RESEARCH ARTICLE

Mutual environmental drivers of the community composition, functional attributes and co-occurrence patterns of bacterioplankton in the composite aquatic ecosystem of Taihu watershed in China

Mingkun Liu[1], Xue Han[1], Jun Tong[2], Huifeng Zhu[2] and Xiaohui Bai[1,*]

[1]State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, P.R.China and [2]Shanghai Municipal Water Supply Control & Monitoring Center, Shanghai, 200002, P.R.China

*Corresponding author: School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, P.R.China.
E-mail: xhbai@sjtu.edu.cn

One sentence summary: This study aimed to determine the environmental and ecological factors influencing the diversity, composition, functional attributes, ecological modules, and central species of planktonic prokaryotic communities in the anthropogenic and composite ecosystem comprising the Taihu Lake, Taipu River and Jinze Reservoir in the Taihu Watershed in China.

Editor: Martin W. Hahn

ABSTRACT

This study aimed to determine the environmental and ecological factors influencing the planktonic prokaryotic community profiles in the composite ecosystem comprising Taihu Lake, Taipu River and Jinze Reservoir in the Taihu Watershed in China. A total of 42 water samples were intermittently collected from different sites in 6 months across four seasons. Physicochemical characteristics of the ecosystem, bacterioplankton diversity and composition, the presence of co-occurrence patterns, and environmental predictors of ecological modules in the bacterioplankton network were determined. The central species played a more important role in regulating the structure and function of the bacterioplankton community and in responding to environmental contamination than the entire community. The relative abundance of the phylum *Proteobacteria* and the class *Betaproteobacteria* varied significantly between months and locations, which were identified as core functional taxa. A non-random co-occurrence pattern and function-driven modular structure were observed in the bacterioplankton co-occurrence network. Dissolved oxygen and ammonium nitrogen were the major and mutual environmental predictors of the bacterioplankton community composition, functional attributes and relative abundance of ecological modules. The results improve our understanding of the impact of anthropogenic contamination on bacterioplankton diversity and biogeochemical cycles and the formulation of strategies for bioremediation of the Taihu Watershed.

Keywords: lake–river–reservoir ecosystem; bacterioplankton; community composition; functional attributes; co-occurrence; ecological module

INTRODUCTION

Human survival and development are inseparable from river ecosystems (Nilsson *et al.* 2005). However, anthropogenic activities have resulted in the eutrophication of these aquatic habitats, thereby disrupting their biogeochemical cycles of nitrogen, phosphorous and other nutrients (Camargo and Alonso 2006; He *et al.* 2011) and degrading their water quality and biodiversity (Dudgeon *et al.* 2006). Being the most diverse and dominant organisms, microbial communities are essential in aquatic ecosystems, and they are directly or indirectly impacted by abiotic factors such as nutrient concentration (Ma *et al.* 2016a), temperature (Staley *et al.* 2015) and water residence time (Read *et al.* 2015). Their importance could be attributed to their ecological function in aquatic ecosystems in which they are the main players in ecological processes, including the fluxes of carbon and nutrients as well as the transformation of chemical pollutants (Falkowski, Fenchel and Delong 2008; Newton *et al.* 2011). Many studies have reported that environmental changes caused by human activities considerably affect the function of microbial communities. For example, crude oil pollution in the Gulf of Mexico enhanced the hydrocarbon biodegradation rates of microbial communities (Hazen *et al.* 2010). Therefore, understanding the environmental drivers of the variation in the composition and function of microbial communities is vital to the assessment of water quality and bioremediation of human-impacted freshwater habitats such as rivers and lakes.

The impacts of environmental change on the functional groups of microbial communities as well as the relationship between community structure and functional diversity have attracted extensive research. Nonetheless, recent studies present no consensus on the factors responsible for the microbial functional groups (Hu *et al.* 2018; Yang *et al.* 2019) and on the relationship between microbial community composition and function in aquatic ecosystems (Freimann *et al.* 2013; Staley *et al.* 2014). Hence, the effects of anthropogenic contamination on the taxonomic and functional attributes of microbial communities in aquatic ecosystems should be further investigated.

Microorganisms within a specific habitat form complex networks through microbe–microbe interactions (Faust and Raes 2012; Ghoul and Mitri 2016), which consequently influence the composition and role of microbial communities and even the function of the entire ecosystem (Deng *et al.* 2016; Ghoul and Mitri 2016, Li *et al.* 2017). Recent studies have used co-occurrence network analysis to investigate the potential microbe–microbe interactions in oceans (Cram *et al.* 2015), rivers (Hu *et al.* 2017) and soils (Jiao *et al.* 2016). These studies suggested that microbial communities generally have non-random co-occurrence patterns and modular structures, but they differed in determining the modular structures (Jiao *et al.* 2016; Hu *et al.* 2017). In addition, microbial taxa strongly co-occur within co-occurrence networks, often called ecological clusters or modules (De Menezes *et al.* 2015), which have multiple implications for the maintenance of soil fertility, decomposition and ecological services in terrestrial environments (Delgado-Baquerizo *et al.* 2018; Shi *et al.* 2018). Despite the importance of ecological modules in understanding the aquatic microbiome, their functional profiles and responses to anthropogenic interference remain poorly understood. More importantly, central or keystone species have been known to be significant in maintaining the structure and function of an entire microbial community (Ma *et al.* 2016b). However, the taxonomic and functional structure of central species and their responses to environmental variables in aquatic ecosystems remain largely unexplored. To help fill the aforementioned gaps in research, this study sought to investigate the environmental drivers of the differences in diversity, composition, functional attributes, ecological modules and central species of the planktonic prokaryotic communities in the composite aquatic ecosystem of China's Taihu Lake, Taipu River and Jinze Reservoir, which have been under long-term anthropogenic pressure.

MATERIALS AND METHODS

Study area, sample collection and physicochemical analysis

The Taihu Lake supplies the water running through the Taipu River located in the Yangtze River Delta in the eastern part of the Taihu Watershed in China. Having a total length of 57.6 km, the Taipu River runs through a densely populated and highly industrialized region before it enters the Jinze Reservoir. The river is a protected source of drinking water for Shanghai. Water from the river undergoes ecological purification in the Jinze Reservoir before it is transported as raw water to the water treatment plant. The river also plays an important role in flood control, irrigation and shipping in the region, but it is heavily polluted with agricultural contaminants from non-point sources, micropollutants from tributaries and wastes from ships (Yao *et al.* 2015), subsequently increasing the demand for water treatment in the Jinze Reservoir.

