The ISME Journal (2012) 6, 1127–1136

www.nature.com/ismej

ORIGINAL ARTICLE

Freshwater bacterioplankton richness in oligotrophic lakes depends on nutrient availability rather than on species–area relationships

Jürg Brendan Logue[1], Silke Langenheder[1], Anders F Andersson[1,5], Stefan Bertilsson[1], Stina Drakare[2], Anders Lanzén[3,4] and Eva S Lindström[1]

[1]*Department of Ecology and Genetics/Limnology, Uppsala University, Uppsala, Sweden;* [2]*Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden;* [3]*Centre for Geobiology, University of Bergen, Bergen, Norway and* [4]*Department of Biology, University of Bergen, Bergen, Norway*

A central goal in ecology is to grasp the mechanisms that underlie and maintain biodiversity and patterns in its spatial distribution can provide clues about those mechanisms. Here, we investigated what might determine the bacterioplankton richness (BR) in lakes by means of 454 pyrosequencing of the 16S rRNA gene. We further provide a BR estimate based upon a sampling depth and accuracy, which, to our knowledge, are unsurpassed for freshwater bacterioplankton communities. Our examination of 22 669 sequences per lake showed that freshwater BR in fourteen nutrient-poor lakes was positively influenced by nutrient availability. Our study is, thus, consistent with the finding that the supply of available nutrients is a major driver of species richness; a pattern that may well be universally valid to the world of both micro- and macro-organisms. We, furthermore, observed that BR increased with elevated landscape position, most likely as a consequence of differences in nutrient availability. Finally, BR decreased with increasing lake and catchment area that is negative species–area relationships (SARs) were recorded; a finding that re-opens the debate about whether positive SARs can indeed be found in the microbial world and whether positive SARs can in fact be pronounced as one of the few 'laws' in ecology.

The ISME Journal (2012) **6,** 1127–1136; doi:10.1038/ismej.2011.184; published online 15 December 2011
Subject Category: microbial population and community ecology
Keywords: bacterial richness; biogeography; freshwater; landscape position; productivity; species–area relationship

Introduction

A primary goal of microbial ecology is to understand the distribution of microbial diversity. To study this distribution, the diversity of microbial organisms needs to be adequately measured. Microbial communities have, however—for methodological shortcomings—been hitherto grossly undersampled, leaving the rare taxa unseen (Curtis and Sloan, 2005; Curtis *et al.*, 2006; Woodcock *et al.*, 2006). As a consequence thereof, reliable estimates of microbial richness have remained difficult to obtain. Novel high-throughput sequencing technologies might, though, provide the tools necessary to dip into this 'rare biosphere' (Pedros-Alio, 2006, 2007; Sogin *et al.*, 2006) and to explore microbial diversity to a greater acuity and depth. Yet, greater insight into the extent of microbial diversity might enable microbial ecologists to employ and test established (macro-ecological) theories of biogeography and community assembly. For at present, there is a vivid debate on whether or not microbial diversity shows explainable patterns of spatial distribution comparable to those of macro-organisms (Fenchel, 2003; Martiny *et al.*, 2006), such as the relationship between species richness and ecosystem productivity (for example, Waide *et al.*, 1999; Mittelbach *et al.*, 2001; Cardinale *et al.*, 2009) or the relationship between the number of species in a given area and the size of that area (species–area relationship, SAR) (for example, Arrhenius, 1921; Gleason, 1922; Rosenzweig, 1995).

At the bottom of this debate on the spatial distribution of microbial diversity lies the endeavour to grasp the underlying mechanisms. It is, by now, generally accepted that microbial diversity and its distribution are governed by a complex interplay of both local and regional drivers (for example,

Correspondence: JB Logue, Department of Ecology and Genetics/Limnology, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden.
E-mail: jb.logue@ebc.uu.se
[5]Present address: Science for Life Laboratory, KTH Royal Institute of Technology, 10044 Stockholm, Sweden.
Received 28 March 2011; revised 23 October 2011; accepted 4 November 2011; published online 15 December 2011

Martiny *et al.*, 2006; Logue and Lindström, 2008). Lakes provide a unique object of research to investigate the relative roles of these drivers, in that their distinct boundaries enable a clear discrimination between dynamics acting from within and from without the ecosystem. Further, recognising lakes as integral components of a region, lake ecologists have begun to implement a basic tenet of landscape ecology, namely that the spatial position of an ecosystem within a landscape influences the properties of that very system. In doing so, Kratz and colleagues (Kratz *et al.*, 1991, 1997; Webster *et al.*, 1996; Soranno *et al.*, 1999; Magnuson and Kratz, 2000; Riera *et al.*, 2000) developed the concept of lake landscape position. One of the metrics used so far to define a lake's position in a landscape is lake chain number (LCN), which measures lake landscape position with respect to lakes connected along a linear chain through primarily surface-flow systems (Soranno *et al.*, 1999).

The relative importance of local and regional drivers for microbial diversity is, however, poorly understood (Logue and Lindström, 2008) and only very few (Yannarell and Triplett, 2005; Crump *et al.*, 2007; Nelson *et al.*, 2009) have turned their attention to the role of a lake's position within a landscape. Besides, bacterioplankton biogeography studies have primarily focused on β- rather than on α-diversity. Thus, the general goal of the present study is to investigate local and regional mechanisms and their respective relevance for driving local bacterioplankton richness (BR) (α-diversity). We hypothesise that BR increases with (i) nutrient availability, (ii) lake area (LA) and catchment area (CA) and (iii) LCN. Regarding hypothesis (i), it has been shown that the availability of and the ratio between nutrients together with the efficiency by which species capture these and convert them into biomass, explain productivity-diversity patterns (Cardinale *et al.*, 2009). Pertaining to hypothesis (ii), the theory of island biogeography (MacArthur and Wilson, 1967) predicts that larger habitat areas support a greater number of species (that is, a positive SAR). An increase in habitat area is assumed to lead to an increase in the likelihood of immigration, to a decrease in the probability of extinction, and to an increase in habitat heterogeneity; all considered drivers of biological diversity. Moreover, the likelihood of immigration into a local community (that is, the lake) from the regional species pool (that is, the catchment) might be elevated with an increase in CA. Hence, both lake and CA are used as proxies for the size of a habitat for bacterioplankton communities. As to hypothesis (iii), it has been shown that the position of a lake within a landscape can dictate the lake's hydrological, geomorphological, and biological properties (for example, Kratz *et al.*, 1997); properties all of which have the potential to influence BR. Thus, with shifts in the relevance of hydrological flowpaths, geomorphological, and biological characteristics, BR might change in dependence of LCN. Furthermore, nutrient availability is thought to increase with increasing CA and CAs become bigger the further down a lake is in a lake chain. Therefore, increasing LCN and therewith-related increases in CA and nutrient availability might yield an increase in BR on the grounds of a positive SAR and productivity–diversity relationship, respectively. Both, the productivity–diversity relationship and the SAR have been reported for microbial communities (Horner-Devine *et al.*, 2003, 2004; Green *et al.*, 2004; Bell *et al.*, 2005; Reche *et al.*, 2005; van der Gast *et al.*, 2006; Smith, 2007), yet none of these studies addressed these relationships by means of high-throughput sequencing technologies and, thus, were able to capture microbial richness to the extent necessary to provide adequate estimations thereof (Curtis and Sloan, 2005; Curtis *et al.*, 2006; Woodcock *et al.*, 2006). Hence, to investigate our hypotheses, we surveyed 14 Swedish lakes sharing origin and climate, using 16S rRNA gene pyrosequencing data.

