Geographic Patterns of Bacterioplankton among Lakes of the Middle and Lower Reaches of the Yangtze River Basin, China

Running title: Bacterial biogeography among interconnected lakes

Chengrong Bai^{a,b}, Jian Cai^{a,b}, Lei zhou^{a,b}, Xingyu Jiang^{a,b}, Yang Hu^a, Jiangyu Dai^c,
Keqiang Shao^a, Xiangming Tang^a, Xiangdong Yang^a, Guang Gao^{a,#}

^a State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography
and Limnology, Chinese Academy of Sciences, Nanjing 210008, P. R. China

^b University of Chinese Academy of Sciences, Beijing, 100049, P. R. China

^c State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering,
Nanjing Hydraulic Research Institute, Nanjing, 210029, P. R. China

[#] Corresponding author: Guang Gao

E-mail guanggao@niglas.ac.cn; *Tel* (+86) 25 86882213; *Fax* (+86) 25 57714759

Keywords: Interconnected lakes, Biogeography, Bacterioplankton, Deterministic
processes, Stochastic processes

Abstract

In aquatic ecosystems, microbial biogeography research is critical for unveiling the mechanisms of microbial community succession. However, little is known about the microbial biogeography among interconnected lakes. To address this deficit, we used high-throughput sequencing to explore geographic patterns and the relative importance of ecological processes that shape these patterns in abundant and rare bacterial subcommunities from 25 lakes across the middle and lower reaches of Yangtze River basin (MLYB, located in southeast China), where most of the lakes are interconnected by river networks. We found that there were significant differences in both abundant and rare bacterial subcommunities between the two lake groups that were far from each other, while there was no difference among the nearby lakes in each group. Both abundant and rare bacteria followed a strong distance-decay relationship, especially for rare bacteria. These findings suggest that although the interconnectivity between lakes breaks the geographical isolation of bacteria, the dispersal capability of bacterial taxa was still limited by geographic distance. We also found that although deterministic processes and stochastic processes together drive the bacterial subcommunities assembly, the stochastic processes (based on adjusted R^2 in redundancy analysis) exhibited a greater influence on bacterial subcommunities. Our results implied that bacterial dispersal among interconnected lakes was more stochastically.

Importance

Unraveling the relative importance of ecological processes regulating microbial community structure is a central goal in microbial ecology. In aquatic ecosystems, microbial communities often occur in spatially structured habitats, where connectivity directly affects dispersal and metacommunity processes. Recent theoretical work suggests that directional dispersal among connected habitats leads to higher variability in local diversity and among-community composition. However, the study of microbial biogeography among natural interconnected habitats is still lacking. The findings of this study revealed interesting phenomena of microbial biogeography among natural interconnected habitats, suggested that the high interconnectivity reduced the spatial heterogeneity of bacteria, and caused the dispersal of bacteria to be more stochastically. This study has provided a deeper understanding of the biogeographic patterns of rare and abundant bacterial taxa and their determined processes among interconnected aquatic habitats.

Introduction

In aquatic systems, bacteria are the major drivers of nutrient regeneration and energy flow (1, 2), and understanding their spatial diversity and community structure is critical for understanding their relationship with ecosystem functioning (3-5). Compared with large animals and plants, bacteria are more complex in their spatial diversities because of their small size, large numbers, and rich biodiversity (6). A typical phenomenon of most bacterial community compositions (BCCs), is that a few species have numerous individuals, while numerous species have relatively few individuals; the latter is often called the “rare biosphere” (7).

Abundant taxa traditionally considered to be the core groups that performed major ecological functions, and these taxa were therefore intensively studied (8, 9). However, more recent studies have suggested that rare bacterial taxa also play a key role in global biogeochemical cycles and ecosystem functions (7, 10-12). Therefore, the biogeography of both the abundant and rare bacteria needs to be explored rigorously.

The major research focus in microbial biogeography is to reveal the biogeographic patterns of microbial taxa, and to explore the driving processes that shape these patterns (9, 13). Although recent high-throughput sequencing (HTS) studies have provided important insights into microbial biogeography (4), there are still many unresolved problems and disputes (14-16).

Debate continues about whether microorganisms have geographic patterns. Early

studies suggested that organisms less than 1 mm long did not have biogeographic patterns for their strong dispersal capabilities (6), and the classic hypothesis was interpreted as ‘everything is everywhere’ (17). However, the recent evidence contradicts this, and shows that BCCs are indeed regularly distributed in time and space (16).

A second debate concerns the processes that shape microbial biogeography. Growing evidence supports that both deterministic (environmental and spatial selection-related) and stochastic (dispersal-related) ecological processes determine the BCCs (13). Deterministic processes mean that the composition and relative abundance of species are determined by certain abiotic and biotic factors (18-20). For example, salinity is the main factor influencing the global diversities and spatial distribution patterns of microorganisms, according to Lozupone and Knight (21) and Caporaso et al. (22). Other studies, in aquatic systems, suggest that deterministic processes such as elevation gradient, salinity, and land use, are the main constraints affecting the high-level heterogeneity of bacterial diversities and biogeographic patterns (9, 23-25). Stochastic processes also can explain the biogeographic patterns of bacteria. For instance, Chen et al. (26) revealed that the planktonic community in four marine ecosystems, around Xiamen island in southeast China, was well explained by stochastic processes. At a large spatial scale, Mo et al. (27) reported that in three subtropical bays of China, stochastic processes has a slightly greater influence on both abundant and rare subcommunities than did environmental selection (deterministic processes).

Abundant and rare bacterial taxa are differentiated from each other by two key

significant differences that may influence their geographic patterns: (1) They have different community structures and relative abundances, and may experience different limitations to dispersal. For example, Mo et al. (27) revealed that rare taxa have a stronger response to dispersal limitations than abundant bacterioplankton subcommunities in subtropical bays of China. The different result also reported by Wu et al. (28), who found that abundant taxa had greater dispersal limitations in the surface layer of northwestern Pacific Ocean. (2) Abundant and rare taxa carry out different ecological strategies, may have different ecological niches, and are likely to have different responses to environmental changes. For instance, Liu et al. (13) suggested that among lakes and reservoirs of China, the geographic pattern of rare bacteria is more likely determined by environmental conditions, but abundant taxa are mostly governed by spatial factors. In addition, Liao et al. (9) reported that, among 21 lakes in the Yungui Plateau in China, rare taxa are subject to different environmental drivers than those affecting abundant taxa.

