



# Quantifying the functional genes of C, N, P, and S cycling in a deep lake: Depth patterns and drivers

Peixuan Zhang<sup>a,b,1</sup>, Minglei Ren<sup>a,c,1</sup>, Weizhen Zhang<sup>a</sup>, Yan Xu<sup>b</sup>, Jianjun Wang<sup>a,\*</sup>

<sup>a</sup> State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

<sup>b</sup> Department of Municipal Engineering, School of Civil Engineering, Southeast University, Nanjing 210096, China

<sup>c</sup> State Environmental Protection Key Laboratory of Aquatic Ecosystem Health in the Middle and Lower Reaches of Yangtze River, Jiangsu Provincial Academy of Environmental Science, Nanjing, China

## ARTICLE INFO

### Keywords:

Functional gene diversity  
Community composition  
Water depth  
Deep Lake

## ABSTRACT

Microbial functional composition is important for biogeochemical cycles, and is usually constrained by taxonomic species pool and natural environmental gradients. However, the distribution of functional genes in deep lakes and the driving factors remains elusive. Here, we quantified the abundance of 71 functional genes relevant to carbon degradation, carbon fixation, methane metabolism, nitrogen cycling, phosphorus cycling, and sulfur cycling in 38 sediments along the water depth ranging from 0 to 90 m in Lugu Lake, China, using Quantitative Microbial Ecology Chip (QMEC) and then explored their water-depth diversity pattern and abiotic and biotic drivers. Functional gene diversity showed a hump-shaped pattern along the depth and peaked at around 50 m which is the low boundary of thermocline layer. There were specific environmental preferences among functional gene subgroups such as nitrogen and sulfur cycling genes preferring to deeper and shallow waters, respectively. The dissimilarity of total functional genes increased with water depth distance indicating a distance-decay relationship. There was a congruence between functional and taxonomic composition by showing the positive correlations between the compositions of functional genes and bacteria or archaea. This phenomenon is consistently observed for the six functional gene subgroups, and the congruence strength was highest for nitrogen cycling while lowest for sulfur cycling. Compared to abiotic factors, biotic factors were more relevant to the functional gene diversity and composition. Biotic factors explained 28.5 % of the variance of functional gene diversity, while water depth and other environmental factors such as water total phosphorus and sediment carbon explained 4.2 % and 3.8 %, respectively. For functional gene composition, biotic factors accounted for 25.2 % of the variance, whereas water depth and other environmental factors contributed to 0.7 % and 4.4 %, respectively. Among all explanatory variables for function genes, bacterial composition had the highest contribution, which was further supported by showing its direct effects of 1.45 and 0.86 on functional diversity and composition, respectively. This study for the first time quantified the microbial functional genes along water depth in deep lakes and provided a more comprehensive understanding of the functional dynamics in microbial communities within aquatic ecosystems.

## 1. Introduction

The functional genes of microbial community have significant importance in aquatic ecosystems as they encode specific metabolic pathways, biological processes, or ecological functions (Madsen, 2011). Focusing on the diversity and composition of functional genes rather than just species level can provide more comprehensive insights into the microbial communities and functions of ecosystems (Fuhrman, 2009).

Understanding the biogeographical patterns of functional gene diversity and composition along natural environmental gradients such as altitude, latitude, and water depth contributes to a better understanding of the structure and functioning of aquatic ecosystems (Broman et al., 2022). Previous research had explored the functional diversity declines towards higher elevations and composition of microbial communities exhibits increasing turnover with larger elevational distances (Picazo et al., 2020). Additionally, resources, temperature, pH and geographical

\* Corresponding author.

E-mail address: [jjwang@niglas.ac.cn](mailto:jjwang@niglas.ac.cn) (J. Wang).

<sup>1</sup> Equally contribution to this work.

location independently affect lake microbial communities (Bastida et al., 2021; Garner et al., 2023; Liu et al., 2020; Schulhof et al., 2020). As a unique aquatic ecosystem, the lake exhibits a close correlation between its water depth gradient and various environmental characteristics, such as temperature, nutrient concentration, and electrical conductivity. Hence, water depth in deep lake provides a strong environmental filter for aquatic microbial communities (Pearman et al., 2022; Wang et al., 2017). However, no previous studies have reported the distribution pattern of functional genes of microbial communities along natural gradients of water depth in deep lakes. This might be due to the considerable depth and harsh conditions of deep lakes, making significant challenges in obtaining samples. Furthermore, past research has been limited by the absence of techniques for high-throughput quantitative analysis of functional genes (Zheng et al., 2018; Zheng et al., 2017).

Water depth often closely covary with physiochemical conditions, and biological community diversity and composition (DeLong et al., 2006; Hewson et al., 2006). Depth gradients could result in various environmental conditions such as light, temperature, oxygen, and nutrients (Liu et al., 2019a). These water depth-related environmental factors might shape the functional gene composition through mechanisms such as environmental filtering (Peralta-Maraver et al., 2018). It should be noted that biotic factors also play a crucial role in shaping the distribution and composition of functional genes (Sadeghi et al., 2021). The species pool and community turnover have potential impacts on functional gene diversity and composition (Mark Vellend, 2005). A diverse species pool provides a broader resource of genes involved in various metabolic pathways and adaptive traits (He et al., 2018). Community variation rates typically change in response to environmental changing, due to the microbial ability to adapt to environment (Jing et al., 2022). The rate of turnover could affect the stability and dynamics of functional gene composition (Craven et al., 2018). However, there are lacks of observations evaluating the influence of both abiotic and biotic factors on the diversity and composition of functional genes, as well as the relative contributions of these factors.

Here, we analyzed microbial community and functional genes of 38 samples along the 0–90 m water depth gradient in Lugu Lake, Yunnan, China. Microbial community was profiled through 16S rRNA high-throughput amplicon, and functional composition represented by 71 cycling genes were quantified using the Quantitative Microbial Ecology Chip (QMEC) (Zheng et al., 2018). These functional genes are grouped by six subgroups including carbon degradation, carbon fixation, methane metabolism, nitrogen cycling, phosphorus cycling, and sulfur cycling. We explored water depth patterns of the six subgroups of functional genes and investigated the relative importance of abiotic and biotic factors in explaining their diversity and composition. Our study aimed to address the following three questions: (i) What are the water depth patterns of functional gene diversity? (ii) How are the relationships between the beta diversity of different functional genes and water depth? (iii) How much is the relative importance of water depth, other environmental factors, and biotic factors in predicting functional gene diversity? Our results provided the comprehensive understanding for the water depth pattern of microbial functional genes involved in biochemical cycling in deep lakes and highlight the distinct role of biotic and abiotic factors in shaping functional diversity. The findings had implications for studying functional gene coupling and microbial interactions within aquatic systems, providing a foundation for future research on microbial functional dynamics.

## 2. Materials and methods

### 2.1. Study area

The study was conducted in Lugu Lake (27°41′–27°45′N, 100°45′–100°50′E), located in Yunnan Province, China, on the southeastern margin of the Tibetan Plateau (Zhao et al., 2023). A total of 38

surface sediment samples were collected along the water depth ranging from 0 to 90 m to investigate the functional genes of bacteria in August 2010. Lugu Lake is one of the deepest plateau freshwater lakes in the region, characterized by maximum water depth of 93.5 m, a mean depth of approximately 40.3 m, a total water surface area of 50.5 km<sup>2</sup>, an elevation of 2685 m, and a catchment area of approximately 171.4 km<sup>2</sup> (Zhao et al., 2019). Unlike many other lakes in the area, Lugu Lake remains unfrozen throughout the year due to its warm temperate semi-enclosed nature (Su et al., 2022). In winter, the water temperature of the lake is uniformly distributed vertically, whereas thermal stratification occurs in other seasons (Wu et al., 2020). The lake provides an opportunity to study microbial functional diversity and the underlying drivers in deep lakes, particularly those influenced by high-altitude conditions. The unique environmental conditions offer valuable insights into microbial functional gene diversity and composition in the deep lake ecosystems, though these characteristics may not be universally applicable to all lakes. Human activities and economic development around Lugu Lake have potentially impact on the microbial functional genes and overall ecosystem health (Liu et al., 2019b). This study focused on the influence of various environmental factors, including potential anthropogenic effects such as nutrient pollution, when analyzing functional gene diversity and composition.

### 2.2. DNA extraction and high-throughput amplicon sequencing

Bacterial and archaeal community composition in sediments was taxonomically profiled using the high-throughput 16S rRNA sequencing technology according to methods described in our previous studies (Wang et al., 2013; Wang et al., 2012a). Briefly, we extracted total environment DNA from about 0.5 g dried sediment using the DNeasy PowerSoil DNA Isolation Kit (Qiagen). The bacterial 16S rRNA genes were amplified using universal primers targeting the V4 region (515F, 5′-GTG YCA GCM GCC GCG GTA A-3′; 806R, 5′-GGA CTA CNV GGG TWT CTA AT-3′) (Lee et al., 2023; Walters et al., 2016), the archaeal 16S rRNA genes using the archaeal-specific primers targeting the V4-V5 region (Arch519F, 5′-CAGCCGCCGCGGTAA-3′ and Arch915R, 5′-GTGCTCCCGCCCAATTCCT-3′) (Wei et al., 2019). These amplified products were then sequenced using the Illumina NovoSeq 6000 platform with 2 × 250 bp paired-end according to the manufacture instruction (Babita et al., 2022). We trimmed the primers from the raw reads and discarded the ones shorter than 200 bp using cutadapt v3.5, then generated amplicon sequence variants (ASVs) of bacteria and archaea using the DADA2 package v1.18 (Callahan et al., 2016). Specifically, we filtered out the sequence with the anonymous bases, the expected number of errors higher than two and its quality score lower than 10, then corrected the sequence based on the modeled error rate, which was estimated from the first 100 million bases from clean sequence. We next dereplicated and merged the paired-end sequencing with zero base mismatch and the minimum overlap of 12 for archaea and 50 bacteria, respectively. The chimera sequences were finally identified and removed using the consensus method (Pauvert et al., 2019). The taxonomy for each ASV of bacteria and archaea was annotated using the naive Bayesian classifier method implemented by DADA2 against the SILVA 138.1 database (Quast et al., 2013). The communities of bacteria and archaea were equally sampled with the lowest sequencing depth for the comparison among samples by using the “rrarefy” function in the R package “vegan” v2.5–7 (Dixon, 2003).

