Research Article

Soil Biogeochemical Cycle Couplings Inferred from a Function-Taxon Network

Bin Ma [iD],[1,2,3] **Erinne Stirling** [iD],[1,2,4] **Yuanhui Liu,**[1,2] **Kankan Zhao** [iD],[1,2] **Jizhong Zhou,**[5] **Brajesh K. Singh,**[6] **Caixian Tang,**[7] **Randy A. Dahlgren** [iD],[8] and **Jianming Xu**[1,2]

[1]*Institute of Soil and Water Resources and Environmental Science, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China*

[2]*Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, Zhejiang University, Hangzhou 310058, China*

[3]*Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 310058, China*

[4]*Acid Sulfate Soils Centre, School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia*

[5]*Institute for Environmental Genomics, Department of Microbiology and Plant Biology, and School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA*

[6]*Global Centre for Land-Based Innovation, Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW 2750, Australia*

[7]*Department of Animal, Plant and Soil Sciences, La Trobe University, Melbourne Campus, Bundoora, VIC 3086, Australia*

[8]*Department of Land, Air and Water Resources, University of California, Davis, 95616 CA, USA*

Correspondence should be addressed to Jianming Xu; jmxu@zju.edu.cn

Soil biogeochemical cycles and their interconnections play a critical role in regulating functions and services of environmental systems. However, the coupling of soil biogeochemical processes with their mediating microbes remains poorly understood. Here, we identified key microbial taxa regulating soil biogeochemical processes by exploring biomarker genes and taxa of contigs assembled from metagenomes of forest soils collected along a latitudinal transect (18° N to 48° N) in eastern China. Among environmental and soil factors, soil pH was a sensitive indicator for functional gene composition and diversity. A function-taxon bipartite network inferred from metagenomic contigs identified the microbial taxa regulating coupled biogeochemical cycles between carbon and phosphorus, nitrogen and sulfur, and nitrogen and iron. Our results provide novel evidence for the coupling of soil biogeochemical cycles, identify key regulating microbes, and demonstrate the efficacy of a new approach to investigate the processes and microbial taxa regulating soil ecosystem functions.

1. Introduction

Elemental fluxes in soils are largely driven by microbially catalyzed, but thermodynamically constrained, redox reactions. Soil biogeochemical cycles are the foundation of ecosystem function and affect nutrient and energy flows that regulate productivity within both terrestrial and aquatic ecosystems [1]. Given the central role of soil microbial communities in regulating global biogeochemical processes, managing soil communities provides a powerful tool to combat several increasingly important global challenges, such as feeding the world's increasing population,

soil pollution, biodiversity loss, and climate change [2]. Despite the critical importance of microorganisms in regulating soil biogeochemical processes, fundamental questions concerning the linkage between specific microbial taxa and biogeochemical functions remain poorly understood, thereby limiting scientific advances. A major challenge is rooted in the fact that the vast majority of soil microbial taxa remain uncharacterized, hindering our efforts to untangle their unresolved roles in biogeochemical functions [3]. Given that different microorganisms perform a myriad of roles in biogeochemical processes, linking biogeochemical functions to specific soil microbial

taxa is critical for improving management and conservation policies to maintain key ecosystem functions.

From the simplest perspective, microbial biogeochemical cycling is composed of redox half-cell reactions. The substrates for coupled half-cells in biologically driven redox reactions are sourced from the environment directly or as waste products of microbial metabolism, thereby constructing linked metabolic systems driving biogeochemical cycles [4, 5]. While biogeochemical cycles logically interact with each other, they are generally studied in isolation. Theoretical coupling among soil biogeochemical cycles can be inferred from thermodynamic modeling [6], as well as some field evidence. For example, methane oxidation is linked with N cycling [7], C utilization rates vary with P bioavailability [8], and ammonia oxidation couples with ferric reduction or thiosulfate reduction under anaerobic conditions [9, 10]. However, the coupling of some biogeochemical cycles is predicted to exist only on a thermodynamics basis [6]. Moreover, in all these approaches, there is little information concerning the specific microbes regulating the coupling of biogeochemical cycles.

Functional genes that encode enzymes for specific redox reactions have the potential to reveal the functioning of specific microbes in various biogeochemical cycles. However, the coupling of biogeochemical cycles is often spatially or temporally separated, hindering the efficiency of coupled oxidation/reduction reactions [11]. Hence, microbial taxa have the potential to greatly enhance the understanding of coupling between biogeochemical cycles when their genomes contain genes for both components of the coupled cycles. Despite pioneering conceptual research [12], there is a paucity of information concerning the connection of microbial community composition to biogeochemical functions. Recent advances in cultivation-independent metagenomic sequencing provide a powerful new approach to link biogeochemical functioning with their marine microbial drivers [13], opening opportunities to explore the coupling of soil biogeochemical processes with their specific microbial taxa.

To examine the coupling of soil biogeochemical functions and their potential microbial drivers, we associated biomarker genes for soil biogeochemical cycles with their corresponding microbial taxa by investigating the functional profiles and taxonomy of contigs assembled from the metagenomes of 45 forest soils from forests along a 4000 km latitudinal transect (18 to 48° N) in eastern China (Figure S1). This dataset contains soils spanning five Köppen climate classifications (Aw, Cfa, Cwa, Dwa, and Dwb) ranging from tropical to warm summer continental. We hypothesize that certain biogeochemical cycle couplings are driven by specific microbial taxa and that these taxa are affected by soil properties. Our results provide critical insights into the complex interconnection among biogeochemical cycles and their microbial drivers in forest soil ecosystems.

2. Results

2.1. Functional Gene Composition and Diversity. Multiple regressions on the distance matrix analysis showed that the variation of soil metagenomes along the latitudinal transect was mainly determined by latitude, humic acid, soil pH, mean annual temperature, and dissolved Al and Fe (Figure 1(a)). The impact of distance on soil metagenomes was less than that of environmental variables (Figure 1(b)). Latitude significantly correlated with most functional genes for biogeochemical processes, potentially via a linkage to soil pH (Figures 1(c) and 1(d)). As the primary agent generating the environmental gradient (Figure S2), latitude correlated negatively with functional gene diversity (richness: $r = 0.43$, $P = 0.003$; H' diversity: $r = 0.42$, $P = 0.004$) (Figure 1(e)); functional gene composition shifted along the transect (Figure 1(f)).

2.2. C, N, P, S, and Fe Biogeochemical Cycles. Carbon-cycling biomarker genes were dominated by aerobic CO oxidation (*coxL*), anaerobic fermentation (*LDH*), and anaerobic C fixation (*KorB*) (15.7, 6.7, and 6.3% of all contigs, respectively; Figure 2(a)). Latitude correlated positively with anaerobic C fixation ($P = 0.01$) and negatively with aerobic C fixation (*PRK*) ($P = 0.01$; Figure 2(b)). Four of the seven biomarker genes had associated contigs assigned to known genera (Figure 2(c)). Of the biomarker genes with identifiable taxa (Figure 2(d)), CO oxidation was associated with the broadest range of taxa (Figure 2(d)).

