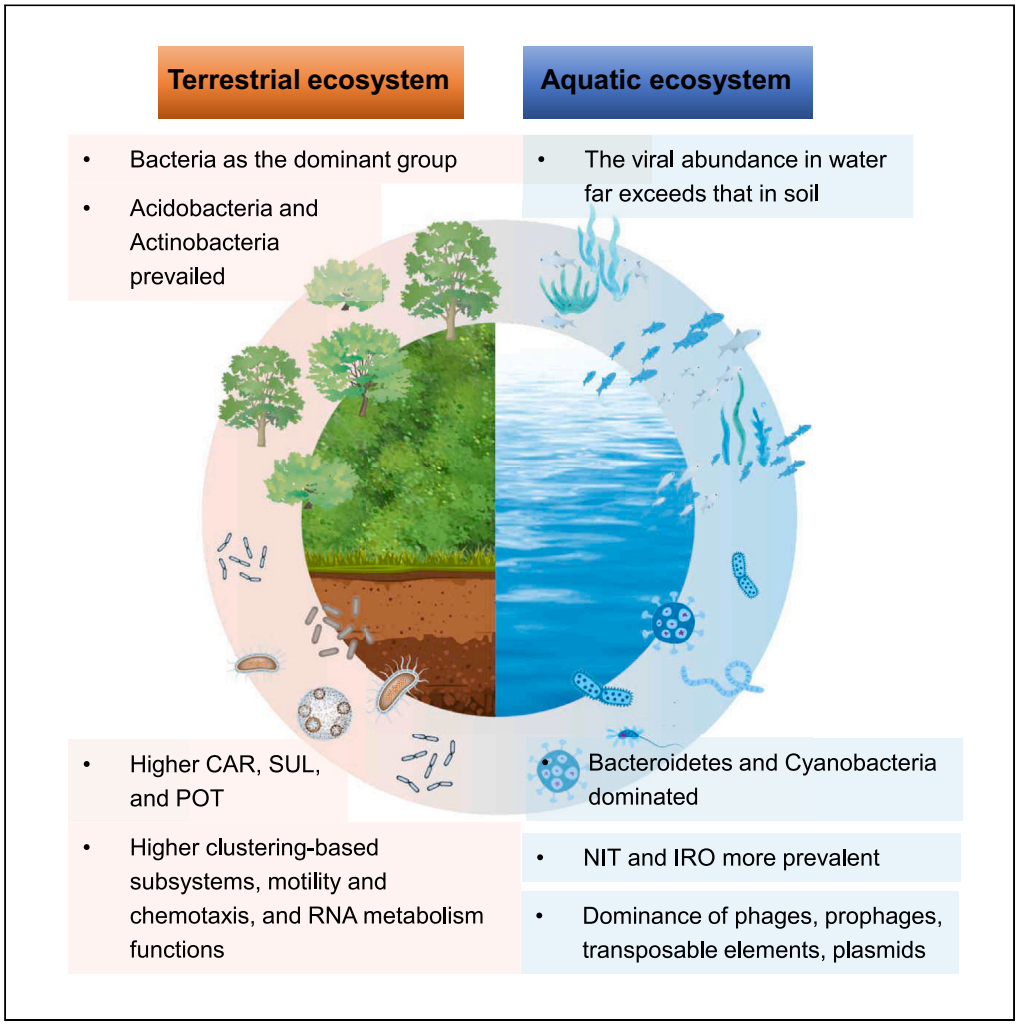
iScience

CellPress
OPEN ACCESS

Article

Metagenomics unravel distinct taxonomic and functional diversities between terrestrial and aquatic biomes



Qi Fu, Kayan Ma, Jiayi Zhao, ..., Xianheng Fu, Dandan Huang, Huaihai Chen

huangdd26@mail.sysu.edu.cn (D.H.)
chenhh68@mail.sysu.edu.cn (H.C.)

Highlights

Water had more diverse archaea, Eukaryota, and virus but less bacteria than soil

Soil functions mainly engaged in membrane transport and regulatory signaling

Water metagenomes harbored more genes related to nitrogen and iron metabolisms

Higher abundance of carbohydrate, sulfur, and potassium cycling in soil metagenomes

Fu et al., iScience 27, 111047
October 18, 2024
https://doi.org/10.1016/j.isci.2024.111047

iScience

CellPress
OPEN ACCESS

Article

Metagenomics unravel distinct taxonomic and functional diversities between terrestrial and aquatic biomes

Qi Fu,[1] Kayan Ma,[1] Jiayi Zhao,[1] Jiaxin Li,[1] Xueying Wang,[1] Meiqi Zhao,[1] Xianheng Fu,[1] Dandan Huang,[1,*] and Huaihai Chen[1,2,*]

SUMMARY

Microbes in terrestrial and aquatic ecosystems play crucial roles in driving ecosystem functions, but currently, there is a lack of comparison regarding their taxonomic and functional diversities. Here, we conducted a global analysis to investigate the disparities in microbial taxonomy and microbial-mediated biogeochemical cycles between terrestrial and aquatic ecosystems. Results showed a higher relative abundance of bacteria, especially Actinobacteria and Acidobacteria, in soil than water metagenomes, leading to a greater proportion of genes related to membrane transport, regulatory, and cellular signaling. Moreover, there was a higher abundance of genes associated with carbohydrate, sulfur, and potassium metabolisms in the soil, while those involved in nitrogen and iron metabolisms were more prevalent in the water. Thus, both soil and water microbiomes exhibited unique taxonomic and functional properties associated with biogeochemical processes, providing valuable insights into predicting and understanding the adaptation of microbes in different ecosystems in the face of climate change.

INTRODUCTION

Microbes are ubiquitous on Earth and play a vital role in various ecosystems, including deserts,[1] forests,[2] tundra,[3] lakes,[4] and marine environments.[5] Geographic distributions of microorganisms are a long-standing and ongoing inquiry in the field of microbiology,[6–8] which provides a starting point to understand ecosystem functioning,[9] such as the global biogeochemical cycling of carbon (C) and nutrients.[10] The Baas-Becking hypothesis, which posits that "everything is everywhere–the environment selects," has long been a foundation for evaluating biodiversity patterns across various biomes. Thus, the contrasting conditions of soils and water may result in distinct microbial taxa and functional diversities between aquatic and terrestrial environments.[11–13]

Metagenomics in international scenarios has been extensively studied for the soil and aquatic microbial communities and their functions,[14] covering a wide range of geographic regions around the globe.[15] Generally, soil microbial communities show marked temporal and spatial heterogeneity,[16,17] while those in the aquatic ecosystems also display considerable variability in their physical and chemical properties.[18,19] For instance, analyses of global topsoil samples have demonstrated that microbiomes exhibit distinct niche adaptability and spatial variations in their relative contribution to global C and nutrient cycling,[20] such as the decomposition of organic residues,[21] mineralization of nutrients, and formation of stable soil organic matter.[22] Similarly, microorganisms that play crucial roles in the C and nutrient cycling of aquatic ecosystems[23] also exhibit large-scale horizontal and vertical patterns worldwide, spanning from the ocean surface to the deep seafloor, where water temperature, salinity, and oxygen content have significant impacts on aquatic microbial diversity.[24] Although large-scale metagenomic studies have explored microbial composition and potential functional patterns worldwide, the direct comparison of microorganisms as well as their functions between soil and water ecosystems remains lacking in international scenarios. Comparing species composition and functionality of microbial communities between aquatic and terrestrial habitats can enhance our understanding of microbial diversity, and ecosystem functions, as well as provide strategies for managing and preserving transitional domains.

Relationships between microbial diversity and functions remain a topic of debate,[25] although the impact of biodiversity on ecosystem stability, productivity, and resilience toward stress and disturbances[26] has long been postulated. Although microbial diversity has been found to exhibit strong correlations with specialized functions such as C and nutrient cycling,[27] emerging evidence suggests a decoupling between microbial taxonomy and function, suggesting the existence of functional redundancy within microbial communities[28–30] meaning that functional similarities might exist among different microbial taxa.[31] Furthermore, the extent of functional redundancy is contingent upon the environmental conditions[32] and the specific type of functions being considered.[30,31] Therefore, it is imperative to investigate the relationship between microbial taxa and functions within diverse functional categories in different environments.

[1]State Key Laboratory of Biocontrol, School of Ecology, Shenzhen Campus of Sun Yat-sen University, Shenzhen, Guangdong 518107, China
[2]Lead contact
*Correspondence: huangdd26@mail.sysu.edu.cn (D.H.), chenhh68@mail.sysu.edu.cn (H.C.)
https://doi.org/10.1016/j.isci.2024.111047

Recently, shotgun metagenomics, a robust tool for the comprehensive exploration of microbial taxonomies and functions within terrestrial and aquatic ecosystems,[33] has greatly expanded our understanding of environmental microorganisms.[34,35] Numerous studies have elucidated the patterns of microbial community and functional diversity as well as their interrelationships at large scales.[6,36] Given that most studies deposited raw data in public gene banks such as MG-RAST,[37] data from these untargeted sequencing approaches encompassing the entire microbial genome offer the capability to conduct a global metagenomic analysis of microbial members in an ecosystem[38] and provide a wealth of data for comprehensive analyses and subsequent comparisons. Here, we hypothesized that soil and aquatic microbial communities display significant taxonomic and functional disparities at the metagenomic level. Particularly, these microbial communities were expected to exhibit distinct characteristics and functional patterns in mediating biogeochemical cycles, such as C and nutrient cycling, within soil and aquatic environments. Based on existing research suggesting a potential decoupling between microbial taxonomy and function,[28–30] we proposed that varying degrees of functional redundancy existed among microorganisms inhabiting soil and aquatic ecosystems. Such redundancy was pivotal in stabilizing and maintaining biogeochemical processes, thereby providing a resilience mechanism that mitigates the impact of environmental changes. To test our hypotheses in this study, we constructed microbial taxonomic and functional datasets based on 933 soil metagenomes and 938 water metagenomes from recent publications in the MG-RAST server. Based on these metagenomes, our objective is to investigate the following aspects: (I) Disparities in taxonomic composition and functional attributes between terrestrial and aquatic metagenomes; (II) Variations in taxonomic composition and functional profiles of microbiomes involved in biogeochemical cycling across terrestrial and aquatic environments; (III) Similarity of microbial taxonomic and functional diversities, as well as their correlations in soil and water microbiomes associated with different biogeochemical cycles.

RESULTS

Metagenomic taxonomy and function between soil and water

Overall, within taxonomic genes annotated in RefSeq databases (Figure 1), the relative abundances of Archaea, Bacteria, Eukaryota, and Virus in the metagenomes differed significantly between soil and water biomes (Figure 2A, $p < 0.05$). Specifically, the relative abundances of Bacteria were significantly higher in the soil (96.45%) than in the water (91.52%). On the contrary, the proportion of Archaea and Eukaryota were approximately 1.73 and 2.05 times greater in water than in soil, respectively. The difference was even greater in the relative abundance of the virus, which was higher in water (1.73%) than in soil (0.06%). To further investigate the difference in taxonomic compositions, the read counts of Archaea, Bacteria, Eukaryota, and Virus taxa were re-normalized to allow for a comparison of the relative abundance at each taxonomic level of phylum, class, or genus between soil and water biomes (Figure 2B). Generally, the profiles of archaeal composition at class levels showed similarities between soil and water biomes, however, significant differences were still present (Figure 2B, $p < 0.05$). For example, soil biomes had greater relative abundances of Halobacteria, Thermoprotei, and Archaeoglobi compared to water biomes. On the other hand, water biomes harbored relatively higher proportions of Methanococci and Methanobacteria. Regarding bacterial composition in the metagenomes, Actinobacteria, Verrucomicrobia, Planctomycetes, and Chloroflexi displayed 2.47–3.61 times higher relative abundance in soil compared to water. In particular, Acidobacteria increased from 0.72% in water to 2.39% in soil. However, the relative abundances of Bacteroidetes, Cyanobacteria, and Proteobacteria, including Alphaproteobacteria, Gammaproteobacteria, and Betaproteobacteria, exhibited an inverse trend with 1.27–2.82 times higher abundance in water than in soil biomes. The compositions of eukaryota at the phylum level were more evenly distributed in water biomes than in soil biomes. Specifically, when compared to water biomes, soil biomes hosted 1.34–3.06 times higher relative abundance of fungi belonging to Ascomycota and Basidiomycota as well as Streptophyta, but a lower relative abundance of metagenomic reads from other animals and plants such as Chlorophyta and Apicomplexa were found in soil ($p < 0.05$). At the family level of viral metagenomes, Microviridae and Siphoviridae showed higher relative abundance, while Myoviridae and Phycodnaviridae exhibited a lower abundance in soil than in water biomes ($p < 0.05$).