Materials and methods

Study sites and sampling

A total of 14 lakes were included in the sampling campaign, all being drainage lakes and situated in the Ånnsjön-area of Jämtland, Sweden (Figure 1). Eight, respectively five, lakes are, in each case, part of a distinct lake chain, both of which converge in Lake Öster-Noren (ÖN).

Sampling took place on 30 and 31 August 2007, when lakes were entirely mixed. Depth-integrated water samples were taken either at the centre of the lake or no <300 m off shore. Depth-integrated sampling was conducted as follows: water was collected every $\frac{1}{2}$, 1, 2, 3, or 5 m over the entire water column, depending on the lake's water depth at the site of sampling. Mixed on site, water samples were subsequently divided into two subsamples: one for water chemistry and one for bacterial abundance and richness analyses. Subsamples were kept cold and in the dark for 1–8 h until further processing in the laboratory.

Lake and CAs were determined and calculated using ArcGIS 9.2 (Environmental Sciences Research Institute, Redlands, CA, USA).

Chemical analyses and bacterial abundance

Chemical analyses. Water chemistry samples were returned cold to the laboratory, whereupon they were divided anew into subsamples. Chemical analyses of total phosphorus (Tot-P), total nitrogen (Tot-N), total organic carbon (TOC) and the ratio between the absorbance measured at 250 nm and at 365 nm (A_{250}/A_{365}) were performed as described previously by Logue and Lindström (2010). Despite all lakes being highly oligotrophic in nature (Tot-P $<10\,\mu g\,l^{-1}$, Tot-N $<0.5\,mg\,l^{-1}$), Tot-P and Tot-N concentrations nevertheless varied distinctively amongst the study lakes with factors 3.4 and 8.4, respectively (see Supplementary Table SI1).

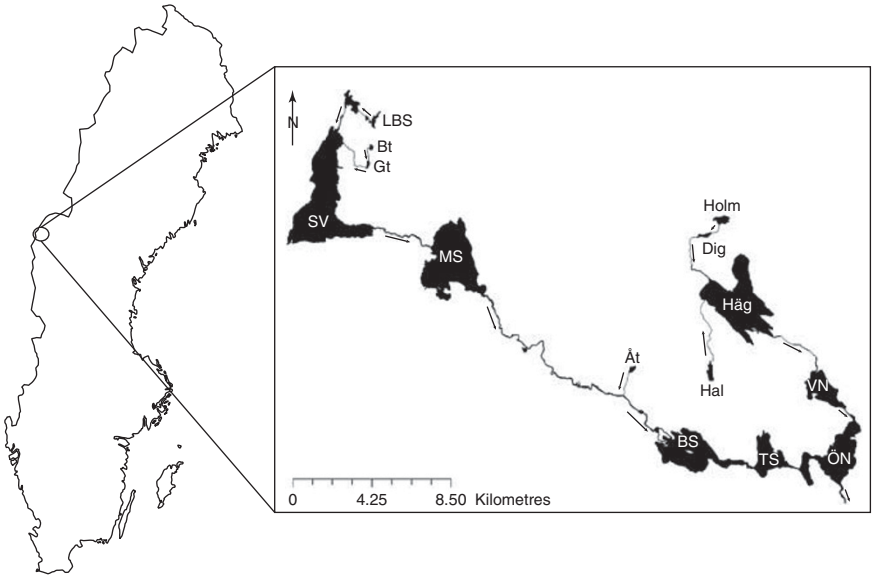

Figure 1 Map visualising the study area, pointing out the 14 study lakes and highlighting the direction of flow. Note that only the connecting streams between the lakes are visualised. Abbreviations of lake names are as in Table 1. Printed with permission from Lantmäteriet (Lantmäteriet Gävle (2010): Permisson I 2010/0058).

Bacterial abundance. A 20-ml subsample was preserved in formaldehyde (final concentration: 4% wv^{-1}) and bacterial abundance was subsequently enumerated by epifluorescence microscopy of acridine-orange-stained cells (Hobbie *et al.*, 1977). Bacterial numbers are based on counts of at least 400 cells in a minimum of 20 microscopic fields.

Nucleic-acid extraction, PCR and 454 pyrosequencing
Bacterioplankton cells were collected onto 0.2 μm. membrane filters (Supor-200 Membrane Disc Filters, 47 mm; Pall Corporation, East Hills, NY, USA), filtering 0.2 l of pre-sieved (225 μm mesh size) lake water. Pre-sieving was carried out to avoid capturing larger particles, such as zooplankton, and did not significantly alter bacterioplankton abundances. Filters were placed into sterile cryogenic vials (Nalgene, Rochester, NY, USA) and stored immediately in liquid nitrogen for transportation to the laboratory, where they were finally kept at −80 °C.

Nucleic-acid extraction. Nucleic-acid extraction was performed following the protocol #3 of the Easy-DNA kit (Invitrogen, Carlsbad, CA, USA) with an extra 0.2 g of 0.1 mm zirconia/silica beads. Extracted nucleic acids were sized and quantified by means of agarose (1%) gel electrophoresis, GelRed staining (Biotium Inc., Hayward, CA, USA) and ultraviolet transillumination.