Nitrogen-cycling biomarker genes were dominated by N assimilation (*glnA*) (20% of all contigs; Figure 3(a)). Latitude correlated positively with anammox ($P < 0.05$), N mineralization ($P < 0.001$), denitrification (*norB*) ($P < 0.05$), and N assimilation ($P = 0.01$), but negatively with nitrate reduction (*modA*) ($P < 0.01$; Figure 3(b)). Seven of the nine biomarker genes had associated contigs assigned to known genera; however, both anammox (*ccoN*) and nitrite oxidation (*narG*) had less than a third of their contigs assigned (Figure 2(c)). Of the biomarker genes with identifiable taxa (Figure 3(d)), Actinobacteria were the most frequently assigned taxa.

Phosphorus-cycling biomarker genes were dominated by substrate phosphorylation (*Ptsl*) and oxidative phosphorylation (*ppk*) (10.9 and 6.2% of all contigs, respectively; Figure S3a) but were not correlated with latitude ($P > 0.05$, Figure S3b). All biomarker genes had at least 23% of contigs assigned to known genera (Figure S3c). Actinobacteria (10 genera) and Proteobacteria (9 genera) played a dominant role in oxidative phosphorylation, whereas Firmicutes (5 genera) contributed to substrate phosphorylation (Figure S3d).

Sulfur-cycling biomarker genes were dominated by dissimilatory sulfate reduction (*Fer4*) (6.3% of all contigs; Figure S4a). Latitude correlated positively with S mineralization (*sseA*) ($P = 0.03$) and negatively with polysulfide reduction (*NrfD*) ($P = 0.03$; Figure S4b). Four of the five biomarker genes had contigs assigned to known taxa (Figure S4c). The biomarker genes with a high proportion of assigned taxa were associated with a wide variety of taxa (Figure S4d), namely, Proteobacteria (26 genera) and Actinobacteria (22 genera).

Iron-cycling biomarker genes were dominated by ferrous oxidation (*Ferritin*) (4.7% of all contigs; Figure S5a). Latitude correlated negatively with both biomarker genes for Fe
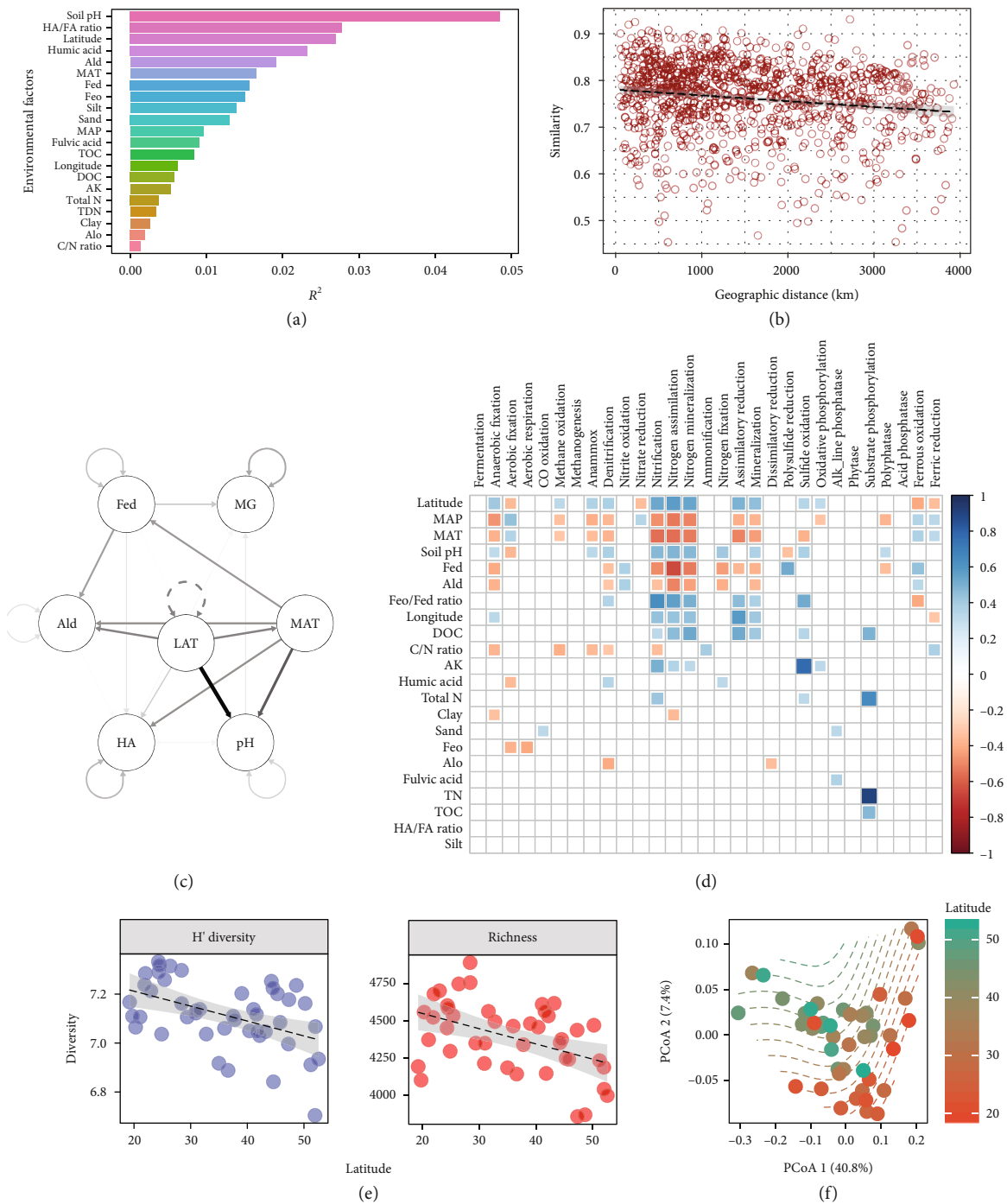
FIGURE 1: Environmental drivers of soil metagenomes. (a) The impact of environmental factors. HA: humic acid; FA: fulvic acid; Ald: dithionite extractable Al; MAT: mean annual temperature; Fed: dithionite extractable Fe; Feo: amorphous sesquioxide Fe; MAP: mean annual precipitation; TOC: total organic carbon; DOC: dissolved organic carbon; AK: available K; TDN: total dissolved N; Alo: amorphous sesquioxide Al. (b) Distance-decay of Bray-Curtis similarity of metagenomes. (c) Structure equation models (SEM) between soil metagenomes (MG) and major environmental/soil drivers, including LAT, MAT, HA, soil pH, Ald, and Fed. (d) Spearman's correlation between biomarker genes of biogeochemical process and environmental/soil factors. (e) Correlation of H' diversity and richness of functional genes in forest soil metagenomes. (f) Principal coordinate analysis (PCoA) of functional genes in forest soil metagenomes; point color represents latitude of sampling sites.

cycling ($P < 0.02$; Figure S5b); the majority of taxa associated with these genes were assigned classifications with the bulk of organisms belonging to Proteobacteria (Figure S5c, d).

