



## LJMU Research Online

Liu, L, Yang, J, Yu, Z and Wilkinson, DM

**The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China.**

<http://researchonline.ljmu.ac.uk/id/eprint/819/>

### Article

**Citation** (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Liu, L, Yang, J, Yu, Z and Wilkinson, DM (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. ISME Journal. ISSN 1751-7370**

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact [researchonline@ljmu.ac.uk](mailto:researchonline@ljmu.ac.uk)

<http://researchonline.ljmu.ac.uk/>

1 **The biogeography of abundant and rare bacterioplankton**  
2 **in the lakes and reservoirs of China**

3 Lemian Liu<sup>1</sup>, Jun Yang<sup>1\*</sup>, Zheng Yu<sup>1</sup> and David M. Wilkinson<sup>2</sup>

4

5 <sup>1</sup>*Aquatic EcoHealth Group, Key Laboratory of Urban Environment and Health, Institute of Urban*  
6 *Environment, Chinese Academy of Sciences, Xiamen, People's Republic of China and* <sup>2</sup>*School of*  
7 *Natural Science and Psychology, Liverpool John Moores University, Byrom Street, Liverpool, UK*

8

9 **Running title:** Biogeography of abundant and rare bacterioplankton

10

11 **Key words:** Bacterioplankton / geographical distribution / microbial diversity / rare biosphere /  
12 high-throughput sequencing

13

14 **Subject Category:** Microbial population and community ecology

15

16 **Correspondence:** Jun Yang, Aquatic Ecohealth Group, Key Laboratory of Urban Environment  
17 and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021,  
18 People's Republic of China.

19 E-mail: [jyang@iue.ac.cn](mailto:jyang@iue.ac.cn)

20 Received 30 June 2014; revised 11 January 2015; accepted 30 January 2015

21

22 **Abstract**

23 Bacteria play key roles in the ecology of both aquatic and terrestrial ecosystems,  
24 however little is known about their diversity and biogeography, especially in the rare  
25 microbial biosphere of inland freshwater ecosystems. Here we investigated aspects of  
26 the community ecology and geographical distribution of abundant and rare  
27 bacterioplankton using high-throughput sequencing and examined the relative  
28 influence of local environmental variables and regional (spatial) factors on their  
29 geographical distribution patterns in 42 lakes and reservoirs across China. Our results  
30 showed that the geographical patterns of abundant and rare bacterial subcommunities  
31 were generally similar, and both of them showed a significant distance-decay  
32 relationship. This suggests that the rare bacterial biosphere is not a random assembly,  
33 as some authors have assumed, and that its distribution is most likely subject to the  
34 same ecological processes that control abundant taxa. However we identified some  
35 differences between the abundant and rare groups as both groups of bacteria showed a  
36 significant positive relationship between sites occupancy and abundance, but the  
37 abundant bacteria exhibited a weaker distance–decay relationship than the rare  
38 bacteria. Our results implied that rare subcommunities were mostly governed by local  
39 environmental variables, whereas the abundant subcommunities were mainly affected  
40 by regional factors. Additionally, both local and regional variables which were  
41 significantly related to the spatial variation of abundant bacterial community  
42 composition were different to those of rare ones, suggesting that abundant and rare  
43 bacteria may have discrepant ecological niches and may play different roles in natural  
44 ecosystems.

45

## 46 Introduction

47 Bacterioplankton are a fundamental component of aquatic ecosystems, having an  
48 extremely high level of genetic diversity, and playing essential roles in global  
49 biogeochemical cycles (Newton *et al.*, 2011). This means that understanding the  
50 diversity of aquatic bacteria and their biogeographical patterns is a major ecological  
51 goal, because it provides an insight into the deeper ecological processes and  
52 mechanisms that underlie and maintain bacterial diversity and ecosystem function  
53 (Hanson *et al.*, 2012). However, unlike the extensive studies of biogeographical  
54 patterns of large animals and plants, works on the spatial scaling of bacterial diversity  
55 is still limited (Preston, 1948; Green and Bohannan, 2006; Martiny *et al.*, 2006;  
56 Ramette and Tiedje, 2007; Logue and Lindström, 2008; Barberán and Casamayor,  
57 2010; Jones *et al.*, 2012; Brown *et al.*, 2014). A further complicating fact is that most  
58 bacterial communities comprise a large number of species. A few of these species are  
59 very abundant, others are moderately abundant, and a large number species, often  
60 called the ‘rare biosphere’, are represented by only a few individuals. This highly  
61 diverse and rare microbial biosphere is largely unexplored (Pedrós-Alió, 2012).  
62 Influential studies interpreted as showing that ‘everything is everywhere’ (developed  
63 by Baas-Becking, 1934) in microbial ecology have made use of the rare biosphere  
64 community as a way of testing large scale biogeographical ideas. For example, using  
65 varied culture conditions to demonstrate the presence of protist taxa in a lake that was  
66 not ecologically suitable for them so that they were present at the site in such low  
67 numbers that they were undetectable by direct microscopy of uncultured samples (see,  
68 for example, Fenchel *et al.*, 1997; Finlay, 1998; Hambright *et al.*, 2015). In the past  
69 few years, a revolution in high-throughput and deep sequencing is allowing a much  
70 more direct identification of the rare biosphere community in an environment  
71 (Pedrós-Alió, 2012; Logares *et al.*, 2014).

72 Recent high-throughput sequencing studies indicated that the ‘rare biosphere’  
73 fraction of the bacterial community follows biogeographical patterns similar to those  
74 of the most abundant members of the community and have distinct spatial distribution  
75 patterns in Arctic Ocean and coastal Antarctic lakes (Galand *et al.*, 2009; Logares *et*  
76 *al.*, 2013). Other studies provided evidence of an occupancy–abundance relationship  
77 for soil and marine abundant bacteria communities (Nemergut *et al.*, 2011), implying  
78 that the dispersal probability of rare bacteria is limited compared to abundant taxa  
79 (Holt *et al.*, 2004; Logue, 2010). In this context inland waters are particularly  
80 interesting as they have significantly more bacterial groups and are more diverse than  
81 marine waters on a global scale (Barberán and Casamayor, 2010). Until now it has  
82 remained unclear if rare bacterial taxa follow similar biogeographical patterns as  
83 abundant taxa in inland lake and reservoir ecosystems.