PCR amplification and template preparation. The bacterial hypervariable regions V3 and V4 of the 16S rRNA gene were PCR amplified using forward primer 341 (5′-CCTACGGGNGGCWGCAG-3′) and individually bar-coded reverse primers 805 (5′-GAC TACHVGGGTATCTAATCC-3′) (note that 454 adaptor (454 Life Sciences, Branford, CT, USA) and seven-base long bar-code sequences are not shown here). Primer 341F matched perfectly to 891 287 of 933 229 bacterial sequences spanning *Escherichia coli* positions 300–400, whereas primer 805R matched perfectly to 727 896 of 815 023 bacterial sequences spanning *E. coli* positions 750–850 (Ribosomal Database Project (RDP), v10.20, http://rdp.cme.msu.edu; Cole *et al.*, 2009). PCR reactions were performed in a 20-μl reaction volume comprising 0.4 U Phusion high-fidelity DNA polymerase (Finnzymes, Espoo, Finland), 1X Phusion HF reaction buffer (Finnzymes), 200 μM of each dNTP (Invitrogen), 500 nM of each primer (Eurofins MWG, Ebersberg, Germany), 0.4 mg ml^{-1} BSA (New England Biolabs, Ipswich, UK) and finally 5–10 ng of extracted nucleic acid. Thermocycling (DNA Engine (PTC-200) Peltier Thermal Cycler; Bio-Rad Laboratories, Hercules, CA, USA) was conducted with an initial denaturation step at 95 °C for 5 min, followed by 25 cycles of denaturation at 95 °C for 40 s, annealing at 53 °C for 40 s and extension at 72 °C for 1 min, and finalised with a 7-min extension step at 72 °C. Four technical replicates were run per sample, pooled after PCR amplification and purified using the Agencourt AMPure XP purification kit (Beckman Coulter Inc., Brea, CA, USA). Nucleic acid yields were checked on a fluorescence microplate reader (Ultra 384; Tecan Group Ltd, Männedorf, Switzerland) employing the Quant-iT PicoGreen dsDNA quantification kit (Invitrogen). Finally, PCR amplicons were pooled in equal proportions to obtain a similar number of 454 pyrosequencing reads per sample.

454 Pyrosequencing. The final amplicon was 454 pyrosequenced with a 454 GS FLX system (454

Life Sciences) at the Norwegian High-Throughput Sequencing Centre, University of Oslo (NSC; Oslo, Norway; http://www.sequencing.uio.no), using Titanium chemistry.

Sequence analyses

Pyrosequence noise filtering and chimera removal were carried out employing a modified version of the PyroNoise flowgram-clustering algorithm (i.e. AmpliconNoise) as described in Quince *et al.* (2011).

Making use of an in-house Perl (http://www.perl.org) script (Andersson *et al.*, 2010), pyrosequencing reads not matching bar-code and/or primer sequences and shorter than 350 bp were further removed and primer sequences subsequently trimmed from both the beginning and the end of each 'good' read. Pyrosequencing reads were thereupon truncated at 350 bp, removing further noise, which is thought to increase towards the 3′ end of a pyrosequencing read (Mardis, 2008). After extracting all non-redundant 454 pyrosequencing reads, these were aligned using the Aligner tool from the RDP's Pyrosequencing Pipeline (http://pyro.cme.msu.edu). Aligned sequences were then clustered into operational taxonomic units (OTUs) at a level of 97% sequence identity, employing the Complete Linkage Clustering tool from the RDP's Pyrosequencing Pipeline. Again, using the in-house Perl script, each 454 pyrosequence was BLASTN (Altschul *et al.*, 1997) searched against a local BLAST database, which comprised 600 316 unique bacterial 16S rRNA gene sequences longer than 1200 bases with good Pintail score downloaded from RDP (v10.22). BLASTN searching was performed using default parameters. 454 pyrosequencing reads have been deposited in the National Center for Biotechnology Information Sequence Read Archive under accession number SRP005457.

Data analyses

For statistical data analyses, sampling efforts (number of sequences obtained per sample) were normalised across the 14 samples because sampling efforts were not equally distributed and samples with the highest number of sequences are expected to yield comparatively higher richness estimates. Normalisation was performed through re-sampling and richness estimates were computed from 22 669 sequences that were randomly drawn from each sample.

BR was estimated using Chao1, ACE and a Bayesian parametric richness estimator (according to Quince *et al.*, 2008). Yet, the richness estimators yielded concurrent results, thus only results for Chao1 are given, which was computed as

$$S_{\mathrm{Chao1}} = S_{\mathrm{obs}} + F_1(F_1 - 1)/2(F_2 + 1),$$

where S_{obs} is the observed number of taxa and F_i the number of taxa with exactly i individuals.

To test for relationships between BR and LCN, LA, CA, and the environment, Spearman's rank correlations were carried out. For use in the correlation analyses, the environmental data-set was first analysed by principal component analysis so as to distil most of the variation of each individual environmental parameter into one single environmental predictor variable (principal component (PC) 1). The environmental parameters Tot-P, Tot-N, TOC, A_{250}/A_{365} (as a proxy for organic matter quality), and conductivity were, therefore, first $\log_{10}$-transformed (except for A_{250}/A_{365}, which was arcsin square-root transformed) and then transformed to z-scores (Legendre and Legendre, 1998). PC 1 (strong loadings from Tot-P, Tot-N, and TOC) and PC 2 (strong loadings from A_{250}/A_{365} and conductivity) explained 54% and 28%, respectively, of the variation within the environmental data-set for all lakes. Only PC 1 was, though, included in the correlation analyses.

Specifically testing the prediction of island biogeography (MacArthur and Wilson, 1967) that larger habitat areas support a greater number of species (that is, a positive SAR) was conducted according to the power-law function generalised by Arrhenius (1921) and Gleason (1922):

$$S = cA^z,$$

where S is the number of species, A is the size of the sampled area, c is an empirically derived species- and location-specific constant (the intercept in log-log space), and z is a measure of the rate of turnover of species across space (the slope of the SAR). With regard to regression analyses, BR, LA, and CA were first $\log_{10}$-transformed.

Moreover, to assess the independent effect of Tot-P, LCN, LA, and CA on BR, partial Spearman's rank correlation analyses were performed. The partial Spearman's rank correlation coefficient ($\rho_{XY \cdot Z}$) measures the degree of association between two variables (Y, X), while controlling for a third one (Z, constant variable).