In aquatic ecosystems, microbial communities often occur in spatially structured habitats, where connectivity directly affects dispersal and metacommunity processes (29). The phenomenon of habitat connectivity occurs in open waters (oceans, bays, and lakes) and in riverine ecosystems. The extent to which habitats within open waters and rivers are connected longitudinally depends on their position and spatial dynamics (30-32). Some recent simulated works suggest that dispersal along connected habitats (e.g., dendritic landscapes) often biased the hydrodynamic direction. This biased connectivity leads to a great spatial heterogeneity of species (32, 33).

In natural conditions, however, some habitats are interconnected with each other without significant bias. Few studies have reported the biogeographic patterns of bacteria among interconnected aquatic habitats, and these studies are focused on the ocean or the marine bays (26, 27). The bacterial biogeography among interconnected lakes is still little known. Therefore, understanding the diversity and biogeographic patterns of bacteria among interconnected lakes will aid our understanding of microbial biogeography, and reveal more information about the ecological processes and mechanisms that underlie and maintain ecosystem function in the terrestrial-aquatic ecosystems.

An intriguing area for study is in the east of China, namely the middle and lower reaches of the Yangtze River Basin (MLYB), which is low and flat, and covered by hundreds of shallow lakes and intricate river networks (34). Some of the nearby lakes are connected by river networks, especially in the wet season. More importantly, with the frequent change of regional water level in the wet season, there is a mutual hydraulic exchange among these lakes. Considering the special relationship among the lake habitats, in our present study we hypothesize that: (1) the high connectivity within and among lakes reduces regional differences of geographic patterns of abundant and rare bacterial taxa, but the bacterial dispersal still limited by geographical distance, (2) the “rare biosphere” with numerous taxa but few individuals in each taxon, are more likely to change in the process of dispersal than are the few taxa with abundant bacteria (35); and (3) in connected lakes, stochastic processes exhibited a greater influence on the geographic pattern of bacteria than do deterministic processes.

Results

The division of abundant and rare taxa, and multivariate abundance cutoff

In total, 720,326 high-quality sequences were recovered, with a mean of $28,813 \pm 4,218$ per sample, which clustered into 9,847 OTUs. At regional and local levels, 468,822 sequences (65.1%), representing 81 (0.8%) OTUs were classified as abundant, while 21,083 (2.9%) sequences, representing 6,788 (68.9%) OTUs were classified as rare (Table S1).

Among all lake samples, no single OTU was abundant ($> 1\%$) in all samples; only two OTUs (including 97,781 sequences) with $> 1\%$ abundance were present in $> 70\%$ of the samples, and 49 OTUs (including 26,319 sequences) were classified as being abundant in only one lake. Among all lake samples, 6,238 OTUs (91.9% of the total) were classified as always rare, and 3,086 OTUs (including 95,443 sequences) with $< 0.01\%$ abundance were present $> 70\%$ of the samples (Fig. S1).

The result of MultiCoLA shown that little variation in both the Spearman and Procrustes correlations after removal of more than 40% of the rare part of the original data and the truncated data sets. In addition, when the number of rare types increased to 5%, the data structure of the truncated matrices exhibited a little variation (Fig. S2). This result indicated that our definition of abundant (34.5%) and rare (2.9%) bacteria are reasonable.

Geographic patterns of the bacterioplankton community

Geographic patterns of abundant and rare bacterial communities had significant positive correlation, both dividing two similar clusters (Fig. S3, $P < 0.01$). We found that seven lakes around Lake Taihu (TH) were clustered together (Cluster1), and the remaining lakes were divided into another cluster (Cluster2). Notably, we found that the two clusters were not divided by the middle and the lower reaches (Fig. 2a, S3). In addition, linear regression analysis showed that with increasing geographic distance, the dissimilarities in both abundant and rare communities were increased. Spearman's correlation analysis gave a correlation coefficient of 0.587 (Mantel test, $P < 0.01$) between geographical distance and the dissimilarity of all bacterial taxa. The Spearman's correlation coefficients for abundant and rare bacterial taxa were 0.586 ($P < 0.01$) and 0.438 ($P < 0.01$), respectively (Fig. 2b). This phenomenon also suggested that lakes that were closer to each other presented more similar BCCs than lakes that were more distant from each other. Moreover, within the geographic distance, the community dissimilarity coefficient of rare taxa was larger than that of abundant taxa, which indicated that the structure of rare taxa changes more markedly with increasing distance than is the case for abundant taxa (Fig. 2b).

The relationship between bacterial mean relative abundance and their sites occupied

The bacterial mean relative abundance, for both abundant and rare taxa, was significantly positively correlated with sites occupied (abundant taxa $r = 0.409$, rare taxa $r = 0.693$, $P <$

0.01) (Fig. 3). The majority of abundant OTUs occupied > 50% of sampling sites, but all rare OTUs occupied < 50% of sampling sites (Fig. 3).

Taxonomic distribution of abundant and rare bacteria

All reads were classified and grouped under 43 phylum-level taxonomic groups (including 2 unknown phyla). Across all samples, *Actinobacteria* and *Proteobacteria* were the two most abundant phyla of bacteria, their average relative abundance accounting for 48.2% and 19.4% of all clean reads, respectively. The next most abundant phyla were: *Verrucomicrobia* (8.1%), *Bacteroidetes* (7.0%), *Planctomycetes* (6.2%), TM7 (4.4%), *Chlorobi* (2.8%), and *Chloroflex* (1.4%). Within the *Proteobacteria* phylum, α -*Proteobacteria* was the most abundant group, accounting for 6.6% of all clean reads; then followed by β -*Proteobacteria*, γ -*Proteobacteria*, TA18, and δ -*Proteobacteria*, their relative abundances were 6.4%, 3.5%, 1.5%, and 1.3%, respectively (Fig. S4).