84 In general, microorganisms, like larger organisms, are influenced by complex and  
85 interacting sets of abiotic and biological processes, leading to variation in their  
86 community distribution patterns at different spatial scales (Green and Bohannan, 2006;  
87 Ramette and Tiedje, 2007; Logue and Lindström, 2008; Jones *et al.*, 2012). For larger  
88 organisms it has been suggested that abundant taxa dominate major ecosystem  
89 processes (e.g. carbon flow and nutrient cycling), while rare taxa can be regarded as  
90 propagule banks and play minor but non-negligible roles but may switch to being  
91 more important abundant taxa as environmental conditions change (Grime, 1998;  
92 Magurran and Henderson, 2003). More recently, similar suggestions have been made  
93 for marine microbial systems (Pedrós-Alió, 2012; Logares *et al.*, 2014), implying that  
94 abundant and rare bacterial taxa may have different ecological responses to  
95 environmental changes. In this study, we hypothesized that the possible controlling  
96 factors of biogeographical distribution of rare bacteria subcommunities were different  
97 from those of abundant bacteria subcommunities in lakes and reservoirs at a large  
98 scale.

99 Bacterial metacommunities are normally assembled by both local environmental  
100 variables (environmental filtering and biotic interactions) and regional factors  
101 (dispersal related processes) (Fenchel, 2003; Nemergut *et al.*, 2011; Liu *et al.*, 2013).  
102 Some studies have showed that local factors are more important (Fierer and Jackson,  
103 2006, Graham and Fine, 2008; Cavender-Bares *et al.*, 2009; Wang *et al.*, 2013), while  
104 other studies hold that bacterial communities are assembled by regional forces  
105 (Leibold *et al.*, 2004; Cottenie, 2005; Martiny *et al.*, 2006). Recently, Lindström and  
106 Langenheder (2012) depicted the relationship between regional and local factors  
107 along gradients of dispersal rate and selective strength. It has been suggested that  
108 nature of bacterial communities are important in the creation of wider biogeographical  
109 patterns as the dispersal potential could affect the connection between community  
110 composition and local and regional factors (Lindström and Langenheder, 2012; Lear  
111 *et al.*, 2014). The most abundant bacteria can disperse readily as there are, by  
112 definition, many more individuals which can potentially be involved in a dispersal  
113 event. In contrast, rare bacteria are less abundant, and so their dispersal rate should be  
114 low compared to abundant taxa. Therefore, we hypothesized that the relative influence  
115 of regional and local factors for abundant bacterioplankton subcommunities were  
116 different from rare subcommunities.

117 In this study, we used high-throughput sequencing to investigate the aquatic  
118 bacterial community along a latitudinal gradient ranging from 24 to 50 °N (over 2700  
119 km) in 42 lakes and reservoirs across China. We sort to determine and compare the  
120 biogeographical patterns and drivers for abundant and rare bacterial subcommunities  
121 at a continental scale. Specifically, we aimed to answer the following key questions:  
122 do abundant and rare taxa show similar or different biogeographical patterns in lakes

123 and reservoirs at a continental scale? Are the controlling factors and their contribution  
124 to the geographical pattern of abundant bacterial subcommunities different from the  
125 rare ones?  
126

## 127 **Material and methods**

### 128 *Study area and sampling*

129 A total of 42 Chinese lakes and reservoirs, which were located between approximately  
130 24 and 50 °N, were included in the sampling campaign (Figure 1, Supplementary  
131 Table S1). Many of these lakes also featured in a study of aquatic testate amoebae by  
132 Ju *et al.* (2014). Field sampling took place from middle July to early August 2012.  
133 Surface water samples (upper 50–200 cm) in the epilimnion were collected at the  
134 center of each lake. All water samples were subsequently divided into two subsamples:  
135 one for water chemistry and the other for bacterial community analyses. All samples  
136 were stored in the dark at 4 °C and returned to the laboratory as soon as possible.  
137

### 138 *Physico-chemical analysis*

139 Water temperature, electrical conductivity, pH, dissolved oxygen, and turbidity of the  
140 epilimnion layer were measured in situ with YSI multi-parameter water quality sonde  
141 (YSI, Yellow Springs, OH, USA). Water depth of sampling site was measured with a  
142 Speedtech SM-5 Depthmate portable sounder (Speedtech Instruments, Great Falls, VA,  
143 USA). Water transparency was determined with a 30 cm Secchi disc. The  
144 concentrations of chlorophyll a (Chl a), total nitrogen and total phosphorus were  
145 measured according to standard methods (Greenberg *et al.*, 1992).  
146

### 147 *DNA extraction, PCR and high-throughput sequencing*

148 A total of 500 mL water samples for bacterioplankton analyses were filtered through a  
149 0.22- $\mu$ m pore size polycarbonate filters (47 mm diameter, Millipore, Billerica, MA,  
150 USA) following Liu *et al.*, (2013). The filters were stored at –80 °C until DNA  
151 extraction. Total DNA was extracted directly from the filter using FastDNA spin kit  
152 (Bio101, Carlsbad, CA, USA) according to the manufacturer's instructions. Total  
153 DNA was sent to the Personal Biotechnology Co., Ltd. in Shanghai, China for  
154 high-throughput sequencing on an Illumina MiSeq instrument (San Diego, CA, USA)  
155 using a paired-end 150-bp sequence read run. A set of primers was used to amplify the  
156 hypervariable V4 region (about 207 bp) of bacterial 16S rRNA gene. In this study, the  
157 forward primer was 5' - AYTGGGYDTAAAGNG - 3', and the reverse primer was 5'  
158 - TACNVGGGTATCTAATCC - 3' (Claesson *et al.*, 2009). Each DNA sample was  
159 individually PCR-amplified in triplicated 25- $\mu$ L reactions included an initial  
160 denaturation at 94°C for 5 min, followed by 25 cycles of 30 s at 94°C, 30 s at 50°C

161 and 30 s at 72°C. At the end of the amplification, the amplicons were subjected to  
162 final 7 min extension at 72°C. Each reaction contained 1 × PCR buffer, 2.5 mM  
163 dNTPs, 0.625 U of Taq DNA polymerase, 10 μM of each primer, and 20 ng of target  
164 DNA.