To evaluate differences in BR and environmental characteristics (Tot-P, Tot-N, TOC, A_{250}/A_{365}, and conductivity) between lakes higher up in the landscape and those lower down, Wilcoxon rank sum tests were performed. To correct for multiple comparisons, sequential Bonferroni adjustments were carried out. Lake area was the criterion based upon which the grouping was carried out (Supplementary Table SI1). Thus, lakes higher up in the landscape include lakes of LCN one and two, whereas lakes lower down in the landscape comprise lakes of LCN three to six.

All statistical data analyses were conducted in R (R, 2008) to which a significance level of 0.05 was applied, unless otherwise stated (see Table 3).

Results

Bacterioplankton richness

In total, 567 093 pyrosequencing reads spanning the V3 and V4 region of the 16S rRNA gene were

Table 1 454 Pyrosequencing effort and bacterioplankton richness estimates at 97% sequence identity

Lake	*Lake ID*	*Total number of sequences*	*Number of sequences after AmpliconNoise*	*Number of "good" sequences*[a]	*Number of unique sequences*[b]	*Number of singletons*[c]	*Number of unique OTUs at the 97% cutoff level*[d]	*Chao1 at the 97% cutoff level*[e]	*ACE at the 97% cutoff level*[e]	*Parametric richness estimator at the 97% cutoff level*[f]
Lång-Björsjön	LBS	37 842	25 263	22 669	1549	858	1240	2080	3523	7052
Bustadtjärnen	Bt	41 323	31 424	28 803	1686	1076	1236	2846	5265	9305
Gravatjärnen	Gt	42 518	32 269	28 682	2338	1648	1639	4171	7932	13 374
Skalsvattnet	SV	36 150	27 273	24 668	1168	497	838	1224	1633	1970
Medstugusjön	MS	41 138	31 294	28 407	1152	508	804	1386	1949	2903
Åtvändtjärnen	Åt	42 225	32 003	28 677	2364	1632	1677	3992	6991	13 317
Bodsjön	BS	40 874	30 726	25 872	1683	956	1217	2455	3841	7704
Tännsjön	TS	43 289	32 872	30 465	1223	566	813	1500	1974	3241
Holmtjärnen	Holm	40 805	30 635	28 793	1048	465	739	1370	1827	2846
Digernästjärnen	Dig	41 590	31 402	28 559	1208	570	882	1544	2423	3640
Häggsjön	Häg	39 562	30 458	27 321	1082	441	732	1097	1324	1539
Hallåstjärnen	Hal	39 546	30 399	29 196	1322	732	919	1893	3123	6593
Väster-Noren	VN	39 423	30 267	26 902	999	432	723	1260	1555	1938
Öster-Noren	ÖN	40 808	31 312	30 009	1003	419	664	1167	1456	1959

Abbreviation: OTUs, operational taxonomic units.
[a]Number of sequences after flowgram pre-filtering (AmpliconNoise) and removal of chimeric and short (<350 bp) sequences.
[b]Number of unique sequences after flowgram pre-filtering (AmpliconNoise), removal of chimeric and short (<350 bp) sequences and truncation of sequences at 350 bp.
[c]Number of singletons after flowgram pre-filtering (AmpliconNoise), removal of chimeric and short (<350 bp) sequences and truncation of sequences at 350 bp.
[d]OTU clustering at 97% sequence identity after normalisation (see *Material and methods*).
[e]Chao1 and ACE richness estimates at 97% sequence identity (after normalisation, see *Material and methods*).
[f]Bayesian parametric richness estimate using an inverse Gaussian distribution (see Quince *et al.*, 2008) (after normalisation, see *Materials and methods*).

retrieved, of which 389 023 remained after the employment of AmpliconNoise and removal of chimeric and short (<350 bp) sequences that is between 22 669 and 30 465 'good' reads per lake (Table 1). The number of unique sequences for each lake ranged from 999 to 2364, with a grand total of 14 250 unique sequences. Grouping these into OTUs at a 97% sequence identity cut-off level resulted in 664–1677 OTUs per sample, with a total of 6309 OTUs (after normalisation, see *Material and methods*). Rarefaction analysis based on OTU clustering (97% sequence identity) suggests that a complete census of the richness in these bacterioplankton communities was not accomplished as the rarefaction curves did not yet reach the plateau phase (Figure 2). Estimates of BR varied between 1097 and 4171 (Chao1) (Table 1).

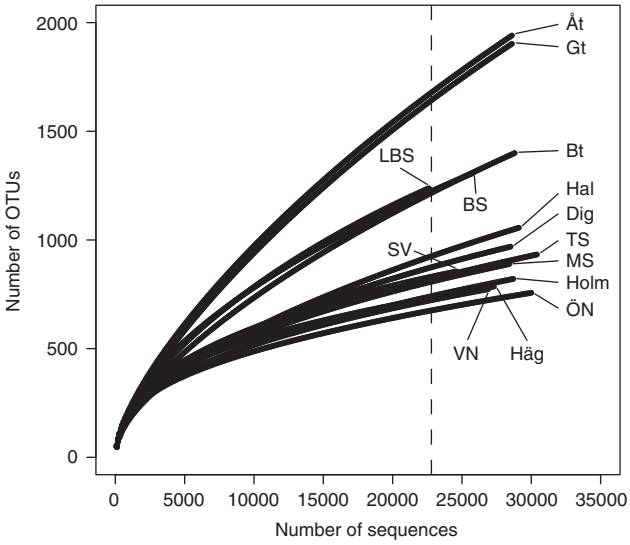

Figure 2 Rarefaction curves for the 14 lake bacterioplankton communities at a level of 97% sequence identity between 16S rRNA gene fragments, covering the hypervariable regions V3 and V4. The dashed line visualises the lowest sampling effort for which all samples were normalised during data analyses. Abbreviations of lake names are as in Table 1.