In the lake–river–reservoir study area, we established seven sampling points: the Taipu Gate (TP) standing in one of the outlets of the Taihu Lake, the Pingwang Bridge (PW), the Luxu Bridge (LX) and the reservoir entrance (RK), which are all situated in the Taipu River. The reservoir central, ecological island and reservoir exit are located inside the Jinze Reservoir (Fig. 1). For 6 intermittent months (between October of 2017 and July of 2018), we completed the collection of samples following the guidance on 'Water quality–Guidance on sampling techniques (HJ 494−2009)'. More specifically, we set three points about 10 m apart on the same cross-section, surface water (about 50 cm) was sampled at each point and was manually pooled into one sample for further analysis. Water samples were transported to the laboratory within 12 h, and stored at 4°C for at most 1 week prior to chemical analysis. Water temperature (WT), pH and dissolved oxygen (DO) were measured *in situ* using a YSI 6920 water quality monitor (YSI Inc., Yellow Springs, Ohio, USA). Total organic carbon (TOC) was estimated using a TOC analyzer (TOC-VPH, Shimadzu, Japan). Analytical methods for other physicochemical variables are listed in Table S1, see online supplementary material. For the bacterioplankton analysis of the water samples, ∼500–1000 mL of the samples were filtered through a 5-μm Durapore membrane filter (Xinya, China) to remove large particulate matter (the volumes for pretreatment depend on the amount of biomass in the water samples) and then were filtered again using 0.22-μm Sterivex-GP filters (Millipore, Bedford, MA) to collect the bacteria. All samples were stored at −80°C until further analysis.

DNA extraction and sequencing

Genomic DNA was extracted from each membrane through the cetyl trimethylammonium bromide (CTAB) extraction protocol. An 800 μL volume of CTAB lysis buffer [100 mM Tris-HCl (pH 8.0), 100 mM sodium-EDTA (pH 8.0), 100 mM phosphate buffer (pH 8.0), 1.5 M NaCl, 1% cetyltrimethylammonium bromide] and ∼0.3 g of glass beads were vibrated for 10 min and put into a grinder (30 Hz, 2 min); 20 μL of proteinase K
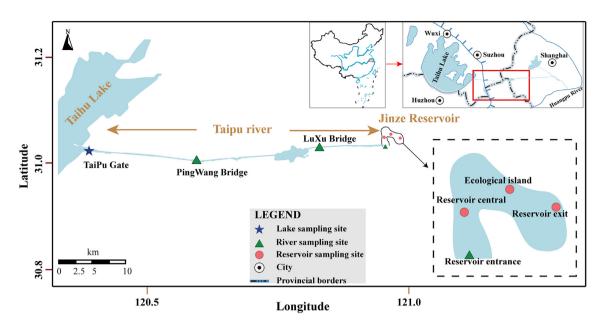
Figure 1. Geographic location of the sampling sites across the three major locations: Taihu Lake, Taipu River and Jinze Reservoir.

was added and incubated at 65°C for 1 h, followed by 800 μL of phenol:chloroform:isoamyl alcohol = 25: 24: 1 (pH > 7.8), and centrifuged at 12 000 rpm for 15 min. An equal volume of chloroform:isoamyl alcohol = 24:1 was added and the mixture was again centrifuged at 12 000 rpm for 15 min. NaAc (1/10 volume, pH = 5.2) was then added along with a pre-cooled equal volume of isopropanol, stored at −20°C for 2 h, and then centrifuged at 12 000 rpm for 10 min. After the supernatant was discarded, 200 μL of 70% (vol/vol) ethanol was added to wash the DNA/RNA pellet. The V3–V4 region of the bacterioplankton 16S rRNA gene was amplified through PCR at 95°C for 2 min, followed by 25 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 30 s and a final extension of 72°C for 5 min, using primers 338F (5′-ACTCCTACGGGAGGCAGCA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′) (Dennis *et al.* 2013). A total of three replicates were performed for each sample to avoid biased amplification. The PCR reactions were performed in triplicate using 20-μL mixtures containing 4 μL of 5 × FastPfu buffer, 2 μL of 2.5 mM dNTPs, 0.8 μL of each primer (5 μM), 0.4 μL of FastPfu polymerase and 10 ng of template DNA. A negative control containing all the reagents except the template DNA was included with each set of reaction mixtures. Three replicates of each sample were then pooled together and purified using AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, USA) according to the manufacturer's instructions and were quantified using QuantiFluor™-ST (Promega, USA). All PCR products were sequenced using Illumina MiSeq platform (Shanghai Majorbio Bio- pharm Technology Co., Ltd., China). The raw reads were deposited into the NCBI Sequence Read Archive database (Accession Number: PRJNA623476).

Data processing and functional annotation

Raw fastq files were demultiplexed and were quality-filtered using QIIME (version 1.30) with the following criteria: (i) the 250-bp reads were truncated at any site receiving an average quality score of Q20 over a 50-bp sliding window, and truncated reads shorter than 50 bp were discarded; (ii) the reads containing more than two nucleotide mismatches during primer matching or ambiguous characters were removed; and (iii) only the sequences that overlap by more than 10 bp were assembled according to their overlap sequence. The reads that could not be assembled were discarded. In order to avoid bias due to sampling depth, subsequent analyses of bacterioplankton community were completed after rarefaction to the same sampling depth (26 319 reads per sample) using QIIME v1.9.0. Operational taxonomic units (OTUs) were clustered with 97% similarity cut-off using UPARSE (version 7.1), and chimeric sequences were identified and removed using UCHIME. The phylogenetic affiliation of each 16S rRNA gene sequence was analyzed through the Ribosomal Database Project (RDP) Classifier (Release11.3) against the Silva SILVA (Release 119) 16S rRNA database with a confidence threshold of 70% (Wang *et al.* 2007).

Statistical analyses

Prior to the analyses, all environmental variables were log10 (x+1) transformed to improve normality, and the sequence abundance of OTUs was transformed by Hellinger distance. The extent of dissimilarity of the environmental variables in different locations was tested through the multi-response permutation procedure (MRPP), analysis of similarity (ANOSIM) and analysis of distance matrices (ADONIS). All these analyses were performed using the R 'vegan' package.

To assess microbial diversity and abundance, the alpha values of OTU richness and Shannon–Wiener index were calculated using QIIME (http://qiime.org/index.html) with 26 319 reads per sample, which was the minimum number of sequences required to normalize differences in sequencing depth. The differences in alpha diversity across multiple months and locations were determined through Tukey's honestly significant difference (HSD) test. The correlation between taxonomic and environmental dissimilarity was analyzed through the Mantel test.