### 2.3. High-throughput quantitative sequencing of microbial functional genes

To quantify the abundance of functional genes that participated in carbon, nitrogen, phosphorus, and sulfur cycling, the extracted DNA was subjected to high-throughput quantitative PCR (HT-qPCR) on a SmartChip Real-time PCR system (Zheng et al., 2018). Detailed information on functional genes is provided in Table S1. The Quantitative

Microbial Ecology Chip (QMEC) comprises 72 primer pairs in total, including 36 designed pairs and 35 published pairs targeting carbon cycling, nitrogen cycling, phosphorus cycling, sulfur cycling, and one pair targeting the bacterial 16S rRNA gene as a reference gene. The QMEC chip provides a more efficient and convenient approach for simultaneously quantify the abundance of functional genes across, large-scale samples, such as 72 DNA samples or 24 samples with three replicates in a single run. Details of these primer pairs could be found from previous studies (Zheng et al., 2017).

#### 2.4. Abiotic and biotic attributes

We collected and analyzed biotic factors and abiotic factors to explain carbon, nitrogen, phosphorus, and sulfur cycling functional genes. In the field, we measured water depth, water temperature, pH, dissolved oxygen, and conductivity of surface water, bottom water, and surface sediments, using a YSI 650 multi-parameter display system equipped with a 600XL probe (Nguyen et al., 2023; Nguyen and Huynh, 2023). In the laboratory, we examined the total nitrogen, total phosphorus,  $\text{HCO}_3^-$  concentration and silicon content of the surface water (Syarif Sukri et al., 2023). For surface sediments, we analyzed total phosphorus, loss on ignition, porosity, water content, 19 types of metal ions (Al, Ba, Be, Ca, Co, Cr, Cu, Fe, Li, K, Mg, Mn, Na, Ni, Pb, Sr, Ti, V, and Zn) and particle size (particle size is divided into five groups:  $< 4 \mu\text{m}$ ,  $4\text{--}16 \mu\text{m}$ ,  $16\text{--}32 \mu\text{m}$ ,  $32\text{--}64 \mu\text{m}$  and  $> 64 \mu\text{m}$ ) (Table S2). Previous studies described detailed methods for measuring and calculating these abiotic factors (Wang et al., 2007). Principal component analysis (PCA) was used for these 19 metal ions to reduce their dimensionality, and then used the first and second axes (PC1 and PC2) as additional environmental parameters (Wang et al., 2012b; Wu et al., 2019; Zhao et al., 2019).

In terms of biotic factors, we used the following variables to instead the biotic characteristics of bacteria and archaea: (i)  $\alpha$ -diversity indices: Shannon diversity index was used to measure the diversity of microbial communities (Kembel et al., 2010); (ii) NMDS1 and NMDS2: The first and second axes of non-metric multidimensional scaling analysis (NMDS) of bacterial and archaeal communities were used to transform the similarities of different bacterial or archaeal samples for distance differences to reveal the community variations among microbial communities (Taguchi and Oono, 2005).

#### 2.5. Statistical analyses

We considered the popular metrics such as species richness, Simpson diversity, and Shannon diversity. Compared to the other two metrics, Shannon diversity is calculated with both species identity and abundance and thus includes the information of species richness and evenness (Wu et al., 2020). Considering the high occurrence of functional genes across samples, we used Shannon diversity to quantify the alpha diversity for functional genes and taxonomic composition of microbes. We further used linear and quadratic models to examine the relationship between the Shannon diversity of functional gene and the water depth (van Strien et al., 2012). The model with the lowest Akaike Information Criterion (AIC) value was chosen as the best model (Spellerberg and Fedor, 2003). The details about the model selection sees previous studies (Hu et al., 2024a; Hu et al., 2024b).

For beta diversity, we used the Bray-Curtis dissimilarity for both taxonomic composition and functional genes. This measure evaluates the differences in community composition between pairs of sites, quantifying the variation in functional genes (Graco-Roza et al., 2022). Water depth distance was quantified using the Euclidean distance. To investigate the relationship among change in functional gene beta diversity and water depth distance, and beta diversity of bacterial and archaeal communities, we employed the Gaussian generalized linear model to analyze the distance-decay relationship (Morlon et al., 2008). Significance was assessed using the Mantel test (Spearman correlation)

with 9,999 permutations. We performed NMDS based on Bray-Curtis dissimilarity of functional genes to visualize change in functional gene composition along the water depth gradient (Ishwaran, 2007).

To study the most important predictors of functional characteristics including functional gene diversity and composition, we employed a machine learning model called random forest using the R package randomForest V4.7.1.1. The functional gene composition was represented by the first axis of NMDS (Non-metric Multidimensional Scaling) based on Bray-Curtis dissimilarity (Biau and Scornet, 2016; Genuer et al., 2015). The dependent variables were functional characteristics, and the explanatory variables included water depth, other environmental factors and biotic factors. Before conducting random forest analysis, we used the Pearson correlation coefficient to assess the statistical dependence between the explanatory variables. Variables with high correlation coefficients (Pearson  $r \geq 0.7$ ) were excluded from the model to reduce the influence of spurious relationships between variables. Then, the random model generated an optimal number of 2,000 trees using cross-validation (Biau and Scornet, 2016). The importance of variables was determined based on the frequency of split selection and the model improvement per split (averaged across all trees) (Ishwaran, 2007). To more effectively compare the various predictor contributions, we converted the scores of importance into percentages relative to the sum of importance values. Additionally, we removed the variables with the lowest contribution and repeated the test until the minimum contribution of each variable exceeded 5%. Additionally, we partitioned the variation in Shannon diversity and composition associated with the explanatory variables into main driver categories (Anderson and Cribble, 1998). We divided the explanatory variables into three parts: water depth, other environmental factors, and biotic factors, and then estimated the proportion of variation explained by these three components.

Finally, we used structural equation modeling (SEM) to examine and quantify the interactive effects of water depth, other environmental factors, and biotic factors on functional gene diversity (Gene.S) or functional gene composition (Gene.N) (Lefcheck, 2016). The method involves hypothesizing a causal path and converting it into a regression equation, which is then evaluated against the data to support or refute the hypothesized pathway. SEM could evaluate the direct and indirect effects of water depth on Gene.S or Gene.N. Composite variables such as other environmental and biotic factors used in SEM models were used to explain the overall impact of various physical and chemical properties or biotic factors of bacteria and archaea (Grace et al., 2016). In addition, we used the scale function to Z-score transform all environmental variables before modeling to enable comparisons between multiple predictors and models. Based on the multiple regression analysis of Gene.S or Gene.N, and the calculation formula of the composite variables of Gene.S or Gene.N, we selected the specific indicators of each composite variables in the SEM model (Table S5). We examined alternative models using AIC and overall model fit statistics based on the hypothesized paths between the full model (Fig. S11). Finally, the model with the minimum AIC value and satisfying the model fitting statistics was selected (Grace et al., 2010). Adequate model fit was determined based on a nonsignificant  $\chi$  test ( $P > 0.05$ ), high comparative fit index ( $CFI > 0.9$ ), low AIC value, and low standardized root mean square residual ( $SRMR < 0.1$ ) (Craven et al., 2018; Hu et al., 2020). Detailed modeling fit indices for all alternative models were presented in Table S6. SEM was implemented using the R package lavaan V0.6.16, which facilitates the construction of multi-latent variable models using path diagrams to explain the underlying relationships in the model.

All statistical analyses and figure production were performed using R statistical software V4.3.1, with the packages such as vegan V2.6.4, randomForest V4.7.1.1, dplyr V1.1.3, ggplot2 V3.4.3 and ggcov V0.9.8.

### 3. Results

Using quantitative sequencing technology and statistical analysis, we identified four main findings: (i) Distinct and significant patterns in the

diversity of functional genes along the water depth gradient; (ii) The composition of functional genes was distinctly separated by the water depth gradient; (iii) There was a gene-taxon congruence between functional gene composition and the taxonomic composition; (iv) The biotic and abiotic drivers of functional genes. These main findings were detailed in separation sections as below.

### 3.1. Distinct patterns of functional gene diversity along the water depth gradient

We found there were distinct and significant patterns in the diversity of functional genes along water depth gradient. The Shannon diversity of overall functional genes showed a significant hump-shaped trend along water depth ( $R_{adj}^2 = 0.44, P < 0.001$ ) (Fig. 1a). However, six subgroups exhibited distinct patterns. Specifically, the diversity of methane metabolism genes ( $R_{adj}^2 = 0.16, P = 0.018$ ) and nitrogen cycling genes ( $R_{adj}^2 = 0.68, P < 0.001$ ) increased first, then decreased when the water depth greater than about 50 m, with a peak at a depth about 50 m (Fig. 1d, e). However, for the sulfur cycling gene ( $R_{adj}^2 = 0.20, P = 0.007$ ), Shannon diversity showed a significant hump-shaped trend along water depth, increasing in the range of 0 ~ 50 m, and then declining after exceeding 50 m, with a peak at 50 m (Fig. 1g). Additionally, carbon degradation genes ( $R_{adj}^2 = 0.11, P = 0.051$ ) and carbon fixation genes ( $R_{adj}^2 = 0.05, P = 0.096$ ) showed a weakly significant hump-shaped and linearly increasing trend, respectively (Fig. 1b, c). The relationship between the phosphorus cycling gene and water depth was nonsignificant ( $P = 0.575$ , Fig. 1f).

### 3.2. The functional gene composition changes along water depth gradient

The composition of functional genes in Lugu Lake was well separated by the water depth gradient, with less variation in deep layers (Fig. 2h-n). Among the functional genes, the composition of nitrogen cycling genes exhibited the most pronounced separation along the water depth gradient (Fig. 2l). As a support for the findings, the composition of the overall functional genes represented by Bray-Curtis dissimilarity was significantly positively correlated with water depth changes ( $r = 0.23, P = 0.002$ ), revealing a distance-decay relationship with water depth (Fig. 2a). Specifically, this relationship was strongest for the nitrogen cycling ( $r = 0.45, P < 0.001$ ), followed by the sulfur cycling ( $r = 0.18, P = 0.008$ ), then methane metabolism ( $r = 0.15, P = 0.002$ ), carbon fixation ( $r = 0.13, P = 0.046$ ), and phosphorus cycling ( $r = 0.12, P = 0.014$ ) (Fig. 2b-g). In addition, phosphorus cycling had the highest initial Bray-