BR, LCN, LA, CA, and the environment

BR did not increase within increasing LCN in either of the two lake chains (Figures 3a and b). In fact, BR decreased with increasing LCN (Spearman's rank correlations; $\rho_{S1} = -0.644$, $P_1 = 0.061$; $\rho_{S2} = -0.725$, $P_2 = 0.103$). In addition, the relationships between BR and LA (Figure 3c) (Spearman's rank correlation; $\rho_S = -0.785$, $P = 0.001$) and between BR and CA (Figure 3d) (Spearman's rank correlation: $\rho_S = -0.543$, $P = 0.048$) were both significantly negative. The relationship between BR and the local environment (PC 1; strong loadings from Tot-P, Tot-N and TOC but see *Material and methods* for an in-depth description), on the other hand, was highly significant (Figure 3e) (Spearman's rank correlation; $\rho_S = 0.622$, $P = 0.020$). The relationship between BR and Tot-P was, too, highly significant with BR increasing with increasing Tot-P concentrations (Figure 3f) (Spearman's rank correlation; $\rho_S = 0.859$, $P = 0.000$) and exemplifies the general trend observed between BR and nutrient concentrations (Tot-N, $\rho_S = 0.660$; TOC, $\rho_S = 0.701$). Finally, the influence of Tot-P on BR remained significant when

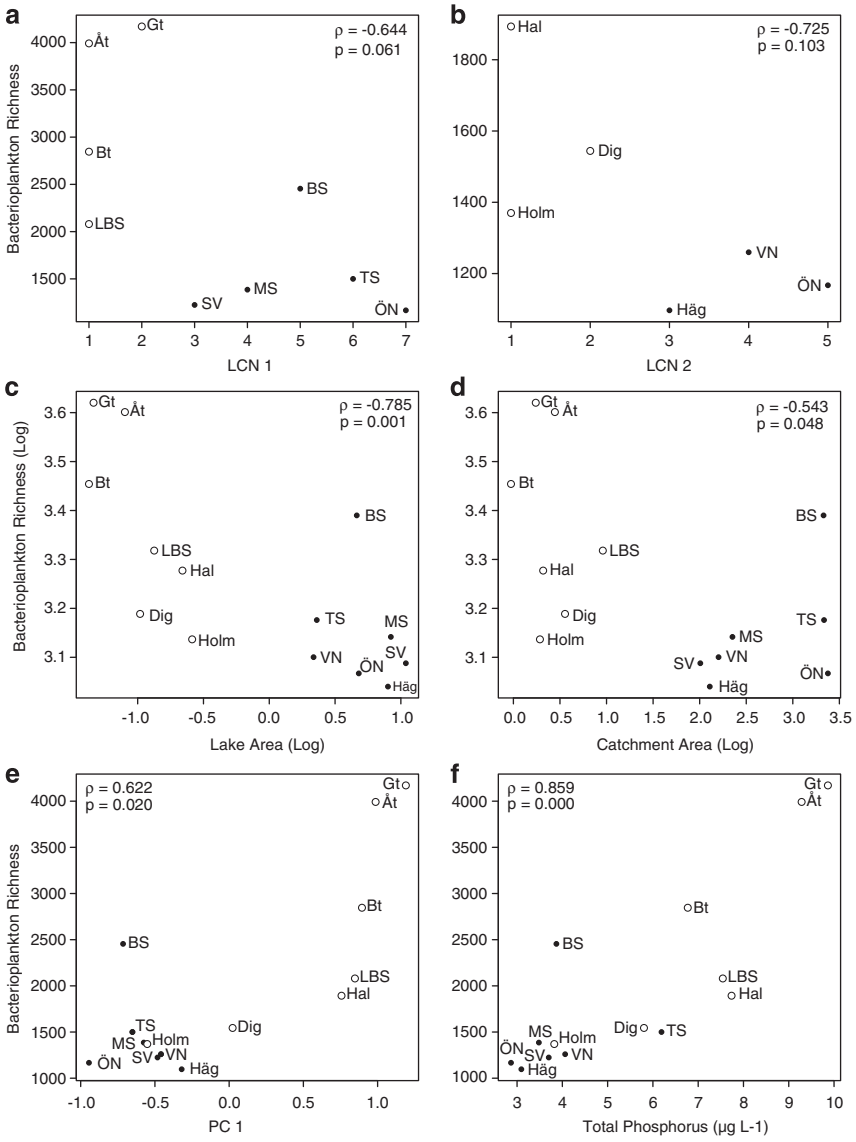

Figure 3 Relationship between BR and (**a**) LCN 1 and (**b**) LCN 2, (**c**) LA, (**d**) CA, (**e**) the environment (PC 1; strong loadings from Tot-P, Tot-N, and TOC but see *Materials and methods* for an in-depth description) and (**f**) Tot-P content. BR is given as the Chao1 richness indices. For visualisation of the relationships between BR and LA and BR and CA, BR, LA, and CA were $\log_{10}$-transformed. Open circles correspond to the lakes high up in the landscape (LCN of one and two), while filled circles refer to lakes lower down in the landscape (LCN of three to six). Abbreviations of lake names are as in Table 1.

controlling for LA, CA, and LCN (Table 2). Yet, the correlation between BR and LCN remained insignificant, whereas the correlations between BR and LA and between BR and CA became insignificant once controlling for Tot-P (Table 2).

The *z*-values for the SARs between BR and LA and between BR and CA were both negative ($z_{LA} = -0.151$, $z_{CA} = -0.078$), however, only the first was significant ($P_{LA} = 0.004$, $P_{CA} = 0.059$).

BR, landscape position, and the environment

BR in lakes higher up in the landscape (lakes of LCN one and two) was significantly higher than that in lakes lower down in the landscape (lakes of LCN three to six) (Wilcoxon rank sum test; $W = 43$, $P = 0.017$). Moreover, several of the environmental parameters measured did also differ amongst these two groups (Table 3).

Discussion

The ready application of novel high-throughput sequencing technologies, such as 454 pyrosequencing, has revealed a tremendous diversity within microbial communities (for example, Sogin *et al*., 2006; Galand *et al*., 2009; Andersson *et al*., 2010; Galand *et al*., 2010; Kirchman *et al*., 2010) and has lead to the discovery and coining of the

Table 2 Results from partial Spearman's rank correlation analyses

Variable Y	*Variable* X	*Variable* Z	$\rho_{XY \cdot Z}$	P
BR	Tot-P	LCN1	0.822	0.000*
BR	LCN1	Tot-P	0.040	0.922
BR	Tot-P	LCN2	0.798	0.022*
BR	LCN2	Tot-P	−0.447	0.387
BR	Tot-P	LA	0.599	0.013*
BR	LA	Tot-P	−0.243	0.406
BR	Tot-P	CA	0.793	0.000*
BR	CA	Tot-P	−0.030	0.920

Abbreviations: BR, bacterioplankton richness; CA, catchment area, LA, lake area; LCN, lake chain number; Tot-P, total phosphorus.
Variable *Z* is the variable that was held constant. The partial correlation coefficient is given as $\rho_{XY \cdot Z}$. Significant *P*-values are labelled with an asterisk.

