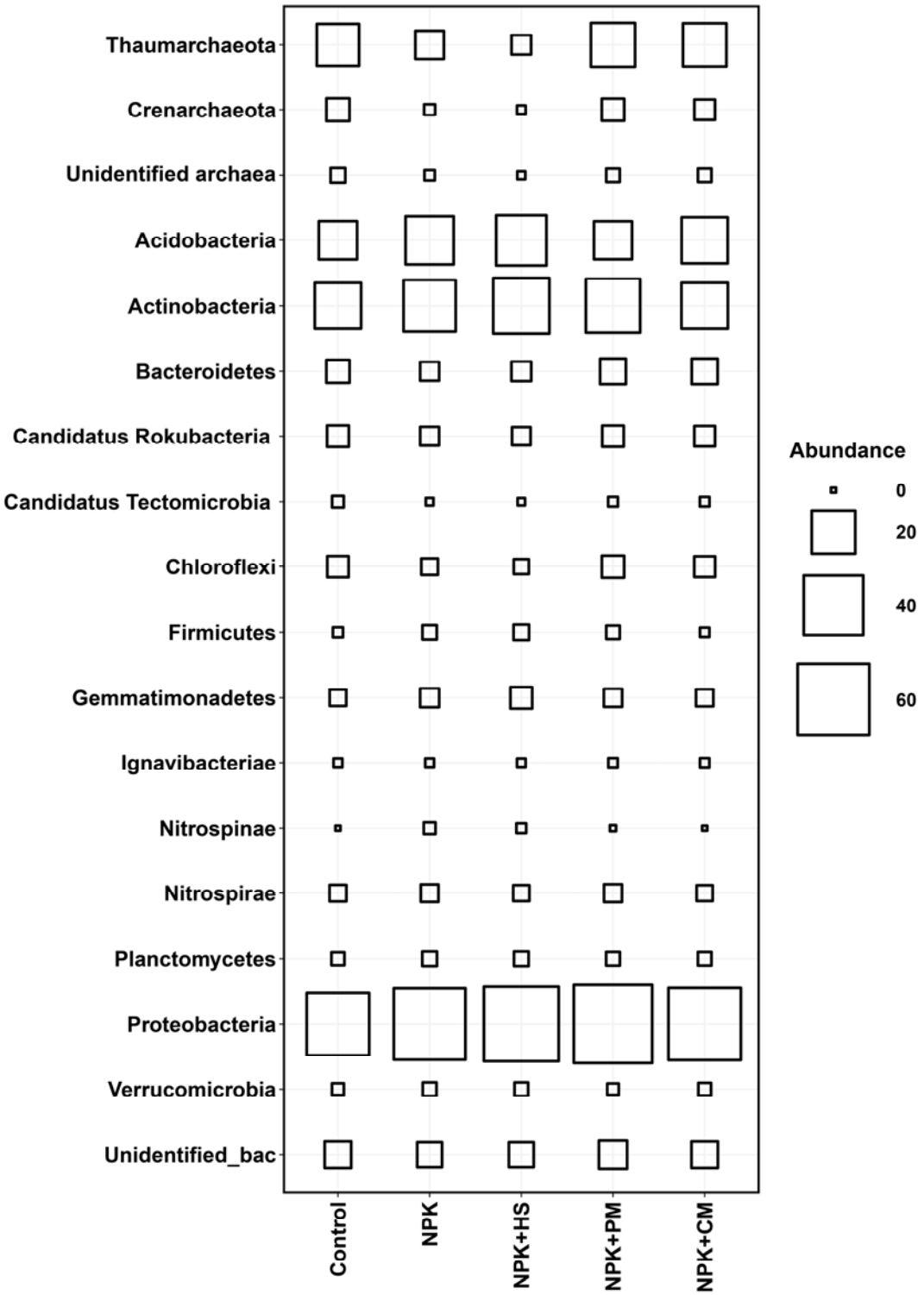

Figure S1.Taxonomic composition of the microbial communities involved in nitrogen cycling at phylum level. Control, non-fertilization; NPK, chemical fertilization; NPK+HS,NPK with wheat straw; NPK+PM, NPK with pig manure;NPK+CM, NPK with cow manure.