To assess the variations in metagenomic functions annotated in the SEED Subsystems database performed by different taxonomic groups between soil and water biomes, we constructed a non-metric multidimensional scaling (NMDS) plot to compare the relative abundance of metagenomic functions attributed to all Archaea, Bacteria, Eukaryota, and Virus (Figure 2C). Generally, viral metagenomic functions were significantly different from those of Archaea, Bacteria, and Eukaryota. The functional profile of soil bacteria was more closely aligned with that of water than other groups, while the differences in viral functions were greatest between soil and water biomes. Besides, the heat maps with dendrograms generated through hierarchical cluster analysis depicted the relative abundance of dominant functions at SEED Subsystem level 3 and provided evidence of these functional distinctions (Figure S1). Specifically, at the functional classification level of viruses, clustering-based subsystems (CBS), Motility and chemotaxis (MOT, e.g., Rhamnolipids), and RNA metabolism (RNA, e.g., rRNA modification, RNA methylation, RNA pseudouridine syntheses, and tRNA modification), the three most abundant functions accounting for 62.61% of the total reads, exhibited higher proportions in the soil compared to water biomes. However, metagenomic reads associated with phages, prophages, transposable elements, plasmids (PPT, e.g., r1t-like streptococcal phages, phage capsid proteins, phage packaging machinery, phage replication) demonstrated a significant dominance of 84.58% within the water biomes (Figures 2D and S1). Meanwhile, it was observed that archaea, bacteria, and eukaryotes in the soil contributed a higher proportion of metagenomic reads involved in membrane transport (MEM, e.g., Periplasmic-binding-protein-dependent Transport System for Glucosides), Regulatory and cellular signaling (RCS, e.g., cAMP signaling in bacteria). On the other hand, soil biomes showed a lower relative abundance of reads involved in the metabolisms of amino acids (AAD, e.g., Putrescine utilization pathways) and Protein (PRO, e.g., Glycine reductase, Sarcosine reductase, and Betaine reductase) compared to water biomes. Furthermore, Venn diagrams revealed a significant overlap of taxonomic composition between soil and water biomes, ranging from 29% for eukaryota to 62% for bacteria (Figure 3A). However, when examining functional aspects, the overlap was considerably

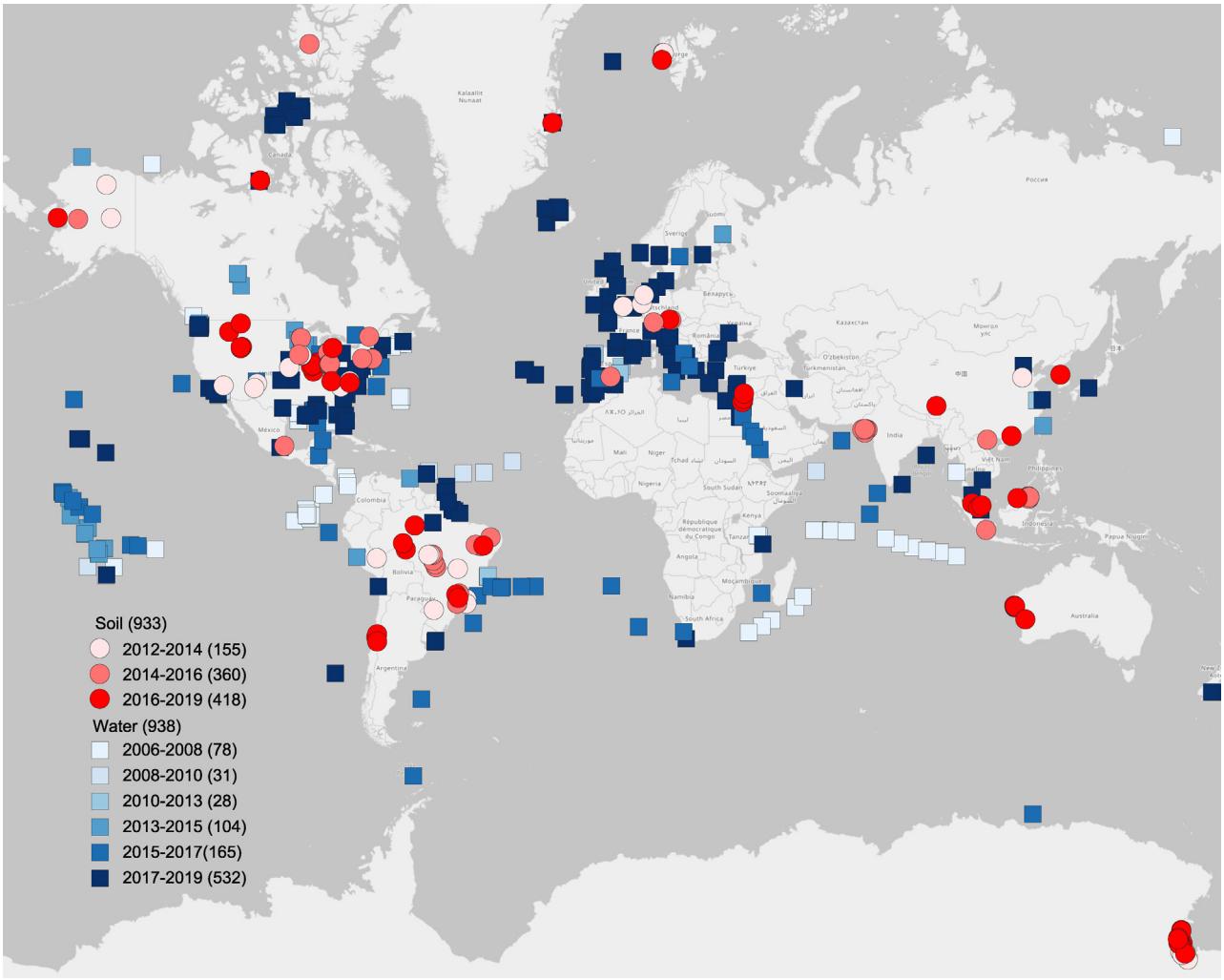


Figure 1. Global distribution of soil and water metagenomes

Locations of 933 soil metagenomes from 55 publications and 938 water metagenomes from 55 publications are used in this study. Legends show groups of publication periods. Sample sizes of each group are given in parentheses.

lower, ranging from 1% for viruses to 7% for archaea and bacteria. Soil microorganisms generally shared 10% of the functional reads and exhibited greater similarity in functional profiles compared to microorganisms in the water only sharing 5% of the functional reads (Figure 3B).

Metagenomic taxonomy and function involved in biogeochemical cycles in soil and water

Boxplots were constructed to compare taxonomic and functional compositions of reads involved in C and nutrient cycling based on Bray-Curtis pairwise distance (Figure 4A). Overall, we observed variations in the levels of taxonomic and functional similarity among metagenomes, with differences being evident across habitats (soil and water) and biogeochemical cycles (CAR, NIT, SUL, and so forth) (Figure 4A). Specifically, for microbiomes in soil and aquatic environments associated with NIT, PHO, and POT, metagenomic functional similarity (48% for NIT, 50% for PHO, and 34% for POT) exhibited higher levels of resemblance than taxonomic similarity (31% for NIT, 30% for PHO, and 21% for POT). Similar trends were found in the comparison of taxonomic (58% for NIT, 56% for PHO, and 51% for POT) and functional similarity (62% for NIT, 60% for PHO, and 59% for POT) in water. Conversely, metagenomes linked to CAR and IRO displayed contrasting patterns of higher similarity of taxonomy (68% in soil and 43% in water for CAR, 49% in soil and 29% in water for IRO) than function (53% in soil and 41% in water for CAR, 38% in soil and 24% in water for IRO) in both soil and water environments. Furthermore, taxonomic similarity among metagenomes related to SUL (58%) was greater than their functional similarity (52%) in the soil environment; however, this trend was reversed in the water environment (31% for taxonomy and 38% for function). To investigate the relationship between taxonomy and function in soil and water, pairwise comparisons of Bray-Curtis similarity were created between functional and taxonomic diversities. Regardless, significant positive correlations between taxonomy and function were found (Figure 4B, $p < 0.001$), with stronger correlations consistently observed in water (r^2 = 0.33–0.68)

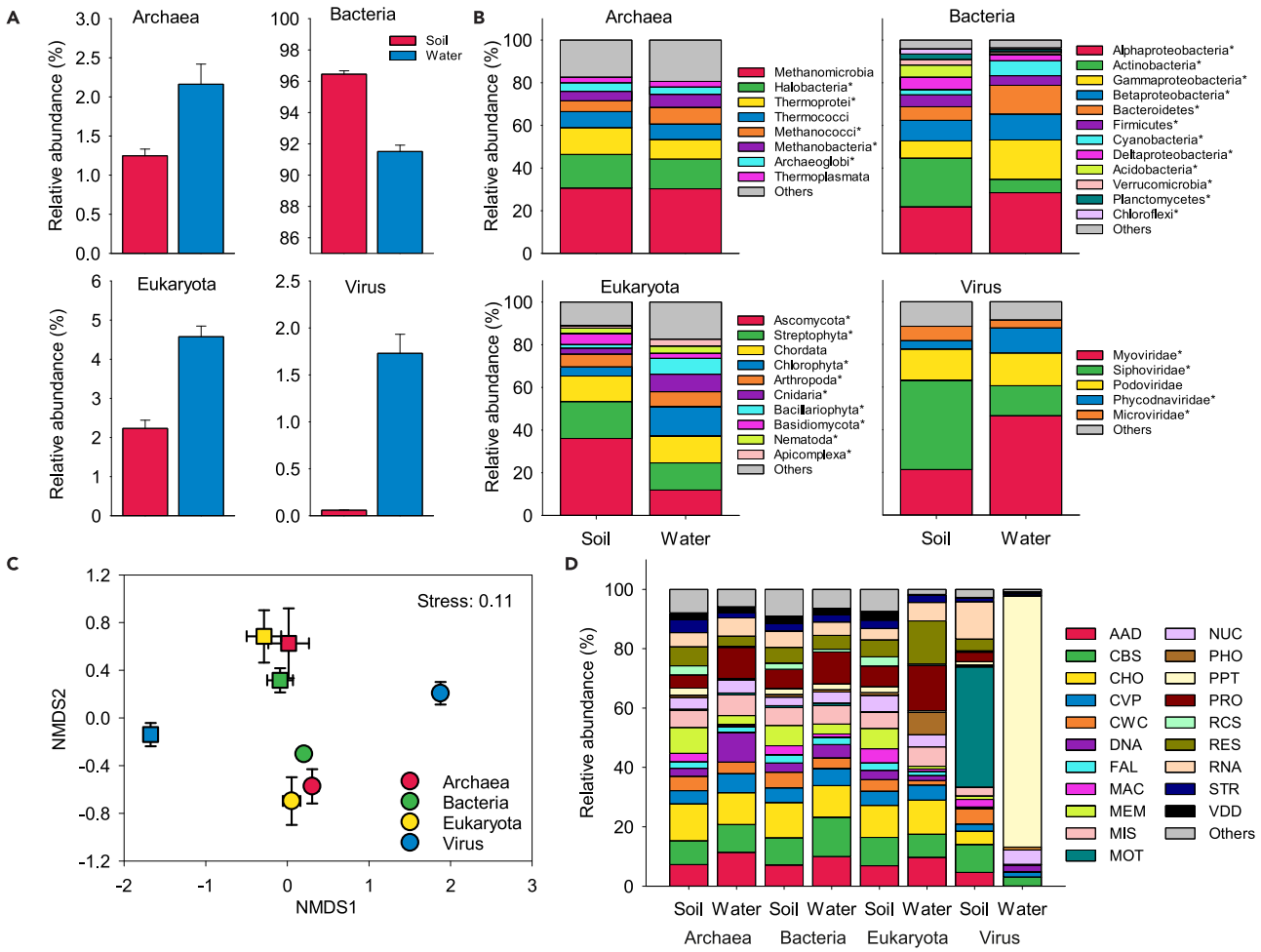


Figure 2. Microbial taxonomic and functional diversities in soil and water

(A) Relative abundance of archaea, bacteria, eukaryota, and viruses in soil and water.

(B) Relative abundance of dominant microbial taxonomic compositions at the phylum or family or genus level (mean >1%).

(C) Non-metric multidimensional scaling (NMDS) shows functional metagenomic beta diversities of all archaea, bacteria, eukaryote, and viruses from soil and water samples. The error bars represent the standard deviation of the data.

(D) Relative abundance of dominant soil and water metagenomes (mean > 1%) annotated in the Subsystems database at level 1 (Function). AAD, amino acids and derivatives; CBS, clustering-based subsystems; CHO, carbohydrates; CVP, cofactors, vitamins, prosthetic; CWC, cell wall, and capsule; DNA, DNA metabolism; FAL, fatty acids, lipids, and isoprenoids; MAC, metabolism of aromatic compounds; MEM, membrane transport; MIS, miscellaneous; MOT, motility and chemotaxis; NUC, nucleosides and nucleotides; PHO, Photosynthesis; PPT, phages, prophages, transposable elements, plasmids; PRO, protein metabolism; RCS, regulation and cell signaling; RES, respiration; RNA, RNA metabolism; STR, stress response; VDD, virulence, disease, and defense. Data are represented as mean ± SEM.

than those in soil (r^2 = 0.23–0.34). In addition, the functions related to SUL (r^2 = 0.33–0.68) and POT (r^2 = 0.34–0.54) generally had stronger correlations than other functions (r^2 = 0.23–0.45) in both soil and water biomes.