Table 3 Testing environmental parameters for differences between the two groups of lakes (high up versus low down in the landscape), applying Wilcoxon rank sum tests

	p_W	*Level of significance*
TOC	0.0006*	0.010
Tot-P	0.007*	0.013
Tot-N	0.009*	0.017
Conductivity	0.209	0.025
A_{250}/A_{365}	0.798	0.050

Abbreviations: A_{250}/A_{365}, ratio between absorbances at 250 nm and 365 nm; TOC, total organic carbon; Tot-N, total nitrogen; Tot-P, total phosphorus.
P-values are Bonferroni corrected (sequential). Significant *P*-values are labelled with an asterisk.

'rare biosphere' (Pedros-Alio, 2006, 2007; Sogin *et al.*, 2006). Here, we investigated what might determine freshwater BR; that is, we explored the relative roles of regional (landscape position, lake and CA) and local (environmental properties) drivers for BR. BR was, therefore, assessed by means of 454 pyrosequencing.

We recorded a highly significant relationship between BR and local lake environmental properties (Figure 3e). Our results indicate that the availability of nutrients of these highly oligotrophic lakes (see Supplementary Table SI1) accounted for most of the variation detected in BR amongst the 14 study lakes (Figure 3f, see also SI2 for phyla-related patterns). The overall availability of limiting resources, such as Tot-P, Tot-N, TOC, and organic matter, is believed to be, in part, accountable for the relationship between productivity and species richness (see Cardinale *et al.*, 2009). Relationships between productivity and species richness have been reported repeatedly for plants and animals, yet the form of this relationship and the underlying mechanisms determining this apparent relationship remain elusive (for example, Mittelbach *et al.*, 2001; Chase and Leibold, 2002; Pärtel *et al.*, 2007). However, species richness is, in general, hypothesised to first increase and then decrease with productivity, producing a hump-shaped relationship (Rosenzweig, 1995; Mittelbach *et al.*, 2001). Evidence from laboratory studies (Bohannan and Lenski, 1997; Kassen *et al.*, 2000) and observations from field surveys (Fisher *et al.*, 2000; Horner-Devine *et al.*, 2003) indicate that microbial diversity can, too, be influenced by productivity. Horner-Devine *et al.* (2003) observed that different microbial groups can differ in response to changes in productivity, exhibiting both unimodal and *U*-shaped relationships. The positive linear relationship between BR and nutrient availability found in our nutrient-poor study lakes (see Supplementary Table SI1) might describe the positive, monotonically increasing phase located on the left side of the expected hump-shaped productivity-richness curve (Figure 3f). Yet, to be able to draw conclusions about the generality of the productivity–diversity relationship across all groups of organisms, a greater number of microbial studies are needed that investigate this relationship over a wider range of trophic and temporal conditions.

Our results, furthermore, show that BR neither increased with increasing LCN nor with increasing LA or CA, in fact, it decreased along the lake chains (Figures 3a–d, see also SI2 for phyla-related patterns). This observation of negative SARs is a finding of particular interest, given that positive SARs are regarded to be one of the few general patterns in ecology (Lawton, 1999). Even though it has been observed that SARs become less pronounced if the body size of the organism is small (Drakare *et al.*, 2006), a handful of previous studies have reported a positive SAR for microbial eukaryotes and bacteria (Green *et al.*, 2004; Horner-Devine *et al.*, 2004; Bell *et al.*, 2005; Reche *et al.*, 2005; van der Gast *et al.*, 2006). Yet, it has to be assumed that this discrepancy in results between these and our study is very likely methodological in nature. Microbial richness estimates in the above-mentioned studies derived either from community-profiling techniques (Green *et al.*, 2004; Bell *et al.*, 2005; Reche *et al.*, 2005), an approach heavily criticised for yielding results that have very little to do with the actual richness (Loisel *et al.*, 2006; Blackwood *et al.*, 2007), or from 16S rRNA clone libraries (Horner-Devine *et al.*, 2004). Both, traditional cloning and sequencing and community profiling methodologies, however, markedly undersample communities (Curtis and Sloan, 2005; Curtis *et al.*, 2006; Woodcock *et al.*, 2006) and, thus, most likely provide less reliable richness estimates compared with the 454-pyrosequencing approach employed in our study. In fact, conducting additional analyses on our data based on richness estimations comparable to that of community-profiling methodologies (that is, employing relative abundance cut-off levels of 1% and 2%), yielded results more similar to those reported by the above-mentioned studies (data not shown). Yet, because Woodcock *et al.* (2006) have shown that greater sampling effort should yield more positive SARs and the fact that

we did not observe this indicates other mechanisms than those causing positive SARs to be of importance for BR in these lakes. Our results suggest that productivity is of greater importance for shaping BR in these oligotrophic lakes rather than are lake or CA. However, another possibility is that the scales employed in our study were not optimal to detect positive SARs. Drakare *et al*. (2006) showed that both the fit and the slope of the SARs investigated were scale-dependent and concluded that mechanisms underlying richness at different scales strongly affect the shape of a SAR. In our study, LA and CA spanned 2–3 and 3–4 orders of magnitude, respectively, which is similar to or even higher than that in previous studies that observed SARs for aquatic bacteria (Green *et al*., 2004; Horner-Devine *et al*., 2004; Bell *et al*., 2005; Reche *et al*., 2005; van der Gast *et al*., 2006) and comparable to work done on larger organisms (for example, Arrhenius, 1921; Gleason, 1922; Rosenzweig, 1995). However, Bell *et al*. (2005), Reche *et al*. (2005) or van der Gast *et al*. (2006), for instance, studied areas and volumes on a much smaller scale compared with the lake and CAs in our study. As microbes can be expected to function on rather small scales, it can be possible that the negative SARs in our study are a result of lake and CAs being too big to enable detection thereof. In any case, we observed a negative SAR at either scale, albeit we cannot say whether or not lake or CA are appropriate in describing the size of a microbial habitat, which possibly needs to be defined and studied at much smaller spatial scales. However, as our study excluded many of the methodological limitations of previous studies reporting positive SARs for microorganisms, it re-opens the debate about whether SARs can indeed be found in the microbial world and, thus, are indeed universally valid, applicable to all organisms from all domains of life.