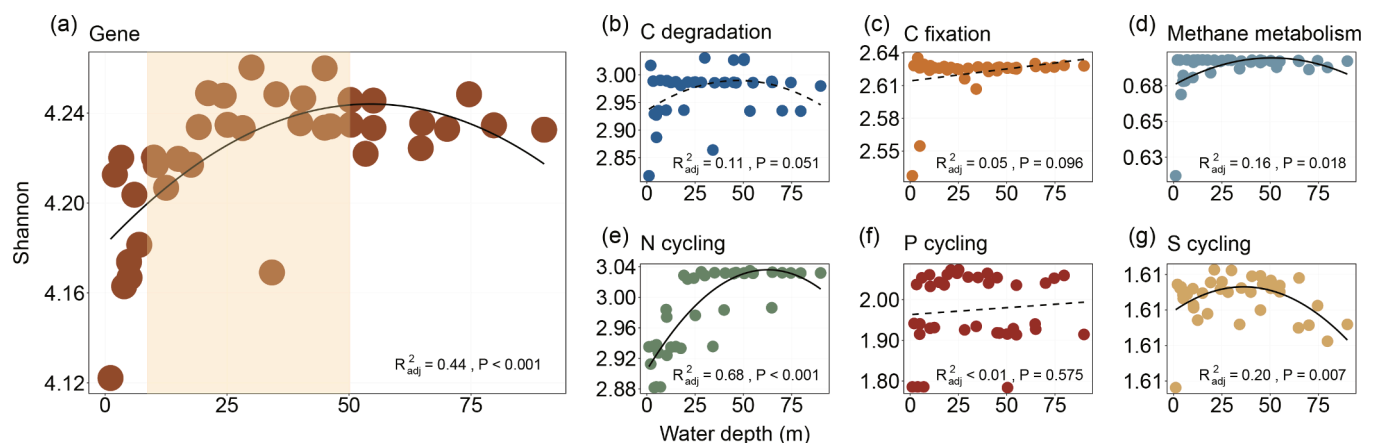
Curtis dissimilarity of 0.103, while sulfur cycling showed the lowest value of 0.028. Notably, nitrogen cycling showed the highest turnover rate with a slope of 0.442. In comparison, the turnover rates of methane metabolism, phosphorus cycling, carbon degradation, sulfur cycling, and carbon fixation were much lower than nitrogen cycling. Specifically, methane metabolism had a turnover rate of 0.158, phosphorus cycling at 0.128, carbon degradation at 0.097, sulfur cycling at 0.080, and carbon fixation at 0.073 (Table. S3).

### 3.3. The relationship between functional and taxonomic composition

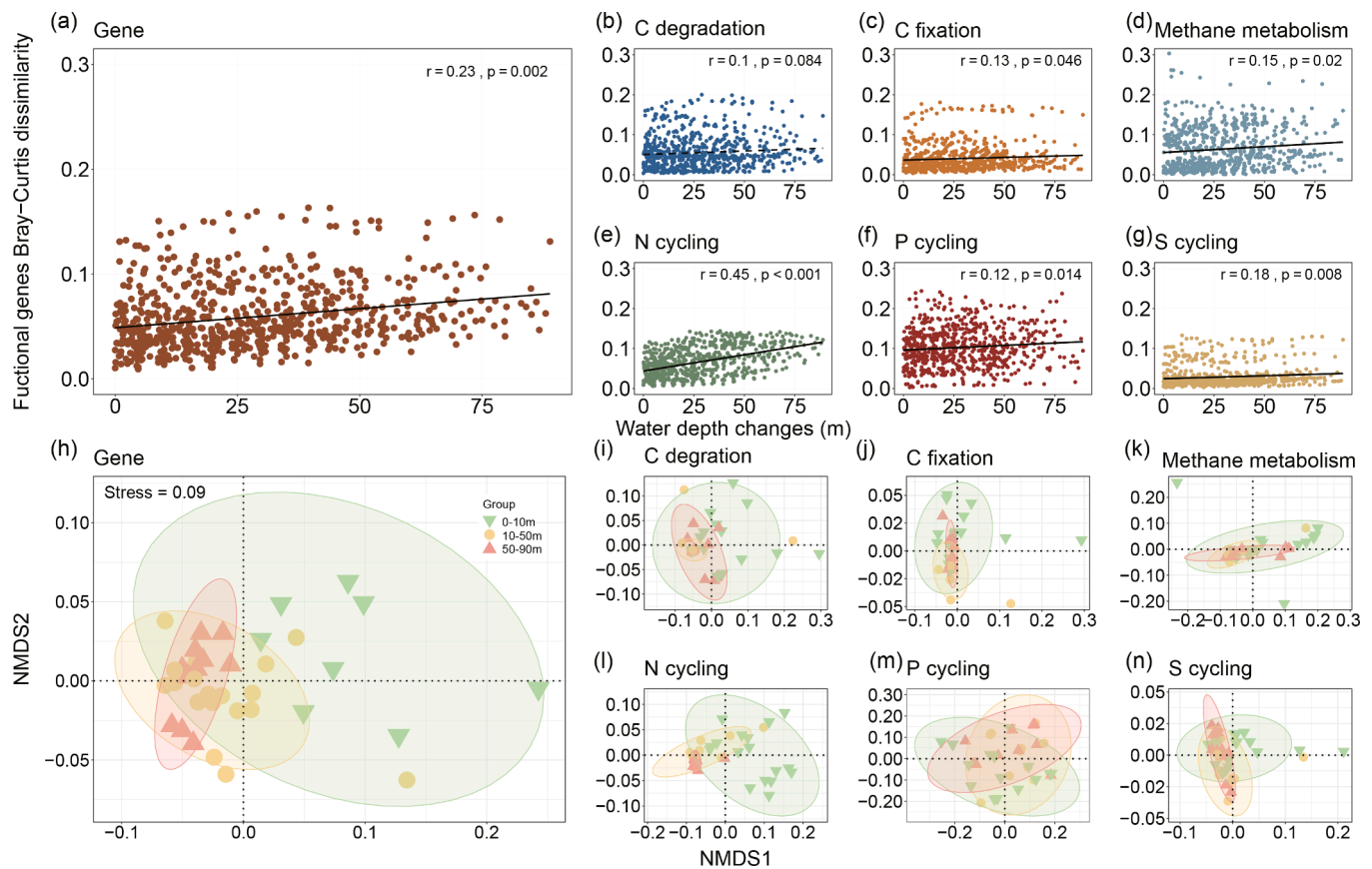
There was a significantly positive correlation between functional gene composition and community composition of bacteria and archaea, indicating a strong gene-taxon congruence of functional genes with the two kingdoms (Fig. 3a-d). Compared with archaea ( $r = 0.65, P < 0.001$ ) (Fig. 3c), the relationship between functional genes and bacteria was higher ( $r = 0.77, P < 0.01$ ) (Fig. 3a). The relationships were also consistently observed between all six subgroups and communities of two kingdoms ( $P < 0.05$ ). Specifically, for bacteria, nitrogen cycling genes had the highest Mantel  $r$  of 0.87 ( $P < 0.001$ ), followed by carbon fixation genes ( $r = 0.59, P < 0.001$ ), then sulfur cycling ( $r = 0.56, P < 0.001$ ), methane metabolism ( $r = 0.54, P < 0.001$ ), carbon degradation ( $r = 0.51, P < 0.001$ ), and phosphorus cycling ( $r = 0.22, P < 0.001$ ) (Fig. 3b and S2). For archaea, nitrogen cycling functional genes also showed the highest Mantel  $r$  of 0.74 ( $P < 0.001$ ), followed by carbon fixation functional genes ( $r = 0.47, P < 0.001$ ), then sulfur cycling ( $r = 0.47, P < 0.001$ ), methane metabolism ( $r = 0.44, P < 0.001$ ), carbon degradation ( $r = 0.43, P < 0.001$ ), phosphorus cycling ( $r = 0.21, P < 0.001$ ) (Fig. 3d and S3). In addition, the slope of dissimilarity in nitrogen cycling gene composition was highest while sulfur cycling showed the lowest for both of bacteria and archaea (Table. S4).

### 3.4. Abiotic and biotic factors driving functional gene diversity and composition

We asked three questions to elucidate the factors driving changes in functional gene. (i) How do functional characteristics including Shannon diversity and composition relate to biotic and abiotic factor? (ii) What is the relative importance of these factors in explaining functional characteristics including the relative abundance, diversity and composition? (iii) What is the interaction between abiotic and biotic factors in explaining functional characteristics? We performed multiple statistical tests, including envfit function analysis, spearman correlation analysis,



**Fig. 1.** Water-depth diversity patterns for functional genes. We considered the Shannon diversity of total functional genes (a) and six subgroups of functional genes (b-g). The yellow shaded area represents the thermocline range. Functional genes were measured by Quantitative Microbial Ecology Chip technology, and the six subgroups are carbon degradation, carbon fixation, methane metabolism, nitrogen cycling, phosphorus cycling and sulfur cycling (Table S1). The relationships between functional gene diversity and water depth were evaluated by linear and quadratic models. The better model was selected based on the lower value of Akaike's information criterion.



**Fig. 2.** Relationships between functional gene composition and water depth. The upper panels (a-g) show the relationships between water depth changes and Bray-Curtis dissimilarity of total functional genes (a) and six subgroups (b-g). The distance decay relationships were evaluated by a linear regression model. The  $r$  value from mantel tests was used to examine correlations between Bray-Curtis dissimilarity of functional genes and water depth changes for 9,999 permutations. The lower panels (h-n) show non-metric multidimensional scaling (NMDS) plots of total functional genes (h) and six subgroups (i-n). Each point in the lower panels represented a sample, which was colored according to water depth, from green (0–10 m) to yellow (10–50 m) and then to red (50–90 m).

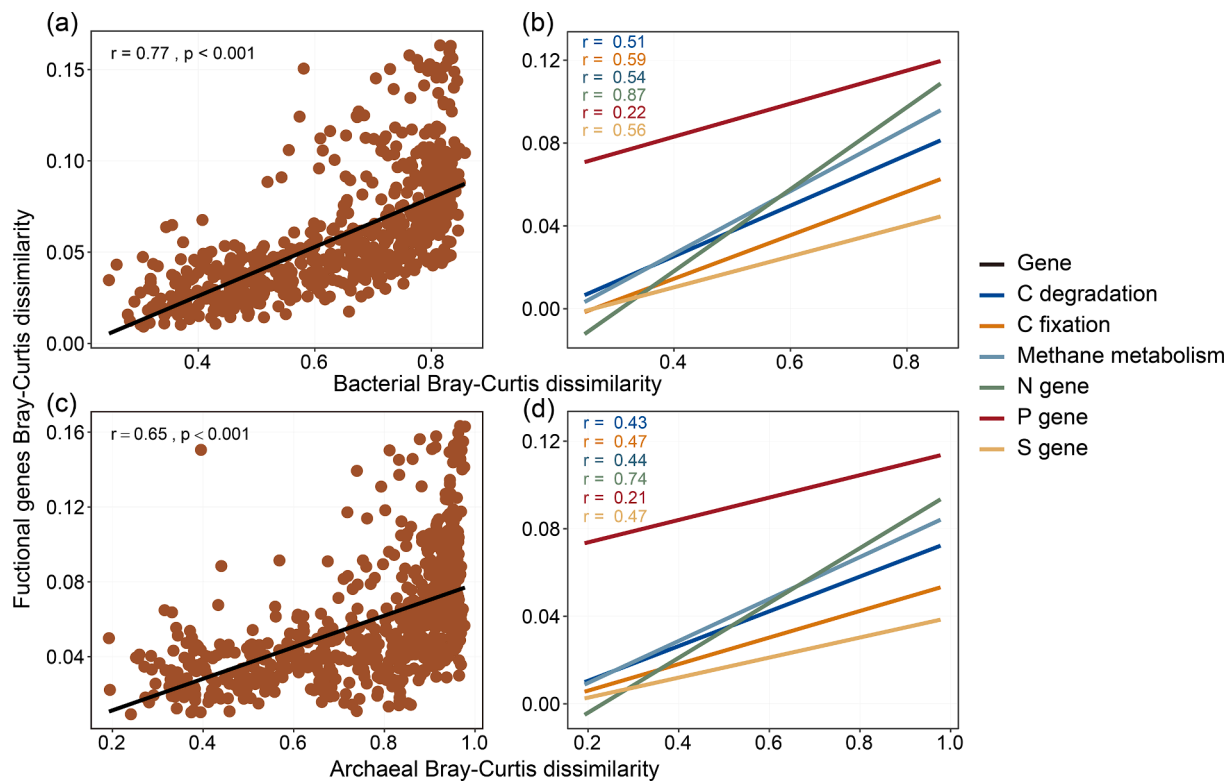
random forest model, variation partitioning analysis, and Mantel test.