Unique microbial taxonomic compositions associated with biogeochemical cycling

Globally, the taxonomic compositions of microbial communities involved in C and nutrient functions exhibited distinct patterns between soil and water metagenomes. The PCoA revealed that the taxonomic structure of microbes related to the metabolisms of carbohydrates (CAR), Nitrogen (NIT), Phosphorus (PHO), Sulfur (SUL), Iron (IRO), and Potassium (POT) in soil metagenomes formed separate clusters from those in water metagenomes (Figure 5A, $p < 0.001$). Among them, the function of IRO displayed the greatest separation from other functions in both soil and water metagenomes, especially in soil environments. Specifically, the relative abundance of major phyla associated with C and nutrient functions, such as Proteobacteria (61%), Actinobacteria (13%), Bacteroidetes (8%), Cyanobacteria (5%), and Firmicutes (4%), and so forth showed a distinct difference between soil and water metagenomes (Figure 5B). Specifically, relative abundances of Acidobacteria, Actinobacteria, Chloroflexi, Deltaproteobacteria, and Firmicutes in terrestrial ecosystems were significantly higher than those in aquatic

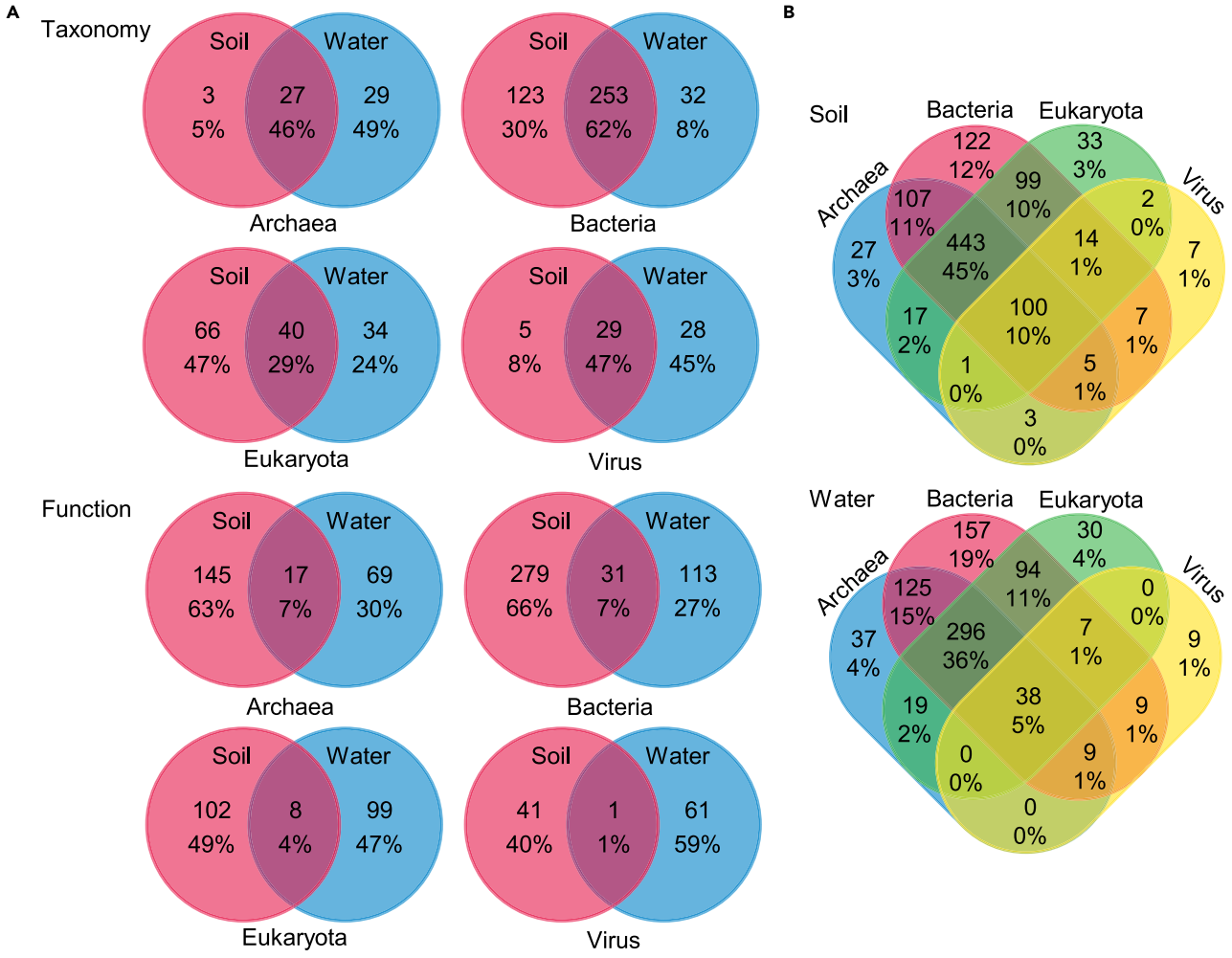


Figure 3. "Shared" taxonomy and function

(A) "Shared" taxonomy and function between soil and water. Venn's diagrams show the overlap of dominant microbial taxonomic compositions and functions in soil and water metagenomes.

(B) "Shared" function between taxonomies. Venn diagram, where areas of overlap indicate the number and percentage of taxonomies with the same function among the overlapping groups.

ecosystems (Figure 5C, $p < 0.05$). Particularly, in comparison to aquatic ecosystems, Acidobacteria and Actinobacteria exhibited a 10-fold and 3-fold increase, respectively, in soil environments. In these phyla, the main genera showing significant differences among habitats included *Solibacter* and *Koribacter* (Acidobacteria), *Mycobacterium*, *Streptomyces*, and *Frankia* (Actinobacteria), and *Bacillus* (Firmicutes) (Table S3). Conversely, Bacteroidetes, Cyanobacteria, Gammaproteobacteria, and Euryarchaeota were relatively predominant in water metagenomes for C and nutrient metabolisms than soil ones (Figure 5C, $p < 0.05$). Notably, there was a significant increase of Bacteroidetes by 79% and Cyanobacteria by 184% in aquatic ecosystems compared to the soil environments, including major genera, marine bacterium *Prochlorococcus* and freshwater bacterium *Synechocystis*, both belonging to Cyanobacteria (Table S3).

The statistical significance of the taxonomic difference between soil and water metagenomes was found to be similar for these six functions (Figure 6A, Pseudo-F = 375–552, $p < 0.001$), contributing to approximately 26.3–31.9% of the beta-diversity variation. Among these functions, microbes involved in POT exhibited the highest taxonomic difference of 31.9%, while those related to SUL showed the lowest difference of 22.4%. Additionally, Venn's diagrams revealed that microbes conducting different functions shared 55–69% of dominant genera between soil and water environments, with CAR-associated species having the highest overlap while NIT-associated species shared the least (Figure 7A). The PERMANOVA Pairwise tests also confirmed the divergence in beta-diversity of microbial taxonomic compositions at the genus level for conducting C and nutrient functions was more significant between soil and water biomes (pseudo-t = 19.4–28.4, $p < 0.001$) than within each biome (pseudo-t = 7.8–23.1, $p < 0.001$) (Table S2). Furthermore, we found that the disparity in the composition of microbe taxa involved in C and nutrient functions within soil biomes (pseudo-t = 14.8–23.1, $p < 0.001$) was more pronounced than in water biomes (pseudo-t = 7.8–13.0, $p < 0.001$). In particular, the taxonomic composition of microbes involved in IRO function exhibited the highest dissimilarity compared to other

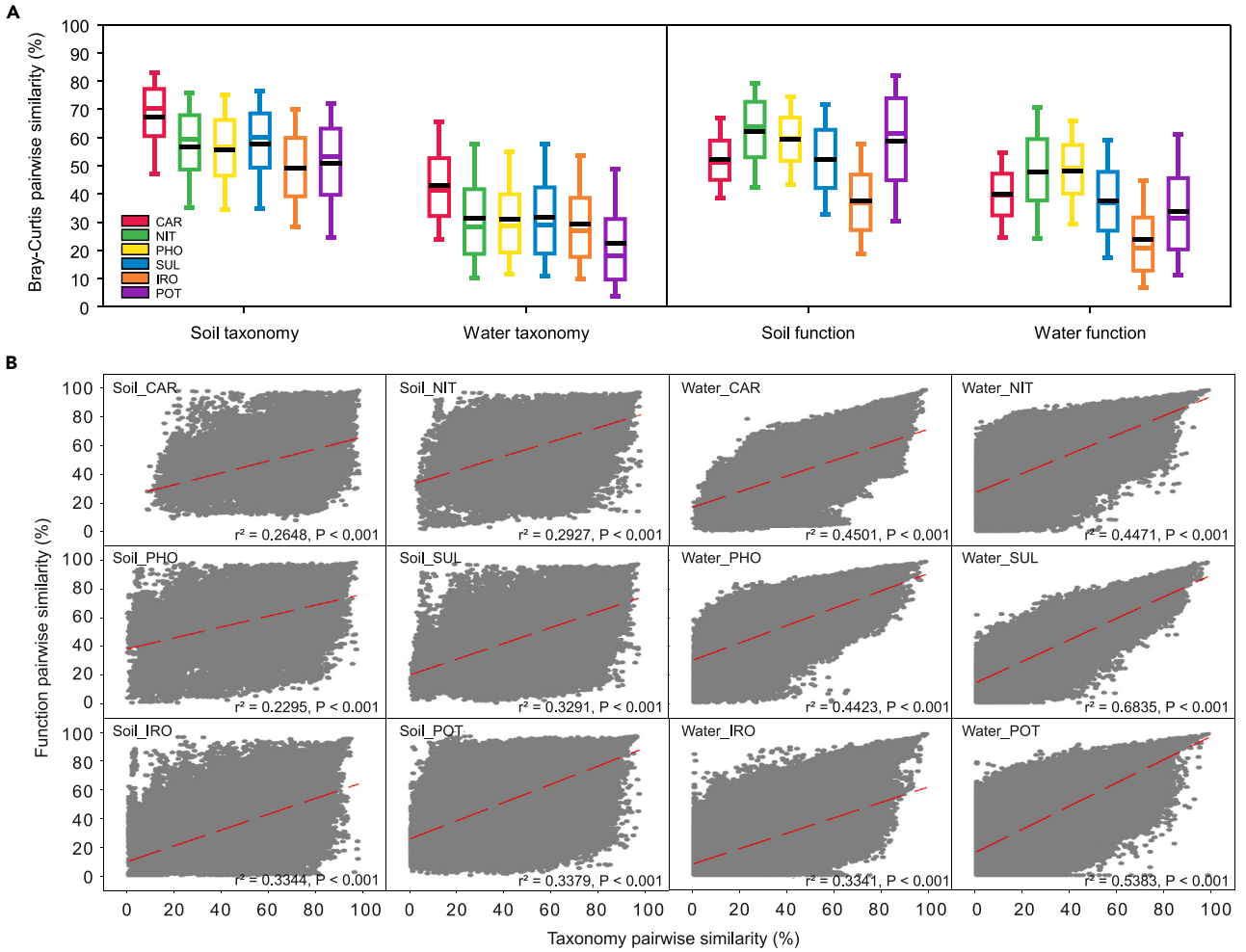


Figure 4. Pairwise similarity of taxonomy and function in soil and water biomes

(A) Soil and water similarity of taxonomy and function. Boxplots and mean values (black line) of pairwise Bray-Curtis similarity of taxonomic compositions at genus levels (Taxonomy) and functional profiles at function levels (Function) for conducting carbohydrate and nutrient functions in soil and water metagenomes.

(B) Relations between taxonomic and functional beta diversities. Pearson's correlations between pairwise Bray-Curtis similarity of microbial taxonomic and functional compositions for conducting carbohydrate and nutrient functions in soil and water metagenomes. The r-squared and *p* values are given. CAR (carbohydrates), NIT (nitrogen metabolism), PHO (phosphorus metabolism), SUL (sulfur metabolism), IRO (iron acquisition and metabolism), and POT (potassium metabolism).

functions in both soil (Pseudo-t = 20.2–23.1, $p < 0.001$) and water metagenomes (Pseudo-t = 8.9–13.0, $p < 0.001$), suggesting that IRO was one of the processes governed by highly specialized microbial communities among tested six biogeochemical cycles. Besides, Venn's diagrams showed that the shared dominant genera involved in different C and nutrient functions accounted for 39% of soil metagenomes and 43% in water metagenomes (Figure 7B).

Distinct microbial functional profiles for biogeochemical processes

Generally, distinct patterns between soil and water metagenomes were also observed in metagenomic functional profiles related to carbohydrate and nutrient metabolisms (Figure 6B). The functions that exhibited the greatest dissimilarity were POT and SUL, with 28.5% (Pseudo-F = 463, $p < 0.001$) and 22.4% (Pseudo-F = 295, $p < 0.001$) of variation contributed by soil and water metagenomes, respectively. The differences in PHO and NIT functions between soil and water were relatively less significant, with an explanation of variation reduced to 13.9% (pseudo-F = 160, $p < 0.001$) and 16.8% (pseudo-F = 222, $p < 0.001$), respectively.