165

#### 166 *Sequence analysis*

167 Raw sequence data were processed using MOTHUR v.1.33.3 (Schloss *et al.*, 2009).  
168 Paired-end reads were merged. Sequences were then quality controlled with the  
169 following settings: any sequences length < 150 or > 300, average quality < 30,  
170 ambiguous bases > 0, homopolymer length > 6 were removed for further analysis.  
171 The remaining sequences were aligned to a reference alignment, and those sequences  
172 that did not align to the correct region were eliminated. To further reduce the noise in  
173 our sequences, we utilized pre-clustering and the resulting sequences were screened  
174 for chimeras using UCHIME (Edgar *et al.*, 2011). We then used Bayesian classifier to  
175 classify those sequences against the Ribosomal Database Project 16S rRNA gene  
176 training set (version 9, <http://rdp.cme.msu.edu>). We required an 80% pseudobootstrap  
177 confidence score (Wang *et al.*, 2007). All Archaea, Eukaryota, chloroplasts,  
178 mitochondria and unknown sequences were culled. Finally, sequences were split into  
179 groups according to their taxonomy and assigned to operational taxonomic units  
180 (OTUs) at a 3% dissimilarity level. The OTUs which contained < 2 reads were not  
181 used to avoid possible biases. For our data analyses, we used a randomly selected  
182 subset of 26322 sequences from each sample to standardize sequencing effort across  
183 samples.

184 In this study, the definition of abundant or rare OTUs combined their local and  
185 regional relative abundance. Locally abundant OTUs were defined as the OTUs with a  
186 representation > 1% within a sample, while locally rare OTUs were defined as having  
187 an abundance < 0.01% within a sample (Galand *et al.*, 2009; Pedrós-Alió, 2012). We  
188 then calculated the average relative abundance of these locally abundant or rare OTUs  
189 across all samples. The OTUs that had a mean relative abundance of > 0.1% were  
190 defined as regionally abundant OTUs, whereas the OTUs with a mean relative  
191 abundance of < 0.001% were defined as regionally rare OTUs (Logares *et al.*, 2014).  
192 The ecological literature usually considers ‘rarity’ to be a continuous variable –  
193 therefore there is always a level of arbitrariness when defining a cutoff point for rarity  
194 in any given study (Gaston, 1994). To reduce this problem we have defined our ‘rare  
195 biosphere’ with reference to other recent publications, to facilitate comparisons  
196 between studies. In addition, to reduce the effect of arbitrary definition of abundant  
197 and rare OTUs, we performed multivariate cutoff level analysis (MultiCoLA) to  
198 systematically estimate how our datasets are affected by the definition of abundant  
199 and rare OTUs (Gobet *et al.*, 2010).

200

201 *Data analyses*

202 Bray-Curtis similarity matrix is considered to be one of the most robust similarity  
203 coefficients for ecological studies (Kent, 2012) and was applied to our community  
204 dataset of bacterial OTU relative abundance. A non-metric multidimensional scaling  
205 (MDS) ordination was used to investigate differences in bacterioplankton  
206 communities between sites (Clarke and Gorley, 2001). Our study area includes several  
207 major climate types, thus five regions were divided based on their climate and  
208 geographical characteristics (Ju *et al.*, 2014). To evaluate the significant differences of  
209 bacterioplankton communities between these five regions, we used the  
210 randomization/permutation procedure analysis of similarities (ANOSIM). The  
211 ANOSIM statistic global R is calculated as the difference of between-group and  
212 within-group mean rank similarities, thus it displays the degree of separation between  
213 groups. Complete separation is indicated by  $R = 1$ , whereas  $R = 0$  suggests no  
214 separation (Clarke and Gorley, 2001). RELATE calculates rank correlation among the  
215 entire, abundant and rare bacterial community similarity matrices by PRIMER 5.0,  
216 and it thus provides a significance test with the matching coefficient pm, which is  
217 equivalent to the Mantel's test (Clarke and Gorley, 2001). We referred to positive and  
218 negative correlation levels between 0.5 and 1 as strong relationships at  $P < 0.01$ .

219 Spearman rank correlations were used to determine the relationships between the  
220 Bray-Curtis similarity of bacterial community and the geographical distance of lakes  
221 or reservoirs, and the relationship between the Euclidean distance of environmental  
222 variables and the geographical distance.

223 A set of regional variables were generated through the use of principal coordinates  
224 of neighbor matrices (PCNM) analysis (Borcard and Legendre, 2002; Legendre *et al.*,  
225 2008) based on the longitude and latitude coordinates of each sampling site. The  
226 normality of the physico-chemical variables were checked using Shapiro-Wilk test  
227 and variables were  $\log(x+1)$  transformed with the exception of pH, to improve  
228 normality and homoscedasticity for multivariate statistical analyses. Canonical  
229 correspondence analysis (CCA) was performed to explore the relationships between  
230 bacterial communities and physico-chemical and PCNM variables. This method was  
231 chosen because preliminary detrended correspondence analysis (DCA) on bacterial  
232 community data revealed that the longest gradient lengths were longer than 3.0,  
233 indicating that the majority of species exhibited unimodal responses to the  
234 environmental variation (Lepš and Šmilauer, 2003). Before the CCA, we used a  
235 forward selection procedure to select local physico-chemical variables and regional  
236 variables using the 'ordiR2step' function from vegan (Blanchet *et al.* 2008). All  
237 non-significant ( $P > 0.05$ ) variables were eliminated in further analyses.