Firstly, we find functional characteristics exhibited a close relationship with a variety of factors, including water depth, bottom temperature, bottom conductivity, bottom pH, total phosphorus in water, grain size, loss on ignition, Bac.nm1, Arc.div, and Arc.nm1 (Table S2, Fig. 4a). For functional and community composition, there were similarity in abiotic factors influencing the composition of bacterial and archaeal communities, including water depth, bottom temperature, surface temperature, surface conductivity, loss on ignition, the first principal component for metal ions, water content in sediment and porosity, but the abiotic factors influencing the composition of functional genes differed and less than them (Fig. S9). Abiotic factors also had a stronger effect on the bacterial and archaea community composition, but their impacts on functional genes were relatively weaker (Fig. S9).

Secondly, abiotic factors including water depth, temperature, and a range of biotic factors were the most effective in elucidating the changes in the relative abundance of functional genes within the deep lake (Fig. 4c). For example, functional genes associated with nitrogen cycling were mainly affected by metal ions concentration, total phosphorus in sediment, porosity and water content in sediment. Genes related to carbon fixation which are mainly responsible for converting  $\text{CO}_2$  into organic compounds, were predominantly driven by metal ions concentration and pH. Meanwhile, the functional genes involved in the carbon degradation were mainly influenced by water content in sediment and metal ions concentration (Fig. 4b-c). Six subgroups of functional genes were influenced by different abiotic factors. For example, nitrogen cycling genes are primarily affected by environmental factors such as

total phosphorus in sediment and water content in sediment, while carbon cycling genes, including carbon degradation and fixation, are significantly affected by bicarbonate and metal ions. Moreover, the explained variation of the nitrogen cycling genes in the random forest model was substantially greater than the other five categories (Fig. 4b).

The random forest model reveals that water depth had highest contribution of 21.50 % for diversity and 21.52 % for composition of functional genes (Fig. 5a-b, S6 and S7). The bacterial community composition also proved to be significant in predicting diversity pattern with 16.52 % and composition pattern with 20.53 % (Fig. 5a, b). In addition, other environmental variables, such as loss on ignition, total phosphorus in water, and bicarbonate also played an important role in predicting the functional characteristics with 16.23 % and 11.95 %, 11.18 % and 9.92 %, 7.23 % and 5.46 %, respectively (Fig. 5, S6 and S7). Regarding functional composition, the biotic factors and water depth were identified as key driving factors for all functional gene composition, followed by other environmental factors (Fig. S4). For example, the total phosphorus in water significantly contributed to the composition of nitrogen cycling and phosphorus cycling gene (Fig. S4). Additionally, we partitioned the variations in functional characteristics into three categories: water depth, biotic factors and other environmental factors (Fig. 5c, d). This model explained 74.2 % of the variance in functional gene Shannon diversity and 72.3 % in their composition. Consistent with above random forest model results, biotic factors had a more pronounced impact on functional genes than water depth and other environmental factors. For functional genes Shannon diversity, biotic factors explained 28.5 % of variation, while water depth and environmental factors only explained 4.2 % and 3.8 % of the variance, respectively



**Fig. 3.** The relationships between functional gene composition and archaeal and bacterial community composition. The upper panels (a-b) show the relationships between bacterial Bray-Curtis dissimilarity and the Bray-Curtis dissimilarity of total functional genes (a) and six subgroups (b). The lower panels (c-d) show the relationships between archaeal Bray-Curtis dissimilarity and the Bray-Curtis dissimilarity of total functional genes (c) and six subgroups (d). Regression of linear relationship based on a linear model was shown with a solid line. Mantel tests were used to examine correlations between differences in functional gene composition and differences in community composition for 9,999 permutations. The  $r$  values of Mantel statistics are shown, with all  $P$  values being less than 0.001. The colors of regression line represented different categories of functional gene. The details of the relationships between bacterial or archaeal Bray-Curtis dissimilarity and the Bray-Curtis dissimilarity of six subgroups are shown in [Figure S2](#) and [Figure S3](#).

([Fig. 5c](#)). In terms of the functional gene composition, biotic factors accounted for 25.2 %, whereas water depth and environmental factors contributed to merely 0.7 % and 4.4 % respectively ([Fig. 5c-d](#)).

Thirdly, we applied structural equation modeling to statistically analyze the hypothesized interaction among water depth, other environmental and biotic factors ([Fig. 6](#), S13 and S14). To summarize the influencing factors of functional gene diversity and composition, we incorporated two metrics (Gene.S and Gene.N) related to genes involving carbon, nitrogen, phosphorus, and sulfur cycling in the model ([Fig. S10](#)). When examining all possible connections among these factors in the model, water depth was included in the final best-fitting SEM and significantly enhanced the model's predictive power for indices of functional gene diversity and composition ([Fig. 6a-d](#)). Our SEM model showed two findings. First, water depth exerted different intensity and direction of effect on diversity and composition of functional genes. After incorporating the water depth variable, the explained variation of the SEM models for diversity and composition improved by 5.5 % and 1.9 %, respectively ([Fig. 6a-d](#)) and the improvement for diversity was more substantial ([Fig. 6a, b](#)). Water depth exerted diverse directional influences for diversity and composition. Secondly, water depth impacted functional genes through both direct and indirect effects on their diversity and composition. The direct effects of water depth on diversity and composition were  $-0.43$  and  $0.31$ . The indirect effects mainly occurred through interactions with biotic and abiotic environmental factors ([Fig. 6c, d](#)). Specifically, the total indirect effects of water depth on diversity and composition were  $0.98$  and  $0.87$ , respectively ([Fig. 6b, d](#), [Table S7](#)). Other environmental factors had a direct effect of  $0.54$  on composition, and strong indirect effects of  $0.53$  and  $0.35$  on diversity and composition through biotic factors ([Fig. 6b, d](#), [Table S7](#)).

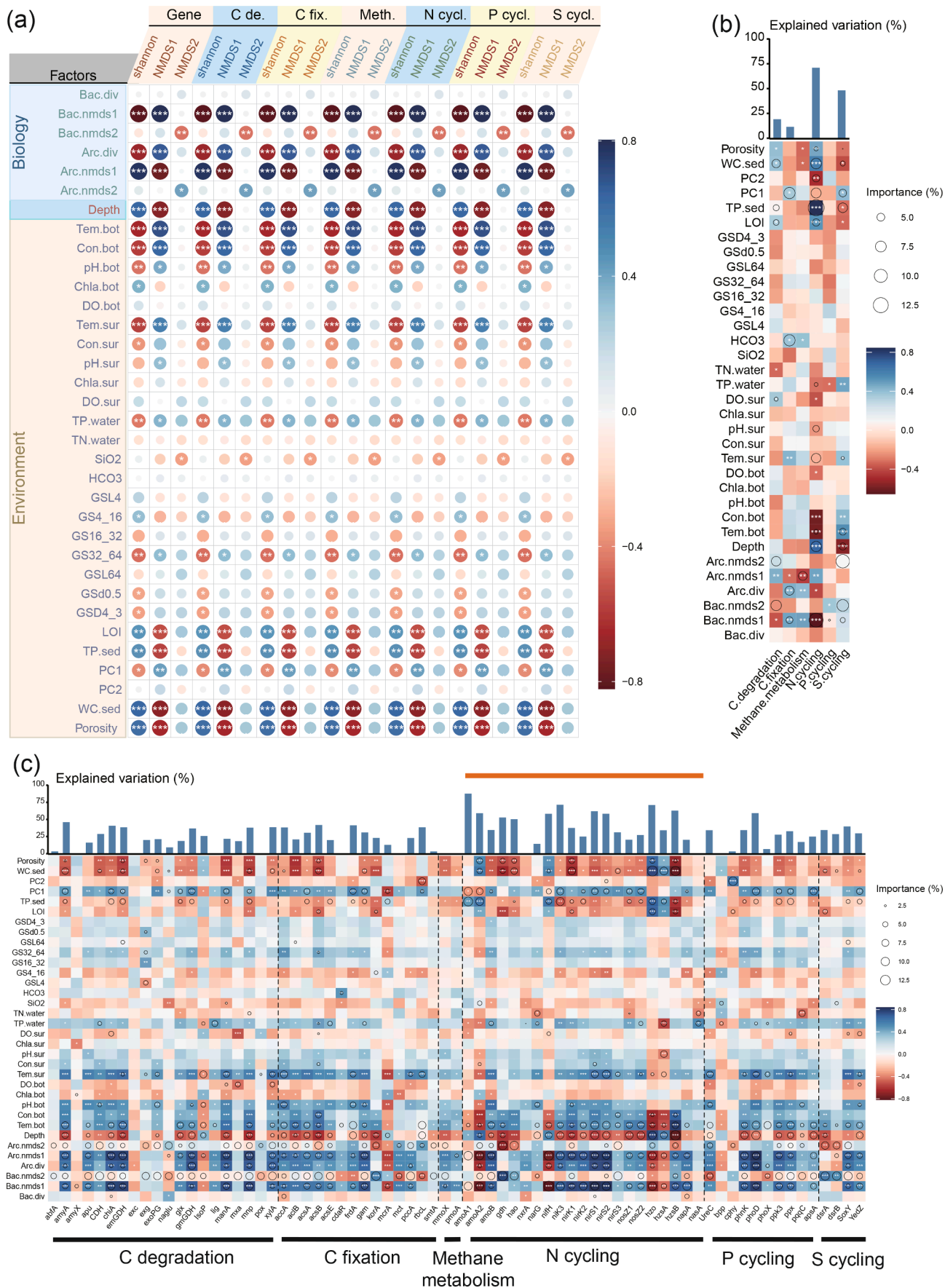
This dual effect of water depth gradient was further statistically supported by variation partitioning analysis and random forest analysis ([Fig. 5](#)).

