Similar community assembly mechanisms underlie similar biogeography of rare and abundant bacteria in lakes on Yungui Plateau, China

Jingqiu Liao,1,a Xiaofeng Cao,1, Jie Wang,1, Lei Zhao,2, Jinhua Sun,3, Dalin Jiang,1, Yi Huang1,*
1State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University, Beijing, China
2Yunnan Key Laboratory of Pollution Process and Management of Plateau Lake – Watershed, Kunming, China
3Chinese Research Academy of Environmental Sciences, Beijing, China

Abstract

The revolution of molecular techniques in the field of microbial ecology not only allows an in-depth exploration of abundant bacteria, but also confirms the existence of “rare biosphere.” However, the mechanisms underlying the assembly process of abundant and rare bacteria in plateau lakes are still poorly understood. In the present study, distribution of abundant and rare bacteria in 21 lakes on Yungui Plateau, China and the relative importance of deterministic and neutral factors to their biogeographic patterns were examined. Results surprisingly showed that similar biogeographic patterns of rare and abundant taxa were generally shaped via similar assembly mechanisms. For both bacterial groups, deterministic process strongly influenced the distribution, while neutral process was not at play. In addition, they both exhibited a weak distance-decay relationship. Nevertheless, environmental drivers of rare and abundant taxa and their importance were not exactly the same. Water physicochemical property was the most dominant driver shaping the biogeographic pattern of rare taxa, followed by lake morphology and watershed land-use, while for abundant taxa, watershed land-use was the most dominant driver and lake morphology did not play a significant role. This study suggested that rare bacteria in high-altitude aquatic ecosystems might experience equally low extinction risk and respond to environmental changes in a similar manner to abundant ones, but their ecological niches and functions were not identical. To fully understand the assembly mechanism of bacterial community, it is necessary to differentiate the community by traits of taxa.

Bacteria are typically diverse and complex organisms crucial to ecosystem function for their vital roles in global biogeochemical cycles (Xia et al. 2010; Newton et al. 2011; Liao et al. 2015a; Cao et al. 2016). The revolution of high-throughput sequencing and increasing sampling efforts have confirmed the existence of “rare bacterial biosphere,” which is composed of massive species with a few individuals (Liu et al. 2015). These rare bacteria exhibit characteristic metabolic activities (Jones and Lennon 2010; Campbell et al. 2011), carry out particular metabolic functions important for the turnover of certain elements (Pester et al. 2010; Pedrós-Alió 2012), and can be regarded as a propagule bank (Grime 1998). Compared to rare bacteria, abundant ones have been explored to a deeper degree. They appear to contribute a major function to ecosystems (Kim et al. 2013). Considering the importance of rare and abundant taxa, it is necessary to unveil their biogeographic patterns, which will provide more useful information for functional and ecological investigations (Herlemann et al. 2011; Hanson et al. 2012; Liu et al. 2015). However, it is still under debate whether rare bacterial taxa share similar biogeography with abundant ones. In order to resolve this dispute and also to interpret a global map of bacterial diversity, more efforts are in demand to characterize the biogeographic patterns of rare and abundant taxa in different environmental settings.
Community assembly, one of the important ecological processes shaping the bacterial biogeography, has implications ranging from ecosystem stability and functions to human health (Zhang et al. 2014). The mechanisms underlying the assembly process is thus an important query in the field of microbial ecology. The growing efforts in understanding these mechanisms mainly center on two schools of thought (Liao et al. 2016a,b; Sloan et al. 2006; Oťteru et al. 2010; Langenheder and Žekely 2011; Liao et al. 2016a,b; Lindström and Östman 2011; Lindström and Langenheder 2012; Nemergut et al. 2013). The first one is the deterministic mechanism, which suggests that niche-based processes, such as environmental filtering and biotic interactions, determine the biogeographic pattern of microbes (Gilbert et al. 2012). Generally, deterministic mechanism predicts that species composition will change with environmental variables (Dumbrell et al. 2010). Local-scale environmental factors, especially water physicochemical properties (e.g. temperature, salinity, pH), were most widely examined variables (Campbell et al. 2011; Ferrenberg et al. 2013; Zhang et al. 2014; Liu et al. 2015). By contrast, the relationship between landscape-scale environmental factors (e.g. land-use and lake morphology) and microbial communities has been evaluated to a lesser degree (Peralta et al. 2012; Stendera et al. 2012). To fully understand the importance of deterministic process, it is necessary to incorporate multiscalar environmental variables. The second school of thought is the neutral mechanism, in which community assembly is simply influenced by neutral factors, such as random reproduction, death, dispersal, and speciation (Vanwonterghem et al. 2014). Neutral mechanism predicts that the dissimilarity of species composition changes with geographic distance between samples, and species abundance follows a unique distribution against species frequency (Hubbell 2001; Dumbrell et al. 2010). The neutral community model (NCM) published by Sloan et al. (2006) is regarded as a testable prediction in properly modelling prokaryote communities and has validated the importance of neutral process in many studies (Ostman et al. 2010; Ayarza and Erjman 2011; Logares et al. 2013; Pholchan et al. 2013).

Obtained contrary conclusions even though the two assembly mechanisms have been explored in many studies of microbial communities, only a few studies tried to address the question whether the biogeographic patterns of abundant and rare taxa are shaped via different assembly mechanisms and they even (Crump et al. 2012; Liu et al. 2015). In the present study, we hypothesize that abundant and rare taxa are assembled via different mechanisms mainly for three reasons. First, abundant and rare species differ significantly in richness, carry out different functions in natural ecosystems, and may have discrepant ecological niches (Kim et al. 2013; Liu et al. 2015). Second, rare and abundant taxa are likely to have different ecological responses to environmental changes (Pedrós-Alió 2012; Logares et al. 2013). Third, rare and abundant taxa possibly experience entirely different extinction risk, as in large organisms, extinction risk is often associated with rarity (Caughley 1994).