Across all samples, the community structures of both abundant and rare taxa at the phylum level have certain differences. (1) The rare taxa have more species than abundant taxa (Fig. 4). (2) The relative abundances of different phyla in abundant and rare taxa were different. The phyla of the abundant community were dominated by a large group (*Actinobacteria*), while the relative abundance of each rare phylum was relatively uniform. The relative abundance of the same phylum in abundant taxa and rare taxa was as follows: *Actinobacteria* (abundant $60.7 \pm 10.4\%$ vs rare $24.9 \pm 8.1\%$ (mean $\pm$ s.e.)), *Verrucomicrobia* ($10.0 \pm 7.6\%$ vs $4.2 \pm 2.0\%$), α -*proteobacteria* ($3.4 \pm 4.4\%$ vs

10.2±3.1%), *β-proteobacteria* (4.4 ± 4.9% vs 9.4 ± 5.3%), *γ-proteobacteria* (2.1± 1.8% vs 6.8 ± 3.1%), *Planctomycetes* (6.4 ± 6.0% vs 8.0 ± 3.7%), *Bacteroidetes* (3.2 ± 3.7% vs 11.1 ± 8.4%), *TM7* (4.5 ± 7.8% vs 2.9 ± 2.5%), *Chlorobi* (3.5 ± 4.0% vs 3.7 ± 2.9%), and *TA18-proteobacteria* (1.3 ± 1.7% vs 2.3 ± 2.2%). Moreover, there were several phyla that showed high relative abundance in rare taxa only, such as *δ-proteobacteria* (5.2 ± 2.9%), *Acidobacteria* (1.9 ± 1.3%), and *Firmicutes* (1.5 ± 0.8%). And a total of 5.4±3.1% groups were occupied by large numbers of low abundance phyla (Fig. 4).

The relative contribution of deterministic and stochastic processes in the assembly of abundant and rare taxa

After analysis, 14 physicochemical variables (Table S2) and 6 significant PCNMs (PCNM1-6, $P < 0.05$) were subject to RDA, CCA, and VPA analyses. The Mantel results revealed that both environmental variables and spatial variables had a significant effect on bacterial sub-communities. Abundant bacterial communities were primarily driven by environmental variables, contrasting with rare bacterial sub-communities which were primarily driven by spatial variables (Table 1).

The RDA ordination showed that the change of the abundant bacterial sub-community was significantly related to two environmental variables (Temp and Chl *a*) and one spatial factor (PCNM1) (by forward model selection) (Fig. 5a). In contrast, the changes in the rare bacterial sub-community taxa were significantly related to the same environmental variables (Temp and Chl *a*) but more spatial factors (PCNM1 and PCNM2) (Fig. 5a).

The result of VPA showed that deterministic processes, namely environmental variables (E) and spatial distance (S), accounted for 35.7% of the abundant bacterial taxa variations; the part accounted for by E and S together was 27.2%. Abundant bacterial taxa were determined by environmental variables, which accounted for 31.8% of the abundant taxa variation (Fig. 5b). For rare bacterial taxa, however, deterministic processes (including E and S) only accounted for 7.2% of the variations. The spatial variable was the dominant driver, explaining 6.4% of the rare taxa variation (Fig. 5b).

In addition, the S alone can explain 3.9% and 1.6% of the abundant and rare variation, respectively. It was notable that a large proportion of bacterial taxa variations were not explained by deterministic processes, especially for rare species.

To explore the effects of the stochastic process on bacterial geographic patterns, the neutral model was employed in our study. The result showed that the neutral model explained a large fraction ($R^2 = 52\%$) of the variation of all bacterial taxa (Fig. 6).

Discussion

Geographic patterns in abundant and rare bacteria

The study of BCCs and their biogeographic patterns in particular areas is a fundamental research area in aquatic microbial ecology (1), in which bacterial biogeography considers the relationship between the biodiversity patterns and geographic distributions. New molecular tools and technologies have revealed more information about the “rare

biosphere” to us, and rare bacteria are now recognized as having a great effect on ecological functions and mechanisms (12, 36). The ecological relationship between abundant bacteria and rare bacteria, especially the correlation between their geographical patterns has become an attractive research area.

Habitat connectivity directly shapes diversity patterns in microcosm metacommunities at different levels (29). Recent theoretical studies of bacterial biogeography have considered both isolated and interconnected habitats. Among isolated habitats, some studies have reported that the division of bacterial biogeography was consistent with geographical regions. For instance, Liu et al. (13) revealed that in lakes and reservoirs in different regions of China, geographical regions were significantly separated for BCCs. Logares et al. (4) also found that the regional bacterial community was mainly structured by geographic origin. Among interconnected habitats, bacterial subcommunity is complicated among the different regions or sites. For example, Tang et al. (1) reported that free-living (FL) and particle-attached (PA) bacteria were not significantly separated in the different areas of Lake Taihu. At a large spatial scale, namely in some marine bays in China, bacterial biogeography among some interconnected habitats was consistent with their geographical area (26, 27, 37).

Patterns of bacterial abundance are suggested to be the outcome of processes that are inherently driven by the structure of the respective habitat network [30, 31]. Our observation provides important insights for exploring bacterial community patterns among interconnected habitats. Our observations have shown that all lakes were

significantly separated as two large clusters at $P = 0.001$ for the bacterioplankton communities. Interestingly, we found that there is a relatively large geographic distance between the lakes in cluster1 and cluster2 (Fig 2a). In this scenario, the dispersal of bacteria between the lakes in cluster1 and cluster2 may be more difficult. A similar phenomenon also reported in some large-distance marine bays in China (27).

However, the ordination of the lakes in each cluster is scattered (Fig 2a). In MLYB, many nearby lakes are connected by intricate river networks, which provide pathways for bacterial dispersal, and break the geographical isolation among lakes. In particular, our samples were collected during the wet season, when high water levels ensured that most lakes were fully connected through the river networks. Furthermore, there was no significant hydrodynamic bias among these lakes, as they are locating in plain areas. In this scenario, the dispersal of bacteria among different lakes is likely to be more randomly. More importantly, contrary to the biased connectivity, the spatial heterogeneity of bacterial diversities among interconnected lakes is reduced (32, 33).

Further, some studies have reported that the rare bacterial biosphere may be significantly correlated with abundant taxa, (rather than being a random collection of taxa) (9, 13, 27). The results of our present study, in particular, the correlation of geographic patterns between abundant and rare bacteria, supported this contention: we found similar biogeographic patterns of rare and abundant taxa (Fig. S3, $r=0.873$, $P < 0.01$). Our present study indicates that even in interconnected lakes, rare bacteria might in a similar manner as abundant ones in responding to dispersal processes or environmental changes

(9, 35).