The relationship between bacterioplankton community composition and environmental factors was investigated through constrained analysis of principal coordinates (CAP) based on Bray–Curtis distance using the 'capscale' function in the R package 'Vegan'. Before the CAP analysis, the variance inflation factor (VIF) was measured to evaluate the collinearity

of different variables (Wang *et al.* 2016a), and the environmental variables with *P*-value >0.05 or with VIF > 10 were removed from the succeeding analysis. VIF values of WT, biological oxygen demand (BOD), total nitrogen (TN), arsenic (As) and sulfate (SO4^{2-}) >10 were removed. The relatively more important variables were chosen through forward selection (Blanchet, Legendre and Borcard 2008). The associations between normalized mean values for significant environmental conditions (X variables) and major bacteria genera (Y variables) were modeled through partial least-squares regression (PLSR) analysis using the "Partial Least Squares(pls)" package. Through linear discriminant analysis (LDA) coupled with effect size measurements (LEfSe) analysis (using all-against-all comparisons; Kruskal–Wallis sum-rank test, *P* < 0.05), the strong association between most major bacteria genera (LDA score ≥ 3.0) and sample months were identified among different samples (Segata and Huttenhower 2011)

The OTUs were annotated through Functional Annotation of Prokaryotic Taxa (FAPROTAX) to determine the functional groups for the analysis of ecological implications (Louca, Parfrey and Doebeli 2016). We summarized the OTUs assigned with main functional annotations and their abundance in our samples. The association between functional groups and bacterioplankton population was identified from phylum to class. The correlation between environmental dissimilarity and the functional structures of bacterioplankton communities was determined through the Mantel test. The relationship between environmental factors and the functional structures of bacterioplankton communities was evaluated through CAP analysis based on Bray–Curtis distance.

The co-occurrence patterns of bacterioplankton taxa were explored through network analysis using igraph packages (Csardi and Nepusz 2006). The co-occurrence of OTUs was checked in at least 20% of the samples and in at least 42 reads. The correlation between two OTUs with corresponding Spearman's correlation coefficient (*r*) > 0.6 and *P* < 0.01 was considered statistically significant (Barberan *et al.* 2012). The ecological modules were identified through the Louvain algorithm, and network visualization and modular analysis were conducted using Gephi (Bastian, Heymann and Jacomy 2009). The relative abundance of each module was computed by averaging the standardized relative abundances (Z-score) of the taxa belonging to each module (Delgado-Baquerizo *et al.* 2018). The ecological function of the modules in the network was then analyzed through FAPROTAX. A total of 10 000 Erdös–Réyni random networks were generated using the igraph package to compare the data with the topology of the real network, with each edge having the same probability of being assigned to any node (Erdös and Rényi 1960).The nodes with having a high connect degree (>100) and low betweenness centrality values (<5000) in a co-occurrence network were defined as central species, similar to the designation of keystone species in other studies (Berry and Widder 2014; Ma *et al.* 2016b). We analyzed the relative abundance of the number of OTUs annotated with specific function within the module or central species compared with total number of OTUs annotated with same function in total microbiome. The environmental predictors of the ecological modules in the bacterioplankton network were identified through random forest analysis (Breiman 2001; Trivedi *et al.* 2016). The random forest model determined the importance of each predictor variable by evaluating the percentage increases in the mean squared error (MSE) of the variables, so that higher MSE percentage values mean more important variables (Liaw and Wiener 2002). The

significance of the models and cross-validated R^2 values was assessed using the A3 package with 5000 permutations of the response variable. Similarly, the significance of each predictor for the response variables was evaluated using the rfPermute package.

RESULTS

Environmental dissimilarity of the lake, river and reservoir

All the physicochemical variables are listed in Table S2, see online supplementary material. A principal component analysis (PCA) plot of environmental variables revealed a clear spatio-temporal pattern of environmental variables in the study area (Fig. 2). The dissimilarity tests based on the Euclidean distance indicate that a significant environmental variation exists among the 6 different months and the three major locations, more prominently so between the 6 different months (Table S3, see online supplementary material). Moreover, we noticed that some environmental variables significantly varied between months or locations (Table S4, see online supplementary material). For example, the concentrations of DO and ammonium nitrogen (NH$_4$-N) were significantly higher (Tukey, *P*-value < 0.05) in the February samples than in others samples (Fig. S1, see online supplementary material), and the concentrations of permanganate index (COD$_{Mn}$), ammonium nitrogen (NH$_4$-N), TOC, petroleum (Pet), stibium (Sb), chloride (Cl) and sulfate (SO$_4$$^{2-}$) were significantly lower (Tukey, *P*-value < 0.05) in the lake samples than in the river and reservoir samples (Fig. S2, see online supplementary material).

Variation in the diversity of bacterioplankton communities

The high-throughput sequencing generated a total number of 2 117 692 sequences, and the number of filtered sequences per sample ranged from 26 319 to 59 113 (Fig. S3A, see online supplementary material), which were grouped into 2328 OTUs at a cutoff of 97% identity after sub-sampling of 26 319 reads per sample. Rarefaction curves indicated that most samples reached or almost reached the saturation stage (Fig. S3B, see online supplementary material), suggesting that the majority of biodiversity was recovered based on current sequencing depth. We calculated the alpha diversity indices including the OTU richness as well as the Shannon–Wiener index. The OTU richness was significantly varied among months and locations (Tukey, *P*-value < 0.05) (Fig. S4A, see online supplementary material). The Shannon–Wiener index did not significantly differ among months or locations (Tukey, *P*-value > 0.05) (Fig. S4B, see online supplementary material). In addition, we assessed the correlation between alpha diversity indices and physicochemical variables (Table S5, see online supplementary material). The OTU richness and Shannon–Wiener index were significantly and negatively correlated with the DO, but they were positively correlated with the levels of nitrate nitrogen (NO$_3$-N) and TN. Significant difference in community composition between month pairs was identified, with a plot of NMDS showing that most samples belonging to the same month were grouped together (Fig. S5A, see online supplementary material). Further results of dissimilarity tests confirmed that the variation of OTU-level community composition in different months was more pro-
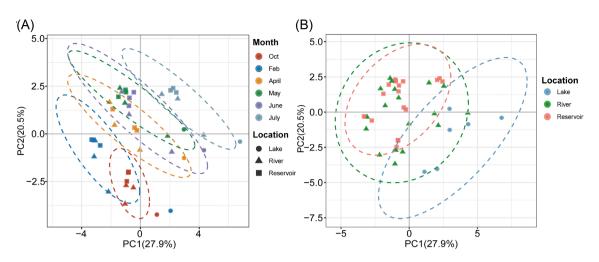
Figure 2. Spatio-temporal pattern of environmental variables in the study area. Principal component analysis with environmental variables in the 6 months (**A**) and in the three locations (**B**). 95% confidence ellipses are shown around the samples grouped based on months and locations.

nounced than in different locations (Table S3). Despite this variation, 647 OTUs (21.79% of total OTUs) were distributed among 6 months (Fig. S6, see online supplementary material). The number of unique OTUs was higher in February (6.88%, 88/1279) than in other months (Fig. S6).