## 4. Discussion

Understanding these patterns is crucial for grasping how microbial functional diversity responds to environmental gradients in deep lake. Our study provides insights into the water depth distribution of functional genes and their drivers in a deep lake, using QMEC quantitative technology. We found that: (i) The functional gene diversity showed a hump-shaped pattern throughout the water depth, reaching a peak at around 50 m in the thermocline region; (ii) Functional gene dissimilarity increased with water depth distance showing a distance-decay relationship; (iii) Biotic factors had more important effects on functional characteristics than water depth and other environmental factors.

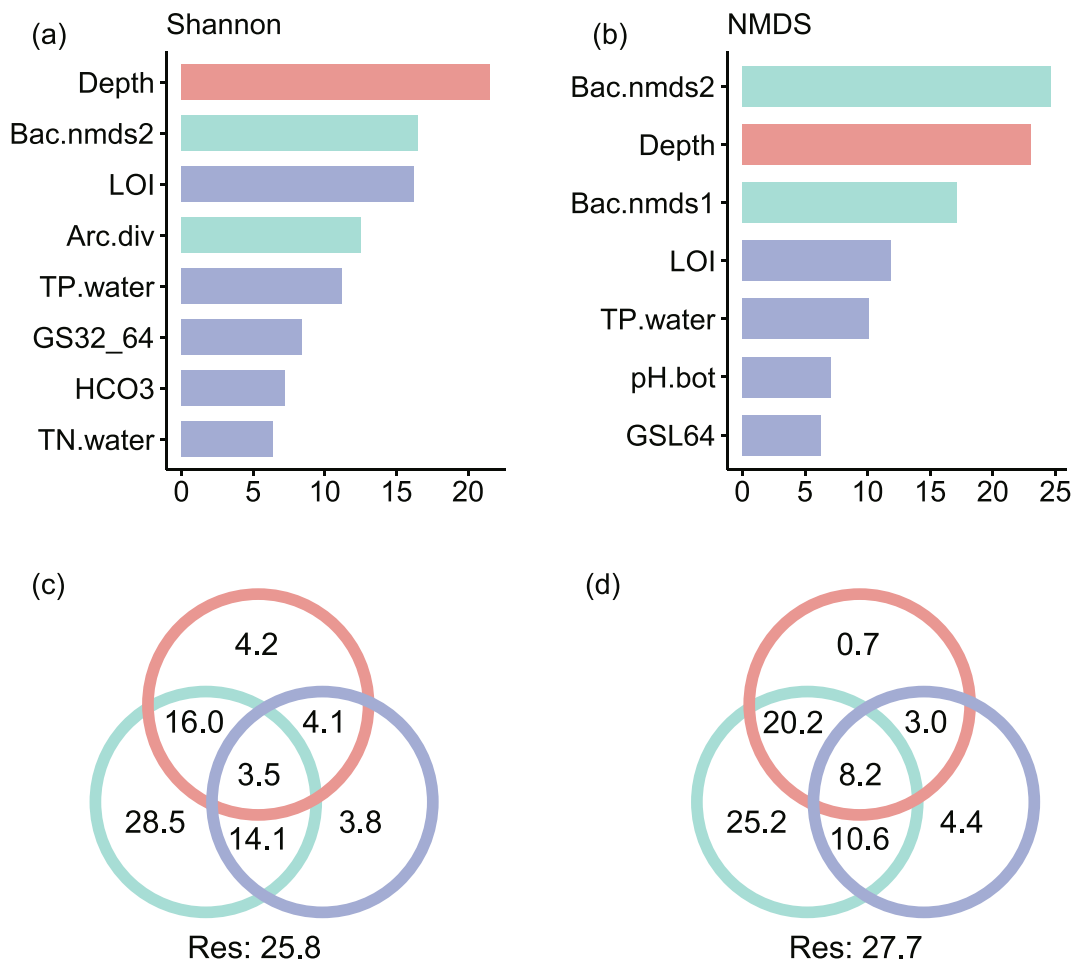
### 4.1. Patterns of functional gene diversity along water depth

The diversity of overall functional genes was highest at approximately 50 m depth, around the thermocline region ([Chang et al., 2022](#); [Ren and Wang, 2022](#)), being consistently with the subgroups including methane metabolism, nitrogen cycling, and sulfur cycling genes. This suggests that microbial processes related to these functional genes were most diverse and active in the thermocline. The thermocline as a distinctive layer in a body of water, is characterized by a rapid temperature decline along the water depth, effectively separating the mixed or surface layer from the colder bottom or deeper layers of water ([Mazumder et al., 1990](#)). This unique environment within the



(caption on next page)

**Fig. 4.** Abiotic and biotic factors underlying the functional gene abundance, diversity and composition. (a) The correlations between abiotic and biotic factors with functional characteristics. Abiotic and biotic factors evaluated in this analysis are provided in Table S2. (b) The impact of abiotic and biotic factors on the abundance of functional gene subgroups at the level of subgroups of functional genes. (c) The impact of abiotic and biotic factors on the abundance of functional gene at the level of individual functional genes. The strength of Spearman correlation is shown with the change in color, and the significance is indicated by Asterisks. The color gradient on the right represents the spearman rank correlation coefficient, with higher positive values (dark blue) indicating stronger positive correlation, and higher negative values (dark red) indicating stronger negative correlation. \*\*\*P<0.001, \*\*P<0.01, \*P<0.05. The importance of abiotic and biotic factors in explaining individual functional gene changes through random forest analysis was represented by circles of different sizes.



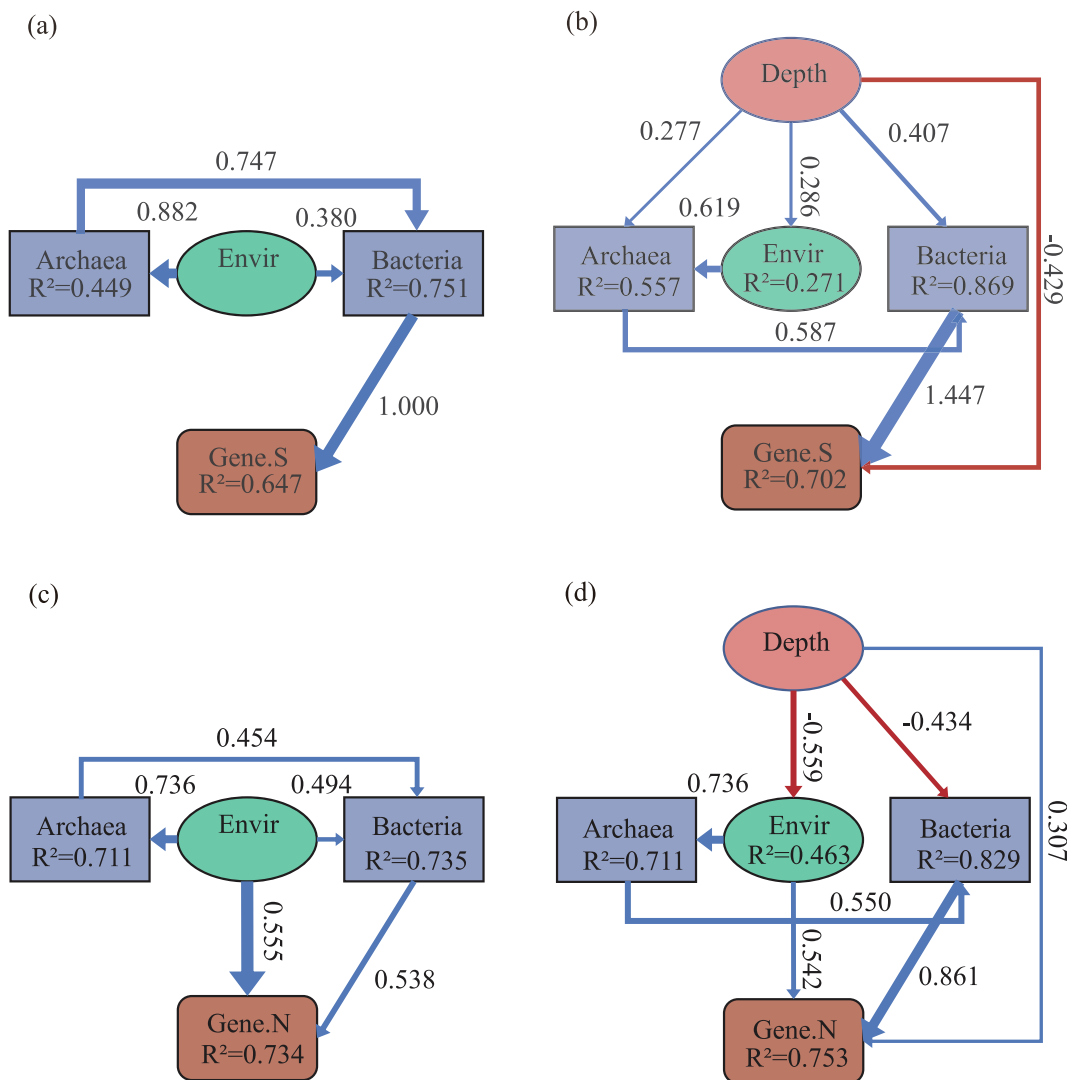
**Fig. 5.** Relative contribution of environmental variables to functional gene Shannon diversity and composition. Random forest analysis identified and quantified significant predictors of Shannon diversity (a) and composition (b) of total functional genes. The first axis of NMDS was used to represent functional gene composition. We selected the explanatory variables with a relative contribution rate > 5 %. Details of variable abbreviations are provided in Table S2. The relative impact of explanatory variables on Shannon diversity (c) and composition (d) of functional genes was determined by variant partitioning analysis (VPA). We classified the explanatory variables into water depth (red circles), biotic factors (green circles), and other environmental factors besides water depth (purple circles) (see Table S2 for details on these three attributes). The numbers in the circles represent the variance explained by the variables.

thermocline offers optimal conditions for the presence and activity of functional genes (Cantin et al., 2011). For example, the oxygen variation is highly dynamic, and the gradients of temperature and redox are significantly steep in the region of thermocline of Lugu Lake (Chang et al., 2022). These factors could affect microbial composition and functional gene activity, which is consistent with other studies, such as in the tropical freshwater Lake Tanganyika, where depth-dependent factors including oxygen and nutrients influence microorganisms involved in biogeochemical cycling (Tran et al., 2021). Additionally, the thermocline usually acts as a barrier to nutrient flow and temperature in aquatic ecosystems (Kurt, 2019), creating distinct niche and resource pools for various microorganisms (Humayoun et al., 2003; Xing et al., 2019). The special niche promotes lots of microbial activities to occur in this region. For example, the bacterial biomass similarly exhibits a hump-shaped distribution as the water depth increases in Lugu Lake,

peaking within the thermocline region (Zhao et al., 2019). The thermocline layer of marine shows the peak of the microbes such as the abundance of heterotrophic bacteria in the eastern tropical North Pacific, based on flow cytometry measurements and 16S rRNA analysis (Ma et al., 2009). Therefore, we infer that the increasing diversity of functional genes around the thermocline was associated with the size of the microbial species pool structured by dynamic environmental change (Lapteva and Sokolova, 2019; Pajares et al., 2017), which was also verified in the subsequent random forest analysis (Fig. 5).