238 To determine the relative contribution of local (in this study, environmental

239 variables) and regional (in this study, spatial factors based on PCNM) variables to the  
240 distribution of bacterial communities, standard and partial Mantel tests were also  
241 performed (Legendre and Legendre, 2012). The similarity matrices of bacterial  
242 community composition were obtained using Bray-Curtis index. The local and  
243 regional matrices were obtained using Euclidean distances with significant variables.  
244 PCNM, CCA, DCA, forward selection procedure and Mantel tests were performed in  
245 the R language environment. Rarefaction curves and richness estimate of Chao 1 and  
246 ACE were calculated in Vegan with R software (R Development Core Team).

247

#### 248 *Accession number*

249 All sequence data from this study have been deposited in the public NCBI database  
250 (<http://www.ncbi.nlm.nih.gov/>) under the accession number SRX525963.

251

## 252 **Results**

### 253 *The estimate of species richness and multivariate abundance cutoff*

254 In total, 10559 bacterial OTUs with 1105524 sequences were obtained in this study,  
255 and the number of bacterial OTUs varied from 816 (Donghaizi Lake) to 2026  
256 (Dongxintunnanpao Lake) per sample (mean = 1399, standard error (SE) = 48, n = 42)  
257 (Supplementary Table S1). The total number of bacterial OTUs (10559) was roughly  
258 equivalent to the number estimated by abundance-based richness estimators such as  
259 Chao 1 ( $10791 \pm 22$ ) and ACE ( $10952 \pm 48$ ). Both the estimated species-accumulation  
260 curves (Supplementary Figure S1) and extrapolated species richness indices (Chao 1  
261 and ACE) (Supplementary Table S2) indicated that the majority of the pelagic  
262 bacterioplankton taxa had been recovered from the studied lakes and reservoirs.

263 At local and regional levels, 143 (1.4%) OTUs with 751588 sequences (68.0%) were  
264 classified as abundant OTUs, while 7598 (72.0%) taxa and 29824 (2.7%) sequences  
265 were rare OTUs (Supplementary Table S2). At local level, however, no OTU was  
266 always abundant (> 1%) in all samples, and only two OTUs (including 143102  
267 sequences) with > 1% abundance were present > 70% of the samples. In addition,  
268 5435 OTUs (including 16623 sequences) were always locally rare (< 0.01%) in all  
269 samples (Supplementary Figure S2).

270 The multivariate cutoff level analysis showed that when the structure of community  
271 data were compared between the truncated and the original matrices, little variation in  
272 data structure was observed up to a removal of 45% of the rare part of the dataset. On  
273 the other hand, when the increasing amount of rare types was > 5%, the data structure  
274 of truncated matrices showed a little variation (Supplementary Figure S3). To fully  
275 capture all bacteria sequences in any environment is still impossible; however our  
276 definitions of abundant (32.0%) and rare (2.7%) bacteria are reasonable and objective

277 within the constraints of current technology.

278

### 279 *Geographical patterns of bacterial community*

280 We found a distinct biogeographical distribution pattern for the rare bacterial taxa,  
281 which showed striking similarities with the geographical patterns of both abundant  
282 subcommunity ( $\rho_m = 0.762$ ,  $P < 0.01$ ) and entire community ( $\rho_m = 0.807$ ,  $P < 0.01$ )  
283 (Figure 2). Interestingly, our five geographical regions were significantly separated at  
284  $P = 0.001$  for the bacterioplankton communities, and the global R among the five  
285 groups was 0.398, 0.372 and 0.464 for the entire bacterial community, the abundant  
286 bacterial subcommunity and the rare bacterial subcommunity, respectively.

287

### 288 *Distance effects on community composition and environmental variables*

289 Overall, the similarity in bacterial community composition between any two lakes or  
290 reservoirs decreased with the increasing of geographic distance (Figure 3). For  
291 example, Spearman correlation analysis gave a correlation coefficient of  $-0.498$  ( $P <$   
292  $0.01$ ) between the similarity of entire bacterial community and geographic distance.  
293 Also, the Spearman correlation coefficients for abundant and rare taxa were  $-0.398$  ( $P$   
294  $< 0.01$ ) and  $-0.507$  ( $P < 0.01$ ) between the subcommunity similarity and geographic  
295 distance, respectively. In addition, lakes and reservoirs that were closer to each other  
296 presented more similar environmental conditions (Spearman correlation coefficient =  
297  $0.137$ ,  $P < 0.01$ , Supplementary Figure S4).

298

### 299 *Abundance–occupancy relationship*

300 Bacterial relative abundance and local occupancy were positively correlated ( $r =$   
301  $0.409$ ,  $P < 0.01$ ) (Figure 4). In addition, 98% of abundant OTUs occupied more than  
302 50% of sites, whereas no rare OTU occupied more than 50% of sites (Supplementary  
303 Figure S5).

304

### 305 *Effects of local and regional factors on bacterioplankton distribution*

306 The CCA ordination showed that four environmental variables (water temperature,  
307 electrical conductivity, turbidity and total nitrogen) and five regional factors  
308 (PCNM1-3 and 13-14) were significantly related to the change of abundant bacterial  
309 subcommunity by forward model selection ( $P < 0.05$ ) (Figure 5). However, the CCA  
310 for the rare bacterial subcommunity showed a different pattern. Three local  
311 environmental factors (electrical conductivity, transparency and depth) and six  
312 regional variables (PCNM2-3, 6, 10 and 13-14) were significantly related to the  
313 variation of rare bacterial community composition ( $P < 0.05$ ) (Figure 5).

314 The Mantel and partial Mantel results revealed that rare bacterial subcommunities  
315 were primarily governed by local environmental factors. However, the regional

316 factors explained more variation in abundant bacterial distribution (Table 1).