Soil metagenomes exhibited higher abundances of CAR, SUL, and POT functions, whereas water metagenomes were more prevalent in NIT and IRO functions (Figure 5D, $p < 0.001$). Regarding the CAR function, soil metagenomes (0.04–4.61%) were significantly greater in nearly all functional categories at SEED Subsystems level 2 than water metagenomes (0.20–4.01%), except CO_2 fixation which was on average 0.14% more abundant in water than soil metagenomes (Table S4, $p < 0.001$). For the SUL function, terrestrial ecosystems displayed a dominance of

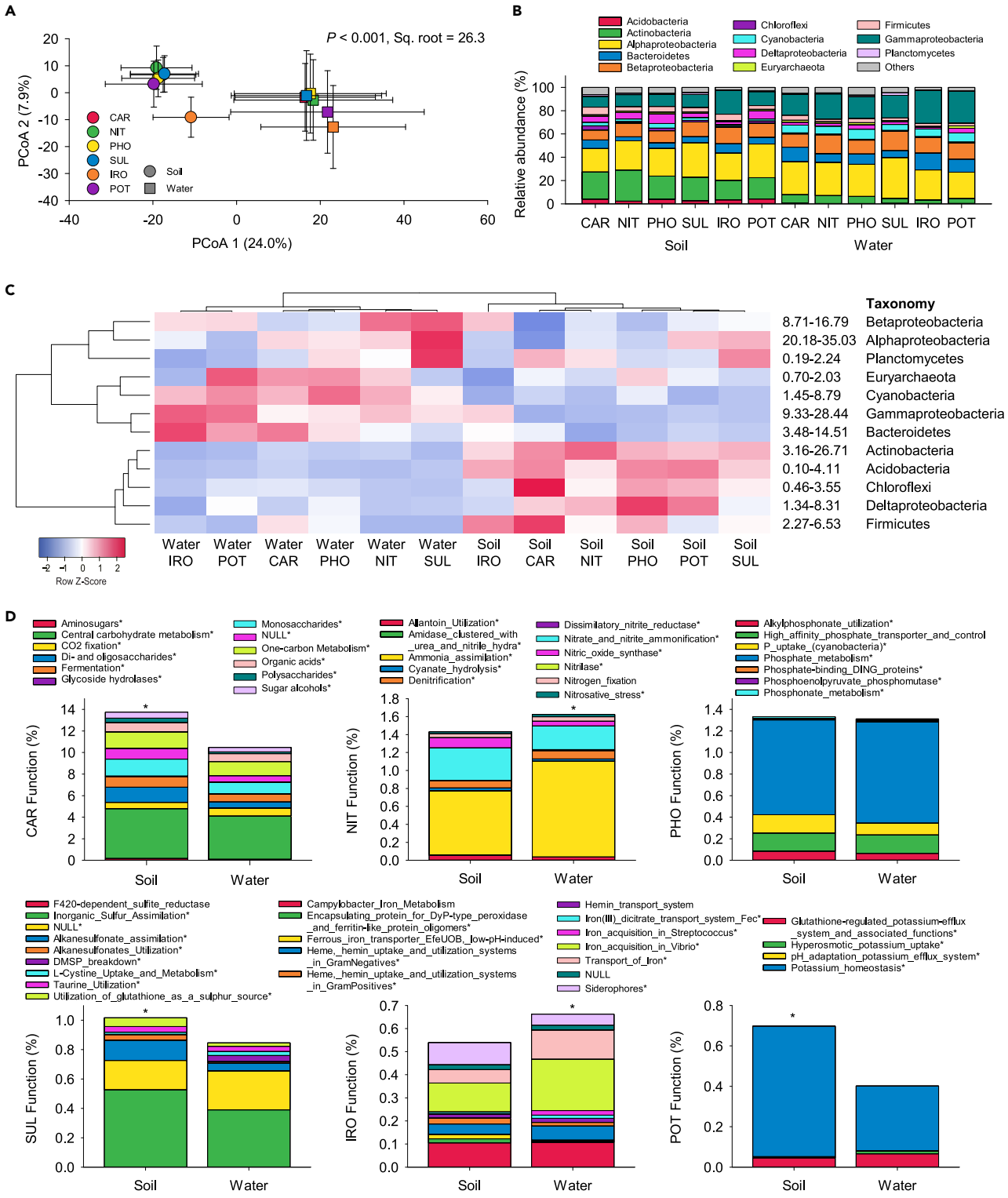


Figure 5. Distinct microbial taxonomic compositions and functional profiles

(A) Principal coordinates analysis (PCoA) showing the beta diversity of microbial taxonomic compositions at genus levels for conducting carbohydrate and nutrient functions in soil and water metagenomes. The error bars represent the standard deviation of data ranges. Variation explained by two principal coordinate dimension is given in parentheses by percentage. *p* values and Sq. root of PERMANOVA are also given. CAR (carbohydrates), NIT (nitrogen metabolism), PHO (phosphorus metabolism), SUL (sulfur metabolism), IRO (iron acquisition and metabolism), and POT (potassium metabolism).

Figure 5. *Continued*

(B) Relative abundance of dominant microbial taxonomic compositions at phylum or class levels (mean >1%) for conducting carbohydrate and nutrient functions in soil and water metagenomes.

(C) Heatmaps showing normalized relative abundance of dominant microbial taxonomic compositions at phylum or class levels (mean >1%) for conducting carbohydrate and nutrient functions in soil and water metagenomes. Dendrograms of hierarchical cluster analysis are shown to group taxonomic phylum or class as well as carbohydrate and nutrient functions in soil and water metagenomes.

(D) Relative abundance of microbial functional categories for conducting carbohydrate (level 2) and nutrient functions (level 3) in soil and water metagenomes. Asterisks indicate significant differences between soil and water metagenomes at $\alpha = 0.05$. Data are represented as mean ± SEM.

alkanesulfonate assimilation (0.14%) and utilization (0.04%), inorganic sulfur assimilation (0.53%), and the utilization of glutathione as a sulfur source (0.06%), but water biomes prevailed in DMSP breakdown (0.04%) and L-cystine uptake and metabolism (0.02%). In terms of the POT function, soil metagenomes exhibited a higher abundance in the function of potassium homeostasis (0.65%), while other functional categories were prevalent in water metagenomes such as Glutathione-regulated potassium-efflux system and associated functions (0.65%) ($p < 0.001$). For the NIT function, water biomes were dominant in Ammonia assimilation (1.07%), while terrestrial ecosystems prevailed in Allantoin utilization (0.05%), Nitrate and nitrite ammonification (0.36%), and nitric oxide synthase (0.11%) ($p < 0.001$). For the IRO function, Iron acquisition in Streptococcus (0.02%) and Vibrio (0.22%) along with Iron (III) dicitrate transport system Fec (0.01%) and Transport of Iron (0.13%) exhibited higher abundance in water than soil metagenomes. Conversely, soil metagenomes displayed greater abundances of genes related to metabolisms for DyP-type peroxidase and ferritin-like protein oligomers (0.02%), Ferrous iron transporter EfeUOB (low-pH-induced) (0.02%), Heme and hemin uptake and utilization systems in Gram Positives (0.03%), and Siderophores (0.10%) ($p < 0.001$). Although the relative abundance of the PHO function did not significantly differ between soil and water metagenomes, its sub-level categories, such as Alkylphosphonate utilization (0.08%), P uptake (cyanobacteria) (0.17%), and phosphonate metabolism (0.02%) were more prevalent in terrestrial ecosystems. Nevertheless, water biomes demonstrated a higher relative abundance in phosphate metabolism (0.94%) ($p < 0.001$).

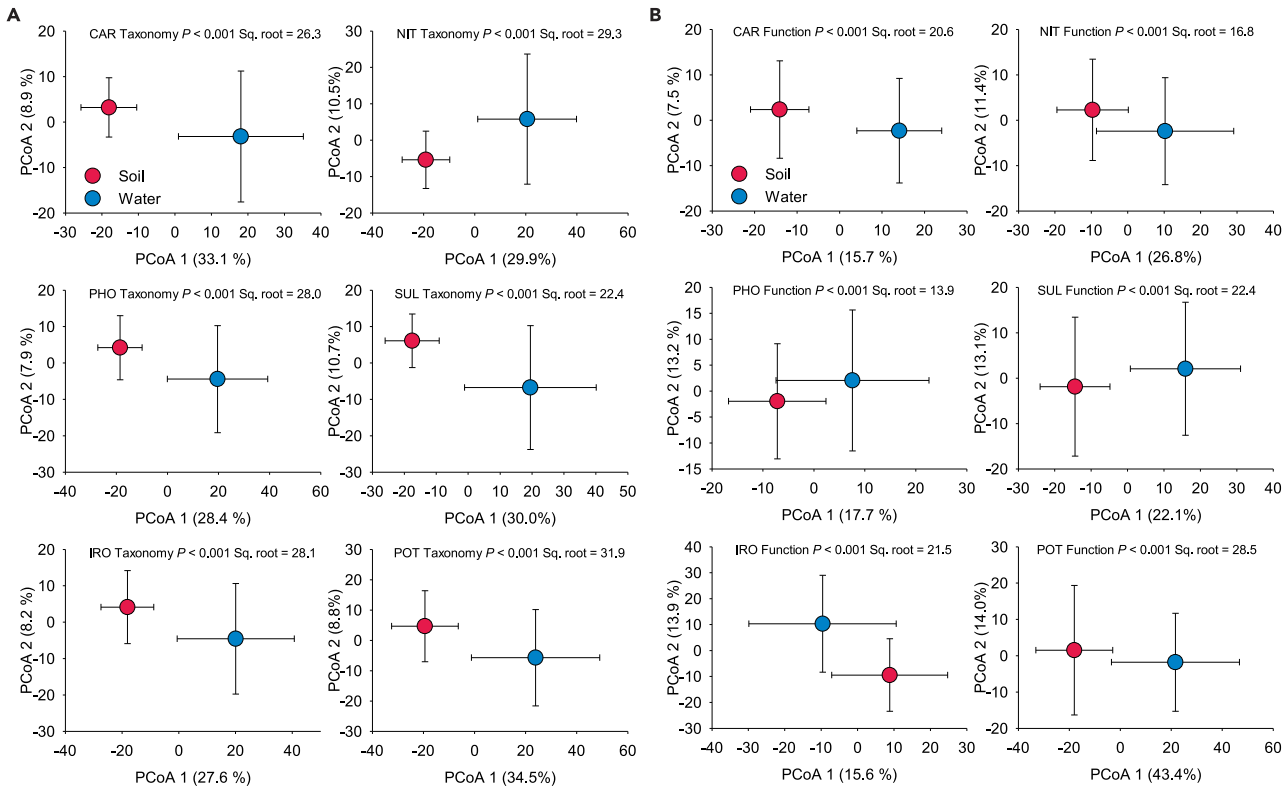


Figure 6. Soil and water beta-diversity of function

(A) PCoA (principal coordinates analysis) showing the beta diversity of microbial taxonomic profiles for conducting carbohydrate and nutrient functions in soil and water metagenomes.

(B) PCoA (principal coordinates analysis) showing the beta diversity of microbial functional profiles at function levels for conducting carbohydrate and nutrient functions in soil and water metagenomes. The error bars represent the standard deviation of data ranges. Variation explained by two principal coordinate dimensions is given in parentheses by percentage. *p* values and Sq. root of PERMANOVA are also given. CAR (carbohydrates), NIT (nitrogen metabolism), PHO (phosphorus metabolism), SUL (sulfur metabolism), IRO (iron acquisition and metabolism), and POT (potassium metabolism). Data are represented as mean ± SEM.

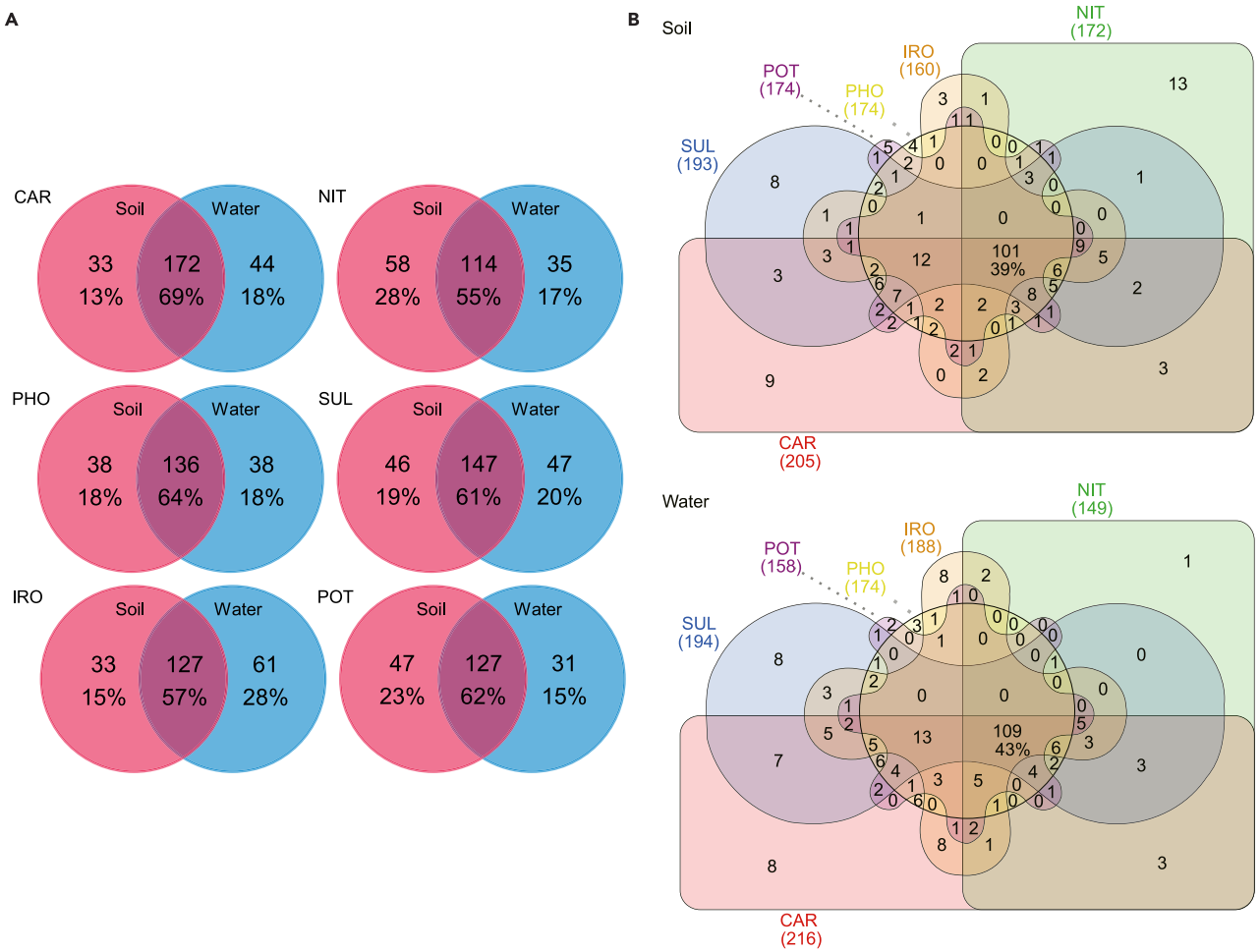


Figure 7. "Shared" taxonomy relating to carbohydrate and nutrient functions

(A) "Shared" taxonomy between soil and water. Venn's diagrams show dominant microbial taxonomic compositions at genus levels (mean >0.1%) for conducting carbohydrate and nutrient functions shared between soil and water metagenomes. CAR (carbohydrates), NIT (nitrogen metabolism), PHO (phosphorus metabolism), SUL (sulfur metabolism), IRO (iron acquisition and metabolism), and POT (potassium metabolism).