We also report that in the MLYB, both abundant and rare bacterial taxa had a significant distance-decay relationship (Fig. 2b). Some recent studies reported that the distance-decay relationship of bacteria occurs among both isolated habitats (9, 13, 23) and connected habitats (27). Bacterial community dissimilarity rises with increasing geographic distance if: (1) species filter by local habitat conditions, or (2) if they are limited in dispersal (38). Among interconnected habitats, species can disperse directly through connecting channels. However, the distance-decay relationship showed in both the present study and others (e.g. Mo et al. (27) and Wei et al. (26)), still revealed a phenomenon of limited dispersal. Notably, consistent with Mo et al. (27), we found that there was no significant increase in environmental variability with geographic distance. This result may be explained by bacterial individuals tending to colonize nearby sites (38). In this scenario, the distance-decay relationship would be directly related to the spatial scales among habitats, and would still emerge among different habitats with the same environmental conditions or niche requirements (39). The geographic distance between the lakes in each cluster was less than that between the three marine bays studied by Mo et al. [26]. Correspondingly, the smaller is the geographic distance between regions, the higher is the similarity of the community composition. The small geographic distance between lakes in each cluster may explain why our results are similar to those of Tang et al. [11]. On the contrary, the geographic distance between the lakes in cluster1 and cluster2 is much larger. The large spatial distance might lead to differences of bacterial

subcommunities between the two regions.

We also found that the community dissimilarity increased more significant for rare bacteria with the increasing of geographic distance than it did for abundant bacteria. Our result was consistent with Mo et al. (27) and Liao et al. (9) who reported that most of the abundant bacteria appeared in most of the sampling sites, while all rare bacteria appeared in only a few sampling sites (Fig. 3). Further, consistent with Mo et al. (27), we found that the biodiversity patterns between abundant and rare taxa are largely different (at phylum level) (Fig. 4). The abundant taxa showed relative few community compositions and high relative abundance, whereas the rare taxa had complex community composition and low relative abundance (Fig. S1, Fig. 4). Liu et al. (13) suggested that rare bacteria with low local abundance have an increased probability of local extinction, and a decreased probability of dispersal.

Controlling factors for geographical distributions of abundant and rare bacterial taxa

Investigating the controlling factors, and their contributions to geographical distributions, of abundant and rare bacterial taxa are fundamental research topics for ecologists. Improved understanding of these topics will help us to reveal the reasons for particular biogeographic patterns, and allow prediction of the community succession in the habitat. Study of relevant deterministic processes may focus on ecological selectors as limiting factors, including nutrients, land use, and habitat connectivity (1, 9, 29). However, spatial

variation in the BCCs is also influenced by stochastic processes, such as dispersal limitation, successional stages, and random effects (40).

Previous research works have highlighted two key areas for which further study is needed (27, 41). First, few studies have investigated the influence of both deterministic and stochastic processes on the biogeographic pattern of bacteria. Second, the unexplored topic of the influence of both deterministic and stochastic processes on the biogeographic pattern of bacteria in interconnected lakes. So, to address these deficits, in our present study, we focused on these unexplored problems, and tried to reveal the rules of bacterial biogeography among interconnected terrestrial waters.

We used ecological statistical methods (forward selection) to verify the correlation between deterministic variables (environmental and spatial variables) and the spatial distribution of abundant and rare bacterial taxa (42). We found that only two environmental factors (Temp, Chl *a*) were significantly related to the spatial distribution of abundant and rare taxa subcommunities (Fig. 5a). These environmental variables exactly reflected the habitat requirements of bacteria in this area. Most of the lakes in the MLYB are suffer from mesotrophic or eutrophic (43), with extremely high phytoplankton and Chl *a* values. The formation of phytoplankton directly provides energy and materials for the growth and metabolism of bacteria (1), so that determines the bacterial biodiversity. In addition, the temperature is critical for maintaining bacterial growth and metabolic activity. Therefore, the change in temperature systematically altered the relationship between biodiversity and ecosystem functioning (44-46). We notice that

although our samples were collected at the same time, the surface water temperature has changed greatly during the rainfall process. Across all lakes, the maximum temperature difference reached 9.55°C (Table S2). Large temperature difference might be another important reason for the bacterial spatial heterogeneity.

We found that both environmental and regional variables were significantly related to community structures of abundant and rare bacterial taxa (Table 1), and that these two variables together explained a large proportion of bacterial distributions (Fig. 5b). The spatial variables explained more variation of rare bacterial community than did the environmental variables (Fig. 5b). In contrast, the environmental variables explained more variation of the abundant bacterial community than did the spatial variables (Fig. 5b). This result was consistent with the Mantel tests for the correlation between community similarity and local environmental and regional factors using Spearman's coefficient (Table 1).

We note that a contrasting result was reported from marine bays by Mo et al. (27). These authors reported that while environmental and spatial variables together explained a large part of the variations in abundant communities, they explained little of the variations in rare communities. They suggested that purely environmental components were slightly less influential contributors than the purely spatial variables in regulating the assembly of both abundant and rare subcommunities.

The discrepancy between our present work and that of Mo et al. [26] may be explained in two aspects. Firstly, we hold that interconnectivity among lakes leads to a frequent

exchange of energy and environmental variables, and reduces the heterogeneity of environment among nearby lakes (Fig. S5). Secondly, with low relative abundance, the rare taxa, were influenced by spatial variables more strongly than were the abundant taxa (Fig. 2b).

In addition, we hold that the interconnectivity among lakes leads to the dispersal of bacteria in different habitats to be more stochastic than deterministic, especially for rare bacteria. Thus, the contribution of environmental and spatial variables to the geographic pattern of bacterial communities were quite low, especially for rare species (Fig. 5b). Our result was consistent with several recent studies on different interconnected marine bays, and their abundant and rare subcommunities (26, 27, 37). However, in contrast to the isolated lakes, which indicated that abundant and rare bacterial taxa were strongly influenced by the deterministic process, and only weakly influenced by the stochastic process (9).

The NCM analysis showed that stochastic process explained a large fraction ($R^2=52\%$) of the variation of all bacterial taxa in the interconnected lakes in the MLYB (Fig. 6), consistent with similar studies on the bacterial subcommunities of interconnected marine bays (26, 27, 37), but contrasting with the situation in isolated lakes (9). These results indicate that the interconnectivity among habitats increases the randomness of bacterial dispersal. Two main reasons could support our result: (1) most of the lakes were interconnected by numerous river networks so that the flow of water among these lakes was not a significant bias; and (2) bacteria cannot avoid adverse conditions or seek

favorable environments, in contrast to larger aquatic animals. Moreover, this result was also consistent with our finding of the no regional differences of bacteria. In addition, there was still a large amount of unexplained variation that could be attributed to other biotic or abiotic variables than the ones mentioned in this paper.