2.3. Linking Biogeochemical Processes to Microbial Taxa Using a Function-Taxa Bipartite Network. The function-taxa bipartite network between genes linked to biogeochemical
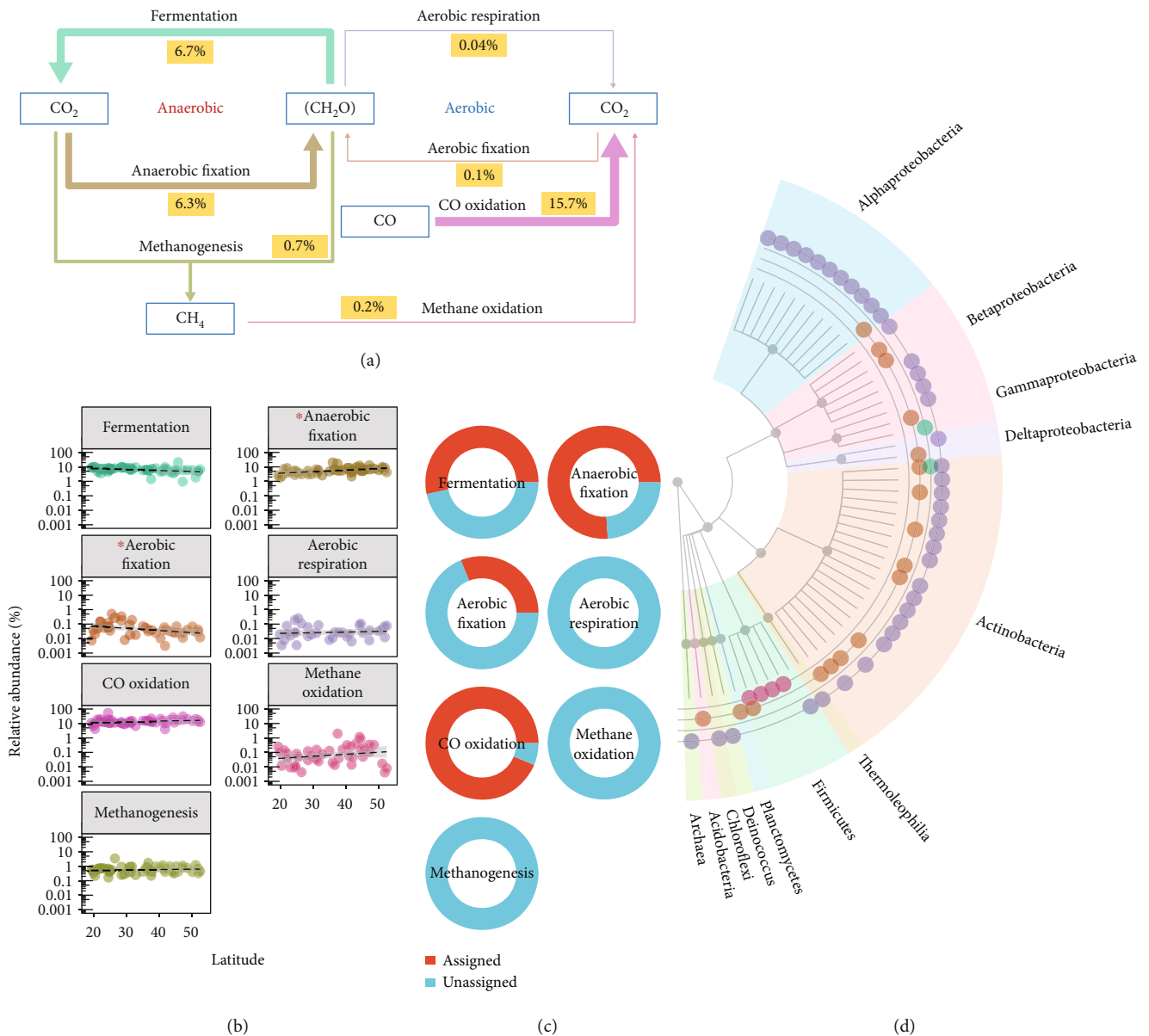
FIGURE 2: Carbon biogeochemical cycle genes in forest soils. (a) Carbon biogeochemical cycle; arrow thickness indicates proportion of biomarker gene contigs for corresponding pathways. (b) Linear relationships between contigs and latitude. Asterisks indicate significant correlation with latitude ($P < 0.05$, $n = 45$). (c) Proportion of contigs assigned to known taxa. (d) Distribution of known taxa.

processes and their associated taxa formed eight clusters (Figure 4). Biogeochemical processes are potentially driven by the taxa present within the same cluster. Therefore, taxa within a cluster are potentially regulating both biogeochemical processes and the coupling processes. Most clusters contained processes from at least two biogeochemical cycles; the maximum number in a cluster was four (C, N, P, and S). There was distinct clustering for processes requiring similar redox states.

Carbon and P cycling presented together in three clusters (Figure 4). The first cluster, between genes linked to aerobic methane oxidation and acid phosphatase cycles, was driven by Proteobacteria (*Anaeromyxobacter* and *Bradyrhizobium*)

and Actinobacteria (*Mycobacterium* and *Nocardia*). The second cluster, between genes linked to aerobic C fixation and phytase, was driven by Firmicutes (*Streptococcus*, *Lactococcus*, and *Enterococcus*). The third cluster, between genes linked to CO oxidation and substrate phosphorylation, was driven by Actinobacteria (*Amycolatopsis* and *Conexibacter*) and Proteobacteria (*Azospirillum* and *Rhodopseudomonas*).