(B) "Shared" taxonomy among functions. Venn's diagrams show dominant microbial taxonomic compositions at genus levels (mean >0.1%) for conducting carbohydrate and nutrient functions shared among different functions in soil and water metagenomes.

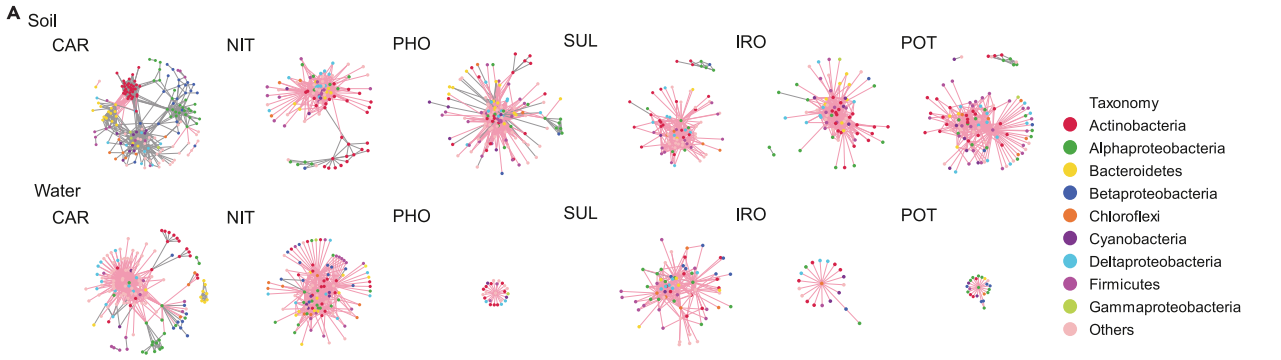
Co-occurrence networks of biogeochemical cycling and associated taxonomies in soil and water

In addition, a network analysis approach was employed to investigate the co-occurrence patterns among different microbiomes. Generally, compared to water metagenomes, microbes in soil metagenomes exhibit larger and more complex networks in CAR, PHO, IRO, and POT functions as indicated by the higher numbers of total nodes, links, average connectivity, average clustering coefficient, average geodesic distance, and modularity (Figure 8A). Specifically, for the CAR function in soil metagenomes, most taxonomic interactions occurred within the same module of phylum or class, such as Actinobacteria, Alphaproteobacteria, Bacteroidetes, and Betaproteobacteria, with a modularity of 0.54, whereas the taxonomic interactions in other functions displayed a lower modularity of 0.18–0.31. Among six pairs of functional networks examined, there was no consistent trend between soil and water metagenomes (Figure 8B). The CAR and POT functions had more total nodes and links in soil compared to water metagenomes, while the IRO function showed less total nodes and links. Additionally, the NIT, PHO, and SUL functions exhibited longer average path distances in soil functional networks when compared with those of water metagenomes.

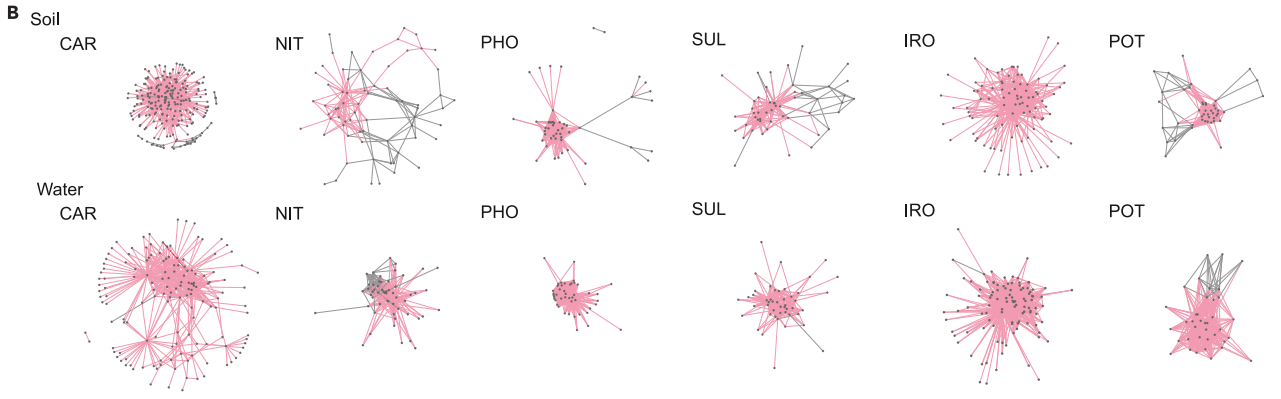
DISCUSSION

Heterogeneity in taxonomic and functional diversity between soil and water

As the most heterogeneous component of the biosphere,[39] soil possesses a vast internal surface area that accommodates a substantial amount of microbial biomass.[40] Among these microbial communities, soil bacteria are widely considered to be by far the most numerous organisms,[8,41] surpassing other soil microorganisms such as archaea, eukaryotes, and viruses in terms of their relative abundance as also



	Soil						Water					
	CAR	NIT	PHO	SUL	IRO	POT	CAR	NIT	PHO	SUL	IRO	POT
Total nodes	175	101	139	82	74	106	146	110	25	75	20	23
Total links	863	528	713	286	298	294	414	418	24	244	19	23
Average degree (avgK)	9.86	10.46	10.26	6.98	8.05	5.55	5.67	7.60	1.92	6.51	1.90	2.00
Average clustering coefficient (avgCC)	0.48	0.38	0.44	0.29	0.17	0.16	0.28	0.01	0.00	0.03	0.00	0.00
Average path distance (GD)	3.39	2.77	2.65	2.20	2.25	2.24	2.89	2.46	1.92	2.45	1.99	1.98
Modularity (No. of modules)	0.54(6)	0.18(4)	0.22(5)	0.25(3)	0.22(4)	0.31(5)	0.35(7)	0.25(4)	1(0)	0.24(3)	1(0)	1(0)



	Soil						Water					
	CAR	NIT	PHO	SUL	IRO	POT	CAR	NIT	PHO	SUL	IRO	POT
Total nodes	222	68	57	52	106	50	148	55	53	47	112	43
Total links	669	167	446	192	398	466	458	281	429	246	719	399
Average degree (avgK)	6.03	4.91	15.65	7.39	7.51	18.64	6.19	10.22	16.19	10.47	12.84	18.56
Average clustering coefficient (avgCC)	0.05	0.23	0.58	0.23	0.26	0.68	0.01	0.33	0.61	0.30	0.29	0.68
Average path distance (GD)	3.18	3.25	2.15	2.41	2.11	1.85	3.08	1.94	1.70	1.97	1.95	1.60
Modularity (No. of modules)	0.34(10)	0.44(5)	0.11(4)	0.25(4)	0.23(6)	0.12(3)	0.34(5)	0.19(4)	1(0)	0.15(4)	0.14(3)	0.09(2)

Figure 8. Co-occurrence networks of taxonomy and function

(A) Co-occurrence networks of microbial taxonomies. Network graphs of taxonomic compositions at genus levels for conducting carbohydrate and nutrient functions in soil and water metagenomes. Node color represents the classification of taxonomic compositions at phylum or class levels. A black edge indicated a positive interaction between two nodes, while a red edge indicated a negative interaction. A summary of key network indexes is given in the table. CAR (carbohydrates), NIT (nitrogen metabolism), PHO (phosphorus metabolism), SUL (sulfur metabolism), IRO (iron acquisition and metabolism), and POT (potassium metabolism).

(B) Co-occurrence networks of microbial functions. Network graphs of functional categories at function levels for conducting carbohydrate and nutrient functions in soil and water metagenomes. A black edge indicated a positive interaction between two nodes, while a red edge indicated a negative interaction. A summary of key network indexes is given in the table.

confirmed in our study (Figure 2A). Besides, water has also been acknowledged as a major bacterial habitat,[42] but our findings indicated that the relative abundance of bacteria in water was significantly lower compared to in soil ($p < 0.05$), accompanied by an increase in the relative abundance of archaea, eukaryotes, and viruses. It was worth noting that viral metagenomes exhibited the most pronounced disparity between soil and water, with approximately 29 times higher relative abundance observed in water than in soil, a finding similar to a previous report on the earth's virome,[43] which is possibly attributable to high viral concentration found in the marine.[43,44]

Based on the metagenomic databases, our analysis revealed significant variations in metagenomic taxonomic compositions between soil and water (Figure 2B). The bacterial community was dominated by a diverse group of Proteobacteria, including Alpha-, Beta-, Delta-, and

Gamma-proteobacteria, possibly due to their wide range of metabolic diversity,[45] including phototrophs, photoheterotrophs, and chemolithotrophs, which contribute to their widespread distribution and greater abundance in aquatic ecosystems.[46,47] Compared to the microbial communities in the watershed, soil microbiomes showed a much higher relative abundance of Acidobacteria and Actinobacteria but harbored relatively lower proportions of Bacteroidetes and Cyanobacteria compared to water microbiomes. Acidobacteria are primarily heterotrophic, with most species being aerobic or microaerophilic.[48] Due to cellular specialization, enzyme stability, and a wide range of nutrient uptake transporters, Acidobacteria communities can thrive in lower pH and stressful soil environments.[49] Although recent studies have confirmed the presence of Actinobacteria populations across different aquatic environments,[50,51] Actinobacteria are traditionally known as inhabitants of the soil[52] that decompose complex mixtures of organic matter from dead animals, plants, and microbes.[53] In contrast, the strict anaerobes, such as most species of Bacteroides[54] and Cyanobacteria[55] that benefit from nanomolar concentrations of oxygen in water, were more abundant in the aquatic environments.

As for metagenomic reads belonging to Eukaryota, the absolute dominance of Ascomycota in all of our soil samples (Figure 2B) aligns with patterns observed in soils elsewhere,[56] which is dictated by their wind dispersal abilities, lifestyles, and functional attributes.[57] Compared to aquatic environments, soil provides a richer source of organic matter[58] with a solid matrix that facilitates the growth and attachment of fungal hyphae for Ascomycota[59] as decomposers or saprophytes.[60] Besides, Streptophyta and Chlorophyta constitute two major lineages of green algae, with the former encompassing land plants,[61] exhibiting contrasting distribution patterns between terrestrial and aquatic ecosystems. Moreover, we found that the differences in virus proportions at the family level between soil and water were also statistically significant (Figure 2B), as the groups of Microviridae and Siphoviridae are predominantly present in the soil environment while being less abundant in the aquatic habitats. Microviridae and Siphoviridae are typical DNA bacteriophages and mainly infect enterobacteria, intracellular parasitic bacteria, and spiroplasma,[62,63] so the increased relative abundance of bacteria may be the cause for higher relative abundance of these viruses in the soil. Conversely, we showed that Myoviridae and Phycodnaviridae exhibited higher relative abundance in water environments ($p < 0.05$). The contractile tails and complex tail structures of Myoviridae may facilitate their movement and survival within environments of dynamic water flows.[64] Additionally, the enrichment of eukaryotic algae, which could be potentially infected by Phycoviridae, may be the reason for their relatively abundant distributions in aquatic ecosystems.[65]

The majority of metagenomic functions were found in both terrestrial and aquatic ecosystems, but their relative abundances varied across different biomes (Figure 2C), especially for viruses (Figure 2D) that are greatly dependent on habitat conditions.[66] Although some functions appeared to be stably shared among Archaea, Bacteria, and Eukaryota, certain genes related to AAD (amino acids and derivatives), MEM (membrane transport), and PRO (protein metabolism) were hyper-variable.[67] For example, genes of MEM were associated with the microbial capability to import or export multiple compounds, facilitating active uptakes of available nutrition from the soil,[68] which may favor their survival in complex terrestrial ecosystems. Meanwhile, dissolved C, N, and P enriched in aquatic ecosystems may support the metabolisms of amino acids and protein for microorganisms,[69] as a study spanning the global ocean microbiomes reported a relatively higher abundance of these metabolic processes as well.[70] Additionally, in accordance with the previous study,[71] we found that specific genes, such as MOT (motility and chemotaxis) in soil and PPT (phages, prophages, transposable elements, and plasmids) in water, were found to be enriched divergently (Figure 2A), in particular, with a high percentage of MOT genes in the soil despite viruses being traditionally considered non-motile entities. In contrast, higher relative abundances of PPT in the water may indicate their potential for adaptation to aquatic environments.[72]