Conclusions

This study has provided a deeper understanding of the biogeographic patterns of rare and abundant bacterial taxa and their determined processes among interconnected aquatic habitats. Our conclusions strongly support our hypotheses, we interestingly found that the high connectivity within and among lakes reduces regional differences of geographic patterns of abundant and rare bacterial taxa, especially among nearby lakes. However, there were significant differences in both abundant and rare bacterial subcommunities between the two lake groups that were far from each other. This phenomenon was consistent with the result that both abundant and rare bacteria followed a strong distance-decay relationship. In addition, the “rare biosphere” with numerous taxa but few individuals in each taxon, were encountered more dispersal limitation and followed a stronger distance-decay relationship. Furthermore, deterministic processes and stochastic processes together drive the bacterial subcommunities assembly, but stochastic processes exhibited a greater influence on the biogeography of bacterial communities among the interconnected aquatic ecosystem.

Materials and methods

Study area and sample collection

With a total area of 18,400 km², the MLYB has more than 600 lakes larger than 1 km² (47). These lakes are vital to both regional economic development and ecological stability (48, 49). However, due to heavy external pollutant inputs from local and upstream areas, most of these lakes are polluted, and many are eutrophic (50). There is a strong hydraulic disturbance and exchange among lakes, due to the area's location in the subtropical climatic zone, with distinct dry and wet seasons.

The MLRB is divided into the middle reaches (MR) and the lower reaches (LR) by the city Hukou (51). Field work was conducted in July 2016. We randomly selected 25 lakes in the MR and the LR from which to collect water samples (Fig. 1, and Table S1). At each lake, one surface water sample (from the top 50cm of the water column) was collected with a 5 L sterilized polypropylene bottles at the lake center. All water samples were transported to the laboratory in dark cooling boxes and processed for physicochemical analyses and bacterial community analyses as soon as possible.

Physicochemical analysis

Water column depth (WD) and Secchi depth (SD) were measured by a water depth gauge (Uwitec, Austria) and Secchi disk, respectively. Water temperature (Temp), pH, electrical conductivity (EC), dissolved oxygen (DO), salinity (Sal), and turbidity of the epilimnion

layer were measured *in situ* with a multiparameter water quality sonde (YSI 6600 V2, Yellow Springs Instruments, U.S.A.). Nine physicochemical parameters were determined using standard methods [40]: total nitrogen (TN), ammonium (NH_4^+), nitrate (NO_3^-), nitrite (NO_2^-); total phosphorus (TP), orthophosphate (PO_4^{3-}); dissolved inorganic carbon (DIC), dissolved organic carbon (DOC), total suspended solids (TSS), and chlorophyll *a* (Chl *a*).

DNA extraction, PCR, and high-throughput sequencing

For bacterioplankton analyses, a 100-500 ml water sample was filtered through a 0.22- μm pore size polycarbonate filter (47 mm diameter, Millipore, Billerica, MA, USA) (52), and then the filter was put into a sterile centrifuge tube and rapidly stored at -80°C . Total DNA was extracted directly from the filters, using FastDNA[®] spin kit for soil (MP Biomedicals, Solon, OH, USA) according to the manufacturer's instructions. After checked the quality and quantity of the amplified DNA fragment (used the NanoDrop ND-1000 UV/Vis spectral photometer), the V4 region of the bacterial 16 ribosomal RNA gene was amplified by primers 515F (5'-GTGCCAGCM-GCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') (53) with 20 μL of DNA. The PCR amplifications for each DNA sample followed these steps: initial denaturation at 94°C for 5 min, followed by 25 cycles of 30 s at 94°C , 30 s at 50°C , and 30 s at 72°C ; the final step was an extension at 72°C for 7 min. Subsequently, the total nucleic acids were sent to BGI Co., Ltd (Wuhan, China) for high-throughput sequencing on an Illumina

MiSeq instrument (San Diego, CA, USA).

Bioinformatics analysis

Raw sequence data were processed using the Quantitative Insights into Microbial Ecology (QIIME V. 1.9.1) pipeline (54). After estimating the quality of raw reads by using FastQC (v0.11.5), the Illumina pair-end reads were assembled using FLASH, with a minimum overlap of 10 bases (55). Then, 16S reads were de-multiplexed and quality-filtered following the pipeline and previous research results (56). Next, we checked and then removed potential chimera sequences with a reference-base method using VSEARCH v1.11.1 (57), which implements the UCHIME algorithm (58).

Bacterial phylotypes were identified and assigned to operational taxonomic units (OTUs, 97% cutoff), based on the 13_8 release of the Greengenes database [48]. After taxonomy classification, representative sequences were aligned and further refined using PyNAST (59, 60), and phylogenetic trees were built using FastTree (61). We removed all archaea, chloroplasts, mitochondria, unassigned sequences, and spurious singletons from the OTU table. Finally, to minimize the sequencing depth of different samples, the high-quality OTU table was randomly resampled based on the sample with the smallest sequencing (18,753 reads) using the QIIME pipeline (54).

Definition of abundant taxa and rare taxa

We defined abundant and rare OTUs based on their relative abundance, considering two

thresholds: on the regional and local scale. At a regional scale, OTUs with a mean relative abundance of $> 0.1\%$ across all samples were defined as abundant taxa, and those with a mean relative abundance of $< 0.001\%$ were defined as rare taxa (4, 13). At the local scale, the thresholds were: abundant if $> 1\%$ within a sample, and rare if $< 0.01\%$ within a sample (7). To reduce the effect of arbitrariness, we used the multivariate cutoff level analysis (MultiCoLA V. 1.4) as described by Gobet (62).

Data analyses

The β -diversity indices were analyzed by QIIME (V. 1.9.1) as part of the OTUs' downstream analysis (54). The geographic patterns of both abundant and rare taxa among the lakes were revealed by non-metric multi-dimensional scaling ordination analysis (NMDS). In addition, the cluster analysis was used to test whether rare bacterial taxa share similar biogeography with abundant ones. For NMDS and Cluster analysis, OTU data were initially subject to Hellinger's transformation (63). The relationships between bacterial community dissimilarity and geographic distance of these lakes were calculated by Spearman's rank correlations. The dissimilarity of the bacterial community was determined by Bray–Curtis dissimilarity matrix, and the geographic distance was determined by geographic coordinates of each lake. To determine if there was a significant difference in bacterial communities with increasing geographic distance, the Mantel test was used.