317

### 318 *Taxonomic distributions of abundant and rare bacteria*

319 Actinobacteria (abundant  $27.8 \pm 1.5\%$  vs. rare  $7.3 \pm 0.6\%$ , mean  $\pm$  SE), Bacteroidetes  
320 ( $14.0 \pm 1.5\%$  vs.  $10.8 \pm 0.9\%$ ), Cyanobacteria ( $8.5 \pm 1.1\%$  vs.  $2.3 \pm 0.2\%$ ), Firmicutes  
321 ( $0.4 \pm 0.2\%$  vs.  $6.8 \pm 0.6\%$ ), Planctomycetes ( $2.8 \pm 0.5\%$  vs.  $1.2 \pm 0.2\%$ ),  
322 Verrucomicrobia ( $21.2 \pm 1.9\%$  vs.  $5.6 \pm 0.4\%$ ), Alpha- proteobacteria ( $5.9 \pm 0.4\%$  vs.  
323  $5.9 \pm 0.5\%$ ), Beta-proteobacteria ( $5.4 \pm 1.0\%$  vs.  $4.1 \pm 0.4\%$ ) and  
324 Gamma-proteobacteria ( $4.1 \pm 0.5\%$  vs.  $5.8 \pm 0.3\%$ ) taxa were the most frequent  
325 groups in both abundant and rare taxa. However, there were more unclassified  
326 bacterial groups ( $9.0 \pm 0.8\%$  vs.  $39.0 \pm 1.1\%$ ) among the rare taxa, and a much higher  
327 number of taxonomic groups were present among rare bacteria compared with  
328 abundant bacteria (Supplementary Figure S6).

329

## 330 **Discussion**

### 331 *Geographical patterns in abundant and rare bacteria*

332 New molecular tools and increasing sampling effort have confirmed the existence of a  
333 'rare biosphere' (Pedrós-Alió, 2012). The cosmopolitan theory proposes that rare  
334 phylotypes (species) are recruited through immigration (dispersal) and that they are  
335 protected from active loss by both viral lysis and predation because of their low  
336 abundance (Pedrós-Alió, 2006) –as in the longstanding 'Everything is Everywhere'  
337 approach to microbial biogeography which was developed mainly using  
338 morpho-species data (Finlay, 1998; 2002). This low loss rate runs counter to mid-20th  
339 century suggestions that the rarest planktonic organisms in a lake were likely to be  
340 most at risk of extinction (Hutchinson, 1964). This raises an important question  
341 because if rare microbes have a lower, or equal, extinction risk to common ones then  
342 this would be an area where the ecology of microbes differs from that of larger  
343 organisms where extinction risk is often associated with rarity (Caughley, 1994).

344 Recent studies using high-throughput sequencing in Arctic Ocean and coastal  
345 Antarctic lakes demonstrated that rare bacteria had distinct spatial distribution  
346 patterns, and showed a similar spatial pattern to the abundant bacterial subcommunity  
347 (Galand *et al.*, 2009; Logares *et al.*, 2013). Our results corroborated those previous  
348 findings, and provided further evidence from subtropical and temperate inland lakes  
349 and reservoirs to support the concept that rare bacteria exhibit biogeographical  
350 patterns. Further, we found both abundant and rare bacterial communities strongly  
351 adhere to distance-decay relationship. This could be because of dispersal and/or  
352 because lakes close to each other tend to have similar environments. In general, these  
353 similar geographical distribution patterns in abundant and rare bacterial

354 subcommunities suggested that the rare bacteria biosphere is not a random collection  
355 of taxa (Galand *et al.*, 2009).

356 However, within this general pattern there were also some interesting differences  
357 between the distributions of abundant and rare bacteria. First, our results showed that  
358 abundant bacteria have a weaker distance–decay relationship ( $r = -0.398$ ) than rare  
359 bacteria ( $r = -0.507$ , Figure 3). Second, the bacteria distribution frequency showed a  
360 significant positive relationship with the OTUs relative abundance in our studied lakes  
361 and reservoirs (Figure 4). These results indicate that the abundant bacteria with high  
362 local abundance have a decreased probability of local extinction and increase  
363 probability of dispersal, thereby resulting in a widespread or ubiquitous distribution.  
364 In fact, bacteria are of a size where long distance aerial dispersal is clearly possible  
365 (Wilkinson *et al.*, 2012) and the more abundant taxa are presumably more likely to  
366 become airborne – perhaps via bubble burst processes associated with wind created  
367 waves (Hamilton and Lenton, 1998). In addition, the commoner taxa are also more  
368 likely to be moved by waterbirds which have been shown to affect the movement of  
369 some zooplankton between lakes (Figuerola *et al.*, 2005). An alternative, or additional,  
370 mechanism is that the abundant bacteria may grow on a wider spectrum of resources  
371 compared with the rare taxa so leading to them being able to reach high population  
372 levels in a wider range of lakes (Hambright *et al.*, 2015). Pedrós-Alió (2006)  
373 hypothesized that rare taxa have a cosmopolitan distribution, due to a low loss rate  
374 and a potentially unlimited dispersal capacity. Our results showed that the hypothesis  
375 that the rare biosphere has a cosmopolitan distribution did not apply in most cases.  
376 Conversely, our results suggest that abundant bacteria tend towards more  
377 cosmopolitan distributions, while the majority of rare bacteria are more likely to show  
378 restricted distributions. This matches the well documented abundancy-occupancy  
379 relationship detailed for larger organisms (Gaston *et al.*, 2000; Holt *et al.*, 2004) –  
380 suggesting similar ‘rules’ govern the ecology of microorganisms and the larger  
381 organisms that have more often been the subject of ecological study.