#### 4.2. Water depth pattern of functional gene composition

The composition of overall functional gene and six subgroups demonstrated a significant distance-decay relationship with water depth. This means that there was a significant increase in the



**Fig. 6.** Structural equation model of functional gene diversity and composition. Shannon diversity and composition of functional genes were represented by Gene.S and Gene.N respectively. Water depth, other environmental factors, bacterial biotic factors and archaeal biotic factors had direct and indirect effects on Gene.S (b) and Gene.N (d). The best-fitting model accounted for the effect of the predictor variables on Gene.S (a, b) or Gene.N (c, d) by excluding (a, c) or including (b, d) water depth variables. Blue and red arrows indicate positive and negative effects, respectively. Arrow widths and accompanying numbers are the relative effects of the modelled relationship (that is, standardized path coefficients). R<sup>2</sup> represents the proportion of variance explained by the variable. Please refer to Table S5 for more details on model fitting.

dissimilarity of functional gene composition as the water depth change. This aligns with the general ecological principle that microbial community often exhibits distance-decay relationships across spatial or environmental distance (Morlon et al., 2008; Nekola and White, 2004; Soininen et al., 2007). The observed distance-decay relationship in functional gene composition could be a result of the varying environmental conditions with water depth change (Raposeiro et al., 2018). Our study shows that a variety of environmental factors such as temperature, dissolved oxygen, and nutrient concentration exhibited similar change along depth gradients consistently with previous study in Lake Lugu (Chang et al., 2022).

Besides, water depth change had distinct effect on initial dissimilarity and turnover rate of functional gene composition across different functional subgroups. Regarding initial dissimilarity, phosphorus cycling genes had the highest value among six subgroups. This suggests that phosphorus cycling genes could respond intensively to environmental change along a narrow range of water depth within the lake surface (Li et al., 2023; Zhang et al., 2023). In contrast, sulfur cycling genes had the lowest initial dissimilarity. Additionally, nitrogen cycling

genes exhibited the highest turnover rates in response to changes in water depth. This pronounced variation in nitrogen cycling genes across different water depths suggests that they were sensitive to water depth and other environmental factors that influence the nitrogen cycling process. This was in line with the results of a recent research of temperature and oxygen stratification shape the distribution of functional nitrogen genes in Lake Alchichica (Pajares et al., 2017). This implies that microbial communities related to nitrogen cycling dynamically adapt to changes in water depth (Zehr and Ward, 2002), further highlighting the crucial role of water depth in determining the ecological functions within aquatic ecosystems.

#### 4.3. The congruence between functional and taxonomic composition

There was clear gene-taxon congruence between their composition. The clear pattern suggests that high beta diversity of microbial communities had a pivotal role in sustaining multiple functions, indicating there was no universal combination of species capable of simultaneously supporting all functions in the Lugu Lake (Mori et al., 2018).

Functional genes had similar gene-taxon congruence across bacterial and archaeal kingdoms. This might imply that functional genes of different microbial kingdoms, although evolutionarily distinct, shared common ecological and environmental drivers and exhibited same functional gene composition patterns (Koonin and Wolf, 2008; Wessén et al., 2011). The similar congruence across kingdoms probably stemmed from that several coexisting microorganisms, despite taxonomically distinct, can encode the same metabolic functions (Louca et al., 2018). Although the identities of bacteria and archaea encoding each function have changed significantly across time, space and evolution process, their functions have been minimally affected (Louca et al., 2016b). This evolutionary link promoted functional composition consistency between the two kingdoms (Allers and Mevarech, 2005).

The six subgroups also exhibited significantly gene-taxon congruence. Notably, this congruence was most pronounced between nitrogen cycling genes and microbial community, which suggests that nitrogen metabolism might be the most critical ecological function in Lugu Lake. This was verified by the roles of *Nitrospira* and *Nitrosarchaeum*, the key species executing ammonia oxidation in the nitrogen cycling, possess a notable abundance advantage over other species in the microbial community of Lugu Lake (Figs. S11 and S12). Their abundance follows a hump-shaped pattern along water depth, peaking near the thermocline (Ren and Wang, 2022; Winter et al., 2009). Their dominance in Lugu Lake caused nitrogen cycling genes to exhibit highest turnover rate in microbial community composition, highlighting the critical role of nitrogen cycling in this lake. In contrast, the congruence for sulfur cycling genes was the lowest. This might be attributed to the lower abundance of sulfur-cycling microorganisms in the ecosystem. Our results showed that the top ten most abundant bacterial and archaeal genus were not involved in sulfur cycling metabolism (Fig. S11). Microbes related to sulfur cycling such as sulfate reduction are less dominant or stable in Lugu Lake, leading to lower congruence between sulfur cycling genes and the broader taxonomic composition of the microbial community.

#### 4.4. Biotic and abiotic factors driving functional genes diversity and composition along water depth

The effect of water depth combined with other environmental and biotic factors was stronger than its pure effect on functional characteristics in Lugu Lake. Water depth, as a multifaceted proxy variable, represents a range of environmental gradients changes like temperature, light, and nutrient levels (Boehrer and Schultze, 2008; Humayoun et al., 2003; Nevalainen, 2012). It directly and indirectly affected microbial communities and functional genes across lake ecosystem. For example, water depth explains a considerable portion of the variation in bacterial communities in Lugu Lake and ecosystem functioning involving the cycling of carbon, nitrogen, phosphorus and sulfur in Issyk Kul Lake and Semiarid Lake (Rojas-Jimenez et al., 2021; Zhang et al., 2021; Zhao et al., 2019). Other environmental factors also exerted a strong impact on the diversity of microbial functional genes. This impact likely arose because microbial communities underwent environmental filtering, where their composition and diversity were filtered and selected by specific environmental conditions (Song et al., 2019). Only those well adapted species could survive and thrive, leading to the aggregation of species with similar functional characteristics (Louca et al., 2016a).

Additionally, other environmental factors impacted microbial communities and functional genes differently (Fig. S9). Functional gene diversity and composition showed relatively stable and a lower susceptibility than microbial community to environmental change. This functional stability might be supported by the hypothesis, functionally redundant presents in microorganisms that executes the carbon, nitrogen, phosphorus, and sulfur cycling (Wagg et al., 2019). Moreover, these functional cycling genes, which are involved in the basic survival and metabolic processes, tended to respond relatively slowly to environmental changes compared to microbial community. Particularly, there was a strong correlation between the nitrogen cycling genes and

environmental factors indicating that the nitrogen cycling gene shows a considerable response to these factors (Yang et al., 2014). This suggests that environmental changes significantly influenced the nitrogen cycling gene more than others. This might be due to the concentration and form of nitrogen in the water environment are affected by many factors, such as inflow materials, biological processes, temperature, pH, and dissolved oxygen levels (Penn et al., 2019). Therefore, the genes controlling the nitrogen cycling rapidly responded to change in environmental factors to effectively regulate nitrogen conversion and utilization (Nelson et al., 2016; Sun et al., 2021).

The six subgroups of functional gene had distinct relationships with environmental factors, indicating that functional genes had specific environmental preferences. For example, nitrogen cycling genes preferred to low temperature, conductivity and deeper depth, however, sulfur cycling genes tended to high temperature, conductivity and shallower depth. The preferences of genes to specific environment may stemmed from the environmental preferences of the corresponding species. Moreover, environmental factors affecting all functional genes were also closely related to individual functional genes. Interestingly, we discovered the functional genes of the same metabolic pathway tended to be driven by the same environmental factors. The individual genes involved in carbon degradation, carbon fixation, phosphorus cycling and sulfur cycling were considerably impacted by similar environmental factors. The similar drivers implies that there was a certain synergy among them (Anantharaman et al., 2016; Song et al., 2022). And these cycling processes were interrelated and coordinated (Raes and Bork, 2008).

Biotic factors had a greater impact than environmental factors on functional characteristics. Among biotic factors, both the size of the species pool and changes in microbial composition affected the diversity and composition of functional genes. This highlighted the intricate link between species interactions and the diversity of functional genes (Escalas et al., 2019). Our SEM results provides further evidence that three variables including water depth, environmental and biotic factors jointly explained the diversity and composition of functional genes. Among these biotic factors, the bacterial community composition was the most pivotal driver for the diversity and composition of each type of functional gene. Interestingly, bacteria potentially had a more influential role in shaping the genetic composition of these biogeochemical cycles in deep lakes. The dominance of bacteria in explaining functional characteristics might be because (i) Bacteria has higher relative abundance and species diversity than archaea in many aquatic ecosystems (Chen et al., 2017; Flemming and Wuertz, 2019), which makes them occupy a more important position in biological community structure and ecological function (Locey and Lennon, 2016; Louca et al., 2016b); (ii) Bacteria maintains a wide array of metabolic, genetic, and physiological capabilities, allowing for metabolic diversification across various niches, which allows bacteria to dominate archaea in most aquatic environments (Chen et al., 2017; Valentine, 2007); (iii) Bacteria exhibits faster growth rates and shorter generation times than archaea, which enables bacteria to quickly respond to environmental changes and expand in ecosystem (Wani et al., 2022); (iv) As a deep lake, Lugu Lake displays obvious seasonal temperature and nutrient stratification (Zhao et al., 2023), leading to considerable differences in the energy sources available to microbial communities across different water layers (Chang et al., 2022). This created an environment of discontinuous stress in Lugu Lake, challenging the resident microbial communities. Therefore, bacteria demonstrates a remarkable capacity to promptly respond to energy availability fluctuations and efficiently utilizes resources as they become accessible in such environments marked by discontinuous energy stress (Valentine, 2007). These capacities enabled bacteria become more dominant under the condition of insufficient supply of cellular energy in deep lakes.