To explore the influence of deterministic processes on BCCs, the canonical ordination

analyses were used in the R environment using the Vegan package. The deterministic processes including environmental and spatial variables (geographic distance). The normal distribution of the environmental variables was checked using the Shapiro–Wilk test by SPSS (version 22), and these factors were $\log(x+1)$ transformed (with the exception of pH) to improve normality and homoscedasticity for multivariate statistical analyses. To avoid collinearity, environmental factors with variance inflation factors (VIF > 20) were deleted. The spatial variables were generated by principal coordinates of neighborhood matrices (PCNM) analysis (64), based on the geographical coordinates of each sample.

The correlations between bacterial communities and environmental and spatial variables were analyzed by suitable ordination analysis method determined by the detrended correspondence analysis (DCA). The result of DCA revealed that the longest gradient lengths for both all and abundant bacteria were < 3, indicating redundancy analysis (RDA) is suitable for both communities; while the longest gradient lengths for rare bacteria was > 4, indicating canonical correspondence analysis (CCA) is suitable for rare bacterial community (65).

Environmental variables and spatial variables that significantly explained parts of the variation in BCCs were selected by forward selection. Based on a Monte Carlo test with 999 permutations, only variables that explained a significant ($p < 0.05$) additional proportion of total variance were included in the subset of forward selected variables (42). In addition, the contributions of environmental variables and spatial variables to

geographic patterns of BCCs were assessed by Variation partitioning analysis (VPA).

To evaluate the potential importance of stochastic processes for BCCs, we used the neutral community model (NCM) (66). This model predicts the relationship between the frequency of detecting an OTU and its relative abundance along taxonomic ranks (37). In the NCM, the parameter Nm determines the correlation between occurrence frequency and regional relative abundance, with N describing the metacommunity size, and m being the immigration rate. The parameter R^2 determines the overall fit to the neutral model (66). In our present study, all computations for this model were performed in R (version 3.4.4) (R Core Team, 2015) (67).

Nucleotide sequence accession number

The 16S gene sequences generated in the present study were submitted to the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov>) under the accession number SRP173893.

Acknowledgments

Our present study was supported by the Major Science and Technology Program for Water Pollution Control and Treatment (2017ZX07203-004), the National Natural Science Foundation of China (41621002, 41790423, 41530753, 41571462), the “One-Three-Five” Strategic Planning of NIGLAS (NIGLAS2017GH05) and the Key Research Program of Frontier Sciences, CAS (QYZDJ-SSW-DQC008).

References

1. Tang X, Chao J, Gong Y, Wang Y, Wilhelm SW, Gao G. 2017. Spatiotemporal dynamics of bacterial community composition in large shallow eutrophic Lake Taihu: High overlap between free-living and particle-attached assemblages. *Limnol Oceanogr* 62:1366-1382.
2. Freimann R, Buergmann H, Findlay SEG, Robinson CT. 2013. Bacterial structures and ecosystem functions in glaciated floodplains: contemporary states and potential future shifts. *ISME J* 7:2361-2373.
3. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* 10:497-506.
4. Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Gobet A, Kooistra WH, Mahe F, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Stoeck T, Santini S, Siano R, Wincker P, Zingone A, Richards TA, de Vargas C, Massana R. 2014. Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* 24:813-21.
5. Xiong JB, Ye XS, Wang K, Chen HP, Hu CJ, Zhu JL, Zhang DM. 2014. Biogeography of the Sediment Bacterial Community Responds to a Nitrogen Pollution Gradient in the East China Sea. *Appl Environ Microb* 80:1919-1925.

- 537 6. Finlay BJ. 2002. Global dispersal of free-living microbial eukaryote species.
538 Science 296:1061-1063.
- 539 7. Pedrós-Alió C. 2012. The rare bacterial biosphere. *Ann Rev Mar Sci* 4:449-66.
- 540 8. Cottrell MT, Kirchman DL. 2003. Contribution of major bacterial groups to
541 bacterial biomass production (thymidine and leucine incorporation) in the
542 Delaware estuary. *Limnol Oceanogr* 48:168-178.
- 543 9. Liao J, Cao X, Wang J, Zhao L, Sun J, Jiang D, Huang Y. 2017. Similar
544 community assembly mechanisms underlie similar biogeography of rare and
545 abundant bacteria in lakes on Yungui Plateau, China. *Limnol Oceanogr*
546 62:723-735.
- 547 10. Caron DA, Countway PD. 2009. Hypotheses on the role of the protistan rare
548 biosphere in a changing world. *Aquat Microb Ecol* 57:227-238.
- 549 11. Campbell BJ, Yu LY, Heidelberg JF, Kirchman DL. 2011. Activity of abundant
550 and rare bacteria in a coastal ocean. *Proc Natl Acad Sci U S A* 108:12776-12781.
- 551 12. Pester M, Bittner N, Deevong P, Wagner M, Loy A. 2010. A 'rare biosphere'
552 microorganism contributes to sulfate reduction in a peatland. *ISME J*
553 4:1591-1602.
- 554 13. Liu L, Yang J, Yu Z, Wilkinson DM. 2015. The biogeography of abundant and
555 rare bacterioplankton in the lakes and reservoirs of China. *ISME J* 9:2068-77.
- 556 14. Fenchel T, Esteban GF, Finlay BJ. 1997. Local versus global diversity of
557 microorganisms: cryptic diversity of ciliated protozoa. *Oikos* 80:220-225.