Microbial community composition varied at phylum and genus levels for all 6 months (Figs S7 and S8, see online supplementary material). The dominant phyla of the whole river were *Actinobacteria* (35.98%), *Proteobacteria* (32.42%), *Cyanobacteria* (14.37%) and *Bacteroidetes* (9.12%) (Fig. S7 and Table S5). The percentage of other phyla was <2.0%. *Betaproteobacteria* was the most abundant class of *Proteobacteria*, comprising 23.00% of the total sequences (Table S6), followed by *Alphaproteobacteria* (5.51%), *Gammaproteobacteria* (2.87%), *Deltaproteobacteria* (1.90%) and *Epsilonproteobacteria* (0.3%). Table S6, see online supplementary material, shows the relative abundance of major bacterial phyla in all samples and the significance of differences between months and locations based on one-way analysis of variance (ANOVA). The relative abundance of *Cyanobacteria*, *Chloroflexi* and *Planctomycetes* had the largest dynamics among all samples, while that of *Actinobacteria*, *Bacteroidetes* and *Proteobacteria* had the lowest. ANOVA showed that the abundance of most bacterial phyla differed significantly between months (P-value < 0.05), except for *Cyanobacteria* (P-value = 0.715) and *Verrucomicrobia* (P-value = 0.221). The relative abundance of most bacterial phyla did not differ significantly between locations, except for *Proteobacteria* (P-value = 0.017) and *Cyanobacteria* (P-value = 0.032). The relative abundance of most bacterial phyla did not differ significantly between locations, except for *Proteobacteria* (P-value = 0.036). Furthermore, *Betaproteobacteria* was the only class of *Proteobacteria* that differed significantly between months (P-value = 0.000) and locations (P-value = 0.035). The major bacteria genera (relative abundance > 0.50%) were *hgcI_clade* (22.90%), *CL500–29_marine_group* (8.69%), *norank_c_Cyanobacteria* (8.41%), *Malikia* (3.87%) and *Limnohabitans* (3.80%) (Fig. S8A). We used the LEfSe analysis to search for the statistically different bacteria genera among the samples from the 6 months (Fig. S8B). The results showed a strong association between most (88.46%, 23/26) major bacteria genera and sample months (Fig. S8B). For example, the abundance of *Malikia*, *Limnohabitans*, *Flavobacterium*, *Rhodoferax*, *Microcystis* and *norank_f_env.OPS_17* were significantly enriched from October to February.

Microbial community associated with environmental factors

The constrained analysis of principal coordinates (CAP) based on Bray–Curtis distance and permutation testing revealed that the selected environmental variables such as DO, NH_4-N, fluorine (F), acrylonitrile butadiene styrene (ABS), NO_3-N, COD_{Mn}, TOC and pH substantially affected the bacterioplankton community (Fig. 3A), and that they altogether substantiated the 61.28% compositional variation in the total microbiome (Table S7, see online supplementary material). The first axis was positively correlated with the DO, NH_4-N, NO_3-N, COD_{Mn}, TOC and pH but was negatively correlated with the ABS and F. The second axis was positively correlated with the ABS, NO_3-N, and pH but was negatively correlated with the DO, NH_4-N, F, COD_{Mn} and TOC. Among the selected variables, the DO and NH_4-N showed the strongest effect on the composition of the total microbiome. The Mantel test showed that the selected environmental variables significantly affected the bacterioplankton community structure (r = 0.370 and P-value < 0.001).

We modeled the covariance between the environmental factors and bacterioplankton taxa at genus levels through PLSR modeling and subsequently identified the major bacteria genera that may be affected by the environmental variables (Fig. 3B). PLSR modeling was initially performed among the genus level with relative abundance >0.5% and was used to select the taxa explaining the >40% variance in the first two components. The DO was strongly and positively associated with *unclassified_f_Rhodobacteraceae* and *Sediminibacterium*. The NH_4-N was significantly and positively associated with *Limnohabitans* and *Rhodoferax*. The COD_{Mn} was strongly and positively associated with *unclassified_f_Comamonadaceae*. The TOC was strongly and positively associated with *unclassified_f_Sporichthyaceae*. Spearman's correlation coefficients generated among these variables and factors revealed significant correlations (P-value < 0.05) (Table S8B, see online supplementary material).

Functional groups of the bacterioplankton communities associated with the environmental variables

A total of 45 bacterioplankton functional groups were identified from our samples using FAPROTAX (Louca, Parfrey and Doebeli 2016). The Mantel test showed that their functional structure
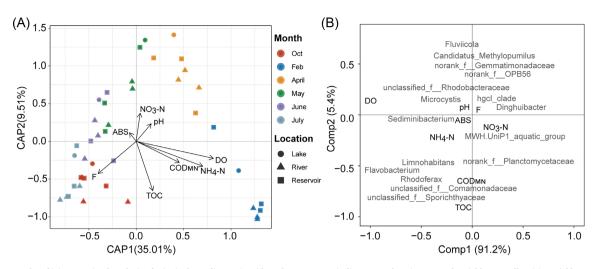
Figure 3. Results of (**A**) constrained analysis of principal coordinates (CAP) based on Bray–Curtis distance and environmental variables as well as (**B**) partial least squares regression (PLSR) correlation of the selected environmental variables (black text) and major bacterial genera (gray text).

was highly correlated with their taxonomic structure ($r = 0.366$, P-value < 0.001) and the selected environmental variables ($r = 0.161$, P-value $= 0.009$).