## 5. Conclusions

Our study revealed the distribution patterns of functional genes in response to water depth gradients in deep lake. Actually, we examined the relationships between the abundance of functional genes and the natural water depth gradients for the first time and had numerous findings. For instance, we observed a clear hump-shaped pattern in the diversity of functional genes across water depth, with a peak in the thermocline region. Functional gene composition was well separated by water depth, and showed a significant distance-decay relationship along the gradient. Our findings also highlighted that the different functional gene subgroups had specific environmental preferences. Water depth emerged as the primary driver of both the diversity and composition of functional genes. Additionally, water depth indirectly influenced functional genes through environmental factors like temperature, conductivity, and metal ions. Moreover, biotic factors such as the species pool size and composition of bacterial and archaeal communities had a more substantial impact on functional gene diversity and composition than abiotic factors. This study enhanced our understanding of functional dynamics in microbial communities, offering a broader perspective on the ecological processes within aquatic ecosystems. The findings had implications for studying functional gene coupling and microbial interactions within aquatic systems, providing a foundation for future research on microbial functional dynamics. In addition, the insights are valuable for predicting how microbial communities might respond to environmental changes and for effectively managing aquatic ecosystems. Future research could aim to validate these results across multiple lakes and diverse aquatic systems to establish broader applicability.

### CRedit authorship contribution statement

**Peixuan Zhang:** Writing – review & editing, Writing – original draft, Resources, Formal analysis, Data curation. **Minglei Ren:** Writing – review & editing, Resources, Methodology, Data curation. **Weizhen Zhang:** Resources, Data curation. **Yan Xu:** Writing – review & editing. **Jianjun Wang:** Writing – review & editing, Formal analysis, Conceptualization.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (42225708, 42372353, 92251304, 92351303, 42002304) and the State Environmental Protection Key Laboratory of Aquatic Ecosystem Health in the Middle and Lower Reaches of Yangtze River (AEHKF2023003). We appreciate Yongguan Zhu and Jianqiang Su for the help in function gene quantification.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2024.112532>.

## References

- Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., et al., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7, 13219.
- Anderson, M.J., Cribble, N.A., 1998. Partitioning the variation among spatial, temporal and environmental components in a multivariate data set. *Aust. J. Ecol.* 23, 158–167.
- Babita, R., Sanwal, P., Arya, M., Chandola, R., Joshi, G.K., 2022. Microbial community structure and functional gene analysis of Keem (a starter culture) native to Jaunsari tribe of Uttarakhand. *Research Square Platform LLC, India*.
- Bastida, F., Eldridge, D.J., García, C., Kenny Png, G., Bardgett, R.D., Delgado-Baquerizo, M., 2021. Soil microbial diversity–biomass relationships are driven by soil carbon content across global biomes. *International Symposium on Microbial Ecology* 15, 2081–2091.
- Biau, G., Scornet, E., 2016. A random forest guided tour. *TEST* 25, 197–227.
- Boehrer, B., Schultze, M., 2008. Stratification of lakes. *Rev. Geophys.* 46, RG2005.
- Broman, E., Izabel-Shen, D., Rodríguez-Gijón, A., Bonaglia, S., Garcia, S.L., Nascimento, F.J.A., 2022. Microbial functional genes are driven by gradients in sediment stoichiometry, oxygen, and salinity across the Baltic benthic ecosystem. *Microbiome* 10, 126.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583.
- Cantin, A., Beisner, B.E., Gunn, J.M., Prairie, Y.T., Winter, J.G., 2011. Effects of thermocline deepening on lake plankton communities. *Can. J. Fish. Aquat. Sci.* 68, 260–276.
- Chang, F., Hou, P., Wen, X., Duan, L., Zhang, Y., Zhang, H., 2022. Seasonal Stratification Characteristics of Vertical Profiles and Water Quality of Lake Lugu in Southwest China. *Water* 14, 2554.
- Chen, X., Andersen, T.J., Morono, Y., Inagaki, F., Jørgensen, B.B., Lever, M.A., 2017. Bioturbation as a key driver behind the dominance of Bacteria over Archaea in near-surface sediment. *Sci. Rep.* 7, 2400.
- Craven, D., Eisenhauer, N., Pearse, W.D., Hautier, Y., Isbell, F., Roscher, C., et al., 2018. Multiple facets of biodiversity drive the diversity–stability relationship. *Nat. Ecol. Evol.* 2, 1579–1587.
- DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U., et al., 2006. Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior. *Science* 311, 496–503.
- Dixon, P., 2003. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930.
- Escalas, A., Hale, L., Voordeckers, J.W., Yang, Y., Firestone, M.K., Alvarez-Cohen, L., et al., 2019. Microbial functional diversity: From concepts to applications. *Ecol. Evol.* 9, 12000–12016.
- Flemming, H.C., Wuerzt, S., 2019. Bacteria and archaea on Earth and their abundance in biofilms. *Nature Review Microbiology* 17, 247–260.
- Fuhrman, J.A., 2009. Microbial community structure and its functional implications. *Nature* 459, 193–199.
- Garner, R.E., Kraemer, S.A., Onana, V.E., Fradette, M., Varin, M.P., Huot, Y., et al., 2023. A genome catalogue of lake bacterial diversity and its drivers at continental scale. *Nature Microbiol.* 8, 1920–1934.
- Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2015. VSURF: an R package for variable selection using random forests. *The R Journal* 7, 19–33.
- Grace, J.B., Anderson, T.M., Olf, H., Scheiner, S.M., 2010. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* 80, 67–87.
- Grace, J.B., Anderson, T.M., Seabloom, E.W., Borer, E.T., Adler, P.B., Harpole, W.S., et al., 2016. Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* 529, 390–393.
- Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A., Alahuhta, J., Altman, J., et al., 2022. Distance decay 2.0—a global synthesis of taxonomic and functional turnover in ecological communities. *Glob. Ecol. Biogeogr.* 31, 1399–1421.
- He, Z., Zhang, P., Wu, L., Rocha, A.M., Tu, Q., Shi, Z., et al., 2018. Microbial Functional Gene Diversity Predicts Groundwater Contamination and Ecosystem Functioning. *Mbio* 9, e02435.
- Hewson, I., Steele, J.A., Capone, D.G., Fuhrman, J.A., 2006. Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. *Limnol. Oceanogr.* 51, 1274–1283.
- Hu, A., Wang, J., Sun, H., Niu, B., Si, G., Wang, J., et al., 2020. Mountain biodiversity and ecosystem functions: interplay between geology and contemporary environments. *International Symposium on Microbial Ecology* 14, 931–944.
- Hu, A., Han, L., Lu, X., Zhang, G., Wang, J., 2024a. Global patterns and drivers of dissolved organic matter across Earth systems: insights from H/C and O/C ratios. *Fundamental Research*.
- Hu, A., Jang, K.S., Tanentzap, A.J., Zhao, W., Lennon, J.T., Liu, J., et al., 2024b. Thermal responses of dissolved organic matter under global change. *Nat Commun* 15, 576.
- Humayoun, S.B., Bano, N., Hollibaugh, J.T., 2003. Depth Distribution of Microbial Diversity in Mono Lake, a Meromictic Soda Lake in California. *Appl. Environ. Microbiol.* 69, 1030–1042.
- Ishwaran, H., 2007. Variable importance in binary regression trees and forests. *Electronic Journal of Statistics* 1, 519–537.
- Jing, X., Prager, C.M., Chen, L., Chu, H., Gotelli, N.J., He, J.S., et al., 2022. The influence of aboveground and belowground species composition on spatial turnover in nutrient pools in alpine grasslands. *Glob. Ecol. Biogeogr.* 31, 486–500.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., et al., 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464.