To verify our hypothesis, bacterial communities in 21 lakes located on Yungui Plateau, China were investigated by 454 pyrosequencing of the 16S ribosomal RNA gene. These lakes are ideal environments to study the biogeography and assembly mechanisms of rare and abundant bacteria, since they contain rich and diverse microbes, and have a large range of physical, chemical, spatial gradients, and different watershed land use types. The importance of deterministic process was evaluated by assessing the relationship between bacterial communities and multi-scale environmental factors including water physicochemical property, lake morphology, and watershed land-use. The importance of neutral process was examined using Sloan et al. (2006)'s neutral model. We aimed to answer the following key questions: do abundant and rare bacterial taxa display similar or different biogeographic patterns? Are these two bacterial groups assembled via different assembly mechanisms? What is the quantified importance of the deterministic and neutral processes underlying their assembly?

Materials and methods

Site and sampling

Sampling sites in the 21 typical freshwater lakes on Yungui Plateau, China (Supporting Information Fig. S1) are characterized by a high altitude location of 1200 m above, subtropical plateau monsoon climate, an average annual rainfall of 1000 mm, low temperature, and high vulnerability to human disturbance (Li et al. 2005). Surface water was collected from each lake using a Ruttner water sampler (Hydro-Bios, Germany, 1000 mL) between August and September 2013. Within each lake, subsamples were taken from four locations with equal volume. Then, they were immediately placed on ice and conveyed to lab for processing. In order to establish a broader species and remove temporal and within-lake spatial variability, four subsamples from each lake were pooled into one (Miller and McMahon 2011; Dadheech et al. 2013). Water samples were pumped through 0.22 μm Sterivex-GP filters (Millipore) and the filters were kept at −80°C until DNA extraction.

DNA extraction, PCR amplification, and 454 pyrosequencing

E.Z.N.A. D5525-01 Water DNA Kit (Omega Bio-tek, Norcross, Georgia, U.S.A.) was used for microbial DNA extraction. The V1–V3 region of the 16S ribosomal RNA gene of bacteria was amplified by primers 27F(5’-AGAGTTTGATCCTGGCTCAG-3’) and 533R(5’-TTACCGCGGGCTGC TGGCAC-3’) in polymerase chain reaction (PCR) (Wu et al. 2012). PCR reactions (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 at 72°C for 5 min) were performed in a 20 μL
mixture 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 mixture of amplicons was ligated to Roche 454 GS FLX+ Titanium platform (Roche 454 Life Sciences, Branford, Connecticut, U.S.A.) after purification by AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, California, U.S.A.) and quantification by QuantiFluor™-ST (Promega, U.S.A.) and platform (Roche 454 Life Sciences, Branford, Connecticut, U.S.A.)

Bioinformatics analysis

MOTHUR version 1.33.3 (Schloss et al. 2009) was used to convert pyrosequencing flowgrams to sequence reads. QIIME version 1.3.0 (Caporaso et al. 2010) was subsequently employed to remove barcodes on the sequences and high quality reads (sequences lengths < 200 bp, quality scores < 25) and ambiguous reads (ambiguous base > 0). To further improve quality, resulting sequences were denoised using DeNoiser version 0.851 (Reeder and Knight 2010) and putative chimeras were detected and excluded with UCHIME (Edgar et al. 2011). All Archaea, Eukaryota, chloroplasts, mitochondria, and unknown sequences were removed. Finally, high-quality sequences were clustered into operational taxonomic units (OTUs) at a 3% dissimilarity level using the average neighbor method and taxonomy was assigned via the SILVA database using k-mer searching method (Pruesse et al. 2007). The OTUs table was rarefied to 7909 reads per sample.

Identification of abundant taxa, rare taxa, habitat generalists, and specialists

The thresholds for regionally abundant and rare taxa were set based on the mean relative abundance distribution and recent publications. Regionally abundant taxa were defined as the OTUs > 0.1% across all samples according to Liu et al. (2015). This value (log<sub>10</sub>0.001 = −3) lies within the outlier area of the mean relative abundance distribution (Supporting Information Fig. S2), which matches previous finding that most microbial communities consist of a few abundant taxa (Pedrós-Alió 2006, 2012; Logares et al. 2013; Liu et al. 2015). Regionally rare taxa were defined as the OTUs with a mean relative abundance of < 0.006% across all samples according to Logares et al. (2013). This value (log<sub>10</sub>0.00006 = −4.2) selects nearly half of the OTUs as rare taxa according to the mean relative abundance distribution in Supporting Information Fig. S2, which also accords with previous finding that most microbial communities are composed of a great many rare taxa (Pedrós-Alió 2006, 2012; Logares et al. 2013; Liu et al. 2015). To confirm the rationality of the cutoffs chosen in defining rare and abundant taxa and to reduce the effect of arbitrariness, multivariate cutoff level analysis (MultiCoLA) was further conducted in MultiCoLA.1.4 based on Gobet et al. (2010) and Liu et al. (2015).

Habitat generalists were defined according to “niche breadth” approach (Levins 1968; Logares et al. 2013). Specialists were defined by both “niche breadth” and INDicator VALues (INDVAL) analysis (Dufrene and Legendre 1997). For further details, please see Extended Experimental Procedures in Supporting Information.

Environmental variables

Latitude and longitude were measured using the Global Positioning System (GPS) during the field work. Environmental variables were considered in terms of water physicochemical property, lake morphology, and watershed land-use. Ammonium-nitrogen (NH₄-N), total Kjeldahl nitrogen (TKN), total nitrogen (TN), soluble reactive phosphorus (SRP), and total phosphorus (TP) were measured following USEPA (1983). Nitrate-nitrogen (NO₃-N) was measured according to APHA (1998). Chemiluminescence detection of permanganate index (CODmn) was measured following the acidic potassium permanganate method (EPAC 2002). Salinity, pH, dissolved oxygen (DO), and temperature were measured in situ using a Horiba W-23XD Multi-Parameter Water Quality Meter (Horiba, Japan). Transparency (SD) was determined by Secchi disk according to Davies-Colley (1988). Classification, coverage of watershed land-use and lake area were obtained using a supervised classification method based on the 2012 SPOT 5 satellite images with a 2.5 m resolution in ENVI 4.7 (http://www.harrisgeospatial.com/). Map of research area (Supporting Information Fig. S1) was created by ArcGIS 10.1 (http://www.esri.com/software/arcgis/arcgis-for-desktop) using the vector data of lake and watershed boundary provided by the Yunnan Key Laboratory of Pollution Process and Management of Plateau Lake-Watershed. Environmental variable dataset was listed in Supporting Information Table S1.