Furthermore, a total of 26 main functional groups with high relative abundance $>0.1\%$ and their co-occurrence in 100% of our samples were further visualized in a heat map plot (Fig. 4). The main functional groups were clustered into four clusters based on Pearson's correlation coefficient. Cluster I (1.56%) and Cluster II (2.00%) were principally related to the degradation of some nutrients and the nitrogen cycle. Cluster III (42.33%) included the bacterioplankton that are chemoheterotrophic (10.75%), aerobic and chemoheterotrophic (7.73%), methylotrophic (2.70%) or photoheterotrophic (1.01%). Cluster III also consisted of animal parasites or symbionts (5.37%), human pathogens causing pneumonia (4.98%), all human pathogens (5.35%), as well as the bacterioplankton involved in methanol oxidation (2.48%), fermentation (1.47%) or dark hydrogen oxidation (0.49%). Cluster IV (53.77%) was associated with the autotrophic bacterioplankton including the cyanobacteria (13.19%) and those that are phototrophic (14.20%), oxygenic and photoautotrophic (13.19%) or photoautotrophic (13.19%). Moreover, the relative abundance of Cluster III was higher in February than in other months. Cluster IV was higher in the reservoir than in other locations, and this may be related to seasonal switching and the physiological adaptations of the bacterioplankton communities.

The association between bacterioplankton taxonomy and functional groups showed that Proteobacteria (especially *Betaproteobacteria*) were the predominant bacterioplankton taxa associated with most functional groups in the samples (Figs S9 and S10 and Table S9, see online supplementary material). Some bacterioplankton taxa possessed specific functions. *Cyanobacteria* was associated with phototrophy, oxygenic photoautotrophy and photoautotrophy, whereas class *Alphaproteobacteria* was linked with ureolysis and intracellular parasitism, and class *Gammaproteobacteria* was involved in hydrocarbon degradation and methanotrophy (Figs S9 and S10 and Table S9).

CAP analysis demonstrated that NH_4-N, F, DO, pH and Sb significantly affected the functional groups, and that NH_4-N was the most influential environmental variable (Table S7). By associating the functional groups with bacterioplankton taxonomy, we found that *Proteobacteria* (especially *Betaproteobacteria*) contributed the largest fraction to most functional groups (Figs S9

and S10). Therefore, the greater influence of NH_4-N on the functional groups may have resulted from the stronger correlation between NH_4-N and *Betaproteobacteria*. The result of Spearman's correlation test among the selected environmental variables and functional groups is shown in Fig. S11, see online supplementary material.

Co-occurrence network analysis

The riverine bacterioplankton taxa were grouped into four ecological modules comprised of groups strongly co-occurring with one another (Fig. S12A, see online supplementary material). We found significant relationships between NH_4-N and the relative abundance of the dominant ecological modules (Fig. S12B, see online supplementary material). Furthermore, we explored the functional profiles of the ecological modules through FAPROTAX analysis, and the results showed that different modules perform various functions (Fig. S13, see online supplementary material). Module I was attributed to methanotrophy, aerobic ammonia oxidation, nitrification and ureolysis, whereas Module II was most predominant in the functional groups related to metabolism and human health, and Module III was associated with chlorate reduction, methanol oxidation and methylotrophy.

We compared the topological properties of the riverine network with the identically sized Erdös–Réyni random network to describe the co-occurrence pattern of the inter-relationships among nodes. The average path length (APL) was 2.611 edges, the clustering coefficient (CC) was 0.563 and the modularity index (MD) was 1.482. The structural properties of the real-world network were greater than those of the random networks (1.906, 0.0968 and 0.0787 for APL, CC and MD, respectively). Therefore, the bacterioplankton network consisted of highly connected OTUs and formed a small-world topology.

Central species are extensively connected taxa that individually or collectively exert, in a guild, a considerable influence on microbiome structure and function irrespective of their abundance across space and time (Banerjee *et al.* 2018). In this study, they were identified as having a high connect degree (>100) and low betweenness centrality values (<5000) in co-occurrence networks. Based on these criteria, a total of 317 OTUs with average relative abundances ranging from 0.0041 to $\sim8.31\%$ accounted
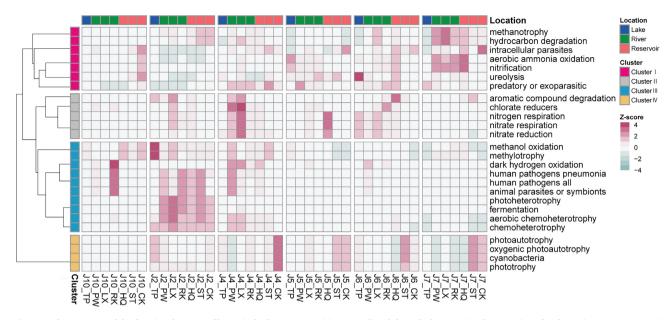
Figure 4. The structure of the functional groups of bacterioplankton communities as predicted through the Functional Annotation of Prokaryotic Taxa (FAPROTAX) analysis. Each row and column of the heatmap corresponds to a single function group and sample, respectively. The row data for each column was Z-score normalized.

for 42.89% of the total reads and were subsequently recognized as the central species (Table S10, see online supplementary material). Taxonomic classification revealed that *Frankiales*, *Burkholderiales*, *Acidimicrobiales* and *Sphingomonadales* were the dominant orders of bacterioplankton in the microbiome, the network and the central species (Fig. S14, see online supplementary material). In addition, the composition of the central species showed significantly corresponding relationships with environmental dissimilarity (Mantel test, $r = 0.374$, $P < 0.001$). Moreover, the selected environmental variables altogether validated the 72.94% composition variation in the central species, the >61.28% total microbiome and the 49.40% bacterioplankton functional groups (Table S7). The Mantel test also revealed that the central species have the strongest correlation with the environmental variables attributed to the central species, total microbiome and functional groups (Table S11, see online supplementary material). The correlation of the central species was also more rigid with the total microbiome and the functional groups than with the environmental variables (Table S11).

We determined the potential environmental predictors of the ecological modules through random forest analysis. DO and NH_4-N were the major predictors of the relative abundance of ecological modules (Fig. 5). As expected, our results revealed that the other environmental variables such as S^{2-}, F and TOC were also useful as predictors, but their relative importance was highly module-dependent (Fig. 5). The result of Spearman's correlation test between the selected environmental variables and the relative abundance of the ecological modules is shown in Table S12, see online supplementary material.