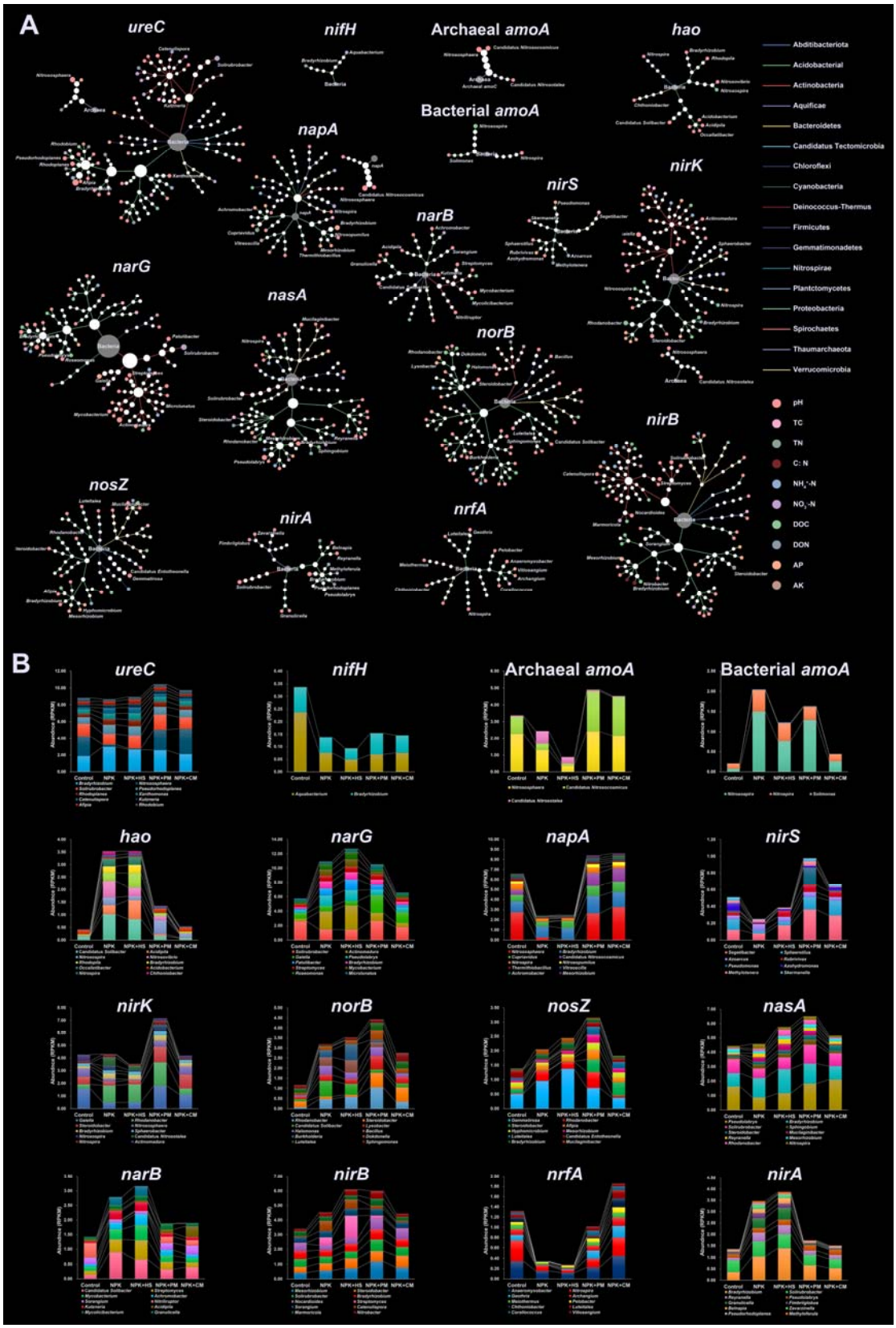

Figure S2.(A) Taxonomic distribution of functional genes involved in N-cycling.

For each gene, the nodes in the tree from root to the end represent the

taxonomic level of kingdom, phylum, class, order, family, genus. The edges were pained with different colors to distinguish different phyla. The genus nodes pained with different colors showed their most correlated soil properties. The dominant genera in each gene were labeled. (B) Stacked bar charts showing the dominant genera of each gene under the five treatments.