- 558 15. Staley JT. 1997. Biodiversity: Are microbial species threatened? Commentary.
559 Curr Opin Biotech 8:340-345.
- 560 16. Martiny JB, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL,
561 Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S,
562 Ovreas L, Reysenbach AL, Smith VH, Staley JT. 2006. Microbial biogeography:
563 putting microorganisms on the map. Nat Rev Microbiol 4:102-12.
- 564 17. Bass-Becking LGM. 1934. Geobiologie of inleiding tot de milieukunde. Van
565 Stockum & Zoon, The Hague, the Netherlands.
- 566 18. Tripathi BM, Stegen JC, Kim M, Dong K, Adams JM, Lee YK. 2018. Soil pH
567 mediates the balance between stochastic and deterministic assembly of bacteria.
568 ISME J 12:1072-1083.
- 569 19. Fargione J, Brown CS, Tilman D. 2004. Community assembly and invasion: An
570 experimental test of neutral versus niche processes. Proc Natl Acad Sci U S A
571 101:414-414.
- 572 20. Crump BC, Hopkinson CS, Sogin ML, Hobbie JE. 2004. Microbial biogeography
573 along an estuarine salinity gradient: Combined influences of bacterial growth and
574 residence time. Appl Environ Microb 70:1494-1505.
- 575 21. Lozupone CA, Knight R. 2007. Global patterns in bacterial diversity. Proc Natl
576 Acad Sci U S A 104:11436-11440.
- 577 22. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh
578 PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of

579 millions of sequences per sample. *Proc Natl Acad Sci U S A* 108:4516-4522.

580 23. Li H, Zeng J, Ren L, Wang J, Xing P, Wu QL. 2017. Contrasting patterns of
581 diversity of abundant and rare bacterioplankton in freshwater lakes along an
582 elevation gradient. *Limnol Oceanogr* doi:10.1002/lno.10518: 1570–1585.

583 24. Hayden CJ, Beman JM. 2016. Microbial diversity and community structure along
584 a lake elevation gradient in Yosemite National Park, California, USA. *Environ*
585 *Microbiol* 18:1782-91.

586 25. Schiaffino MR, Lara E, Fernandez LD, Balague V, Singer D, Seppey CC,
587 Massana R, Izaguirre I. 2016. Microbial eukaryote communities exhibit robust
588 biogeographical patterns along a gradient of Patagonian and Antarctic lakes.
589 *Environ Microbiol* 18:5249-5264.

590 26. Chen W, Pan Y, Yu L, Yang J, Zhang W. 2017. Patterns and Processes in Marine
591 Microeukaryotic Community Biogeography from Xiamen Coastal Waters and
592 Intertidal Sediments, Southeast China. *Front Microbiol* 8:1912.

593 27. Mo Y, Zhang W, Yang J, Lin Y, Yu Z, Lin S. 2018. Biogeographic patterns of
594 abundant and rare bacterioplankton in three subtropical bays resulting from
595 selective and neutral processes. *ISME J* doi:10.1038/s41396-018-0153-6:2198–
596 2210.

597 28. Wu W, Logares R, Huang B, Hsieh C-h. 2017. Abundant and rare picoeukaryotic
598 sub-communities present contrasting patterns in the epipelagic waters of marginal
599 seas in the northwestern Pacific Ocean. *Environmental Microbiology* 19:287-300.

- 600 29. Carrara F, Altermatt F, Rodriguez-Iturbe I, Rinaldo A. 2012. Dendritic
601 connectivity controls biodiversity patterns in experimental metacommunities. *Proc*
602 *Natl Acad Sci U S A* 109:5761-5766.
- 603 30. Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF,
604 Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A. 2004. The
605 metacommunity concept: a framework for multi-scale community ecology. *Ecol*
606 *Lett* 7:601-613.
- 607 31. Benda LEE, Poff NL, Miller D, Dunne T, Reeves G, Pess G, Pollock M. 2004.
608 The Network Dynamics Hypothesis: How Channel Networks Structure Riverine
609 Habitats. *BioScience* 54:413.
- 610 32. Altermatt F, Fronhofer EA. 2018. Dispersal in dendritic networks: Ecological
611 consequences on the spatial distribution of population densities. *Freshwater Biol*
612 63:22-32.
- 613 33. Carrara F, Rinaldo A, Giometto A, Altermatt F. 2014. Complex interaction of
614 dendritic connectivity and hierarchical patch size on biodiversity in river-like
615 landscapes. *Am Nat* 183:13-25.
- 616 34. Yin H, Liu G, Pi J, Chen G, Li C. 2007. On the river–lake relationship of the
617 middle Yangtze reaches. *Geomorphology* 85:197-207.
- 618 35. Logares R, Lindstrom ES, Langenheder S, Logue JB, Paterson H, Laybourn-Parry
619 J, Rengefors K, Tranvik L, Bertilsson S. 2013. Biogeography of bacterial
620 communities exposed to progressive long-term environmental change. *ISME J*

621 7:937-48.

622 36. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. 2009. Ecology of the rare
623 microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* 106:22427-32.

624 37. Liu M, Yu Z, Yu X, Xue Y, Huang B, Yang J. 2017. Invasion by Cordgrass
625 Increases Microbial Diversity and Alters Community Composition in a Mangrove
626 Nature Reserve. *Front Microbiol* 8:2503.

627 38. Bell T. 2010. Experimental tests of the bacterial distance-decay relationship.
628 *ISME J* 4:1357-65.

629 39. Hubbell SP. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*
630 (MPB-32) (Monographs in Population Biology). Princeton University Press,
631 USA.

632 40. Liu L, Yang J, Yu X, Chen G, Yu Z. 2013. Patterns in the composition of
633 microbial communities from a subtropical river: effects of environmental, spatial
634 and temporal factors. *PLoS One* 8:e81232.

635 41. Wang JJ, Shen J, Wu YC, Tu C, Soininen J, Stegen JC, He JZ, Liu XQ, Zhang L,
636 Zhang EL. 2013. Phylogenetic beta diversity in bacterial assemblages across
637 ecosystems: deterministic versus stochastic processes. *ISME J* 7:1310-1321.

638 42. Blanchet FG, Legendre P, Borcard D. 2008. Forward selection of explanatory
639 variables. *Ecology* 89:2623-2632.

640 43. Wu JL, Zeng HA, Yu H, Ma L, Xu LS, Qin BQ. 2012. Water and Sediment
641 Quality in Lakes along the Middle and Lower Reaches of the Yangtze River,

642 China. *Water Resour Manag* 26:3601-3618.

643 44. Garcia FC, Bestion E, Warfield R, Yvon-Durocher G. 2018. Changes in
644 temperature alter the relationship between biodiversity and ecosystem functioning.
645 *Proc Natl Acad Sci U S A* 115:10989-10999.

646 45. Birgander J, Olsson PA, Rousk J. 2018. The responses of microbial temperature
647 relationships to seasonal change and winter warming in a temperate grassland.
648 *Global Change Biol* 24:3357-3367.

649 46. Rosel S, Allgaier M, Grossart HP. 2012. Long-Term Characterization of
650 Free-Living and Particle-Associated Bacterial Communities in Lake Tiefwaren
651 Reveals Distinct Seasonal Patterns. *Microbial Ecol* 64:571-583.