DISCUSSION

Environmental drivers of the diversity and composition of bacterioplankton communities

We observed a clear spatio-temporal environment dissimilarity in the study area (Fig. 1 and Table S3), similar to other aquatic environments disturbed by anthropogenic interference (Hu *et al.* 2017; Wang *et al.* 2019). Our study showed that higher nitrogen

load and lower DO supported the greater alpha diversity of bacterioplankton communities particularly in the Taipu River, compared with the Taihu Lake (Fig. S3 and Table S4). The significant correlations among the DO, TN and alpha diversity in our study could be attributed to sewage inflow (Yao *et al.* 2015), which could augment the nitrogen and other nutrients in the water (Table S2), assist in the growth of some microorganisms (Chrzanowski, Sterner and Elser 1995) and consume the DO (Table S2).

Taihu Lake, Taipu River and Jinze Reservoir represent three aquatic ecosystems with distinctive environmental characteristic and anthropogenic stressors. However, the temporal variation in bacterial communities was more pronounced than that of spatial variation. This result was consistent with previous findings in Jiulong River (Hu *et al.* 2017) and Ganjiang River (Wang *et al.* 2016b), which could be associated with the temporal environment dissimilarity. With respect to the regulation of community composition, several dominant environmental factors were identified, including DO, NH_4-N, F, ABS, NO_3-N, TOC, COD_{Mn} and pH. In addition, these environmental factors can also impact the distribution of dominant bacterial phyla. For example, the relative abundance of *Proteobacteria* and *Cyanobacteria* and their major environmental drivers, NH_4-N and pH, has the same distribution between months and locations. Our findings indicate that a deterministic process of environmental filtering shapes the diversity and composition of bacterioplankton communities in aquatic environments.

Functional profiles of the bacterioplankton communities and their environmental drivers

Previous studies have suggested that the functional profiles of microbial communities deepen our understanding of the influences of environmental dissimilarity on ecosystem processes in aquatic environments (Raes *et al.* 2011; Wang *et al.* 2011; Staley *et al.* 2014; Meziti *et al.* 2016). In this study, we analyzed the functional profiles of bacterioplankton communities through FAPROTAX analysis, and found 26 main functional groups that could be clustered into 4, based on the Pearson's
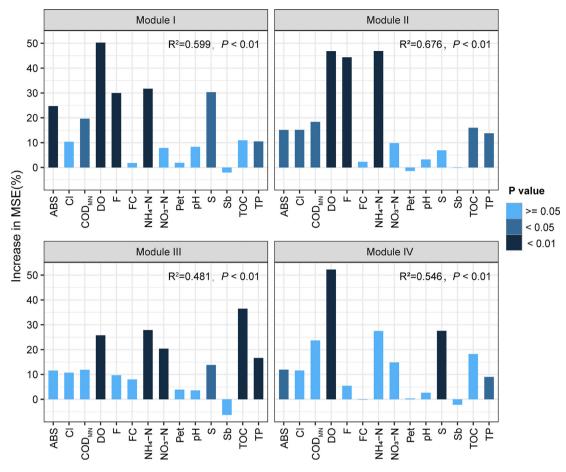
Figure 5. Results of the random forest analysis identifying the significant environmental predictors of the relative abundance of ecological modules I–IV of the bacterioplankton co-occurrence network.

correlation coefficient (Fig. 4). Cluster IV (53.77%), with highest relative abundance among the bacterioplankton communities, largely executed by the phylum *Cyanobacteria* (Fig. S9), substantiating the indispensability of the phylum in executing the functions of microbial communities. Chemoheterotrophy and aerobic chemoheterotrophy were the most dominant functional groups in Cluster III. Chemoheterotrophic bacteria usually serve as decomposers that help in alleviating crude oil pollution and recycling organic matter (Kämpfer *et al.* 1993). Moreover, our study showed that bacterioplankton could become sources of pathogens that may cause diseases and threaten human health. Other functions in Cluster I and II were mostly related to the nitrogen cycle.

We revealed potential composition-related variation of community function in our samples, with NH_4-N being the most influential environmental variable (Table S7). Specifically, the majority of Clusters I, II and III showed higher abundance in rivers (Fig. 4), with Proteobacteria including *Betaproteobacteria* and *Gammaproteobacteria* contributing chiefly to those functional groups. This was consistent with previous studies in polluted rivers (Yang *et al.* 2019), where coupling of community composition and functions was detected. Furthermore, CAP analysis also indicated that composition and functional attributes were affected by similar environmental factors (especially DO and NH_4-N). Hence, the composition and functional attributes of the bacterioplankton communities consistently responded to the changes in their environment.

Ecological drivers of the co-occurrence network in the aquatic ecosystem

Co-occurrence patterns reflect the nature of microbial interactions (Barberan *et al.* 2012). The comparison of our network with an identically sized random network showed a non-random co-occurrence pattern of bacterioplankton communities in the lake–river–reservoir environment, affirming previous studies on the importance of deterministic processes in the structure of bacterioplankton communities (Jiao *et al.* 2016; Hu *et al.* 2017).

Despite the importance of ecological modules within the network for ecosystem function, studies on the functional profiles of ecological modules as well as the relationship between environmental variables and the relative abundance of ecological modules in aquatic ecosystems are still inadequate. In this study, the OTU nodes of the co-occurrence network were grouped into four major ecological modules (Fig. S12A). The results of FAPROTAX analysis demonstrated that the functional groups of the bacterioplankton communities were dominated by specific ecological modules (Fig. S13). Thus, the robust co-occurrence correlations in the bacterioplankton ecological network indicate strong ecological linkages among the communities. For instance, module I consisted of some aerobic ammonia-oxidizing bacteria belonging to the order Nitrosomonadales and genus *Nitrospira* and the methanotrophs of Verrucomicrobia and order Methylococcales (Table S10). Aerobic ammonia-oxidizing bacteria oxidize ammonia to hydroxylamine using

ammonia monooxygenase (AMO) (Hooper *et al.* 1997), whereas the methanotrophs of Gammaproteobacteria oxidize methane using methane monooxygenase (MMO) (Hakemian and Rosenzweig 2007), but other studies showed that AMO-associated bacteria could oxidize methane (Arp and Stein 2003), and that MMO-related bacteria could oxidize ammonia to hydroxylamine (Stein and Klotz 2011). Some ammonia-oxidizing bacteria, such as *Nitrosospira* sp., can also degrade organic nitrogen compounds through ureolysis to produce additional ammonia (Burton and Prosser 2001). The genomes of the anaerobic ammonium-oxidizing bacteria and the methanotrophs of *Proteobacteria* and *Verrucomicrobia* encode hydroxylamine oxidoreductase (HAO)-like proteins that likely oxidize hydroxylamine to nitrite (Ren, Roy and Knowles 2000; Nyerges and Stein 2009; Maalcke *et al.* 2014).Thus, these previous studies imply that the anaerobic ammonium-oxidizing bacteria and the methanotrophs in our samples are also important in nitrification, and that our findings validated the non-random and function-driven co-occurrence patterns that may exist in aquatic bacterioplankton communities.