Variable selection was conducted by Pearson’s correlation analysis using SPSS version 17 to avoid the arching effect in canonical correspondence analysis (CCA) and redundancy analysis (RDA) (Supporting Information Tables S2, S3). Among the water physicochemical variables, NO₃-N and TKN were excluded for their significant correlation with TN (r₁, r₂ > 0.700, p < 0.01) and SRP was excluded for its low content (Supporting Information Table S1). CODmn was excluded for its significant correlation with TP (r > 0.7, p < 0.01). Eight variables—salinity, pH, DO, temperature, SD, NH₄-N, TN, and TP—were selected. Among the watershed land-use variables, forestland had a significant negative correlation with agricultural land (r = −0.921, p < 0.01), urban land (r = −0.823, p < 0.01) and barren land (r = −0.565, p < 0.01). Agricultural land, grassland, water body, urban land, and barren land were selected for further analyses.

Neutral model

The neutral community model (NCM) developed by Sloan et al. (2006, 2007) has demonstrated that stochastic community assembly is associated with a characteristic relationship
between abundance and frequency of microbes detected in multiple local communities (Baptista et al. 2008). This model employs the least-squares method to generate the best fit distribution curve with associated values for the distribution parameters $\alpha$ and $\beta$ against regional relative abundance ($p_i$) for each OTU$_{0.03}$. $\alpha$ and $\beta$ are functions of $p_i$ and $N_T/m$, where $N_T$ is the number of sequence in the local community and $m$ is the rate of the individuals immigrating from the source community into the local community. The lowest relative abundance detected for OTU$_{0.03}$ in the present study was 1/7909, which was defined as the detection limit (d) of this model. The proportion of variability $R^2$ which quantified the fit level of detection frequency to the model was calculated according to Ostman et al. (2010)’s method. Model calculations were performed using MATLAB 7.10.0 (Mathworks, Natick, Massachusetts, U.S.A.).

**Statistical analysis**

Good’s coverage index and Chao1 richness were calculated using MOTHUR 1.33.3 (Schloss et al. 2009). The dissimilarity matrix of bacterial composition was determined by Bray–Curtis distance. The geographical distance was calculated from latitude and longitude coordinates using Euclidean distance in R language environment (R Development Core Team 2008). Spearman’s rank correlation between the dissimilarity of bacterial community and the geographical distance were measured using SPSS version 17. Principal coordinate analysis (PCoA) by CANOCO for Windows Version 5.0 was employed to visually represent the dissimilarity of bacterial composition among samples. Linear regression analysis and Mantel test were performed in R language environment (R Development Core Team 2008). Based on the lengths of gradient, CCA and RDA by CANOCO for Windows Version 5.0 were carried out to quantify the effects of selected environmental variables on biological variation of rare and abundant taxa, respectively. The contribution of water physiochemical property (P), lake morphology (M), and watershed land-use (L) to the distribution of rare and abundant taxa was evaluated using variance partitioning analysis (VPA) by CANOCO for Windows Version 4.5 (Plant Research International, Wageningen, The Netherlands). The significance test was carried out by Monte Carlo permutation (999 times).

**Result**

**Bacterial abundance and multivariate abundance cutoff**

In total, 166,089 high quality sequences were recovered from 21 water samples and were grouped into 1635 OTU$_{0.03}$. Good’s coverage ranged from 97.26% to 99.47%, showing that majority of the bacterial communities was well-captured by this sequencing effort (Supporting Information Table S4). Among the total OTU$_{0.03}$, 173 (10.58%) abundant taxa with 126,971 (76.45%) sequences and 744 (45.50%) rare taxa with 4643 (2.8%) sequences were identified. MultiCoLA showed that little variation in the data set structure and main patterns of community was observed up to a removal of 30% of the rare types when comparing the data structure between the original and truncated matrices (Supporting Information Fig. S3b,c). Meanwhile, when the number of rare types increased to > 5%, the data structure of truncated matrices exhibited a little variation (Supporting Information Fig. S3e,f). These results indicated that the cutoffs for defining abundant taxa (23.55%) and rare taxa (2.8%) were appropriate.

**Biogeographic patterns and taxonomic compositions of rare and abundant taxa**

Abundant taxa occupied the entire range of sites, while 80.7% of rare taxa distributed in only one or two sites and none of them occupied more than six sites (Fig. 1). Both generalists (24 out of 173, 13.9%) and specialists (23 out of 173, 13.3%) were present among abundant taxa, whereas only specialists were detected among rare taxa (42 out of 744, 5.6%) (Fig. 1). The relative abundance of abundant taxa ($r = 0.328$, $p < 0.001$) was less positively correlated to local occupancy compared to rare taxa ($r = 0.481$, $p < 0.001$).

Similar biogeographic patterns of rare and abundant taxa were observed in PCoA plots, both showing considerable heterogeneity and four similar clusters of samples (Fig. 2a,b). LSH and YH closely clustered together in both plots, indicating a similar composition of rare taxa and also abundant taxa. In-depth analysis of taxonomic composition identified 26 phyla among entire bacteria (Supporting Information Table S5). *Proteobacteria* and *Bacteroidetes* predominated within both abundant and rare taxa. All of the rare taxa except for *WCHB1-60* were present within each phylum, while abundant taxa were only detected in 10 phyla—*Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Firmicutes*, *Cyanobacteria*, *Verrucomicrobia*, *Planctomycetes*, *Chloroflexi*, *Armatimonadetes*, and *Deinococcus-Thermus* (Supporting Information Fig. S5, Table S5).