382

### 383 *Controlling factors for geographical distributions of abundant and rare bacteria*

384 An important ecological aspect for the understanding of biogeography is to determine  
385 the possible controlling factors and their contribution to spatial variation of bacterial  
386 community. Several local factors are known to affect bacteria species coexistence,  
387 these include temperature (Liu *et al.*, 2013), water chemistry (Berdjeb *et al.*, 2011),  
388 quality and quantity of dissolved organic matter (Judd *et al.*, 2006), and grazing  
389 (Pernthaler, 2005). However, spatial variation in the bacterial community is also  
390 influenced by regional factors such as dispersal limitation, mass effects and random  
391 demographics (Liu *et al.*, 2013; Wang *et al.*, 2013).

392 Our CCA analyses for abundant and rare bacterial subcommunities yielded

393 different patterns because both communities were significantly related to different  
394 local and regional factors. A similar finding was reported by Kim *et al.* (2013) from an  
395 activated sludge bioreactor. They demonstrated that the temporal dynamics of  
396 abundant and rare bacterial taxa were constrained by different environmental  
397 variables. Hence, it is clear that abundant and rare bacterial taxa may have different  
398 ecological niches, and the majority of abundant and rare taxa may play different roles  
399 in our investigated lake and reservoir ecosystems. The few abundant taxa account for  
400 most of the biomass and carbon cycling, whereas some rare species may be important  
401 for the cycling of certain elements (Pedrós-Alió, 2012). An example from a peatland  
402 indicated that rare taxon *Desulfosporosinus* could contribute most of the sulfate  
403 reduction in the peat (Pester *et al.*, 2010). Moreover, abundant and rare bacteria may  
404 have different ecological strategies. The resources that some abundant bacteria use  
405 may be more abundant and widespread; alternatively, some abundant bacteria, in  
406 general, may be able to grow on a wider spectrum of resources. In contrast, some  
407 bacteria may be adapted to use only a few specific substrates, or to episodic situations  
408 of high nutrient abundance. For example, Teira *et al.* (2007) incubated natural  
409 bacterial assemblages in microcosms with different amendments of polycyclic  
410 aromatic hydrocarbons (PAHs). During these experiments, when the PAHs were  
411 added, *Cycloclasticus* (a hydrocarbon-degrading gammaproteo bacterium) became an  
412 abundant member of the community. Once the PAHs had disappeared, the  
413 *Cycloclasticus* population reduced and returned to the rare biosphere. Therefore, rare  
414 microbial taxa can be regarded as a propagule bank (c.f. Grime, 1998). They may  
415 become an abundant member of the community, when the right environmental  
416 conditions are met or when the abundance of abundant microbes decreases drastically  
417 or even becomes extinct (Pedrós-Alió, 2012; Logares *et al.*, 2014).

418 A major theme of ecological research is quantitatively characterizing the relative  
419 contribution of local versus regional factors. Partial Mantel tests provide a good  
420 solution to compare the relative importance of local and regional factors (Lindström  
421 and Langenheder, 2012). In this study, we found abundant bacterial subcommunities  
422 were mostly governed by regional factors, whereas the rare bacterial subcommunities  
423 were mainly structured by local environmental variables. Hanson *et al.* (2012)  
424 examined 19 previous studies, and found that most studies suggested local factors had  
425 a greater effect on microbial composition than regional processes. More recently,  
426 Wang *et al.* (2013) illustrated that the dominance of local factors might occur when the  
427 selective strength of local habitat conditions is above a conceptual threshold. In the  
428 present study, our samples spanned a continental scale and comprised contrasting  
429 environmental conditions such as oligotrophic and eutrophic waters, different habitat  
430 types (e.g. lakes vs. reservoirs) and altitudes (Ju *et al.*, 2014). However, the regional  
431 factors seem to be more important than local factors for the biogeographical pattern of

432 abundant bacterial taxa in the lakes and reservoirs. This could be because some  
433 environmental variables that were related to spatial distribution of abundant bacterial  
434 subcommunity were not measured here. Another possible mechanism for this result is  
435 that the populations of abundant bacterial taxa exhibit massive immigration and this  
436 can result in more community variability being related to spatial variables that occur  
437 independently of any measured environmental variables (Lindström and Langenheder,  
438 2012). Wang *et al* (2013) examined the relative influence of deterministic (local  
439 factors) and stochastic (regional factors) processes for bacterial communities from  
440 various habitats. They suggested that stochastic processes may overwhelm  
441 deterministic process, when regional species pools were characterized by  
442 environmental generalists. Unsurprisingly, most abundant bacteria appeared to be  
443 habitat generalists in our study, given that 98% of abundant OTUs occupied more than  
444 50% of sites. Similarly, Lear *et al* (2014) found that the regional factors explained the  
445 largest proportion of variation of bacterial community composition in the ponds that  
446 are connected to each other by ephemeral channels. These authors suggested that large  
447 numbers of immigrant taxa are received by each pond which could enhance the  
448 connection between community composition and regional factors.

449

## 450 **Conclusions**

451 Our results demonstrated that the rare bacterioplankton subcommunity had a distinct  
452 biogeographical pattern in the studied Chinese lakes and reservoirs, which was  
453 reasonably similar to that of the abundant bacteria. However, rare and abundant  
454 bacterial subcommunities were significantly related to different local and regional  
455 variables. Moreover, local processes and factors play the most important role in  
456 structuring rare bacterial assemblages, with regional factors explaining more variation  
457 in abundant bacterial distribution. In this study, we suggest that dispersal potentials  
458 and ecological roles of both rare and abundant bacteria are potentially different.  
459 Therefore, to obtain a comprehensive or full understanding and modeling of bacterial  
460 community biogeography, both groups should be distinguished in future studies.

461

## 462 **Conflict of Interest**

463 The authors declare no conflict of interest.

464

## 465 **Acknowledgements**

466 We thank Dr. Zhengwen Liu at Nanjing Institute of Geography and Limnology,  
467 Chinese Academy of Sciences for providing part of physicochemical data. This work  
468 was funded by the National Basic Research Program of China (2012CB956103), the  
469 National Natural Science Foundation of China (31370471), the Natural Science

470 Foundation for Distinguished Young Scholars of Fujian Province (2012J06009), and  
471 the Knowledge Innovation Program of the Chinese Academy of Sciences  
472 (IUEQN201305). We thank our referees for insightful comments which help improve  
473 the clarity of this paper.