Nitrogen and S cycling occurred together in two clusters (Figure 4). The first association, between genes linked to anammox and dissimilatory sulfate reduction, was driven by Actinobacteria (*Thermobispora*, *Pseudonocardia*, *Brachybacterium*, and *Frankia*), whereas the second
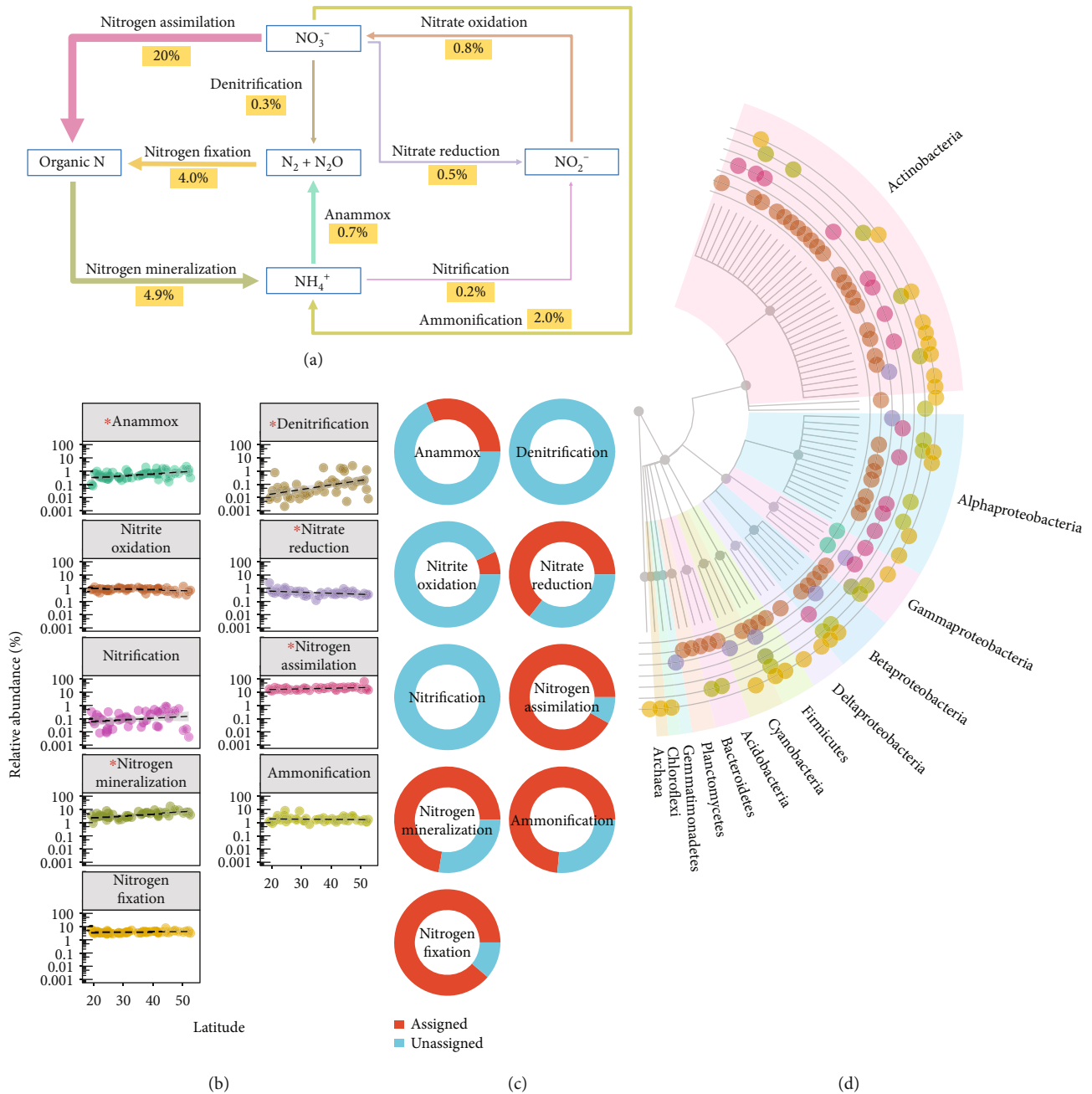
FIGURE 3: Nitrogen biogeochemical cycle genes in forest soils. (a) Nitrogen biogeochemical cycle; arrow width indicates proportion of biomarker gene contigs for corresponding pathways. (b) Linear relationships between contigs and latitude. Asterisks indicate significant correlation with latitude ($P < 0.05$, $n = 45$). (c) Proportion of contigs assigned to known taxa. (d) Distribution of known taxa.

association, between genes linked to N assimilation and assimilatory sulfate reduction, was driven by Firmicutes (*Bacillus*). Finally, genes linked to nitrification, denitrification, and ferric reduction presented together in one cluster; this relationship was driven by Proteobacteria (*Ralstonia*, *Burkholderia*, and *Brevundimonas*).

3. Discussion

Our results provide novel empirical evidence that there is coupling between the biogeochemical cycles that play signif-

icant roles in regulating soil microbial functions in forest ecosystems. The results support the hypothesis that certain biogeochemical cycle couplings are associated with specific microbial taxa. Identified soil microbial functions are consistent with typical latitudinal biodiversity gradients and indicate that these gradients are regulated by environmental variables such as precipitation and soil pH [14]. Of the measured environmental variables, soil pH, a well-known predictor of species richness in soil bacterial and fungal communities [15, 16], was a significant regulator of genes associated with key biogeochemical cycles.
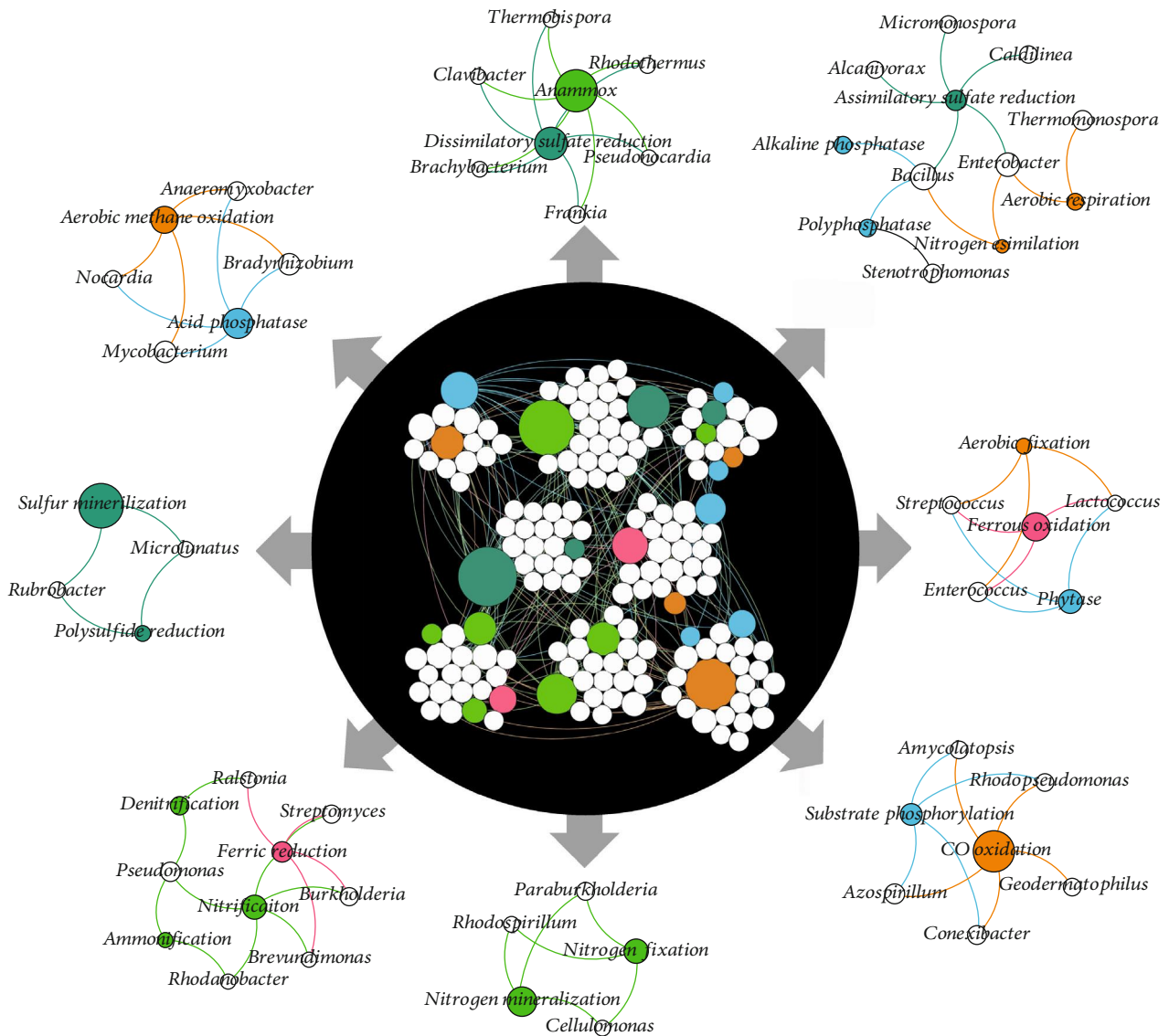
FIGURE 4: Function-taxa bipartite network for biogeochemical cycling processes. Subnetworks were induced from the functional nodes within each biogeochemical module and their coassociating taxa.