![Fig. 1. Abundance-occupancy relationship of abundant and rare taxa.](image)

**Table S4.** Among the total OTUs 0.03, 173 (10.58%) well-captured by this sequencing effort (Supporting Information). Showing that majority of the bacterial communities was well-captured by this sequencing effort (Supporting Information Fig. S3b,c). Meanwhile, when the number of rare types increased to > 5%, the data structure of truncated matrices exhibited a little variation (Supporting Information Fig. S3e,f). These results indicated that the cutoffs for defining abundant taxa (23.55%) and rare taxa (2.8%) were appropriate.
Relation between community composition and geographic distance

Linear regression analysis showed that the community dissimilarity of rare taxa had little relationship with the geographic distance ($R^2 = 0.04$, Fig. 3a). Similarly, the community dissimilarity of abundant taxa were weakly associated with the geographic distance ($R^2 = 0.15$, Fig. 3b). Spearman’s correlation analysis gave a correlation coefficient of 0.19 ($p = 0.000$) and 0.24 ($p = 0.000$) between the dissimilarity of rare and abundant taxa and geographic distance, respectively. Moreover, in Mantel tests neither of the correlation between the dissimilarity of each group and geographic distance was tight (rare taxa, $r = 0.18$, $p < 0.001$; abundant taxa, $r = 0.33$, $p < 0.001$). Overall, the community dissimilarity of rare and abundant taxa was both weakly correlated to geographic distance, indicating a significantly minor distance-decay relationship.

The importance of neutral process in governing community assembly

NCM which evaluates the importance of neutral factors in the assembly process of bacterial community has been separately conducted on rare and abundant taxa. The best-fit neutral curve indicated the relationship between detection frequency and mean relative abundance ($p_i$) of OTU$_{0.03}$ purely driven by neutral process. For the entire bacterial communities, the product $N_{0.03}m$ was 589 generated by this model. Since the number of sequences in each sample was 7909, the $m$ value was estimated to be 0.07. Results showed that the observed data for rare taxa displayed large departure from the best-fit neutral curve (Fig. 4a, $R^2 = -0.57$; negative $R^2$ values occur when there is no fit to the model) (Logares et al. 2013). Thus, neutral process was not at play in the community assembly of rare taxa. As for the abundant taxa, the detected frequency also didn’t follow the frequency predicted by NCM (Fig. 4b, $R^2 = -0.38$), indicating that neutral process was not important for the community assembly of abundant taxa, either.

The importance of deterministic process in governing community assembly

CCA revealed that among the selected water physicochemical factors, six variables—salinity, DO, TP, TN, SD, pH—significantly influenced the distribution of rare taxa (Table 1). Both lake morphology and watershed land-use were important drivers as well (Table 1). All the land-use types—urban land, agricultural land, barren land, grassland, water body—were identified as significant variables for rare taxa ($p < 0.05$). By contrast, only variables of water physicochemical property and watershed land-use revealed by RDA Fig. 2. PCoA plots of (a) rare taxa and (b) abundant taxa in the 21 plateau lakes. (a) PCo Axis 1 explained 8.18% of the variation of rare taxa; PCo Axis 2 explained 7.55% of the variation. Yellow circle, red circle, green circle, and blue circle indicated four clusters of samples. (b) PCo Axis 1 explained 19.79% of the variation of abundant taxa; PCo Axis 2 explained 11.44% of the variation. Yellow circle, red circle, green circle, and blue circle indicated four clusters of samples.
Fig. 3. Spearman’s rank correlations between the Bray–Curtis dissimilarity of (a) rare taxa and (b) abundant taxa and geographical distance.
had a significant impact on abundant taxa. Among variables of water physicochemical property, salinity, DO, and TN were responsible for shaping their distribution (Table 1). Except for water body, all the land-use types played an important role (Table 1). Only the significant environmental variables were further included in VPA, to specifically evaluate the independent influence of water physicochemical property (P), lake morphology (M) and watershed land-use (L), and to better differentiate the importance of deterministic process between the two bacterial groups.

VPA showed that water physicochemical property was the most dominant driver governing rare taxa, independently explaining 28.9% of the variation (Fig. 5). Watershed land-use was another important driver, independently explaining

![Fig. 4. Detection frequencies of different OTUs as a function of mean relative abundance for (a) rare taxa and (b) abundant taxa. The solid black line was the best-fit (least square error) neutral curve.](image)
25.2% of the variation, followed by lake morphology (4.6%). A total of 36.9% of the variation of rare taxa could not be explained. For abundant taxa, watershed land-use was the most dominant driver (23.6%) followed by water physiochemical property (12.9%). The interaction between water physiochemical property and watershed land-use considerably influenced the distribution of abundant taxa, explaining 9.8% of the variation. A total of 53.7% of the variation of abundant taxa could not be explained. These results indicated that although deterministic process was critical to both rare and abundant taxa, the importance of each environmental driver differed between them.

**Discussion**

“Rare biosphere” is of great importance in metabolic and ecological functions (Grime 1998; Pedrós-Alió 2012). Rare bacteria have been explored in both natural environments such as Arctic Ocean (Galand et al. 2009) and coastal Antarctic lakes (Logares et al. 2013), and artificial environments such as wastewater treatment reactors (Kim et al. 2013). However, plenty of issues remain unaddressed. For example, what are the spatial distributions of rare bacterial taxa in different environmental settings? How does the biogeography of rare bacterial taxa differ from that of abundant ones? What mechanism is underlying the assembly of rare taxa, and how does it distinct from the mechanism for abundant taxa? This study focusing on rare and abundant taxa across multiple plateau lakes has provided critical information to answer these questions.