474

475 Supplementary information can be found on ISMEJ's website.

476

## 477 **References**

478 Baas-Becking LGM. (1934). Geobiologie of Inleiding Tot de Milieukunde. WP Van Stockum &  
479 Zoon: The Hague, The Netherlands.

480 Barberán A, Casamayor EO. (2010). Global phylogenetic community structure and  $\beta$ -diversity  
481 patterns in surface bacterioplankton metacommunities. *Aquat Microb Ecol* **59**: 1–10.

482 Berdjeb L, Ghiglione JF, Jacquet S. (2011) Bottom-up versus top-down control of hypo- and  
483 epilimnion free-living bacterial community structures in two neighboring freshwater lakes.  
484 *Appl Environ Microbiol* **77**: 3591–3599.

485 Blanchet FG, Legendre P, Borcard D. (2008). Forward selection of explanatory variables. *Ecology*  
486 **89**: 2623–2632.

487 Borcard D, Legendre P. (2002). All-scale spatial analysis of ecological data by means of principal  
488 coordinates of neighbor matrices. *Ecol Model* **153**: 51–68.

489 Brown MV, Ostrowski M, Grzymski J, Lauro FM. (2014). A trait based perspective on the  
490 biogeography of common and abundant marine bacterioplankton clades. *Mar Genom* **15**:  
491 17–28.

492 Caughley G. (1994). Directions in conservation biology. *J Animal Ecol* **63**: 215–244.

493 Cavender-Bares J, Kozak H, Fine PVA, Kembel SW. (2009). The merging of community  
494 ecology and phylogenetic biology. *Ecol Lett* **12**: 693–715.

495 Claesson MJ, O'Sullivan O, Wang Q, Nikkila J, Marchesi JR, Smidt H *et al.* (2009). Comparative  
496 analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community  
497 structures in the human distal intestine. *PLoS ONE* **8**: e6669.

498 Clarke KR, Gorley RN. (2001). PRIMER v5: User manual/tutorial. PRIMER-E: Plymouth, UK.

499 Cottenie K. (2005). Integrating environmental and spatial processes in ecological community  
500 dynamics. *Ecol Lett* **8**: 1175–1182.

501 Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and  
502 speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

503 Fenchel T. (2003). Biogeography for bacteria. *Science* **301**: 925–927.

504 Fenchel T, Esteban GF, Finlay BJ. (1997) Local versus global diversity of microorganisms: cryptic  
505 diversity of ciliate protozoa. *Oikos* **80**: 220–225

506 Fierer N, Jackson RB. (2006). The diversity and biogeography of soil bacterial communities. *Proc*  
507 *Natl Acad Sci USA* **103**: 626–631.

508 Figuerola J, Green AJ, Michot TC. (2005). Invertebrate eggs can fly: evidence of  
509 waterfowl-mediated gene flow in aquatic invertebrates. *Am Nat* **165**: 274–280.

510 Finlay BJ. (1998). The global diversity of protozoa and other small species. *Int J Paristol* **28**:  
511 29–48.

512 Finlay BJ. (2002). Global dispersal of free-living microbial eukaryote species. *Science* **296**:  
513 1061–1063.

514 Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. (2009). Ecology of the rare microbial  
515 biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* **106**: 22427–22432.

516 Gaston KJ. (1994). Rarity. Chapman & Hall: London, UK.

517 Gaston KJ, Blackburn TM, Greenwood JJD, Gregory RD, Quinn RM, Lawton JH. (2000).  
518 Abundance-occupancy relationships. *J Appl Ecol* **37** (Suppl. 1): 39–59.

519 Gobet A, Quince C, Ramette A. (2010). Multivariate cut off level analysis (MultiCoLA) of large  
520 community datasets. *Nucleic Acids Res* **38**: e155.

521 Graham CH, Fine PVA. (2008). Phylogenetic beta diversity: linking ecological and evolutionary  
522 processes across space in time. *Ecol Lett* **11**: 1265–1277.

523 Green J, Bohannan BJM. (2006). Spatial scaling of microbial biodiversity. *Trends Ecol Evol* **21**:  
524 501–507.

525 Greenberg AE, Clesceri LS, Eaton AD. (1992). Standard Methods for the Examination of Water  
526 and Wastewater. American Public Health Association: Washington (DC), USA.

527 Grime JP. (1998). Benefits of plant diversity to ecosystems: intermediate, filter and founder effects.  
528 *J Ecol* **86**: 902–910.

529 Hambright KD, Beyer JE, Easton JD, Zamor RM, Easton AC, Hallidayschult TC. (2015). The  
530 niche of an invasive marine microbe in a subtropical freshwater impoundment. *ISME J* **9**:  
531 256–264.

532 Hamilton WD, Lenton .TM. (1998). Spora and Gaia: how microbes fly with their clouds. *Etho*  
533 *Ecol Evol* **10**: 1–16.

534 Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic  
535 patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

536 Holt AR, Warren PH, Gaston KJ. (2004). The importance of habitat heterogeneity, biotic  
537 interactions and dispersal in abundance–occupancy relationships. *J Anim Ecol* **73**: 841–851.

538 Hutchinson GE. (1964). The lacustrine microcosm reconsidered. *Am Sci* **52**: 334–341.

539 Jones SE, Cadkin TA, Newton RJ, McMahon KD. (2012). Spatial and temporal scales of aquatic  
540 bacterial beta diversity. *Front Microbiol* **3**: 1–10.

541 Ju LH, Yang J, Liu LM, Wilkinson DM. (2014). Diversity and distribution of freshwater testate  
542 amoebae (Protozoa) along latitudinal and trophic gradients in China. *Microb Ecol* **68**:  
543 657–670.