Soil pH is an edaphic variable sensitive to latitude due to the influence of rainfall and temperature on weathering (Figure S2) as soils in high-rainfall regions tend to have greater levels of acidity, leaching of base cations and humification of organic matter [17]. This relationship is clearly evident in eastern China where pH increases by ~5 units with latitude [18]. In addition to mediating soil community composition [19], soil pH is known to influence specific microbial-mediated biogeochemical processes, such as microbial denitrification and S mineralization. Consistent with our results, denitrification rates were shown to decrease with decreasing pH [18], whereas higher soil pH increases S mineralization rates [20]. In addition, our results indicate that polysulfide reduction was negatively correlated with latitude; polysulfide reduction is mainly driven by *Clostridium* [21], which is an acidophilic genus adapted to the low-pH region of our study. The role of pH in soil

biogeochemical reactions is multifaceted. Bacteria, for example, are often constrained to relatively tight optimal pH ranges [19], and the expression of certain genes is similarly constrained to specific pH ranges [22]. Soil pH affects the strategies that microbes can use to acquire nutrients and extremes of pH require microbes to spend additional energy and resources maintaining their cell's physicochemical integrity [23].

Notably, a prevalence of anoxic microenvironments was inferred by the high proportion of CO oxidation, fermentation, and anaerobic fixation genes across all sampling sites even though all soils were considered well drained and oxygenated. It is possible that oxic degradation of soil organic matter under conditions of low gas permeability (e.g., sites within soil structural units or surrounding recently dead roots) caused a prevalence of transient anoxic microniches (hotspots/hot moments) [24]. Although CO is toxic to many

organisms, numerous cultured and molecular ecological approaches have revealed an unexpectedly diverse group of soil bacteria that have the ability to use CO as an energy source, including members of phyla Proteobacteria and Actinobacteria [25]. The prevalent role of facultative anaerobes is supported by the similar proportions of oxidative and substrate phosphorylation genes, indicating both anaerobic and aerobic generation of energy [26]. An abundance of dissimilatory S reduction genes also suggests a preponderance of anoxic soil conditions as dissimilatory sulfate reactions are typically facilitated by sulfate-reducing microbes in anaerobic environments [27].

Nitrogen cycling is well known to be affected by edaphic properties; N mineralization, anammox, nitrification, and denitrification marker genes were all significantly correlated with latitude in our study. Although N mineralization has previously been negatively associated with latitude [28], we found that *GDH2* increased with latitude. The N assimilation marker gene similarly increased with latitude. Increased relative abundance of *GDH2* may be associated with lower-quality organic matter at high latitudes as glutamate dehydrogenase activity has been shown to decrease as the C : N ratio increases [29, 30]; increased activation energy requirements in colder climates may also lead to an increase in gene expression [31]. Marker genes associated with anammox similarly increased with latitude; however this may have been in response to increasing pH and associated Fe limitation [32]. The denitrification marker gene also increased with latitude; however, this is in contrast with established literature in which denitrification increases with temperature and therefore tends to decrease with latitude [33, 34].

This study provides novel evidence that biogeochemical processes are coupled via associations among individual taxa possessing multiple functional genes, as observed from the function-taxa bipartite network. The most frequent potential association between biogeochemical cycles was between P and C. Methane oxidation was previously shown to correlate positively with P concentration in soil/sediment [7] and acid-phosphatase genes have been shown to be present in all methanotrophic bacteria [7]. One of the predicted driver taxa, *Bradyrhizobium*, was reported to enhance acid phosphatase activity in arbuscular mycorrhizal fungi [35] and has a known association with methanotrophic bacteria in rice paddy soils [36]. Similarly, a positive response was previously observed between P concentration and C fixation in seawater [31]. Another predicted driver, *Streptococcus*, is associated with phytase production and mineralization of phosphate [37]. Among the predicted regulating genera, only *Rhodopseudomonas* appears to be associated with CO and CO_2 metabolism [38].

Prominent coupling also appears to occur between N and S cycling and between N and Fe cycling. Of the relationships between these biogeochemical cycles, sulfate reduction is known to couple with anammox when removing N from wastewater. The predicted microbial drivers *Thermobispora*, *Pseudonocardia*, *Brachybacterium*, and *Frankia* have been implicated in anammox reactions [39], but none of these bacteria appear to be involved in S cycling. Curiously, *Ralstonia* has been shown to use thiosulfate during anaerobic

ammonia oxidation but was directly associated with ferric reduction and denitrification in our cooccurrence network [10]. However, it is possible for one cycle to affect another cycle simply via its limited availability in soil, as with the association between assimilatory sulfate reduction and P nutrition, whereby S deficiency increases the activity of polyphosphatase in some microorganisms [40]. Of the drivers predicted to affect coupling between genes involved in N- and S-cycling processes, *Bacillus* was associated with the highest number of pathways. *Bacillus* affects the growth of other microbes in culture through assimilatory sulfate reduction processes [41]. Other genera in the cluster containing *Bacillus* and *Enterobacter* are implicated in dissimilatory S reduction and have the ability to efficiently reduce nitrate to ammonium [42]. Coupling between ferric reduction and ammonium oxidation is reported in both paddy [43] and upland [44] soils, but our predicted microbial drivers have not been reported to drive ferric reduction process.