Rare taxa exhibited a biogeographic pattern similar to that of abundant taxa (Fig. 2). This similarity implied that the rare bacterial biosphere was not a random collection of taxa (Liu et al. 2015). More importantly, it suggested that rare taxa might respond to environmental changes in a similar manner (Logares et al. 2013) and experience equally low extinction risk as abundant ones (Liu et al. 2015). This scenario of microbial ecology largely differed from that of macro-organism ecology where extinction risk was often associated with rarity (Caughley 1994). The low extinction risk of rare taxa might result from their low abundance, which has protected them from active loss by both viral lysis and predation (Pedrós-Alió 2006). As the abundance of any given bacterium is a result of balance between its growth rate and loss rate, the growth rate of rare taxa should be very low inferred from their low abundance and extinction risk. Previous finding that a large fraction of the rare biosphere was made up of very slowly growing or even non-growing cells validated this inference (Pedrós-Alió 2006). Interestingly, these viable yet relatively inactive populations could grow into abundant taxa, and vice versa, when environmental conditions change or disturbance occurs (Finlay 2002; Campbell et al. 2011; Pedrós-Alió 2012; Sjöstedt et al. 2012). The observation of similar biogeographic patterns of rare and abundant taxa was consistent with previous conclusions from Galand et al. (2009), Logares et al. (2013), and Liu et al. (2015), but contrasted to that from Kim et al. (2013). These different conclusions possibly resulted from different environment settings under study. The first three studies were all carried out in natural conditions, while the latter two studies were conducted in artificial environments.

### Table 1. Summarize effects of selected abiotic variables.

| Variables                      | Explains % | Pseudo-$F_p$ | $p$ | Explains % | Pseudo-$F_p$ | $p$ |
|--------------------------------|------------|--------------|-----|------------|--------------|-----|
| **Water physicochemical variables** |            |              |     |            |              |     |
| Salinity                       | 7.1        | 1.5          | 0.004 | 12.7       | 2.8          | 0.002 |
| DO                             | 6.9        | 1.4          | 0.016 | 9.8        | 2.1          | 0.018 |
| TP                             | 6.6        | 1.3          | 0.008 | 6.6        | 1.3          | 0.194 |
| TN                             | 6.0        | 1.2          | 0.036 | 8.8        | 1.8          | 0.046 |
| SD                             | 6.1        | 1.2          | 0.016 | 4.8        | 1.0          | 0.458 |
| pH                             | 5.8        | 1.2          | 0.040 | 3.7        | 0.7          | 0.790 |
| NH$_4$-N                       | 5.4        | 1.1          | 0.126 | 5.5        | 1.1          | 0.280 |
| Temperature                    | 5.3        | 1.1          | 0.118 | 7.3        | 1.5          | 0.108 |
| **Lake morphological variables** |            |              |     |            |              |     |
| Area                           | 6.4        | 1.3          | 0.010 | 4.0        | 0.8          | 0.682 |
| **Watershed land-use variables** |            |              |     |            |              |     |
| Urban land                     | 7.0        | 1.4          | 0.006 | 13.7       | 3.0          | 0.010 |
| Agricultural land              | 6.4        | 1.3          | 0.002 | 18.8       | 4.4          | 0.002 |
| Barren land                    | 5.7        | 1.1          | 0.020 | 10.7       | 2.3          | 0.018 |
| Grassland                      | 5.8        | 1.2          | 0.014 | 10.3       | 2.2          | 0.02  |
| Water body                     | 5.8        | 1.2          | 0.026 | 4.80       | 1.0          | 0.462 |

Significant values ($p < 0.05$) are in bold.
environments, while later one was proceeded in artificial environment—activated sludge wastewater treatment bioreactor. Rare taxa in natural environments might more likely exhibit a similar biogeographic pattern to abundant taxa. In general, the similar biogeographic patterns of rare and abundant taxa were shaped via similar assembly mechanisms. Neutral effects on the community assembly of both groups were not detected (Fig. 4), indicating that stochastic events such as chance and random dispersal were unimportant in their assembly. The significantly weak distance-decay relationship of rare and abundant taxa further confirmed the unimportance of neutral process (Fig. 3), since in Hubbell’s neutral theory, community dissimilarity was predicted to increase along spatial (distance) gradients owing to dispersal limitation (Chase and Myers 2011). By contrast, deterministic factors exhibited significant importance (Table 1; Fig. 5), indicating that environmental condition has strongly selected rare and abundant taxa. The importance of deterministic mechanism implied that most species in these two groups were not ecologically equivalent and had strict requirement for the environmental conditions. However, Liu et al. (2015) draw the conclusion that the abundant sub-communities were mainly governed by stochastic process. This contrary conclusion might result from the different proportion of generalists detected among abundant taxa, since our previous study demonstrated that the assembly of generalists was strongly governed by neutral process (Liao et al. 2016a). In Liu et al. (2015)’s study, most abundant bacteria appeared to be habitat generalists, while in the present study, only 24 out of 173 (13.87%) abundant taxa were detected as generalists. These findings corroborated Wang et al. (2013)’s inference that the importance of neutral factors for the assembly of abundant taxa is influenced by the proportion of generalists among them. The importance of deterministic process could be overwhelmed by that of neutral process, when regional abundant species pools were characterized by much more generalists. In addition, the range of research zone and habitat type might be two additional factors affecting the importance of neutral process to abundant taxa. In Liu et al. (2015)’s study, 42 sites were under consideration, which exhibited larger geographic gradient. The majority of lakes are shallow lake (water depth < 10 m) which spanned at a continental scale. However, in the present study, only 21 lakes were studied and some of them are deep lakes with a water depth > 30 m.

Although deterministic process was critical to both rare and abundant taxa, the type of environmental drivers and their individual importance differed between them (Table 1; Fig. 5). It suggested that ecological niches and functions of abundant and rare bacteria were not exactly identical. Water physicochemical property, which was identified as an important environmental driver to bacterial communities in many ecosystems (Liao et al. 2015b), was the most dominant driver governing rare taxa. This was consistent with Liu et al. (2015)’s finding. Significant physicochemical variables for rare taxa were identified as salinity, DO, TP, TN, SD, and pH, while these for abundant taxa were only salinity, DO and TN (Table 1). Drastic change in these variables might strongly alter the abundance and composition of rare and abundant taxa. As the common significant factors for both groups, salinity, DO and TN have been frequently demonstrated to be important stressor on bacterial composition (Lozupone and Knight 2007; Liu et al. 2011; Kim et al. 2013). The reasons why TP, SD, and pH significantly affected the distribution of the rare taxa rather than abundant taxa were uncertain and warranted further investigation.