Contrary to our expectations, lakes higher up in the landscape (lakes of LCN one and two) had a significantly higher BR than lakes lower down (lakes of LCN three to six) (Table 3). Local environmental properties (nutrient concentrations) behaved in a similar way, showing differences between the two groups of lakes (Figure 3f, Table 3). The finding that landscape position and local environmental conditions co-varied suggests that the landscape dictates environmental properties, which then directly structure local BR. The partial Spearman's rank correlation analyses further support this finding (Table 2). The only exception to this general trend was Lake Bodsjön, which had a high BR despite its location further down in the lake chain.

The application of high-throughput sequencing technologies, such as 454 pyrosequencing, to surveys of microbial diversity has revealed an enormous number of microbial taxa (for example, Sogin *et al*., 2006). Yet, studies employing such technologies have almost exclusively been conducted assessing microbial communities in marine aquatic ecosystems (Sogin *et al*., 2006; Galand *et al*., 2009, 2010; Andersson *et al*., 2010; Kirchman *et al*., 2010). Thus, this study marks a first exploration of BR in natural freshwater systems, adopting a 454-pyrosequencing approach. As such, we found that a subsample of 200 millilitres of lake water can contain an average of 1009 unique bacterioplankton OTUs at a level of 97% sequence identity—a substantially higher figure compared with, for instance, the value of approximately 160 genomes from relatively nutrient-rich freshwater systems proposed by Ritz *et al*. (1997) in their classic experiments on DNA re-association kinetics. For comparison only, Galand *et al*. (2010), Andersson *et al*. (2010) and Kirchman *et al*. (2010) detected on average 1721, 1510, and 962 bacterioplankton OTUs in marine and brackish aquatic systems, respectively. It should, however, be noted that all three covered the V6 region on the 16S rRNA gene, whereas we targeted the V3 and V4 regions. It has been shown that the choice of primers and other PCR conditions, such as amplicon length, polymerase, annealing temperature, or cycle number affected the estimation of microbial species richness (Liu *et al*., 2008; Engelbrektson *et al*., 2010; Wu *et al*., 2010). Besides, none of the three studies mentioned above applied a method to account for pyrosequencing errors, such as AmpliconNoise (Quince *et al*., 2011), and, hence, counteracted an overestimation of microbial species richness owing to 454 pyrosequencing (Quince *et al*., 2009; Reeder and Knight, 2009). These differences render a direct comparison difficult but nonetheless suggest that BR of marine and nutrient-poor freshwater systems can be similarly high. Finally, despite providing an estimate of BR based on a sampling depth and accuracy, which, to our knowledge, are unsurpassed for freshwater bacterioplankton communities, rarefaction analyses showed that the curves (Figure 2) did not approach an asymptote, indicating that more sequences need to be retrieved to census our microbial communities at large. However, because our statistical analyses are based on Spearman's rank correlation analyses, the actual richness estimates are not of importance with regard to the conclusions drawn here but rather the ranking amongst the samples. Therefore and because the individual rarefaction curves are, in general, well separated from each other in a way that extrapolating the curves would not produce a different ranking, we believe that the results obtained in this study would still be the same even if greater sequencing depths had been applied.

In conclusion, we found that BR significantly increased with increasing nutrient availability, indicating a dependency between richness and productivity. However, BR did neither increase with LCN nor with LA or CA. Thus, there is no indication of a positive SAR. Finally, BR was significantly higher in lakes higher up in the landscape compared with lakes further down, most likely because of

higher nutrient concentrations in lakes higher up in the landscape. It, hence, seems as if the landscape or region leaves a mark on the local environmental characteristics of the lakes, which in turn exert great influence on local lake BR.

Acknowledgements

We gratefully acknowledge the assistance of Ines Kohler with regard to sample collection in Jämtland. Also, we are obliged to Jan Johansson for physico-chemical water and bacterial abundance analyses. We, furthermore, thank Ramiro Logares and Friederike Heinrich for initial help with 454 pyrosequencing sample preparations. We are thankful to Jérôme Comte and Lars Tranvik for thoughtful and constructive comments on an earlier version of the manuscript. Finally, we greatly appreciate the financial support from the Helge Ax:son Johnsons foundation granted to JB Logue, from the Olsson-Borghs foundation granted individually to S Langenheder, ES Lindström and JB Logue, from Formas granted individually to A Andersson, S Bertilsson, and S Langenheder, from the Carl Tryggers foundation granted to S Drakare and from the Swedish Research Council granted individually to S Bertilsson and ES Lindström.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Andersson AF, Riemann L, Bertilsson S. (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4**: 171–181.

Arrhenius O. (1921). Species and area. *J Ecol* **9**: 95–99.

Bell T, Ager D, Song JI, Newman JA, Thompson IP, Lilley AK *et al.* (2005). Larger islands house more bacterial taxa. *Science* **308**: 1884.

Blackwood CB, Hudleston D, Zak DR, Buyer JS. (2007). Interpreting ecological diversity indices applied to terminal restriction fragment length polymorphism data: Insights from simulated microbial communities. *Appl Environ Microbiol* **73**: 5276–5283.

Bohannan BJM, Lenski RE. (1997). Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. *Ecology* **78**: 2303–2315.

Cardinale BJ, Hillebrand H, Harpole WS, Gross K, Ptacnik R. (2009). Separating the influence of resource 'availability' from resource 'imbalance' on productivity-diversity relationships. *Ecol Lett* **12**: 475–487.

Chase JM, Leibold MA. (2002). Spatial scale dictates the productivity-biodiversity relationship. *Nature* **416**: 427–430.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al.* (2009). The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

Crump BC, Adams HE, Hobbie JE, Kling GW. (2007). Biogeography of bacterioplankton in lakes and streams of an arctic tundra catchment. *Ecology* **88**: 1365–1378.

Curtis TP, Head IM, Lunn M, Woodcock S, Schloss PD, Sloan WT. (2006). What is the extent of prokaryotic diversity? *Philos Trans R Soc B-Biol Sci* **361**: 2023–2037.

Curtis TP, Sloan WT. (2005). Exploring microbial diversity - a vast below. *Science* **309**: 1331–1333.

Drakare S, Lennon JJ, Hillebrand H. (2006). The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol Lett* **9**: 215–227.

Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H *et al.* (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* **4**: 642–647.

Fenchel T. (2003). Biogeography for bacteria. *Science* **301**: 925–926.

Fisher MM, Klug JL, Lauster G, Newton M, Triplett EW. (2000). Effects of resources and trophic interactions on freshwater bacterioplankton diversity. *Microb Ecol* **40**: 125–138.