This study demonstrates that the potential coupling of biogeochemical processes with microbial taxa capable of generating functional genes facilitates multielement transformations within a wide range of forest soils. However, the coupling of biogeochemical processes among different taxa, which is often spatially or temporally separated, cannot be discovered with this approach. Moreover, unraveling coupled interactions between biogeochemical processes and the mitigating microbial taxa sheds light on elucidating the importance and role of the unculturable microbial 'dark matter.' Furthermore, a significant limitation of this work is that the metagenomic predicted potential couplings of biogeochemical processes are not experimentally validated due to a large proportion of microbial 'dark matter.' In addition, while using marker genes to assess the rate-limiting step of each process gives us an estimation of community capacity to process biogeochemical reactions, it is possible that an analysis of alternative marker genes would show an alternative response to edaphic gradients.

While the use of undisturbed sites from a single land use type provides a more constrained approach for deciphering the effects of edaphic gradients on soil microbial processes (as opposed to human-induced effects), we cannot know if these relationships hold true for other land uses. In addition to increasing the variety of landscapes sampled, future focus could employ element probes, such as stable isotope probes, for validating coupling of biogeochemical processes through labeling active microbial taxa for different pathways. Additional work could also involve investigation of soil metagenomes from other ecosystems for the comparison of driving taxa between ecosystems. Overall, this study highlights a potential avenue to enhance simulation modeling of soil biogeochemical processes to inform controls of various soil functions and potential management of key taxa to achieve specific soil functions.

4. Materials and Methods

4.1. Experimental Design. To explore potential biogeochemical couplings from surveying soil metagenomes, we collected

soil samples from 45 sampling sites in eastern China covering a latitude of 18° 48' N to 48° 36' N (Figure S1, Table S1). To minimize the anthropological influence on soil microbial functions, all sites were located within natural forest reserves. Topsoil samples (0-15 cm; the dominant rooting zone) were collected from $100 \, \mathrm{m} \times 100 \, \mathrm{m}$ plots. From each plot, we collected three analytical sample replicates, each a composite of five soil cores. The methods for measuring edaphic variables have been described previously [45]. In brief, soil pH, soil texture, organic carbon, and available potassium were determined according to the protocols outlined by the Agricultural Chemistry Committee of China. Total nitrogen was determined using a Flash 2000 NC Analyzer (Thermo Scientific, MA, USA). Sesquioxides (Ald and Fed) were extracted in the dark with dithionite-citrate solution buffered with NH_4-oxalate (pH 3.0; amorphous sesquioxides) or $NaHCO_3$ (amorphous sesquioxides–Alo and Feo) and measured using atomic absorption spectrometry (ContrAA 700, Jena, Germany). The data for mean annual precipitation and temperature were sourced from WorldClim (http://worldclim.org).

4.2. Metagenomic Dataset. Details of the metagenomic data acquisition have been previously described [18]. In brief, shotgun sequencing of metagenomic DNA produced a total of ~1.5 billion paired-end reads (read length = 150 bp). The raw shotgun sequencing reads were preprocessed using ngsShoRT v2.1 [46], and whole-genome *de novo* assemblies for each sample were performed using IDBA-UD; open read frame prediction and annotation were performed using Prodigal v2.50 [47]. The resulting protein translations were assigned by comparisons to Pfam 31.0 using HMMER 3 [48] and KEGG release 84.0 using GhostKOALA [49]. The number of contigs greater than 500 bp in length was 253,807; they had an N50 of 1409 and a maximum length of 160699; mean coverage was 12.9 and ranged from 4.7 to 27.4 for individual samples.

4.3. Biogeochemical Functional Gene Analysis. Biogeochemical functional gene analyses focused on C, N, P, S, and Fe cycling processes. The genetic potential for C, N, and S cycling in the soil microbial community was analyzed using biomarker genes as reported [50] with modifications as follows. If the biomarker genes reported were not found in all 45 metagenomes annotated using the KEGG database, these biomarker genes were replaced with equivalent genes annotated using the Pfam database. The following maker genes were replaced in this manner: methanogenesis biomarker gene K14084 with PF06253, N_2 fixation biomarker gene K02588 with PF00142, ammonification biomarker gene K05904 with PF01077, S oxidation biomarker gene K17227 with PF08770, dissimilatory sulfate reduction biomarker gene K00394 with PF13187, and polysulfide reduction biomarker gene K08352 with PF14589.

Marker genes were also selected for P and Fe biogeochemical cycles. We used PF03767 as the biomarker gene for acid phosphatase, K01077 for alkaline phosphatase, K01083 for phytase, K00937 for oxidative phosphorylation, K08483 for substrate phosphorylation, and PF00719 for

polyphosphatase. For the Fe biogeochemical cycle, we used PF00210 and PF01794 as biomarker genes for ferrous oxidation and ferric reduction as catalyzed by microorganisms, respectively. Biomarker genes used in this study are compiled in Table S2. The proportions of contigs with their corresponding marker genes for each biogeochemical pathway are shown in pathway maps (Figures 2 and 3, S2-S4 inclusive).

4.4. Biomarker Gene Taxonomic Profiles. Contig taxa were assigned using CLARK [51]. Biomarker gene taxonomic profiles were generated from the corresponding biomarker gene contig taxa. Unassigned taxa indicate that the contigs could not be assigned to a known taxon using CLARK. Phylogenetic trees for taxa involving a biogeochemical cycle were generated with hieratical phylogenetic relationships from kingdom to order.

4.5. Function-Taxon Bipartite Network. The function-taxon bipartite network was constructed from the function-taxon relationships as assigned above. The network modules were clustered using the modularity calculation in Gephi with $r = 1$ [52]. Module subnetworks were induced between functional nodes and their connecting taxa.

4.6. Statistical Analysis. All statistical analyses were carried out using R version 3.5.0 [53]. The impact of environmental factors on functional gene composition was estimated by multiple regression on distance matrices by ecodist::MRM in R [54]. Functional gene diversity was assessed using the Shannon-Weiner 'H' diversity index; gene richness was measured using the number of genes found in corresponding samples; trend significance was established by fitting a generalized linear model with stats::glm in R. Functional composition dissimilarity was analyzed using Bray-Curtis dissimilarity and visualized using principle coordinate analysis (PCoA) with the 'vegan' package in R [55]. The effect of latitude on relative gene abundance was established using stats::glm, as above, for each biogeochemical cycle.