Distinct taxonomic and functional compositions involved in biogeochemical cycling between soil and water

Microbial communities play a crucial role in regulating global biogeochemical cycles depending on habitat types.[36] Specifically, the abundance of CAR (carbohydrates), SUL (sulfur metabolism), and PHO (phosphorus metabolism) genes were enriched in soils, while functions of NIT (nitrogen metabolism) and IRO (iron acquisition and metabolism) were relatively dominant in aquatic environments (Figure 5D). Within the CAR function, sub-level functional classifications indicated consistently less related genes in aquatic metagenomes compared to terrestrial ones, which was consistent with our previous work in the intertidal zone of a sea island.[73] Possible reasons might be that C turnover is lower under high salinity,[74] and that anaerobic microbes metabolize C slower than aerobic microbes.[75] The abundance of SUL-related genes in the soil was higher than water because microbial anabolism primarily relies on the assimilation of inorganic sulfate for sulfur (S) acquisition,[76] which was confirmed by the dominance of inorganic S assimilation observed in soil environments based on our analyses. Besides, the soil microbiome exhibited a higher relative abundance of alkanesulfonate assimilation and utilization, potentially benefiting from sulfonate enrichment in terrestrial ecosystems.[77] By contrast, dimethylsulfoniopropionate (DMSP) is a key C and S resource for marine microbial growths,[78] which is synthesized from methionine in algae and bacteria,[79] leading to the prevalence of DMSP breakdown genes in water environments. Likewise, the genes related to POT function, most classified into potassium homeostasis involved in membrane transport at sub-level classification[80] greatly dependent on environmental osmolarity,[81] were found to be more abundant in the soil.[82] However, for the genes involved in NIT, those classified into ammonia assimilation emerge as the predominant N pathway,[36] particularly in the water. In contrast, the presence of genes related to allantoin utilization, nitrate and nitrite ammonification, and nitric oxide synthase was elevated in soil environments. These phenomena were similar to the trend in a study of arid soils[83] showing that habitat conditions are deemed essential environmental variables for the interactive effects between C and N cycling.[36] Fe(II) and Fe(III), which interchange under varying redox conditions prevalent in soil and water environments.[84] Iron's bioavailability drastically varies between soil and aquatic environments, largely dictated by pH levels, the presence of chelating agents, and redox conditions.[85] In soil, iron often exists in forms that are less bioavailable than in aquatic environments, leading to distinct microbial strategies to acquire and utilize iron, influencing community composition.[86,87] Vibrio

species have evolved a wide array of Fe transport systems including the secretion and uptake of high-affinity Fe-binding compounds (siderophores), as well as transport systems for Fe bound to host complexes.[88] These mechanisms enable the microbes to compete for this essential element in the freshwater, estuarine, and marine systems.[89]

Microbial functional groups can also serve as a partial indicator of disparity between soil and water, thereby complementing functional abundance. Our results revealed a distinct taxonomic composition of microbes involved in biogeochemical cycles between soil and water biomes (Figures 5B and 5C). Specifically, Acidobacteria, Actinobacteria, Chloroflexi, Deltaproteobacteria, and Firmicutes had higher relative abundances in the soil compared to water, while Bacteroidetes, Cyanobacteria, and Gammaproteobacteria displayed the opposite trends. These taxonomic groups play significant yet different roles in biogeochemical cycles. For example, Acidobacteria demonstrate the capability to metabolize recalcitrant C substrates,[90] actively participating in diverse carbohydrate breakdown, utilization, and biosynthesis through carbohydrate-active enzymes.[86,91] Besides, they possess a comprehensive repertoire of genes for catalyzing the metabolism of both inorganic and organic N sources.[87] Metagenomic analyses also indicate the potential for Acidobacteria to release siderophores to scavenge Fe from soil minerals by the formation of Fe(III) complexes.[48] Actinobacteria are renowned for their proficiency in the degradation of plant residues,[92] contributing to the global C cycle through plant biomass breakdowns.[93] Additionally, Actinobacteria have been found to be proficient in phosphate[94] and potassium solubilization,[95] as well as siderophore production.[96] However, Bacteroidetes are thought to exhibit a specialization in the degradation of algae-derived ocean polysaccharides[97] with a pivotal role in the mineralization of complex organic substrates such as polysaccharides and proteins.[98] Cyanobacteria could fundamentally regulate the cycling of C, N, S, and Fe through their involvements in primary production and oxygen generation,[99,100] such as those carrying the genes of assimilatory nitrite to ammonium and N fixation pathways.[36] Therefore, our results further showed the differences (Figures 5A and 6A) and similarities (Figure 7A) of the taxonomic composition of microbes associated with C and nutrient cycling between soil and water metagenomes, which was also dependent on the type of biogeochemical cycling.

Similar to this study, Pearson's correlation analysis revealed positive associations between the similarities of function and taxonomy engaged in C and nutrient cycling in both soil[29] and water metagenomes.[101] Despite the heterogeneity and discontinuity of the soil compared to water in aquatic ecosystems,[102] we still revealed that soil metagenomes shared more functional compositions than water ones (Figure 3B), suggesting a potentially higher level of functional redundancy among soil than water microbes. Furthermore, the degrees of functional redundancy, indicated by the relationship between pairwise similarities of metagenomic taxonomy and function, were observed to depend on biogeochemical cycles as well (Figure 4). For example, for the taxonomic composition of microbes associated with CAR, NIT, and PHO in the soil and water environments, there were lower correlation coefficients than other biogeochemical cycles, such as IRO and POT, suggesting a divergence of potential functional redundancy between certain functions. As C is essential for the formation of organic molecules in almost all life forms,[103] the microbial groups associated with CAR can exhibit great taxonomic diversity depending on the forms and sources of C present in various environments.[104] While Fe is abundant in the environment, its bioavailable form (such as Fe(II)) is relatively limited,[105] which constrains the diversity of microbial groups involved in the metabolisms of Fe,[63] such as specialized capabilities for Fe utilization and transformation.[106,107]

Microbial co-occurrence network analysis provides a unique and robust tool for understanding the interactions within microbial communities across different environments. The larger and more complex nature of the soil microbial networks suggests that these communities have to contend with a more diverse array of nutritional sources and ecological niches.[108] In the soil environment, certain functions such as CAR exhibit high modularity, indicating that some microbial communities may have established tight correlations for the collective utilization of resources. These modular networks could prove beneficial for sustaining ecosystem functions and offer better resilience to environmental change.[109] However, compared to aquatic metagenomes, soil microbial networks demonstrate longer average path distances in non-carbon-related functions such as NIT, PHO, and SUL cycling, potentially suggesting more complex pathways for nutrient transformations and functional complementarity among microorganisms.[110] Moreover, the relatively simplified network structures of aquatic microbial communities may reflect moderate nutrient levels and less environmental heterogeneity, allowing for stable and less complex community structures.[111]

Limitations of the study

The extraction efficiency of DNA from environmental samples might affect the comparability between terrestrial and aquatic metagenomes. But all the studies followed standardized procedures for both DNA extraction and sequencing analysis, so we redeemed the observed variations in results were thus likely to be attributable to the differences in environmental sampling rather than to the DNA extraction methods themselves in this research. Future work should investigate the contribution of DNA extraction efficiency to the differences in their microbial compositions.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact: Huaihai Chen (chenhh68@mail.sysu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data reported in this article will be shared by the lead contact upon request. This article does not report original code. Any additional information required to reanalyze the data reported in this article is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was financially funded by the National Natural Science Foundation of China (42277282, 42307476), Basic and Applied Basic Research Foundation of Guangdong Province (2022A1515010861, 2023A1515011540, 2022A1515110368), and Shenzhen Science and Technology Program (JCYJ20220530150201003).

AUTHOR CONTRIBUTIONS

Qi Fu: Writing – original draft, methodology, and investigation. Kayan Ma: Writing – review and editing, and validation. Jiayi Zhao: Methodology, formal analysis, and data curation. Jiaxin Li: Methodology, formal analysis, and data curation. Xueying Wang: Visualization and validation. Meiqi Zhao: Visualization and validation. Xian-Heng Fu: Writing – review and editing and validation. Dandan Huang: critically assessed and interpreted the findings, and writing – review and editing. Huaihai Chen: Conceived the study, secured the research funding, critically assessed and interpreted the findings, and writing – review and editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data collection
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111047.

Received: May 30, 2024
Revised: August 28, 2024
Accepted: September 23, 2024
Published: September 26, 2024

REFERENCES

1. Zhang, K., Shi, Y., Cui, X., Yue, P., Li, K., Liu, X., Tripathi, B.M., and Chu, H. (2019). Salinity Is a Key Determinant for Soil Microbial Communities in a Desert Ecosystem. mSystems *4*, 10–1128.
2. Lladó, S., Lopez-Mondejar, R., and Baldrian, P. (2017). Forest Soil Bacteria: Diversity, Involvement in Ecosystem Processes, and Response to Global Change. Microbiol. Mol. Biol. Rev. *81*.
3. Zak, D.R., and Kling, G.W. (2006). Microbial community composition and function across arctic tundar landscape. Ecology *87*, 1659–1670.
4. Liu, L., Yang, J., Yu, Z., and Wilkinson, D.M. (2015). The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. ISME J. *9*, 2068–2077.
5. Amend, A., Burgaud, G., Cunliffe, M., Edgcomb, V.P., Ettinger, C.L., Gutiérrez, M.H., Heitman, J., Hom, E.F.Y., Ianiri, G., Jones, A.C., et al. (2019). Fungi in the Marine Environment: Open Questions and Unsolved Problems. mBio *10*, e01189-18.
6. Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H., and Caporaso, J.G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc. Natl. Acad. Sci. USA *109*, 21390–21395.
7. Tisthammer, K.H., Cobian, G.M., and Amend, A.S. (2016). Global biogeography of marine fungi is shaped by the environment. Fungal Ecol. *19*, 39–46.
8. Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., and Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. Science *359*, 320–325.
9. Zhang, Y., Ji, P., Wang, J., and Zhao, F. (2016). RiboFR-Seq: a novel approach to linking 16S rRNA amplicon profiles to metagenomes. Nucleic Acids Res. *44*, e99.
10. Zheng, Q., Hu, Y., Zhang, S., Noll, L., Böckle, T., Dietrich, M., Herbold, C.W., Eichorst, S.A., Woebken, D., Richter, A., and Wanek, W. (2019). Soil multifunctionality is affected by the soil environment and by microbial community composition and diversity. Soil Biol. Biochem. *136*, 107521.
11. Tecon, R., and Or, D. (2017). Cooperation in carbon source degradation shapes spatial self-organization of microbial consortia on hydrated surfaces. Sci. Rep. *7*, 43726.
12. Singer, D., Seppey, C.V.W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., Blandenier, Q., Debroas, D., de Groot, G.A., de Vargas, C., et al. (2021). Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. Environ. Int. *146*, 106262.
13. Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S., et al. (2017). Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nat. Ecol. Evol. *1*, 0091.
14. Rout, A.K., Dixit, S., Tripathy, P.S., Rout, S.S., Parida, S.N., Parida, P.K., Sarkar, D.J., Kumar Das, B., Singh, A.K., and Behera, B.K. (2024). Metagenomic landscape of sediments of river Ganga reveals microbial diversity, potential plastic and xenobiotic degradation enzymes. J. Hazard Mater. *471*, 134377.
15. Grossart, H.-P., Van den Wyngaert, S., Kagami, M., Wurzbacher, C., Cunliffe, M., and Rojas-Jimenez, K. (2019). Fungi in aquatic ecosystems. Nat. Rev. Microbiol. *17*, 339–354.
16. Nunan, N. (2017). The microbial habitat in soil: Scale, heterogeneity and functional consequences. J. Plant Nutr. Soil Sci. *180*, 425–429.
17. Behera, B.K., Chakraborty, H.J., Patra, B., Rout, A.K., Dehury, B., Das, B.K., Sarkar, D.J., Parida, P.K., Raman, R.K., Rao, A.R., et al. (2020). Metagenomic Analysis Reveals Bacterial and Fungal Diversity and Their Bioremediation Potential From Sediments