- Koonin, E.V., Wolf, Y.I., 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719.
- Kurt, H., 2019. Depth distribution of microbial diversity in lakes. *Freshwater Microbiology* 225–262.
- Lapteva, N.A., Sokolova, E.A., 2019. Vertical Distribution of Bacteria in Forest Lakes of Karelia. *Biol. Bull.* 46, 1302–1307.
- Lee, H.B., Jeong, D.H., Cho, B.C., Park, J.S., 2023. Comparative analyses of eight primer sets commonly used to target the bacterial 16S rRNA gene for marine metabarcoding-based studies. *Front. Mar. Sci.* 10.
- Lefcheck, J.S., 2016. PIECEWISESEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol. Evol.* 7, 573–579.
- Li, Y., Yu, X., Liu, H., Gong, Z., 2023. The spatio-temporal distribution of alkaline phosphatase activity and phoD gene abundance and diversity in sediment of Sancha Lake. *Sci. Rep.* 13.
- Liu, Y., Chen, C., Yang, S., 2019b. Assessment of Anthropogenic Impact versus Climate Change on the Succession of the Diatom Community in Lugu Lake (Yunnan-Guizhou Plateau, China) Using the Sedimentary Record of Geochemical Elements. *Water* 11, 655.
- Liu, J., Meng, Z., Liu, X., Zhang, X.-H., 2019a. Microbial assembly, interaction, functioning, activity and diversification: a review derived from community compositional data. *Marine Life Science & Technology* 1, 112–128.
- Liu, Y., Ren, Z., Qu, X., Zhang, M., Yu, Y., Zhang, Y., et al., 2020. Microbial community structure and functional properties in permanently and seasonally flooded areas in Poyang Lake. *Sci. Rep.* 10, 4819.
- Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences* 2016; 113: 5970-5975.
- Louca, S., Jacques, S.M.S., Pires, A.P.F., Leal, J.S., Srivastava, D.S., Parfrey, L.W., et al., 2016a. High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* 1, 0015.
- Louca, S., Parfrey, L.W., Doebeli, M., 2016b. Decoupling function and taxonomy in the global ocean microbiome. *Science* 353, 1272–1277.
- Louca, S., Polz, M.F., Mazel, F., Albright, M.B.N., Huber, J.A., O'Connor, M.I., et al., 2018. Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* 2, 936–943.
- Ma, Y., Zeng, Y., Jiao, N., Shi, Y., Hong, N., 2009. Vertical distribution and phylogenetic composition of bacteria in the Eastern Tropical North Pacific Ocean. *Microbiol Research* 164, 624–633.
- Madsen, E.L., 2011. Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* 22, 456–464.
- Mazumder, A., Taylor, W.D., McQueen, D.J., Lean, D.R.S., 1990. Effects of Fish and Plankton on Lake Temperature and Mixing Depth. *Science* 247, 312–315.
- Mori, A.S., Isbell, F., Seidl, R., 2018. beta-Diversity, Community Assembly, and Ecosystem Functioning. *Trends Ecol. Evol.* 33, 549–564.
- Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., et al., 2008. A general framework for the distance-decay of similarity in ecological communities. *Ecol. Lett.* 11, 904–917.
- Nekola, J.C., White, P.S., 2004. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* 26, 867–878.
- Nelson MB, Martiny AC, Martiny JBH. Global biogeography of microbial nitrogen-cycling traits in soil. *Proceedings of the National Academy of Sciences* 2016; 113: 8033-8040.
- Nevalainen, L., 2012. Distribution of benthic microcrustaceans along a water depth gradient in an Austrian Alpine lake – Sedimentary evidence for niche separation. *Limnologia* 42, 65–71.
- Nguyen, T.G., Huynh, N.T.H., 2023. Evaluating Surface Water Quality Using Indexes of Water Quality and Plankton Diversity. *Civil Engineering Journal* 9, 1187–1202.
- Nguyen, G.T., Thi Diem, M.L., Huynh, N.T.H., 2023. Pollution and Risk Level Assessment of Pollutants in Surface Water Bodies. *Civil Engineering Journal* 9, 1851–1862.
- Pajares, S., Merino-Ibarra, M., Macek, M., Alcocer, J., 2017. Vertical and seasonal distribution of picoplankton and functional nitrogen genes in a high-altitude warm-monomictic tropical lake. *Freshw. Biol.* 62, 1180–1193.
- Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., et al., 2019. Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecol.* 41, 23–33.
- Pearman, J.K., Thomson-Laing, G., Thomson-Laing, J., Thompson, L., Waters, S., Reyes, L., et al., 2022. The Role of Environmental Processes and Geographic Distance in Regulating Local and Regionally Abundant and Rare Bacterioplankton in Lakes. *Front. Microbiol.* 12.
- Penn, J.L., Weber, T., Chang, B.X., Deutsch, C., 2019. Microbial ecosystem dynamics drive fluctuating nitrogen loss in marine anoxic zones. *Proceedings of National Academy of Sciences of the United States of America* 116, 7220–7225.
- Peralta-Maraver, I., Galloway, J., Posselt, M., Arnon, S., Reiss, J., Lewandowski, J., et al., 2018. Environmental filtering and community delineation in the streambed ecotone. *Sci. Rep.* 8, 15871.
- Picazo, F., Vilmi, A., Aalto, J., Soininen, J., Casamayor, E.O., Liu, Y., et al., 2020. Climate mediates continental scale patterns of stream microbial functional diversity. *Microbiome* 8, 92.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596.
- Raes, J., Bork, P., 2008. Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* 6, 693–699.
- Raposeiro, P.M., Saez, A., Giralt, S., Costa, A.C., Gonçalves, V., 2018. Causes of spatial distribution of subfossil diatom and chironomid assemblages in surface sediments of a remote deep island lake. *Hydrobiologia* 815, 141–163.
- Ren, M., Wang, J., 2022. Phylogenetic divergence and adaptation of Nitrososphaeria across lake depths and freshwater ecosystems. *International Symposium on Microbial Ecology* 16, 1491–1501.
- Rojas-Jimenez, K., Araya-Lobo, A., Quesada-Perez, F., Akerman-Sanchez, J., Delgado-Duran, B., Ganzert, L., et al., 2021. Variation of bacterial communities along the vertical gradient in Lake Issyk-Kul, Kyrgyzstan. *Environmental Microbiology Reports* 13, 337–347.
- Sadeghi, J., Chaganti, S.R., Shahraki, A.H., Heath, D.D., 2021. Microbial community and abiotic effects on aquatic bacterial communities in north temperate lakes. *Sci. Total Environ.* 781, 146771.
- Schulhof, M.A., Allen, A.E., Allen, E.E., Mladenov, N., McCrow, J.P., Jones, N.T., et al., 2020. Sierra Nevada mountain lake microbial communities are structured by temperature, resources and geographic location. *Mol. Ecol.* 29, 2080–2093.
- Soininen, J., McDonald, R., Hillebrand, H., 2007. The distance decay of similarity in ecological communities. *Ecography* 30, 3–12.
- Song, W., Liu, J., Qin, W., Huang, J., Yu, X., Xu, M., et al., 2022. Functional Traits Resolve Mechanisms Governing the Assembly and Distribution of Nitrogen-Cycling Microbial Communities in the Global Ocean. *Mbio* 13, e03832.
- Song, H.K., Shi, Y., Yang, T., Chu, H., He, J.S., Kim, H., et al., 2019. Environmental filtering of bacterial functional diversity along an aridity gradient. *Sci. Rep.* 9, 866.
- Spellerberg, I.F., Fedor, P.J., 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the ‘Shannon–Wiener’ Index. *Glob. Ecol. Biogeogr.* 12, 177–179.
- Su, K., Wang, Q., Li, L., Cao, R., Xi, Y., 2022. Water quality assessment of Lugu Lake based on Nemerow pollution index method. *Sci. Rep.* 12.
- Sun, R., Wang, F., Hu, C., Liu, B., 2021. Metagenomics reveals taxon-specific responses of the nitrogen-cycling microbial community to long-term nitrogen fertilization. *Soil Biol. Biochem.* 156.
- Syarif Sukri, A., M S, Karama R, Nasrul, Talanipa R, Kadir A., et al., 2023. Utilization Management to Ensure Clean Water Sources in Coastal Areas. *Journal of Human, Earth, and Future* 4, 23–35.
- Taguchi, Y.H., Oono, Y., 2005. Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics* 21, 730–740.
- Tran, P.Q., Bachand, S.C., McIntyre, P.B., Kraemer, B.M., Vadeboncoeur, Y., Kimirei, I.A., et al., 2021. Depth-discrete metagenomics reveals the roles of microbes in biogeochemical cycling in the tropical freshwater Lake Tanganyika. *International Symposium on Microbial Ecology* 15, 1971–1986.
- Valentine, D.L., 2007. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat. Rev. Microbiol.* 5, 316–323.
- van Strien, A.J., Soldaat, L.L., Gregory, R.D., 2012. Desirable mathematical properties of indicators for biodiversity change. *Ecol. Ind.* 14, 202–208.
- Vellend, M., 2005. Species Diversity and Genetic Diversity: Parallel Processes and Correlated Patterns. *Am. Nat.* 166, 199–215.
- Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E.E., Van Der Heijden, M.G.A., 2019. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nat. Commun.* 10.
- Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al., 2016. Improved Bacterial 16S rRNA Gene (V4 and V4–5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* 1, e00009–e15.
- Wang, J., Soininen, J., He, J., Shen, J., 2012a. Phylogenetic clustering increases with elevation for microbes. *Environ. Microbiol. Rep.* 4, 217–226.
- Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J.C., et al., 2013. Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *International Symposium on Microbial Ecology* 7, 1310–1321.
- Wang, Q., Yang, X., Hamilton, P.B., Zhang, E., 2012b. Linking spatial distributions of sediment diatom assemblages with hydrological depth profiles in a plateau deep-water lake system of subtropical China. *Fottea* 12, 59–73.
- Wang, Y., Zhang, R., He, Z., Van Nostrand, J.D., Zheng, Q., Zhou, J., et al., 2017. Functional Gene Diversity and Metabolic Potential of the Microbial Community in an Estuary-Shelf Environment. *Front. Microbiol.* 8.
- Wani, A.K., Akhtar, N., Sher, F., Navarete, A.A., Américo-Pinheiro, J.H.P., 2022. Microbial adaptation to different environmental conditions: molecular perspective of evolved genetic and cellular systems. *Arch. Microbiol.* 204.
- Wei, S., Cui, H., Zhang, Y., Su, X., Dong, H., Chen, F., et al., 2019. Comparative evaluation of three archaeal primer pairs for exploring archaeal communities in deep-sea sediments and permafrost soils. *Extremophiles* 23, 747–757.
- Wessén, E., Söderström, M., Stenberg, M., Bru, D., Hellman, M., Welsh, A., et al., 2011. Spatial distribution of ammonia-oxidizing bacteria and archaea across a 44-hectare farm related to ecosystem functioning. *International Symposium on Microbial Ecology* 5, 1213–1225.
- Winter, C., Kerros, M.-E., Weinbauer, M.G. 2009. Seasonal changes of bacterial and archaeal communities in the dark ocean: Evidence from the Mediterranean Sea. *Limnol. Oceanogr.* 54, 160–170.
- Wu, K., Zhao, W., Wang, Q., Yang, X., Zhu, L., Shen, J., et al., 2019. The Relative Abundance of Benthic Bacterial Phyla Along a Water-Depth Gradient in a Plateau Lake: Physical, Chemical, and Biotic Drivers. *Front. Microbiol.* 10, 1521.
- Wu, K., Zhao, W., Li, M., Picazo, F., Soininen, J., Shen, J., et al., 2020. Taxonomic dependency of beta diversity components in benthic communities of bacteria, diatoms and chironomids along a water-depth gradient. *Sci. Total Environ.* 741, 140462.
- Xing, P., Tao, Y., Luo, J., Wang, L., Li, B., Li, H., et al., 2019. Stratification of microbiomes during the holomictic period of Lake Fuxian, an alpine monomictic lake. *Limnol. Oceanogr.* 65.

- Yang, Y., Gao, Y., Wang, S., Xu, D., Yu, H., Wu, L., et al., 2014. The microbial gene diversity along an elevation gradient of the Tibetan grassland. *International Symposium on Microbial Ecology* 8, 430–440.
- Zehr, J.P., Ward, B.B., 2002. Nitrogen Cycling in the Ocean: New Perspectives on Processes and Paradigms. *Appl. Environ. Microbiol.* 68, 1015–1024.
- Zhang, W., Chen, R., Meng, F., Yuan, H., Geng, M., Cheng, L., et al., 2021. Ecosystem functioning is linked to microbial evenness and community composition along depth gradient in a semiarid lake. *Ecol. Ind.* 132.
- Zhang, T., Wang, J., Zhou, S., Chen, Y., Li, D., 2023. Spatio-temporal dynamic diversity of bacterial alkaline phosphatase *phoD* gene and its environmental drivers in sediments during algal blooms: A case study of shallow Lake Taihu. *J. Environ. Manage.* 336.
- Zhao, W., Hu, A., Ni, Z., Wang, Q., Zhang, E., Yang, X., et al., 2019. Biodiversity patterns across taxonomic groups along a lake water-depth gradient: Effects of abiotic and biotic drivers. *Sci. Total Environ.* 686, 1262–1271.
- Zhao, B., Hu, Y., Yu, H., Chen, S., Xing, T., Guo, S., et al., 2023. A method for researching the eutrophication and N/P loads of plateau lakes: Lugu Lake as a case. *Sci. Total Environ.* 876, 162747.
- Zheng, B.X., Hao, X.L., Ding, K., Zhou, G.W., Chen, Q.L., Zhang, J.B., et al., 2017. Long-term nitrogen fertilization decreased the abundance of inorganic phosphate solubilizing bacteria in an alkaline soil. *Sci. Rep.* 7, 42284.
- Zheng, B., Zhu, Y., Sardans, J., Penuelas, J., Su, J., 2018. QMEC: a tool for high-throughput quantitative assessment of microbial functional potential in C, N, P, and S biogeochemical cycling. *Sci. China Life Sci.* 61, 1451–1462.