652 47. Xu M, Dong X, Yang X, Chen X, Zhang Q, Liu Q, Wang R, Yao M, Davidson T,
653 Jeppesen E. 2017. Recent Sedimentation Rates of Shallow Lakes in the Middle
654 and Lower Reaches of the Yangtze River: Patterns, Controlling Factors and
655 Implications for Lake Management. *Water* 9:1-18.

656 48. Cui L, Gao C, Zhao X, Ma Q, Zhang M, Li W, Song H, Wang Y, Li S, Zhang Y.
657 2013. Dynamics of the lakes in the middle and lower reaches of the Yangtze River
658 basin, China, since late nineteenth century. *Environ Monit Assess* 185:4005-18.

659 49. Yao XW, Wang ZQ, Wang H. 2015. Impact of Urbanization and Land-Use Change
660 on Surface Climate in Middle and Lower Reaches of the Yangtze River,
661 1988-2008. *Adv Meteorol* doi:10.1155/2015/395094:1-10.

662 50. Qin B. 2002. Approaches to Mechanisms and Control of Eutrophication of

- 663 Shallow Lakes in the Middle and Lower Reaches of the Yangze River. *J Lake Sci*
664 14:193-202.
- 665 51. Wang L, Zheng H, Zhao H, Robinson BE. 2017. Nitrogen balance dynamics
666 during 2000-2010 in the Yangtze River Basin croplands, with special reference to
667 the relative contributions of cropland area and synthetic fertilizer N application
668 rate changes. *PLoS One* 12:e0180613.
- 669 52. Liu J, Fu B, Yang H, Zhao M, He B, Zhang XH. 2015. Phylogenetic shifts of
670 bacterioplankton community composition along the Pearl Estuary: the potential
671 impact of hypoxia and nutrients. *Front Microbiol* 6:1-13.
- 672 53. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh
673 PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of
674 millions of sequences per sample. *Proc Natl Acad Sci U S A* 108:4516-4522.
- 675 54. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK,
676 Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D,
677 Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder
678 J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J,
679 Knight R. 2010. QIIME allows analysis of high-throughput community
680 sequencing data. *Nat Methods* 7:335-336.
- 681 55. Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to
682 improve genome assemblies. *Bioinformatics* 27:2957-2963.
- 683 56. Bokulich NA, Joseph CM, Allen G, Benson AK, Mills DA. 2012. Next-generation

- sequencing reveals significant bacterial diversity of botrytized wine. PLoS One
7:e36357.
57. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. VSEARCH: a versatile
open source tool for metagenomics. PeerJ 4:e2584.
58. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves
sensitivity and speed of chimera detection. Bioinformatics 27:2194-2200.
59. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T,
Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA
gene database and workbench compatible with ARB. Appl Environ Microb
72:5069-5072.
60. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R.
2010. PyNAST: a flexible tool for aligning sequences to a template alignment.
Bioinformatics 26:266-267.
61. Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing Large Minimum
Evolution Trees with Profiles instead of a Distance Matrix. Mol Biol Evol
26:1641-1650.
62. Gobet A, Quince C, Ramette A. 2010. Multivariate Cutoff Level Analysis
(MultiCoLA) of large community data sets. Nucleic Acids Res 38:e155.
63. Legendre P, Gallagher ED. 2001. Ecologically meaningful transformations for
ordination of species data. Oecologia 129:271-280.
64. Borcard D, Legendre P. 2002. All-scale spatial analysis of ecological data by

- 705 means of principal coordinates of neighbour matrices. *Ecol Model* 153:51-68.
- 706 65. Lepš J, Šmilauer P. 2003. *Multivariate Analysis of Ecological Data using*
- 707 *CANOCO* doi:doi.org/10.1017/CBO9780511615146. Cambridge University
- 708 Press.
- 709 66. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. 2006. Quantifying
- 710 the roles of immigration and chance in shaping prokaryote community structure.
- 711 *Environ Microbiol* 8:732-40.
- 712 67. Team RC. 2015. *R: A language and environment for statistical computing*. Vienna:
- 713 R Foundation for Statistical Computing. <https://www.r-project.org/>.
- 714

Table. 1 Mantel tests for the correlations between community similarity and local environmental and spatial variables (using Spearman's coefficient)

Effects variables	All bacteria	Abundant bacteria	Rare bacteria
Environmental variables	0.420 ***	0.350 ***	0.349 ***
Spatial variables	0.385 ***	0.321 ***	0.464 ***

To calculate the similarity matrix, the Euclidean distance was used. Significance was tested based on 999 permutations. *** $P < 0.001$

Figure Legends

Fig. 1 Location of the 25 lakes in the middle and lower reaches of the Yangtze River Basin, China

Fig. 2 (a) The non-metric multi-dimensional scaling (NMDS) for bacterial communities in the 25 lake sampling sites. **(b)** Spearman's rank correlations between the Bray-Curtis dissimilarity of bacterial communities and geographic distance (n is the number of comparisons). All: all bacterial taxa; Abundant: abundant taxa; Rare: rare taxa.

Fig. 3 Abundance-occupancy relationship of bacterial taxa. Spearman's rank correlation between mean relative abundance of abundant bacteria (upper panel) and rare bacteria (lower panel), and number of samples occupied (n is the number of OTUs)

Fig. 4 Heat map showing community composition of abundant bacterial taxa compared with rare bacterial taxa in 25 lakes of MLYB, China

Fig. 5 (a) Ordination plots showing all, the abundant, and the rare BCCs in relation to significant local environmental variables and spatial variables; significance $P < 0.05$, $P < 0.01$, $P < 0.001$. **(b)** Variation in all, the abundant, and the rare bacterioplankton

740 metacommunity explained by environmental and spatial and variables. E|S indicates pure
 741 environmental variables, S|E pure spatial variables, E&S share explained variation.
 742 Unexplained = 1-E|S-S|E-E&S. All, Abundant, and Rare represent entire taxa, abundant
 743 taxa, and rare taxa, respectively.

Fig. 6 Fit of the neutral community model (NCM) of community assembly. The predicted occurrence frequencies are shown for all bacterial taxa. The solid black lines indicate the best fit to the NCM as in Sloan et al. (2006). *NM* indicates metacommunity size times immigration, R^2 indicates the fit to the *NM*