Exploring the relationships between microbial ecological modules and environmental variables may significantly improve our understanding of the interactions between microbial communities and abiotic factors by highlighting the most crucial characteristics of the communities (De Menezes *et al.* 2015). The results of the random forest analysis and correlation test showed that NH_4-N was the key environmental driver of the relative abundance of ecological modules within the co-occurrence network. Hence, the variation of NH_4-N concentration may have led to the drastic changes in the co-occurrence network of the bacterioplankton communities. For instance, the relative abundance of module II increased with NH_4-N concentration. Module II was predominant in functional groups related to the degradation of organic matter, fermentation, aerobic chemoheterotrophy and chemoheterotrophy (Fig. S13), which are the more advantageous processes when nitrogen concentration increases. As expected, other environmental predictors were strongly associated with the relative abundance of the bacterioplankton ecological clusters (Fig. 5), similar to the findings in many previous studies (Delgado-Baquerizo *et al.* 2018; Liu *et al.* 2018).

Our results showed that most compositional variation in the central species could be attributed to the selected environmental variables (Table S7), and that they have significant correlations with the total microbiome and the bacterioplankton functional groups (Table S11). These findings further suggest that impacts of environmental dissimilarity on bacterioplankton community composition and functional attributes may be mediated by central species. We also found that almost all dominant orders of the central species were significantly correlated with the DO, NH_4-N and F (Table S13), indicating that central species rapidly respond to environmental changes. Central species are considered to be the drivers of microbial communities (Banerjee *et al.* 2018), exerting their influence on the communities by controlling ecosystem processes. In this study, the central species were the dominant bacterioplankton taxa of photoheterotrophy, methanotrophy, nitrification, aerobic ammonia oxidation, hydrocarbon degradation and fermentation (Fig. S13). The influence of central species on community ecosystem functioning may be more pronounced if a narrow process is used instead of a broad one (Banerjee *et al.* 2018). A narrow process, such as ammonia oxidation, consists of a single step done by a small group of specialized microorganisms. An example of a broad process is organic matter decomposition.

Our study area is located in Taihu Watershed, which is threatened by multiple anthropogenic stressors (especially sewage and cargo ship) with increasing frequency (Yao *et al.* 2015). Historical data (Fig. S15) and our research results showed that long-term nutrient loading (especially nitrogen enrichment) resulted in significant changes of environmental variables along the river and has certain time characteristics. On the one hand, our findings revealed the impacts of environmental variations on bacterioplankton communities. On the other hand, we found significant correlations between the relative abundance of major bacteria genera and environmental variables, which formed the reference for developing microbial indicators for water quality assessment in aquctic environment.

CONCLUSIONS

Our findings suggest that environmental quality can extensively impact bacterioplankton diversity, community composition and function. Moreover, four ecological modules constitute the bacterioplankton co-occurrence network in the Taipu River; the main ecological modules have specific ecological functions. Our results further suggest that DO and NH_4-N are the critical and mutual environmental factors that drive the succession of bacterioplankton communities and that modify the relative abundance of their ecological modules in a composite aquatic ecosystem, such as that of Taihu Lake, Taipu River and the Jinze Reservoir. Apart from these two factors, F can also substantially affect the relative abundance of the main bacterioplankton taxa. Furthermore, the function of bacterioplankton taxa influences the co-occurrence patterns of the community, and central species play an important role in regulating bacterioplankton community composition and functional structure. These findings provide additional references on the ecological consequences of anthropogenic interference on aquatic ecosystems. With this study, we hope to contribute to the environmental management and restoration particularly of the Taihu Watershed.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSEC* online.

ACKNOWLEDGEMENTS

REFERENCES

Arp DJ, Stein LY. Metabolism of inorganic N compounds by ammonia-oxidizing bacteria. *Crit Rev Biochem Mol Biol* 2003;**38**:471–95.

Banerjee S, Schlaeppi K, van der Heijden *et al.* Keystone taxa as drivers of microbiome structure and functioning. *Nat Rev Microbiol* 2018;**16**:567–76.

Barberan A, Bates ST, Casamayor EO *et al.* Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* 2012;**6**:343–51.

Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *ICWSM* 2009;**8**:361–2.

Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front Microbiol* 2014;**5**:1–14.

Blanchet FG, Legendre P, Borcard D. Forward selection of explanatory variables. *Ecology* 2008;**89**:2623–32.

Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.

Burton SAQ, Prosser JI. Autotrophic ammonia oxidation at low pH through urea hydrolysis. *Appl Environ Microbiol* 2001;**67**:2952–7.

Camargo JA, Alonso Á. Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: a global assessment. *Environ Int* 2006;**32**:831–49.

Chrzanowski TH, Sterner RW, Elser JJ. Nutrient enrichment and nutrient regeneration stimulate bacterioplankton growth. *Microb Ecol* 1995;**29**:221–30.

Cram JA, Xia LC, Needham DM *et al*. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. *ISME J* 2015;**9**:2573–86.

Csardi G, Nepusz T. The igraph software package for complex network research. *Inter Journal Complex Syst* 2006;**1695**:1–9.

Delgado-Baquerizo M, Reith F, Dennis PG *et al*. Ecological drivers of soil microbial diversity and soil biological networks in the Southern Hemisphere. *Ecology* 2018;**99**:583–96.

De Menezes AB, Prendergast-Miller MT, Richardson AE *et al*. Network analysis reveals that bacteria and fungi form modules that correlate independently with soil parameters. *Environ Microbiol* 2015;**17**:2677–89.

Deng Y, Zhang P, Qin Y *et al*. Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation. *Environ Microbiol* 2016;**18**:205–18.

Dennis KL, Wang Y, Blatner NR *et al*. Adenomatous polyps are driven by microbe-instigated focal inflammation and are controlled by IL-10–producing T cells. *Cancer Res* 2013;**73**:5905–13.

Dudgeon D, Arthington AH, Gessner MO *et al*. Freshwater biodiversity: importance, threats, status and conservation challenges. *BIORev* 2006;**81**:163–82.