of River Ganga and Yamuna in India. Front. Microbiol. *11*, 556136.
18. Boenigk, J., Wodniok, S., Bock, C., Beisser, D., Hempel, C., Grossmann, L., Lange, A., and Jensen, M. (2018). Geographic distance and mountain ranges structure freshwater protist communities on a European scale. Metabarcoding Metagenomics *2*, e21519.
19. Debroas, D., Domaizon, I., Humbert, J.-F., Jardillier, L., Lepère, C., Oudart, A., and Taïb, N. (2017). Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. FEMS Microbiol. Ecol. *93*, fix023.
20. Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., et al. (2018). Structure and function of the global topsoil microbiome. Nature *560*, 233–237.
21. Cotrufo, M.F., Wallenstein, M.D., Boot, C.M., Denef, K., and Paul, E. (2013). The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: Do labile plant inputs form stable soil organic matter? Global Change Biol. *19*, 988–995.
22. Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J., Encinar, D., Berdugo, M., Campbell, C.D., and Singh, B.K. (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. Nat. Commun. *7*, 10541.
23. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., et al. (2015). Structure and function of the global ocean microbiome. Science *348*, 1261359.
24. Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., Welch, D.B.M., Martiny, J.B.H., Sogin, M., Boetius, A., and Ramette, A. (2011). Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. PLoS One *6*, e24570.
25. Isobe, K., Ise, Y., Kato, H., Oda, T., Vincenot, C.E., Koba, K., Tateno, R., Senoo, K., and Ohte, N. (2020). Consequences of microbial diversity in forest nitrogen cycling: diverse ammonifiers and specialized ammonia oxidizers. ISME J. *14*, 12–25.
26. Tardy, V., Mathieu, O., Lévêque, J., Terrat, S., Chabbi, A., Lemanceau, P., Ranjard, L., and Maron, P.A. (2014). Stability of soil microbial structure and activity depends on microbial diversity. Environ. Microbiol. Rep. *6*, 173–183.
27. Singh, B.K., Quince, C., Macdonald, C.A., Khachane, A., Thomas, N., Al-Soud, W.A., Sørensen, S.J., He, Z., White, D., Sinclair, A., et al. (2014). Loss of microbial diversity in soils is coincident with reductions in some specialized functions. Environ. Microbiol. *16*, 2408–2420.
28. Hester, E.R., Jetten, M.S.M., Welte, C.U., and Lücker, S. (2019). Metabolic Overlap in Environmentally Diverse Microbial Communities. Front. Genet. *10*, 989.
29. Chen, H., Ma, K., Lu, C., Fu, Q., Qiu, Y., Zhao, J., Huang, Y., Yang, Y., Schadt, C.W., and Chen, H. (2022). Functional redundancy in soil microbial community based on metagenomics across the globe. Front. Microbiol. *13*, 878978.
30. Louca, S., Polz, M.F., Mazel, F., Albright, M.B.N., Huber, J.A., O'Connor, M.I., Ackermann, M., Hahn, A.S., Srivastava, D.S., Crowe, S.A., et al. (2018). Function and functional redundancy in microbial systems. Nat. Ecol. Evol. *2*, 936–943.
31. Chen, H., Ma, K., Huang, Y., Fu, Q., Qiu, Y., Lin, J., Schadt, C.W., and Chen, H. (2022). Lower functional redundancy in "narrow" than "broad" functions in global soil metagenomics. Soil *8*, 297–308.
32. Fetzer, I., Johst, K., Schäwe, R., Banitz, T., Harms, H., and Chatzinotas, A. (2015). The extent of functional redundancy changes as species' roles shift in different environments. Proc. Natl. Acad. Sci. USA *112*, 14888–14893.
33. Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. *35*, 833–844.
34. Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S.G., and Alvarez-Cohen, L. (2015). High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. mBio *6*, e02288-14.
35. Martinez-Hernandez, F., Fornas, O., Lluesma Gomez, M., Bolduc, B., de la Cruz Peña, M.J., Martínez, J.M., Anton, J., Gasol, J.M., Rosselli, R., Rodriguez-Valera, F., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat. Commun. *8*, 15892.
36. Nelson, M.B., Martiny, A.C., and Martiny, J.B.H. (2016). Global biogeography of microbial nitrogen-cycling traits in soil. Proc. Natl. Acad. Sci. USA *113*, 8033–8040.
37. Keegan, K.P., Glass, E.M., and Meyer, F. (2016). MG-RAST, a metagenomics service for analysis of microbial community structure and function. Methods Mol. Biol. *1399*, 207–233.
38. Taş, N., de Jong, A.E., Li, Y., Trubl, G., Xue, Y., and Dove, N.C. (2021). Metagenomic tools in microbial ecology research. Curr. Opin. Biotechnol. *67*, 184–191.
39. Kuzyakov, Y., and Blagodatskaya, E. (2015). Microbial hotspots and hot moments in soil: concept & review. Soil Biol. Biochem. *83*, 184–199.
40. Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. Nat. Rev. Microbiol. *15*, 579–590.
41. Delgado-Baquerizo, M., Reich, P.B., Khachane, A.N., Campbell, C.D., Thomas, N., Freitag, T.E., Abu Al-Soud, W., Sørensen, S., Bardgett, R.D., and Singh, B.K. (2017). It is elemental: soil nutrient stoichiometry drives bacterial diversity. Environ. Microbiol. *19*, 1176–1188.
42. Flemming, H.-C., and Wuertz, S. (2019). Bacteria and archaea on Earth and their abundance in biofilms. Nat. Rev. Microbiol. *17*, 247–260.
43. Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., and Kyrpides, N.C. (2016). Uncovering Earth's virome. Nature *536*, 425–430.
44. Welsh, R.M., Zaneveld, J.R., Rosales, S.M., Payet, J.P., Burkepile, D.E., and Thurber, R.V. (2016). Bacterial predation in a marine host-associated microbiome. ISME J. *10*, 1540–1544.
45. Huang, Y., Shi, W., Fu, Q., Qiu, Y., Zhao, J., Li, J., Lyu, Q., Yang, X., Xiong, J., Wang, W., et al. (2023). Soil development following glacier retreat shapes metagenomic and metabolomic functioning associated with asynchronous C and N accumulation. Sci. Total Environ. *892*, 164405.
46. Kersters, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P., and Stackebrandt, E. (2006). Introduction to the Proteobacteria. In The prokaryotes: a handbook on the biology of bacteria, 5The prokaryotes: a handbook on the biology of bacteria (Springer), pp. 3–37.
47. Polymenakou, P.N., Bertilsson, S., Tselepides, A., and Stephanou, E.G. (2005). Bacterial community composition in different sediments from the Eastern Mediterranean Sea: a comparison of four 16S ribosomal DNA clone libraries. Microb. Ecol. *50*, 447–462.
48. Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., van Veen, J.A., and Kuramae, E.E. (2016). The Ecology of Acidobacteria: Moving beyond Genes and Genomes. Front. Microbiol. *7*, 744.
49. Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S.C., Treusch, A.H., Eck, J., and Schleper, C. (2003). Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. Mol. Microbiol. *50*, 563–575.
50. Girão, M., Ribeiro, I., and Carvalho, M.F. (2022). Actinobacteria from Marine Environments: A Unique Source of Natural Products. In Natural Products from Actinomycetes: Diversity, Ecology and Drug Discovery (Springer), pp. 1–45.
51. López-Pérez, M., Haro-Moreno, J.M., Iranzo, J., and Rodriguez-Valera, F. (2020). Genomes of the "Candidatus Actinomarinales" order: highly streamlined marine epipelagic actinobacteria. mSystems 5, 10–1128. https://doi.org/10.1128/msystems. 01041-20.
52. Selim, M.S.M., Abdelhamid, S.A., and Mohamed, S.S. (2021). Secondary metabolites and biodiversity of actinomycetes. J. Genet. Eng. Biotechnol. *19*, 72.
53. Hazarika, S.N., and Thakur, D. (2020). Actinobacteria. In Beneficial microbes in agro-ecology (Elsevier), pp. 443–476.
54. Ballesté, E., and Blanch, A.R. (2010). Persistence of Bacteroides species populations in a river as measured by molecular and culture techniques. Appl. Environ. Microbiol. *76*, 7608–7616.
55. Charpy, L., Casareto, B.E., Langlade, M.J., and Suzuki, Y. (2012). Cyanobacteria in Coral Reef Ecosystems: A Review. J. Mar. Biol. *2012*, 259571.
56. Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., and Smith, M.E. (2014). Global diversity and geography of soil fungi. Sci. Technol. Humanit. *346*, 1256688.
57. Egidi, E., Delgado-Baquerizo, M., Plett, J.M., Wang, J., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., and Singh, B.K. (2019). A few Ascomycota taxa dominate soil fungal communities worldwide. Nat. Commun. *10*, 2369.
58. Lehmann, J., and Kleber, M. (2015). The contentious nature of soil organic matter. Nature *528*, 60–68.
59. Boer, W., Folman, L.B., Summerbell, R.C., and Boddy, L. (2005). Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiol. Rev. *29*, 795–811.
60. Jia, T., Wang, R., Fan, X., and Chai, B. (2018). A comparative study of fungal community

structure, diversity and richness between the soil and the phyllosphere of native grass species in a copper tailings dam in Shanxi Province, China. Appl. Sci. *8*, 1297.
61. Yang, Z., Ma, X., Wang, Q., Tian, X., Sun, J., Zhang, Z., Xiao, S., De Clerck, O., Leliaert, F., and Zhong, B. (2023). Phylotranscriptomics unveil a Paleoproterozoic-Mesoproterozoic origin and deep relationships of the Viridiplantae. Nat. Commun. *14*, 5542.
62. Brentlinger, K.L., Hafenstein, S., Novak, C.R., Fane, B.A., Borgon, R., McKenna, R., and Agbandje-McKenna, M. (2002). Microviridae, a family divided: isolation, characterization, and genome sequence of ϕMH2K, a bacteriophage of the obligate intracellular parasitic bacterium Bdellovibrio bacteriovorus. J. Bacteriol. *184*, 1089–1094.
63. Prigent, M., Leroy, M., Confalonieri, F., Dutertre, M., and DuBow, M.S. (2005). A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. Extremophiles *9*, 289–296.
64. Tamakoshi, M., Murakami, A., Sugisawa, M., Tsuneizumi, K., Takeda, S., Saheki, T., Izumi, T., Akiba, T., Mitsuoka, K., Toh, H., et al. (2011). Genomic and proteomic characterization of the large Myoviridae bacteriophage ϕTMA of the extreme thermophile Thermus thermophilus. Bacteriophage *1*, 152–164.
65. Van Etten, J.L., Graves, M.V., Müller, D.G., Boland, W., and Delaroque, N. (2002). Phycodnaviridae–large DNA algal viruses. Arch. Virol. *147*, 1479–1516.
66. Louca, S., Parfrey, L.W., and Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. Science *353*, 1272–1277.
67. Chen, H., Ma, K., Huang, Y., Yao, Z., and Chu, C. (2021). Stable soil microbial functional structure responding to biodiversity loss based on metagenomic evidences. Front. Microbiol. *12*, 716764.
68. Trivedi, P., Anderson, I.C., and Singh, B.K. (2013). Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. Trends Microbiol. *21*, 641–651.
69. Li, Z., Wang, Y., Li, J., Liu, F., He, L., He, Y., and Wang, S. (2016). Metagenomic analysis of genes encoding nutrient cycling pathways in the microbiota of deep-sea and shallow-water sponges. Mar. Biotechnol. *18*, 659–671.
70. Haggerty, J.M., and Dinsdale, E.A. (2017). Distinct biogeographical patterns of marine bacterial taxonomy and functional genes. Global Ecol. Biogeogr. *26*, 177–190.
71. Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., et al. (2008). Functional metagenomic profiling of nine biomes. Nature *452*, 629–632.
72. Cotta, S.R., Cadete, L.L., Van Elsas, J.D., Andreote, F.D., and Dias, A.C.F. (2019). Exploring bacterial functionality in mangrove sediments and its capability to overcome anthropogenic activity. Mar. Pollut. Bull. *141*, 586–594.
73. Zhang, J., Fu, Q., Huang, Y., Fan, Y., Liang, M., Chen, H., and Yu, S. (2022). Negative impacts of sea-level rise on soil microbial involvement in carbon metabolism. Sci. Total Environ. *838*, 156087.
74. Setia, R., Gottschalk, P., Smith, P., Marschner, P., Baldock, J., Setia, D., and Smith, J. (2013). Soil salinity decreases global soil organic carbon stocks. Sci. Total Environ. *465*, 267–272.
75. McDonough, L.K., Santos, I.R., Andersen, M.S., O'Carroll, D.M., Rutlidge, H., Meredith, K., Oudone, P., Bridgeman, J., Gooddy, D.C., Sorensen, J.P.R., et al. (2020). Changes in global groundwater organic carbon driven by climate change and urbanization. Nat. Commun. *11*, 1279.
76. Zuo, X., Sun, S., Wang, S., Yue, P., Hu, Y., Zhao, S., Guo, X., Li, X., Chen, M., Ma, X., et al. (2023). Contrasting relationships between plant-soil microbial diversity are driven by geographic and experimental precipitation changes. Sci. Total Environ. *861*, 160654.
77. Wu, B., Liu, F., Fang, W., Yang, T., Chen, G.-H., He, Z., and Wang, S. (2021). Microbial sulfur metabolism and environmental implications. Sci. Total Environ. *778*, 146085.
78. Raina, J.-B., Dinsdale, E.A., Willis, B.L., and Bourne, D.G. (2010). Do the organic sulfur compounds DMSP and DMS drive coral microbial associations? Trends Microbiol. *18*, 101–108.
79. McNeil, S.D., Nuccio, M.L., and Hanson, A.D. (1999). Betaines and related osmoprotectants. Targets for metabolic engineering of stress resistance. Plant Physiol. *120*, 945–950.
80. Nakai, R., Abe, T., Takeyama, H., and Naganuma, T. (2011). Metagenomic analysis of 0.2-μm-passable microorganisms in deep-sea hydrothermal fluid. Mar. Biotechnol. *13*, 900–908.
81. Elabed, H., González-Tortuero, E., Ibacache-Quiroga, C., Bakhrouf, A., Johnston, P., Gaddour, K., Blázquez, J., and Rodríguez-Rojas, A. (2019). Seawater salt-trapped Pseudomonas aeruginosa survives for years and gets primed for salinity tolerance. BMC Microbiol. *19*, 142.
82. Csonka, L.N. (1989). Physiological and genetic responses of bacteria to osmotic stress. Microbiol. Rev. *53*, 121–147.
83. Lacerda-Júnior, G.V., Noronha, M.F., Cabral, L., Delforno, T.P., de Sousa, S.T.P., Fernandes-Júnior, P.I., Melo, I.S., and Oliveira, V.M. (2019). Land Use and Seasonal Effects on the Soil Microbiome of a Brazilian Dry Forest. Front. Microbiol. *10*, 648.
84. Schroth, A.W., Crusius, J., Sholkovitz, E.R., and Bostick, B.C. (2009). Iron solubility driven by speciation in dust sources to the ocean. Nat. Geosci. *2*, 337–340.
85. Wandersman, C., and Delepelaire, P. (2004). Bacterial iron sources: from siderophores to hemophores. Annu. Rev. Microbiol. *58*, 611–647.
86. Kalam, S., Basu, A., Ahmad, I., Sayyed, R.Z., El-Enshasy, H.A., Dailin, D.J., and Suriani, N.L. (2020). Recent Understanding of Soil Acidobacteria and Their Ecological Significance: A Critical Review. Front. Microbiol. *11*, 580024.
87. Eichorst, S.A., Trojan, D., Roux, S., Herbold, C., Rattei, T., and Woebken, D. (2018). Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. Environ. Microbiol. *20*, 1041–1063.
88. Payne, S.M., Mey, A.R., and Wyckoff, E.E. (2016). Vibrio iron transport: evolutionary adaptation to life in multiple environments. Microbiol. Mol. Biol. Rev. *80*, 69–90.
89. Johnson, C.N. (2013). Fitness Factors in Vibrios: a Mini-review. Microb. Ecol. *65*, 826–851.
90. Castro, H.F., Classen, A.T., Austin, E.E., Norby, R.J., and Schadt, C.W. (2010). Soil microbial community responses to multiple experimental climate change drivers. Appl. Environ. Microbiol. *76*, 999–1007.
91. Männistö, M.K., Rawat, S., Starovoytov, V., and Häggblom, M.M. (2012). Granulicella arctica sp. nov., Granulicella mallensis sp. nov., Granulicella tundricola sp. nov. and Granulicella sapmiensis sp. nov., novel acidobacteria from tundra soil. Int. J. Syst. Evol. Microbiol. *62*, 2097–2106.
92. Bao, Y., Dolfing, J., Guo, Z., Chen, R., Wu, M., Li, Z., Lin, X., and Feng, Y. (2021). Important ecophysiological roles of non-dominant Actinobacteria in plant residue decomposition, especially in less fertile soils. Microbiome *9*, 84.
93. Lewin, G.R., Carlos, C., Chevrette, M.G., Horn, H.A., McDonald, B.R., Stankey, R.J., Fox, B.G., and Currie, C.R. (2016). Evolution and Ecology of *Actinobacteria* and Their Bioenergy Applications. In Annual Review of Microbiology, *70*, S. Gottesman, ed., pp. 235–254.
94. Purushotham, N., Jones, E., Monk, J., and Ridgway, H. (2018). Community structure of endophytic actinobacteria in a New Zealand native medicinal plant Pseudowintera colorata (Horopito) and their influence on plant growth. Microb. Ecol. *76*, 729–740.
95. Boubekri, K., Soumare, A., Mardad, I., Lyamlouli, K., Ouhdouch, Y., Hafidi, M., and Kouisni, L. (2022). Multifunctional role of Actinobacteria in agricultural production sustainability: A review. Microbiol. Res. *261*, 127059.
96. Wang, W., Qiu, Z., Tan, H., and Cao, L. (2014). Siderophore production by actinobacteria. Biometals *27*, 623–631.
97. Zheng, R., Cai, R., Liu, R., Liu, G., and Sun, C. (2021). *Maribellus comscasis* sp. nov., a novel deep-sea *Bacteroidetes* bacterium, possessing a prominent capability of degrading cellulose. Environ. Microbiol. *23*, 4561–4575.
98. Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M., Teeling, H., Amann, R.I., and Schweder, T. (2014). Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* '*Gramella forsetii*' KT0803. ISME J. *8*, 1492–1502.
99. Schad, M., Konhauser, K.O., Sánchez-Baracaldo, P., Kappler, A., and Bryce, C. (2019). How did the evolution of oxygenic photosynthesis influence the temporal and spatial development of the microbial iron cycle on ancient Earth? Free Radic. Biol. Med. *140*, 154–166.
100. Sánchez-Baracaldo, P., Bianchini, G., Wilson, J.D., and Knoll, A.H. (2022). Cyanobacteria and biogeochemical cycles through Earth history. Trends Microbiol. *30*, 143–157.
101. Chen, H., Ma, K., Huang, Y., Yang, Y., Ma, Z., and Chu, C. (2021). Salinity drives functional and taxonomic diversities in global water metagenomes. Front. Microbiol. *12*, 719725.
102. Vos, M., Wolf, A.B., Jennings, S.J., and Kowalchuk, G.A. (2013). Micro-scale determinants of bacterial diversity in soil. FEMS Microbiol. Rev. *37*, 936–954.
103. Olah, G.A., Prakash, G.K.S., and Goeppert, A. (2011). Anthropogenic chemical carbon cycle for a sustainable future. J. Am. Chem. Soc. *133*, 12881–12898.
104. Garland, J.L., and Mills, A.L. (1991). Classification and characterization of

heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. Appl. Environ. Microbiol. *57*, 2351–2359.

105. Pii, Y., Borruso, L., Brusetti, L., Crecchio, C., Cesco, S., and Mimmo, T. (2016). The interaction between iron nutrition, plant species and soil type shapes the rhizosphere microbiome. Plant Physiol. Biochem. *99*, 39–48.
106. Li, Y., and Ma, Q. (2017). Iron acquisition strategies of Vibrio anguillarum. Front. Cell. Infect. Microbiol. *7*, 342.
107. Ellermann, M., and Arthur, J.C. (2017). Siderophore-mediated iron acquisition and modulation of host-bacterial interactions. Free Radic. Biol. Med. *105*, 68–78.
108. DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. *72*, 5069–5072.
109. Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. *35*, 7188–7196.
110. Clarke, K., and Gorley, R. (2015). Getting started with PRIMER v7. PRIMER-E: Plymouth, Plymouth Marine Lab. *20*.
111. Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., and Wishart, D.S. (2016). Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. *44*, W147–W153.
112. Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. *42*, D553–D559.
113. Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. *42*, D206–D214.
114. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. *44*, D457–D462.
115. Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. *43*, D261–D269.
116. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. *44*, D286–D293.
117. Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., and Tiedje, J.M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. *37*, D141–D145.
118. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.
119. Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinf. *16*, 169.
120. Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. BMC Bioinf. *13*, 113.
121. Zhou, J., Deng, Y., Luo, F., He, Z., and Yang, Y. (2011). Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. mBio *2*, e00122-11.
122. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Soil and water metagenomes	MG-RAST	https://www.mg-rast.org/
Software and algorithms		
PRIMER 7	Statistical software for ecological data, or any other data	https://www.primer-e.com/our-software/primer-version-7/
R	Statistical software for data science	https://www.r-project.org

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Our study does not include experimental models or study participants typical in the life sciences.

METHOD DETAILS

Data collection

To build up our database, we obtained the available shotgun metagenomic data from the MG-RAST server. We chose the metagenomes that have been released from peer-reviewed publications to ensure data quality. We used the Web of Science database for our literature survey of peer-reviewed publications in 2019, and applied search terms, such as "soil metagenome", "water metagenome", "shotgun sequencing", and "MG-RAST", etc., with the following criteria[1]: environmental DNA of soil or water was sequenced using shotgun sequencing without pre-amplification of target genes[2]; studies directly deposited the soil or water metagenomic sequences in the MG-RAST database[3]; soil or water metagenome data can be publicly accessible in the MG-RAST database. Altogether, we included 933 soil metagenomes and 938 water metagenomes sampled globally from 55 and 55 peer-reviewed papers, respectively (Figure 1; Table S1). We further extracted detailed information of each metagenome from MG-RAST metadata and publications, such as Study ID, MG-RAST ID, publication, bp (base pair), reads, latitude (LAT) and longitude (LONG). The DNA extraction methods employed across the referenced studies may vary, but it was important to note that all the studies followed standardized procedures for both DNA extraction and sequencing analysis. The observed variations in results were thus likely to be attributable to the differences in environmental sampling rather than to the DNA extraction methods themselves.

Based on the Study ID and/or MG-RAST ID information, we downloaded data of microbial taxonomic compositions and functional categories at each classification level. Specifically, databases of RefSeq[112] (domain, phylum, class, order, family, and genus levels) and SEED Subsystems[113] (1, 2, 3, and function levels) were loaded as taxonomic compositions and functional profiles, respectively. The analysis was performed on the MG-RAST server with default settings (maximum e-value cutoff was 1e−5, minimum identity cutoff was 60%, and minimum alignment length was 50). To investigate microbial compositions and their functions, in the "Analysis" tab of the MG-RAST server, we chose "SEED Subsystem" as the source, selected "function" as the level, and added the taxonomic domain that we wanted to investigate (Archaea, Bacteria, Eukaryota, Virus) in the "taxonomy filter" with "RefSeq" as the source and "domain" as the level. Besides, we downloaded the related taxonomic information (relative abundances at the levels of domain, phylum, class, order, family, and genus) to analyze the microbial diversity in the soil and water samples. Additionally, to explore microbial taxonomic compositions responsible for conducting carbohydrate and nutrient cycling functions, in the "Analysis" function of the MG-RAST server, we selected "RefSeq" as the source and 'genus' as the level, and added the functional category in the 'function filter' that we were interested in SEED Subsystems level 1, including CAR (carbohydrates), NIT (nitrogen metabolism), PHO (phosphorus metabolism), SUL (sulfur metabolism), IRO (iron acquisition and metabolism), POT (potassium metabolism). We detected CAR as the most abundant functional category relating to the carbohydrate metabolism in SEED Subsystems level 1. We chose these five nutrient functional categories in SEED Subsystems level 1 to represent typical nutrient cycles in biogeochemistry. We also downloaded the relative abundances of all functional categories in SEED Subsystems at 1, 2, 3, and function levels to assess functional profiles of carbohydrate and nutrient functions in soil and water metagenomes.

We chose SEED Subsystems for functional databases instead of KEGG Orthology (KO),[114] Clusters of Orthologous Groups (COG),[115] and Non-supervised Orthologous Groups (NOG)[116] because SEED Subsystems had a diverse functional classification at level 1 of interest, especially those concerning carbohydrate and nutrient functions. We selected RefSeq database instead of the traditional ribosomal RNA databases, such as RDP (Ribosomal Database Project),[117] Greengenes,[108] or Silva LSU/SSU[109] databases because taxonomic reads in the RefSeq database were comparable to functional ones in the SEED Subsystems database.

QUANTIFICATION AND STATISTICAL ANALYSIS

In order to remove bias in sequencing depths among different studies, we divided reads of each taxonomic or functional category by the total reads in the downloaded data to standardize the data to relative abundance. To compare the correlations of taxonomic and functional

compositions related to carbohydrate and nutrient functions in soil and water, the pairwise similarities of function and taxonomy were presented as boxplots. Based on the relative abundance of taxonomic compositions at the genus level and functional profiles at the function level, we constructed a pairwise Bray-Curtis similarity matrix in PRIMER 7 (Plymouth Routines in Multivariate Ecological Research Statistical Software, v7.0.13, PRIMER-E Ltd, UK).[110] Based on pairwise Bray-Curtis similarity matrix, we used non-metric multidimensional scaling (NMDS), PCoA (principal coordinates analysis), and one-factor PERMANOVA (Permutational multivariate analysis of variance) of the main test (pseudo-F statistics) and pairwise test (pseudo-t statistics) with *P* values and Sq. root reported to analyze the beta-diversity of taxonomic compositions and functional profiles for conducting carbohydrate and nutrient functions in soil and water metagenomes. Besides, an assessment of the associations between functional and taxonomic diversities was performed by calculating the Spearman's Rho correlation coefficient in PRIMER 7. We used HeatMapper[111] to construct heat maps showing the normalized relative abundance of dominant taxonomic compositions at phylum or class levels for microbiome conducting carbohydrate and nutrient functions in soil and water metagenomes with dendrograms of hierarchical cluster analysis to group taxonomy and function based on 'Average Linkage' as the clustering method and 'Pearson's Correlation' as the distance measurement method. Similarly, the dominant functions (level 3) of soil and water microbes were visualized in the same way to present the difference between the soil and water microbiome functional potential. Besides, we performed group significance analyses in QIIME[118] to compare the relative abundances of dominant taxonomic compositions at the phylum/class and genus levels for conducting carbohydrate and nutrient functions in soil and water metagenomes at the Bonferroni-corrected P values < 0.05 of a Kruskal–Wallis non–parametric ANOVA test for multiple comparisons. We used InteractiVenn[119] to construct Venn's diagrams to visualize the number of shared dominant genus between soil and water metagenomes or among different functions.

To examine the potential interactions of microbial taxa that conduct carbohydrate and nutrient functions in soil and water metagenomes across the globe, we used the Molecular Ecological Network Analyses Pipeline (http://ieg4.rccc.ou.edu/MENA/)[120,121] to perform co-occurrence network analysis. We first standardized the data to meet the requirements of the pipeline and uploaded the data to construct the network with default settings, including[1] only keeping the species present in more than half of all samples[2]; filling with 0.01 in blanks with paired valid values[3]; taking the logarithm with the recommended similarity matrix of Pearson's correlation coefficient[4]; calculating the order to decrease the cutoff from the top using Poisson regression distribution only. A default cutoff value (similarity threshold, S_t) for the similarity matrix was generated to assign a link between the pair of species. Then, the global network properties, the individual nodes' centrality, and the module separation and modularity calculations were run based on default settings using greedy modularity optimization. Network files were exported and visualized using Cytoscape software.[122]