Fig. 5. Contribution of environmental divers to the distribution of (a) rare taxa and (b) abundant taxa. P indicates water physicochemical property; M indicates lake morphology; L indicates watershed land-use. Diagram (a) represents the biological variation of rare taxa partitioned into the independent effects of P, M and L, interaction between the any two factors (P × M, P × L, and M × L), interaction of all three factors (P × M × L), and the unexplained part. The geometric areas of circles were proportional to the respective percentages of explained variation.
investigation. The possible explanations were: most abundant taxa in the present study did not function in phosphorous transformation or had sufficient intracellular storage of phosphorus; most of them were not photoheterotrophs or photoautotrophs restricted by light limitation, or had various strategies to minimize damage of UV radiation caused by high transparency; they were not sensitive to environmental pH and could tolerate its significant deviation. As an important stressor on bacteria (Hartman et al. 2008; Stendera et al. 2012), watershed land-use was identified as the most important driver for abundant taxa. It suggested that abundant taxa might be more sensitive to emerging land use practices than rare taxa. Lake morphology and water body in lake catchment had a significant impact on rare taxa rather than abundant taxa. It was possibly that rare taxa could more sensitively respond to changes in habitat space and water resource, or have more likely exhibited the species–area relationship, which is characterized by the increase of the observed number of species with increasing sample area (O’Dwyer and Green 2010). The unexplained biological variation in Fig. 5 might be caused by additional unmeasured but important factors such as dissolved organic matter, which could influence aquatic bacterial communities by providing carbon source (Langenheder and Ragnarsson 2007); water depth, which has proven to significantly drive the distribution and diversity of bacteria along the water column via vertical stratification of limnological properties (Yu et al. 2014a,b, 2015); and temporal factor, as long-term environmental changes could promote the assembly of bacterial communities (Logares et al. 2013).

Conclusions

This study has provided a better understanding of rare and abundant bacteria across multiple freshwater plateau lakes. It demonstrated that rare taxa exhibited similar biogeographic pattern to abundant taxa, and their patterns were generally shaped via similar assembly mechanisms. For both groups, deterministic process played an important role, while neutral process was not at play. Water physicochemical property was the most dominant driver in governing rare taxa, while watershed land-use had the strongest impact on abundant taxa. To fully understand the assembly mechanism of bacterial communities, it is suggested to distinguish bacteria by the traits of taxa, e.g., richness and habitat specialization, and to include more important factors, e.g., water depth and time series in future studies.

References

APHA. 1998. Standard methods for the examination of water and wastewater, 20th ed, p. 1023. APHA, AWWA, WEF.

Ayarza, J. M., and L. Erijman. 2011. Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. Microb. Ecol. 61: 486–495. doi:10.1007/s00248-010-9762-y

Baptista, J. D. C., R. J. Davenport, T. Donnelly, and T. P. Curtis. 2008. The microbial diversity of laboratory-scale wetlands appears to be randomly assembled. Water Res. 42: 3182–3190. doi:10.1016/j.watres.2008.03.013

Campbell, B. J., L. Yu, J. F. Heidelberg, and D. L. Kirchman. 2011. Activity of abundant and rare bacteria in a coastal ocean. Proc. Natl. Acad. Sci. USA 108: 12776–12781. doi:10.1073/pnas.1101405108

Cao, X., J. Wang, J. Liao, J. Sun, and Y. Huang. 2016. The threshold responses of phytoplankton community to nutrient gradient in a shallow eutrophic Chinese lake. Ecol. Indic. 61: 258–267. doi:10.1016/j.ecolind.2015.09.025

Caporaso, J. G., and others. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7: 335–336. doi:10.1038/nmeth.f.303

Caughley, G. 1994. Directions in conservation biology. J. Anim. Ecol. 63: 215–244. doi:10.2307/5542

Chase, J. M., and J. A. Myers. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366: 2351–2363. doi:10.1098/rstb.2011.0063

Crump, B. C., L. A. Amaral-Zettler, and G. W. Kbling. 2012. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. ISME J. 6: 1629–1639. doi:10.1038/ismej.2012.9

Dadheech, P. K., G. Glöckner, P. Casper, K. Kotut, C. J. Mazzoni, S. Mbedi, and L. Krienitz. 2013. Cyanobacterial diversity in the hot spring, pelagic and benthic habitats of a tropical soda lake. FEMS Microbiol. Ecol. 85: 389–401. doi:10.1111/1574-6941.12128

Davies-Colley, R. J. 1988. Measuring water clarity with a black disk. Limnol. Oceanogr. 33: 616–623. doi:10.4319/lo.1988.33.4.0616

Dufrene, M., and P. Legendre. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecol. Monogr. 67: 345–366. doi:10.2307/2963459

Dumbrell, A. J., M. Nelson, T. Helgason, C. Dytham, and A. H. Fitter. 2010. Relative roles of niche and neutral processes in structuring a soil microbial community. ISME J. 4: 337–345. doi:10.1038/ismej.2009.122

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194–2200. doi:10.1093/bioinformatics/btr381

EPAC (Environmental Protection Agency of China). 2002. Standard methods for the examination of water and waste water, 4th ed, pp. 200–284. Chinese Environmental Science Press. Ferrenberg, S., and others. 2013. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. ISME J. 7: 1102–1111. doi:10.1038/ismej.2013.11
Finlay, B. J. 2002. Global dispersal of free-living microbial eukaryote species. Science 296: 1061–1063. doi:10.1126/science.1070710