Data Availability

The sequencing datasets from this study are available in the Public National Center for Biotechnology Information (NCBI) database under BioProject accession number PRJNA293484.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

B. Ma and J.M. Xu designed the study. B. Ma and K.K. Zhao performed the metagenomic analysis. B. Ma performed the network analysis. K.K. Zhao performed the data visualization. B. Ma, E. Stirling, Y.H. Liu, J.Z. Zhou, B. Singh, C.X. Tang, R.A. Dahlgren, and J.M. Xu contributed to the scientific discussion and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Supplementary Materials

Figure S1 Location of sampling sites. Figure S2: metadata for sampling sites along the latitudinal transect. Figure S3: genes associated with phosphorus biogeochemical cycling in forest soils. Figure S4: genes associated with sulfur biogeochemical cycling in forest soils. Figure S5: genes associated with iron biogeochemical cycling in forest soils. Table S1: GPS coordinates for all sampling sites Table S2: biomarker genes of biogeochemical processes. *(Supplementary Materials)*

References

[1] D. A. Wardle, R. D. Bardgett, J. N. Klironomos, H. Setälä, W. van der Putten, and D. H. Wall, "Ecological linkages between aboveground and belowground biota," *Science*, vol. 304, no. 5677, pp. 1629–1633, 2004.

[2] T. W. Crowther, K. E. O. Todd-Brown, C. W. Rowe et al., "Quantifying global soil carbon losses in response to warming," *Nature*, vol. 540, no. 7631, pp. 104–108, 2016.

[3] H. M. Serna-Chavez, N. Fierer, and P. M. van Bodegom, "Global drivers and patterns of microbial abundance in soil," *Global Ecology and Biogeography*, vol. 22, no. 10, pp. 1162–1172, 2013.

[4] S. J. Hallam, N. Putnam, C. M. Preston et al., "Reverse methanogenesis: testing the hypothesis with environmental genomics," *Science*, vol. 305, no. 5689, pp. 1457–1462, 2004.

[5] V. J. Orphan, C. H. House, K. U. Hinrichs, K. McKeegan, and E. DeLong, "Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis," *Science*, vol. 293, no. 5529, pp. 484–487, 2001.

[6] D. L. Sparks, "Advances in coupling of kinetics and molecular scale tools to shed light on soil biogeochemical processes," *Plant and Soil*, vol. 387, no. 1-2, pp. 1–19, 2015.

[7] A. J. Veraart, A. K. Steenbergh, A. Ho, S. Y. Kim, and P. L. E. Bodelier, "Beyond nitrogen: the importance of phosphorus for CH 4 oxidation in soils and sediments," *Geoderma*, vol. 259-260, pp. 337–346, 2015.

[8] K. S. Watkins-Brandt, R. M. Letelier, Y. H. Spitz, M. J. Church, D. Böttjer, and A. E. White, "Addition of inorganic or organic phosphorus enhances nitrogen and carbon fixation in the oligotrophic North Pacific," *Marine Ecology Progress Series*, vol. 432, pp. 17–29, 2011.

[9] L.-J. Ding, X. L. An, S. Li, G. L. Zhang, and Y. G. Zhu, "Nitrogen loss through anaerobic ammonium oxidation coupled to iron reduction from paddy soils in a chronosequence," *Environmental Science & Technology*, vol. 48, no. 18, pp. 10641–10647, 2014.

[10] G. X. Li, H. Li, K. Q. Xiao, and P. Bao, "Thiosulfate reduction coupled with anaerobic ammonium oxidation byRalstoniasp. GX3-BWBA," *ACS Earth and Space Chemistry*, vol. 4, no. 12, pp. 2426–2434, 2020.

[11] P. G. Falkowski, T. Fenchel, and E. F. Delong, "The microbial engines that drive Earth's biogeochemical cycles," *Science*, vol. 320, no. 5879, pp. 1034–1039, 2008.

[12] R. D. Stewart, M. D. Auffret, A. Warr et al., "Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen," *Nature Communications*, vol. 9, no. 1, pp. 870–870, 2018.

[13] K. Anantharaman, J. A. Breier, and G. J. Dick, "Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau spreading center," *The ISME Journal*, vol. 10, no. 1, pp. 225–239, 2016.

[14] M. R. Willig, D. M. Kaufman, and R. D. Stevens, "Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis," *Annual Review of Ecology, Evolution, and Systematics*, vol. 34, no. 1, pp. 273–309, 2003.

[15] L. Tedersoo, M. Bahram, S. Põlme et al., "Global diversity and geography of soil fungi," *Science*, vol. 346, no. 6213, article 1256688, 2014.

[16] N. Fierer and R. B. Jackson, "The diversity and biogeography of soil bacterial communities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 3, pp. 626–631, 2006.

[17] Y. Wei, X. Wu, R. Zeng, C. Cai, and Z. Guo, "Spatial variations of aggregate-associated humic substance in heavy-textured soils along a climatic gradient," *Soil and Tillage Research*, vol. 197, p. 104497, 2020.

[18] B. Ma, Z. Dai, H. Wang et al., "Distinct biogeographic patterns for archaea, bacteria, and fungi along the vegetation gradient at the continental scale in eastern China," *mSystems*, vol. 2, no. 1, 2017.

[19] M. Bahram, F. Hildebrand, S. K. Forslund et al., "Structure and function of the global topsoil microbiome," *Nature*, vol. 560, no. 7717, pp. 233–237, 2018.

[20] A. F. F. R. C. Jaggi, M. S. Aulakh, and R. Sharma, "Temperature effects on soil organic sulphur mineralization and elemental sulphur oxidation in subtropical soils of varying pH," *Nutrient Cycling in Agroecosystems*, vol. 54, no. 2, pp. 175–182, 1999.

[21] Y. Takahashi, K. Suto, and C. Inoue, "Polysulfide reduction by Clostridium relatives isolated from sulfate-reducing enrichment cultures," *Journal of Bioscience and Bioengineering*, vol. 109, no. 4, pp. 372–380, 2010.

[22] J. Bárta, T. Melichová, D. Vaněk, T. Picek, and H. Šantrůčková, "Effect of pH and dissolved organic matter on the abundance of nirK and nirS denitrifiers in spruce forest soil," *Biogeochemistry*, vol. 101, no. 1-3, pp. 123–132, 2010.

[23] C. L. Lauber, M. Hamady, R. Knight, and N. Fierer, "Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale," *Applied and Environmental Microbiology*, vol. 75, no. 15, pp. 5111–5120, 2009.

[24] J. Tian, N. He, L. Hale et al., "Soil organic matter availability and climate drive latitudinal patterns in bacterial diversity from tropical to cold temperate forests," *Functional Ecology*, vol. 32, no. 1, pp. 61–70, 2018.

[25] G. M. King and C. F. Weber, "Distribution, diversity and ecology of aerobic CO-oxidizing bacteria," *Nature Reviews Microbiology*, vol. 5, no. 2, pp. 107–118, 2007.

[26] K. A. Hunt, J. M. Flynn, B:. Naranjo, I. D. Shikhare, and J. A. Gralnick, "Substrate-level phosphorylation is the primary source of energy conservation during anaerobic respiration

of Shewanella oneidensis strain MR-1," *Journal of Bacteriology*, vol. 192, no. 13, pp. 3345–3351, 2010.

[27] G. Muyzer and A. J. M. Stams, "The ecology and biotechnology of sulphate-reducing bacteria," *Nature Reviews Microbiology*, vol. 6, no. 6, pp. 441–454, 2008.