Erdős P, Rényi A. On the evolution of random graphs. *Publ Math Inst Hungar Acad Sci* 1960;**5**:17–61.

Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science* 2008;**320**:1034–39.

Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol* 2012;**10**:538.

Freimann R, Burgmann H, Findlay SE *et al*. Bacterial structures and ecosystem functions in glaciated floodplains: contemporary states and potential future shifts. *ISME J* 2013;**7**:2361–73.

Ghoul M, Mitri S. The ecology and evolution of microbial competition. *Trends Microbiol* 2016;**24**:833–45.

Hakemian AS, Rosenzweig AC. The biochemistry of methane oxidation. *Annu Rev Biochem* 2007;**76**:223–41.

Hazen TC, Dubinsky EA, DeSantis TZ *et al*. Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* 2010;**330**:204–8.

He B, Kanae S, Oki T *et al*. Assessment of global nitrogen pollution in rivers using an integrated biogeochemical modeling framework. *Water Res* 2011;**45**:2573–86.

Hooper AB, Vannelli T, Bergmann DJ *et al*. Enzymology of the oxidation of ammonia to nitrite by bacteria. *Antonie Van Leeuwenhoek* 1997;**71**:59–67.

Hu A, Ju F, Hou L *et al*. Strong impact of anthropogenic contamination on the co-occurrence patterns of a riverine microbial community. *Environ Microbiol* 2017;**19**:4993–5009.

Hu A, Li S, Zhang L *et al*. Prokaryotic footprints in urban water ecosystems, A case study of urban landscape ponds in a coastal city, China. *Environ Pollut* 2018;**242**:1729–39.

Jiao S, Liu Z, Lin Y *et al*. Bacterial communities in oil contaminated soils, Biogeography and co-occurrence patterns. *Soil Biol Biochem* 2016;**98**:64–73.

Kämpfer P, Steiof M, Becker PM *et al*. Characterization of chemoheterotrophic bacteria associated with the in situ bioremediation of a waste-oil contaminated site. *Microb Ecol* 1993;**26**:161–88.

Liaw A, Wiener M. Classification and regression by random Forest. *R news* 2002;**2**:18–22.

Liu YR, Delgado-Baquerizo M, Bi L *et al*. Consistent responses of soil microbial taxonomic and functional attributes to mercury pollution across China. *Microbiome* 2018;**6**:183.

Li X, Meng D, Li J *et al*. Response of soil microbial communities and microbial interactions to long-term heavy metal contamination. *Environ Pollut* 2017;**231**:908–17.

Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. *Science* 2016;**353**:1272–77.

Maalcke WJ, Dietl A, Marritt SJ *et al*. Structural basis of biological NO generation by octaheme oxidoreductases. *J Biol Chem* 2014;**289**:1228–42.

Ma B, Wang H, Dsouza M *et al*. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J* 2016b;**10**:1891–901.

Ma L, Mao G, Liu J *et al*. Spatial-temporal changes of bacterioplankton community along an exhorheic river. *Front Microbiol* 2016a;**7**:250.

Meziti A, Tsementzi D, Ar Kormas K *et al*. Anthropogenic effects on bacterial diversity and function along a river-to-estuary gradient in Northwest Greece revealed by metagenomics. *Environ Microbiol* 2016;**18**:4640–52.

Newton RJ, Jones SE, Eiler A *et al*. A guide to the natural history of freshwater lake bacteria. Microbiol. *Mol Biol Rev* 2011;**75**:14–49.

Nilsson C, Reidy CA, Dynesius M *et al*. Fragmentation and flow regulation of the world's large river systems. *Science* 2005;**308**:405–8.

Nyerges G, Stein LY. Ammonia co-metabolism and product inhibition vary considerably among species of methanotrophic bacteria. *FEMS Microbiol Lett* 2009;**297**:131–6.

Raes J, Letunic I, Yamada T *et al*. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol Syst Biol* 2011;**7**:437.

Read DS, Gweon HS, Bowes MJ *et al*. Catchment-scale biogeography of riverine bacterioplankton. *ISME J* 2015;**9**:516–26.

Ren T, Roy R, Knowles R. Production and consumption of nitric oxide by three methanotrophic bacteria. *Appl Environ Microbiol* 2000;**66**:3891–7.

Segata N, Huttenhower C. Toward an efficient method of identifying core genes for evolutionary and functional microbial phylogenies. *PLoS One* 2011;**6**:e24704.

Shi Y, Li Y, Xiang X *et al*. Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain. *Microbiome* 2018;**6**:27.

Staley C, Gould TJ, Wang P *et al*. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. *Front Microbiol* 2014;**5**:414.

Staley C, Gould TJ, Wang P *et al*. Species sorting and seasonal dynamics primarily shape bacterial communities in the Upper Mississippi River. *Sci Total Environ* 2015;**505**:435–45.

Stein LY, Klotz MG. Nitrifying and denitrifying pathways of methanotrophic bacteria. *Biochem Soc Trans* 2011;**39**:1826–31.

Trivedi P, Delgado-Baquerizo M, Trivedi C *et al*. Microbial regulation of the soil carbon cycle, evidence from gene–enzyme relationships. *ISME J* 2016;**10**:2593–604.

Wang J, Zheng Y, Hu H *et al*. Coupling of soil prokaryotic diversity and plant diversity across latitudinal forest ecosystems. *Sci Rep* 2016a;**6**:19561.

Wang P, Chen B, Yuan R *et al*. Characteristics of aquatic bacterial community and the influencing factors in an urban river. *Sci Total Environ* 2016b;**569-570**:382–9.

Wang P, Zhao J, Xiao H *et al*. Bacterial community composition shaped by water chemistry and geographic distance in an anthropogenically disturbed river. *Sci Total Environ* 2019;**655**:61–9.

Wang Q, Garrity GM, Tiedje JM *et al*. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;**73**: 5261–7.

Wang SY, Sudduth EB, Wallenstein MD *et al*. Watershed urbanization alters the composition and function of stream bacterial communities. *PLoS One* 2011;**6**:e22972.

Yang Y, Gao Y, Huang X *et al*. Adaptive shifts of bacterioplankton communities in response to nitrogen enrichment in a highly polluted river. *Environ Pollut* 2019;**245**:290–9.

Yao H, Qian X, Yin H *et al*. Regional risk assessment for point source pollution based on a water quality model of the Taipu River, China. *Risk Anal* 2015;**35**:265–77.