Galand, P. E., E. O. Casamayor, D. L. Kirchman, and C. Lovejoy. 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. Proc. Natl. Acad. Sci. USA 106: 22427–22432. doi:10.1073/pnas.0908284106

Gilbert, J. A., and others. 2012. Defining seasonal marine microbial community dynamics. ISME J. 6: 298–308. doi:10.1038/ismej.2011.107

Gobet, A., C. Quince, and A. Ramette. 2010. Multivariate Cutoff Level Analysis (MultiCoLA) of large community data sets. Nucleic Acids Res. 38: e155–e155. doi:10.1093/nar/gkq545

Grime, J. P. 1998. Benefits of plant diversity to ecosystems: Immediate, filter and founder effects. J. Ecol. 86: 902–910. doi:10.1046/j.1365-2745.1998.00306.x

Hanson, C. A., J. A. Fuhrman, M. C. Horner-Devine, and J. B. Martiny. 2012. Beyond biogeographic patterns: Processes shaping the microbial landscape. Nat. Rev. Microbiol. 10: 497–506. doi:10.1038/nrmicro2795

Hartman, W. H., C. J. Richardson, R. Vilgalys, and G. L. Bruland. 2008. Environmental and anthropogenic controls over bacterial communities in wetland soils. Proc. Natl. Acad. Sci. USA 105: 17842–17847. doi:10.1073/pnas.0808254105

Hartemink, D. P., M. Labrenz, K. Jürgens, S. Bertilsson, J. J. Waniek, and A. F. Andersson. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 5: 1571–1579. doi:10.1038/ismej.2011.41

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography (MPB-32), v. 32. Princeton Univ. Press.

Jones, S. E., and J. T. Lennon. 2010. Dormancy contributes to the maintenance of microbial diversity. Proc. Natl. Acad. Sci. USA 107: 5881–5886. doi:10.1073/pnas.0912765107

Kim, T. S., J. Y. Jeong, G. F. Wells, and H. D. Park. 2013. General and rare bacterial taxa demonstrating different temporal dynamic patterns in an activated sludge bioreactor. Appl. Microbiol. Biotechnol. 97: 1755–1765. doi:10.1007/s00253-012-4002-7

Langenheder, S., and H. Ragnarsson. 2007. The role of environmental and spatial factors for the composition of aquatic bacterial communities. Ecology 88: 2154–2161. doi:10.1890/06-2098.1

Langenheder, S., and A. J. Székely. 2011. Species sorting and neutral processes are both important during the initial assembly of bacterial communities. ISME J. 5: 1086–1094. doi:10.1038/ismej.2010.207

Levins, R. 1968. Evolution in changing environments: Some theoretical explorations (No. 2). Princeton Univ. Press.

Li, C., S. Zhang, and X. Ye. 2005. Problems of development and protection of lakes on Yunnan Plateau (In Chinese). Yunnan Geol. 24: 462–470. doi:1004-1885(2005)05-462-09

Liao, J., J. Wang, D. Jiang, M. C. Wang, and Y. Huang. 2015a. Long-term oil contamination causes similar changes in microbial communities of two distinct soils. Appl. Microbiol. Biotechnol. 99: 10299–10310. doi:10.1007/s00253-015-6880-y

Liao, J., J. Wang, and Y. Huang. 2015b. Bacterial community features are shaped by geographic location, physicochemical properties and oil contamination of soil in main oil fields of China. Microb. Ecol. 70: 380–389. doi:10.1007/s00440-015-0572-0

Liao, J., X. Cao, L. Zhao, J. Wang, Z. Gao, M. C. Wang, and Y. Huang. 2016a. The importance of neutral and niche processes for bacterial community assembly differs between habitat generalists and specialists. FEMS Microbiol. Ecol. 92: fiw174. doi:10.1093/femslec/fiw174

Liao, J., L. Zhao, X. Cao, J. Sun, Z. Gao, J. Wang, D. Jiang, H. Fan, Y. Huang. 2016b. Cyanobacteria in lakes on Yungui Plateau, China are assembled via niche processes driven by water physicochemical property, lake morphology and watershed land-use. Sci. Rep. 6: 36357. doi:10.1038/srep36357

Lindström, E. S., and Ö. Östman. 2011. The importance of dispersal for bacterial community composition and functioning. PloS One 6: e25883–e25883. doi:10.1371/journal.pone.0025883

Lindström, E. S., and S. Langenheder. 2012. Local and regional factors influencing bacterial community assembly. Environ. Microbiol. Rep. 4: 1–9. doi:10.1111/j.1758-2229.2011.00257.x

Liu, L., J. Yang, and Y. Zhang. 2011. Genetic diversity patterns of microbial communities in a subtropical riverine ecosystem (Jiulong River, southeast China). Hydrobiologia 678: 113–125. doi:10.1007/s10750-011-0834-x

Liu, J., J. Yang, Z. Yu, and D. M. Wilkinson. 2015. The bio-geo-ography of abundant and rare bacterioplankton in the lakes and reservoirs of China. ISME J. 9: 2068–2077. doi:10.1038/ismej.2015.29.

Logares, R., and others. 2013. Biogeography of bacterial communities exposed to progressive long-term environmental change. ISME J. 7: 937–948. doi:10.1038/ismej.2012.168

Lozupone, C. A., and R. Knight. 2007. Global patterns in bacterial diversity. Proc. Natl. Acad. Sci. USA 104: 11436–11440. doi:10.1073/pnas.0611525104

Margulis, M., and others. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380. doi:10.1038/nature03959

Miller, T. R., and K. D. McMahon. 2011. Genetic diversity of cyanobacteria in four eutrophic lakes. FEMS Microbiol. Ecol. 78: 336–348. doi:10.1111/j.1574-6941.2011.01162.x

Nemerow, D. R., and others. 2013. Patterns and processes of microbial community assembly. Microbiol. Mol. Biol. Rev. 77: 342–356. doi:10.1128/MMBR.00051-12
Newton, R. J., S. E. Jones, A. Eiller, K. D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. Microbiol. Mol. Biol. Rev. 75: 14–49. doi:10.1128/MMBR.00028-10