[28] Y. Liu, N. He, X. Wen, G. Yu, Y. Gao, and Y. Jia, "Patterns and regulating mechanisms of soil nitrogen mineralization and temperature sensitivity in Chinese terrestrial ecosystems," *Agriculture, Ecosystems & Environment*, vol. 215, pp. 40–46, 2016.

[29] D. Geisseler, T. A. Doane, and W. R. Horwath, "Determining potential glutamine synthetase and glutamate dehydrogenase activity in soil," *Soil Biology and Biochemistry*, vol. 41, no. 8, pp. 1741–1749, 2009.

[30] M. Takriti, B. Wild, J. Schnecker et al., "Soil organic matter quality exerts a stronger control than stoichiometry on microbial substrate use efficiency along a latitudinal transect," *Soil Biology and Biochemistry*, vol. 121, pp. 212–220, 2018.

[31] M. Olofsson, J. Egardt, A. Singh, and H. Ploug, "Inorganic phosphorus enrichments in Baltic Sea water have large effects on growth, carbon fixation, and N2 fixation by Nodularia spumigena," *Aquatic Microbial Ecology*, vol. 77, no. 2, pp. 111–123, 2016.

[32] M. L. S. Suhadolnik, P. S. Costa, G. M. Castro, F. P. Lobo, and A. M. A. Nascimento, "Comprehensive insights into arsenic- and iron-redox genes, their taxonomy and associated environmental drivers deciphered by a meta-analysis," *Environment International*, vol. 146, p. 106234, 2021.

[33] G. Pinay, B. Gumiero, E. Tabacchi et al., "Patterns of denitrification rates in European alluvial soils under various hydrological regimes," *Freshwater Biology*, vol. 52, no. 2, pp. 252–266, 2007.

[34] E. Tan, W. Zou, Z. Zheng et al., "Warming stimulates sediment denitrification at the expense of anaerobic ammonium oxidation," *Nature Climate Change*, vol. 10, no. 4, pp. 349–355, 2020.

[35] I. Stancheva, M. Geneva, M. Hristozkova et al., "Response of Vigna unguiculata grown under different soil moisture regimes to the dual inoculation with nitrogen-fixing bacteria and arbuscular mycorrhizal fungi," *Communications in Soil Science and Plant Analysis*, vol. 48, no. 12, pp. 1378–1386, 2017.

[36] Z. Bao, T. Okubo, K. Kubota et al., "Metaproteomic identification of diazotrophic methanotrophs and their localization in root tissues of field-grown rice plants," *Applied and Environmental Microbiology*, vol. 80, no. 16, pp. 5043–5052, 2014.

[37] J. R. M. de Lacerda, T. F. da Silva, R. E. Vollú, J. M. Marques, and L. Seldin, "Generally recognized as safe (GRAS) Lactococcus lactis strains associated with Lippia sidoides Cham. are able to solubilize/mineralize phosphate," *Springerplus*, vol. 5, no. 1, p. 828, 2016.

[38] G. S. Joshi, C. E. Bobst, and F. R. Tabita, "Unravelling the regulatory twist–regulation of CO2 fixation in Rhodopseudomonas palustris CGA010 mediated by atypical response regulator(s)," *Molecular Microbiology*, vol. 80, no. 3, pp. 756–771, 2011.

[39] J. Guo, Y. Peng, S. Wang et al., "Pathways and organisms involved in ammonia oxidation and nitrous oxide emission," *Critical Reviews in Environmental Science and Technology*, vol. 43, no. 21, pp. 2213–2296, 2013.

[40] J. Aravind, T. Saranya, and P. Kanmani, "Optimizing the production of polyphosphate from acinetobacter towneri," *Global Journal of Environmental Science and Management*, vol. 1, no. 1, pp. 63–70, 2015.

[41] A. Marietou, H. Røy, B. B. Jørgensen, and K. U. Kjeldsen, "Sulfate transporters in dissimilatory sulfate reducing microorganisms: a comparative genomics analysis," *Frontiers in Microbiology*, vol. 9, 2018.

[42] C. Huang, Z. L. Li, F. Chen et al., "Microbial community structure and function in response to the shift of sulfide/nitrate loading ratio during the denitrifying sulfide removal process," *Bioresource Technology*, vol. 197, pp. 227–234, 2015.

[43] W. H. Yang, K. A. Weber, and W. L. Silver, "Nitrogen loss from soil through anaerobic ammonium oxidation coupled to iron reduction," *Nature Geoscience*, vol. 5, no. 8, pp. 538–541, 2012.

[44] B. Ding, Z. Chen, Z. Li, Y. Qin, and S. Chen, "Nitrogen loss through anaerobic ammonium oxidation coupled to iron reduction from ecosystem habitats in the Taihu estuary region," *Science of the Total Environment*, vol. 662, pp. 600–606, 2019.

[45] B. Ma, H. Wang, M. Dsouza et al., "Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China," *ISME Journal*, vol. 10, no. 8, pp. 1891–1901, 2016.

[46] C. Chen, S. S. Khaleel, H. Huang, and C. H. Wu, "Software for pre-processing Illumina next-generation sequencing short read sequences," *Source Code for Biology and Medicine*, vol. 9, no. 1, p. 8, 2014.

[47] D. Hyatt, G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser, "Prodigal: prokaryotic gene recognition and translation initiation site identification," *BMC Bioinformatics*, vol. 11, no. 1, p. 119, 2010.

[48] R. D. Finn, P. Coggill, R. Y. Eberhardt et al., "The Pfam protein families database: towards a more sustainable future," *Nucleic Acids Research*, vol. 44, no. D1, pp. D279–D285, 2016.

[49] M. Kanehisa, Y. Sato, and K. Morishima, "BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences," *Journal of Molecular Biology*, vol. 428, no. 4, pp. 726–731, 2016.

[50] T. Llorens-Marès, S. Yooseph, J. Goll et al., "Connecting biodiversity and potential functional role in modern euxinic environments by microbial metagenomics," *ISME Journal*, vol. 9, no. 7, pp. 1648–1661, 2015.

[51] R. Ounit, S. Wanamaker, T. J. Close, and S. Lonardi, "CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers," *BMC Genomics*, vol. 16, no. 1, p. 236, 2015.

[52] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks," in *International AAAI Conference on Weblogs and Social Media*, San Jose, CA, USA, 2009.

[53] R Core Team, "R: a language and environment for statistical computing," in *R Foundation for Statistical Computing*, Vienna, Austria, 2018.

[54] S. C. Goslee and D. L. Urban, "TheecodistPackage for dissimilarity-based analysis of ecological data," *Journal of Statistical Software*, vol. 22, no. 7, 2007.

[55] J. Oksanen, F. G. Blanchet, M. Friendly et al., "vegan: community ecology package," *R package*, 2019, http://www.CRAN.R-project.org/package=vegan.