544 Judd KE, Crump BC, Kling GW. (2006). Variation in dissolved organic matter controls bacterial  
545 production and community composition. *Ecology* **87**: 2068–2079.

546 Kent M. (2012). Vegetation Description and Data Analysis. 2<sup>nd</sup> ed. Wiley-Blackwell: Chichester,

547 UK.

548 Kim T, Jeong J, Wells GF, Park H. (2013). General and rare bacterial taxa demonstrating different  
549 temporal dynamic patterns in an activated sludge bioreactor. *Appl Microbiol Biotechnol* **97**:  
550 1755–1765.

551 Lear G, Bellamy J, Case BS, Lee JE, Buckley HL. (2014). Fine-scale spatial patterns in bacterial  
552 community composition and function within freshwater ponds. *ISME J* **8**: 1715–1726.

553 Legendre P, Bordard D, Peres-Neto P. (2008). Analyzing or explaining betadiversity? Comment.  
554 *Ecology* **89**: 3238–3244.

555 Legendre P, Legendre L. (2012). Numerical Ecology. 3<sup>rd</sup> Edition. Elsevier: Amsterdam, The  
556 Netherlands.

557 Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF *et al.* (2004). The  
558 metacommunity concept: a frame work for multi-scale community ecology. *Ecol Lett* **7**:  
559 601–613.

560 Lepš J, Šmilauer P. (2003). Multivariate Analysis of Ecological Data Using CANOCO. Cambridge  
561 University Press: Cambridge, UK.

562 Lindström ES, Langenheder S. (2012). Local and regional factors influencing bacterial community  
563 assembly. *Environ Microbiol Rep* **4**: 1–9.

564 Liu LM, Yang J, Yu XQ, Chen GJ, Yu Z. (2013). Patterns in the composition of microbial  
565 communities from a subtropical river: Effects of environmental, spatial and temporal factors.  
566 *PLoS ONE* **8**: e81232.

567 Logares R, Lindström ES, Langenheder S, Logue JB, Paterson H, Laybourn-Parry J *et al.* (2013).  
568 Biogeography of bacterial communities exposed to progressive long-term environmental  
569 change. *ISME J* **7**: 937–948.

570 Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R *et al.* (2014). Patterns of rare and  
571 abundant marine microbial eukaryotes. *Curr Biol* **24**: 813–821.

572 Logue JB, Lindström ES. (2008) Biogeography of bacterioplankton in inland waters. *Freshw Rev*  
573 **1**: 99–114

574 Logue JB. (2010). Factors influencing the biogeography of bacteria in freshwaters—a  
575 metacommunity approach. Ph.D. thesis, Uppsala University, Uppsala, Sweden.

576 Magurran AE, Henderson PA. (2003). Explaining the excess of rare species in natural species  
577 abundance distributions. *Nature* **422**: 714–716.

578 Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al.* (2006).  
579 Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

580 Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK *et al.* (2011). Global  
581 patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**: 135–144.

582 Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. (2011). A guide to the natural history  
583 of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**: 14–49.

584 Pedrós-Alió C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**:  
585 257–263.

586 Pedrós-Alió C. (2012). The rare bacterial biosphere. *Annu Rev Mar Sci* **4**: 449–466.

587 Pernthaler J. (2005). Predation on prokaryotes in the water column and its ecological implications.  
588 *Nat Rev Microbiol* **3**: 537–546.

589 Pester M, Bittner N, Deevong P, Wagner M, Loy A. (2010). A ‘rare biosphere’ microorganism  
590 contributes to sulfate reduction in a peatland. *ISME J* **4**: 1591–1602.

591 Preston FW. (1948). The commonness and rarity of species. *Ecology* **29**: 254–283.

592 Ramette A, Tiedje JM. (2007). Biogeography: an emerging cornerstone for understanding  
593 prokaryotic diversity, ecology, and evolution. *Microb Ecol* **53**: 197–207.

594 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing  
595 MOTHUR: open-source, platform-independent, community-supported software for describing  
596 and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

597 Teira E, Lekunberri I, Gasol JM, Nieto-Cid M, Álvarez-Salgado XA, Figueiras PG. (2007).  
598 Dynamics of the hydrocarbon-degrading *Cycloclasticus* bacteria during mesocosm-simulated  
599 oil spills. *Environ Microbiol* **9**: 2551–2562.

600 Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naïve Bayesian classifier for rapid assignment  
601 of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

602 Wang J, Shen J, Wu Y, Tu C, Soininen J, Stegen JC *et al.* (2013). Phylogenetic beta diversity in  
603 bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J* **7**:  
604 1310–1321.

605 Wilkinson DM, Koumoutsaris S, Mitchell EAD, Bey I. (2012). Modelling the effects of size on  
606 the aerial dispersal of microorganisms. *J. Biogeogr* **39**: 89–97.

607

608 **Figure Legends**

609 **Figure 1** Location of the 42 sampling sites in China. FJ (included 5 reservoirs) –  
610 Fujian Province, southeast China; CJ (9 lakes) – the lower and middle reaches of  
611 Changjiang River, China; ECC (6 lakes) – east central China, IM (13 lakes) – Inner  
612 Mongolia, north China; NEC (9 lakes) – northeast China.

613

614 **Figure 2** MDS ordination for bacterioplankton communities from 42 lakes and  
615 reservoirs of China. All – all bacterial taxa, abundant – abundant taxa, rare – rare taxa.  
616 For region abbreviations see Figure 1.

617

618 **Figure 3** Spearman's rank correlations between the Bray-Curtis similarity of  
619 bacterioplankton community and geographical distance (n is the number of  
620 comparison).

621

622 **Figure 4** Spearman's rank correlation between median of bacterial OTU relative  
623 abundance and number of sites occupied (n is the number of OTUs).

624

625 **Figure 5** CCA ordination showing the bacterial community composition in relation to  
626 significant local environmental variables and regional geographical factors ( $P < 0.05$ ).  
627 WT – water temperature, EC – electrical conductivity, Turb – turbidity, TN – total  
628 nitrogen.