O’Dwyer, J. P., and J. L. Green. 2010. Field theory for biogeography: A spatially explicit model for predicting patterns of biodiversity. Ecol. Lett. 13: 87–95. doi:10.1111/j.1461-0248.2009.01404.x

Ofir, I. D., M. Lunn, T. P. Curtis, G. F. Wells, C. S. Criddle, C. A. Francis, and W. T. Sloan. 2010. Combined niche and neutral effects in a microbial wastewater treatment community. Proc. Natl. Acad. Sci. USA 107: 15345–15350. doi:10.1073/pnas.1000604107

Östman, Ö., S. Drakare, E. S. Kritzberg, S. Langenheder, J. B. Logue, and E. S. Lindström. 2010. Regional invariance among microbial communities. Ecol. Lett. 13: 118–127. doi:10.1111/j.1461-0248.2009.01413.x

Pedrós-Alió, C. 2006. Marine microbial diversity: Can it be determined? Trends Microbiol. 14: 257–263. doi:10.1016/j.tim.2006.04.007

Pedrós-Alió, C. 2012. The rare bacterial biosphere. Annu. Rev. Mar. Sci. 4: 449–466. doi:10.1146/annurev-marine-120710-100948

Peralta, A. L., J. W. Matthews, D. N. Flanagan, and A. D. Kent. 2012. Environmental factors at dissimilar spatial scales influence plant and microbial communities in restored wetlands. Wetlands 32: 1125–1134. doi:10.1007/s13157-012-0343-3

Pester, M., N. Bittner, P. Deevong, M. Wagner, and A. Loy. 2010. A ‘rare biosphere’ microorganism contributes to sulfate reduction in a peatland. ISME J. 4: 1591–1602. doi:10.1038/isemj.2010.75

Pholchan, M. K., D. C. Baptista, R. J. Davenport, W. T. Sloan, and T. P. Curtis. 2013. Microbial community assembly, theory and rare functions. Front. Microbiol. 4: 68. doi:10.3389/fmicb.2013.00068

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner. 2007. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35: 7188–7196. doi:10.1093/nar/gkm864

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available from http://www.R-project.org.

Reeder, J., and R. Knight. 2010. Rapidly denoising pyrosequencing amplicons read by exploiting rank-abundance distributions. Nat. Methods 7: 668–669. doi:10.1038/nmeth0910-668b

Schloss, P. D., and others. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75: 7537–7541. doi:10.1128/AEM.01541-09

Sjöstedt, J., P. Koch-Schmidt, M. Pontarp, B. Canbäck, A. Tunlid, P. Lundberg, Å. Hagström, and L. Riemann. 2012. Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. Appl. Environ. Microbiol. 78: 1361–1369. doi:10.1038/isemj.2015.29

Sloan, W. T., M. Lunn, S. Woodcock, I. M. Head, S. Nee, and T. P. Curtis. 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ. Microbiol. 8: 732–740. doi:10.1111/j.1462-2920.2005.00956.x

Sloan, W. T., S. Woodcock, M. Lunn, I. M. Head, and T. P. Curtis. 2007. Modeling taxabundance distributions in microbial communities using environmental sequence data. Microb. Ecol. 53: 443–455. doi:10.1007/s00248-006-9141-x

Stendera, S., R. Adrian, N. Bonada, M. Cañedo-Argüelles, B. Hugueny, K. Januschke, F. Pletterbauer, and D. Hering. 2012. Drivers and stressors of freshwater biodiversity patterns across different ecosystems and scales: A review. Hydrobiologia 696: 1–28. doi:10.1007/s10750-012-1183-0

USEPA. 1983. Methods for chemical analysis of water and wastes. U.S. Environmental Protection Agency, Cincinnati, OH. EPA-600/-4-79-020.

Vanwringerhem, L., P. D. Jensen, P. G. Dennis, P. Hugenholtz, K. Rabaey, and G. W. Tyson. 2014. Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. ISME J. 8: 2015–2028. doi:10.1038/isemj.2014.50

Wang, J., and others. 2013. Phylogenetic beta diversity in bacterial assemblages across ecosystems: Deterministic versus stochastic processes. ISME J. 7: 1310–1321. doi:10.1038/isemj.2013.30

Wu, S., G. Wang, E. R. Angert, W. Wang, W. Li, and H. Zou. 2012. Composition, diversity, and origin of the bacterial community in grass carp intestine. PLoS One 7: e30440. doi:10.1371/journal.pone.0030440

Xia, S., and others. 2010. Bacterial community structure in geographically distributed biological wastewater treatment reactors. Environ. Sci. Technol. 44: 7391–7396. doi:10.1021/es101554m

Yu, Z., J. Yang, S. Amalfitano, X. Yu, and L. Liu. 2014a. Effects of water stratification and mixing on microbial community structure in a subtropical deep reservoir. Sci. Rep. 4: 5821. doi:10.1038/srep05821

Yu, Z., J. Zhou, J. Yang, X. Yu, and L. Liu. 2014b. Vertical distribution of diazotrophic bacterial community associated with temperature and oxygen gradients in a subtropical reservoir. Hydrobiologia 741: 69–77. doi:10.1007/s10750-012-1310-2

Yu, Z., J. Yang, L. Liu, W. Zhang, and S. Amalfitano. 2015. Bacterioplankton community shifts associated with epipelagic and mesopelagic waters in the Southern Ocean. Sci. Rep. 5: 12897. doi:10.1038/srep12897
Zhang, Y., Z. Zhao, M. Dai, N. Jiao, and G. J. Herndl. 2014. Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. Mol. Ecol. 23: 2260–2274. doi:10.1111/mec.12739

Acknowledgment
This work was supported by the National Major Science and Technology Program for Water Pollution Control and Treatment (No. 2012ZX07501002-06).

Conflict of Interest
None declared.

Submitted 27 June 2016
Revised 15 August 2016
Accepted 27 September 2016

Associate editor: Rebecca Vega Thurber