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

Galand PE, Potvin M, Casamayor EO, Lovejoy C. (2010). Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *ISME J* **4**: 564–576.

Gleason HA. (1922). On the relation between species and area. *Ecology* **3**: 158–162.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M *et al.* (2004). Spatial scaling of microbial eukaryote diversity. *Nature* **432**: 747–750.

Hobbie JE, Daley RJ, Jasper S. (1977). Use of nuclepore filters for counting bacteria by fluorescence microscopy. *Appl Environ Microbiol* **33**: 1225–1228.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM. (2004). A taxa-area relationship for bacteria. *Nature* **432**: 750–753.

Horner-Devine MC, Leibold MA, Smith VH, Bohannan BJM. (2003). Bacterial diversity patterns along a gradient of primary productivity. *Ecol Lett* **6**: 613–622.

Kassen R, Buckling A, Bell G, Rainey PB. (2000). Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* **406**: 508–512.

Kirchman DL, Cottrell MT, Lovejoy C. (2010). The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ Microbiol* **12**: 1132–1143.

Kratz TK, Benson BJ, Blood ER, Cunningham GL, Dahlgren RA. (1991). The influence of landscape position on temporal variability in 4 North-American ecosystems. *Am Nat* **138**: 355–378.

Kratz TK, Webster KE, Bowser CJ, Magnuson JJ, Benson BJ. (1997). The influence of landscape position on lakes in northern Wisconsin. *Freshwater Biol* **37**: 209–217.

Lawton JH. (1999). Are there general laws in ecology? *Oikos* **84**: 177–192.

Legendre P, Legendre L. (1998). *Numerical ecology*. Elsevier: Amsterdam, The Netherlands.

Liu ZZ, DeSantis TZ, Andersen GL, Knight R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* **36**: e120.

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

Logue JB, Lindström ES. (2010). Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. *ISME J* **4**: 729–738.

Loisel P, Harmand J, Zemb O, Latrille E, Lobry C, Delgenes JP *et al.* (2006). Denaturing gradient electrophoresis (DGE) and single-strand conformation polymorphism (SSCP) molecular fingerprintings revisited by simulation and used as a tool to measure microbial diversity. *Environ Microbiol* **8**: 720–731.

MacArthur RH, Wilson EO. (1967). *The theory of island biogeography*. Princeton University Press: Princeton, New Jersey.

Magnuson JJ, Kratz TK. (2000). Lakes in the landscape: approaches to regional limnology. In: Williams WD (eds). *International Association of Theoretical and Applied Limnology*, Vol 27, Pt 1, Proceedings E Schweizerbart'sche Verlagsbuchhandlung: Stuttgart, pp 74–87.

Mardis ER. (2008). Next-generation DNA sequencing methods. *Ann Rev Genomics Hum Genet* **9**: 387–402.

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

Mittelbach GG, Steiner CF, Scheiner SM, Gross KL, Reynolds HL, Waide RB *et al.* (2001). What is the observed relationship between species richness and productivity? *Ecology* **82**: 2381–2396.

Nelson CE, Sadro S, Melack JM. (2009). Contrasting the influences of stream inputs and landscape position on bacterioplankton community structure and dissolved organic matter composition in high-elevation lake chains. *Limnol Oceanogr* **54**: 1292–1305.

Pärtel M, Laanisto L, Zobel M. (2007). Contrasting plant productivity-diversity relationships across latitude: the role of evolutionary history. *Ecology* **88**: 1091–1097.

Pedros-Alio C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.

Pedros-Alio C. (2007). Dipping into the rare biosphere. *Science* **315**: 192–193.

Quince C, Curtis TP, Sloan WT. (2008). The rational exploration of microbial diversity. *ISME J* **2**: 997–1006.

Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM *et al.* (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**: 639–641.

Quince C, Lanzén A, Davenport RJ, Turnbaugh PJ. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**: 38 : 1–18.

R (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Reche I, Pulido-Villena E, Morales-Baquero R, Casamayor EO. (2005). Does ecosystem size determine aquatic bacterial richness? *Ecology* **86**: 1715–1722.

Reeder J, Knight R. (2009). The 'rare biosphere': a reality check. *Nat Methods* **6**: 636–637.

Riera JL, Magnuson JJ, Kratz TK, Webster KE. (2000). A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, USA. *Freshwater Biol* **43**: 301–318.

Ritz K, Griffiths BS, Torsvik VL, Hendriksen NB. (1997). Analysis of soil and bacterioplankton community DNA by melting profiles and reassociation kinetics. *FEMS Microbiol Lett* **149**: 151–156.

Rosenzweig ML. (1995). *Species diversity in space and time*. Cambridge University Press: Cambridge.

Smith VH. (2007). Microbial diversity-productivity relationships in aquatic ecosystems. *Fems Microbiol Ecol* **62**: 181–186.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored ''rare biosphere''. *Proc Nat Acad Sci USA* **103**: 12115–12120.

Soranno PA, Webster KE, Riera JL, Kratz TK, Baron JS, Bukaveckas PA *et al.* (1999). Spatial variation among lakes within landscapes: ecological organization along lake chains. *Ecosystems* **2**: 395–410.

van der Gast CJ, Jefferson B, Reid E, Robinson T, Bailey MJ, Judd SJ *et al.* (2006). Bacterial diversity is determined by volume in membrane bioreactors. *Environ Microbiol* **8**: 1048–1055.

Waide RB, Willig MR, Steiner CF, Mittelbach G, Gough L, Dodson SI *et al.* (1999). The relationship between productivity and species richness. *Ann Rev Ecol Syst* **30**: 257–300.

Webster KE, Kratz TK, Bowser CJ, Magnuson JJ, Rose WJ. (1996). The influence of landscape position on lake chemical responses to drought in northern Wisconsin. *Limnol Oceanograph* **41**: 977–984.

Woodcock S, Curtis TP, Head IM, Lunn M, Sloan WT. (2006). Taxa-area relationships for microbes: the unsampled and the unseen. *Ecol Lett* **9**: 805–812.

Wu JY, Jiang XT, Jiang YX, Lu SY, Zou F, Zhou HW. (2010). Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol* **10**: 255.

Yannarell AC, Triplett EW. (2005). Geographic and environmental sources of variation in lake bacterial community composition. *App Environ Microbiol* **71**: 227–239.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)