



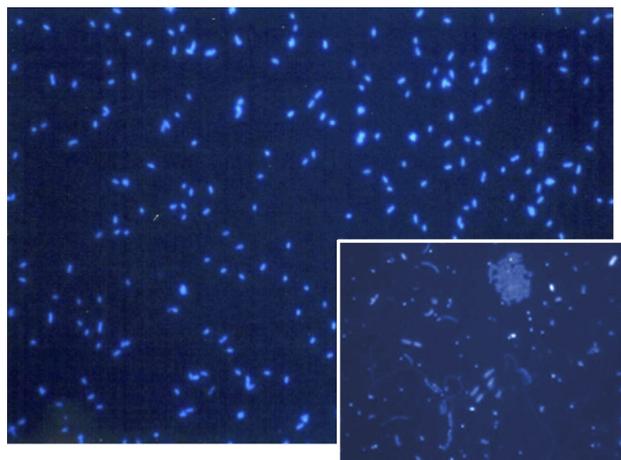
FEATURE ARTICLE

# Global phylogenetic community structure and $\beta$ -diversity patterns in surface bacterioplankton metacommunities

Albert Barberán\*, Emilio O. Casamayor

Department of Continental Ecology-Limnology, Centre d'Estudis Avançats de Blanes — CSIC,  
Accés Cala St. Francesc 14, 17300 Blanes, Spain

**ABSTRACT:** We aimed to identify phylogenetic community patterns in abundant planktonic bacteria (*Alpha*-, *Beta*-, and *Gammaproteobacteria*, *Actinobacteria*, *Cyanobacteria*, and *Bacteroidetes*) from a worldwide range of surface waters (lakes and seas—34 sites and ca. 4500 16S rRNA gene sequences). At each site we assessed the number of observed bacterial groups and the genetic relatedness of the most abundant groups through a community phylogenetic meta-analysis approach in order to (1) explore which potential ecological processes were consistent with the observed phylogenetic patterns in community assembly and (2) disentangle the effects of space and environment in  $\beta$ -diversity patterns for the different bacterial groups. Inland waters had significantly more bacterial groups and were more diverse than marine waters. Marine habitats showed a higher percentage of clustered sites than lakes, and bacterial communities were more closely related than expected by chance. Phylogenetic  $\beta$ -diversity analyses revealed different patterns to both salt composition (marine vs. inland salt lakes) and salt concentration for the dominant bacteria. We observed that while  $\beta$ -diversity patterns for *Bacteroidetes* were mostly shaped by salinity concentration, patterns in *Alphaproteobacteria* and *Gammaproteobacteria* were controlled by salt composition. *Actinobacteria*, *Betaproteobacteria* and *Sphingobacteria* were largely absent from marine habitats and from saline continental sites. In general and despite the lack of contextual metadata, environmental similarity was more relevant than spatial distribution for bacterial  $\beta$ -diversity patterns. However, we detected a geographic signal for some inland waters' groups (i.e. *Actinobacteria*, *Beta*-, and *Gammaproteobacteria*). Overall, the analyses indicated differences among phylogenetic groups and reflected patterns upon which further exploration of community assembly theory could be based.



Marine (background) and freshwater (inset) bacterioplankton look similar under the microscope but present very different dominant bacteria. The ecological drivers guiding assemblage composition are, however, intriguing and challenging for microbial ecologists today.

Photos: L. Alonso

**KEY WORDS:** Bacterioplankton · Community ecology · Metacommunity · Biogeography · Global distribution · Community assembly · Phylogeny · 16S rRNA gene

Resale or republication not permitted without  
written consent of the publisher

## INTRODUCTION

Biogeography explores how and why biological diversity changes along geographical scales. In the case of microorganisms, this topic had until recently received very little attention; the traditional statement 'everything is everywhere, but the environment selects' (Baas-Becking 1934) summarized what was expected for micro-

\*Email: abarberan@ceab.csic.es

organisms, i.e. on the one hand, high dispersal rates due to large population sizes and short generation times (Fenchel 2003) and, on the other hand, environmental determinism in agreement with classical niche-assembly theories. Nonetheless, endemism, dispersal limitation and stochasticity have also been recently shown in different microbial groups (e.g. Whitaker et al. 2003). In this context, the timely debate between neutral and niche views on species distribution offers many opportunities to test new hypotheses with microbial communities. The neutral theory of biodiversity (Hubbell 2001) assumes that the abundance of species is driven by random dispersal and stochastic extinction, and species are assumed to be ecologically identical. Conversely, the niche-assembly view looks at environmental adaptations to explain the abundance and distribution of species (Hutchinson 1961).

The emerging field of metacommunity ecology tries to combine the neutral and niche views, recognizing different dispersal and niche-based processes (see a recent review by Leibold et al. 2004). Metacommunity ecology focuses on sets of local communities linked by the potential dispersal of individuals. Two opposing forces potentially shape local community structure: different local environmental factors could lead to divergence of the communities, while high dispersal rates could homogenize the connected communities. Different types of metacommunities have been proposed depending on the relative importance of dispersal and niche-based processes (Cottenie 2005), and each type conceptualizes different complex dynamics of potential importance for the metacommunity. The neutral model (NM) assumes that random processes drive species abundances. Conversely, the species-sorting (SS) view emphasizes environmental adaptations to explain the abundance and distribution of species. Under the patch-dynamics model (PD), community composition is defined by a tradeoff between dispersal and competitive ability, with no habitat quality differences among patches. Finally, the mass-effect model (ME) mainly relies on the effect of immigration on local dynamics. In the case of aquatic ecosystems, these concepts have been applied to zooplankton (e.g. Cottenie & De Meester 2004) and recently to bacterioplankton community ecology (e.g. Lindström & Logue 2008).

For bacteria, many worldwide comparable 16S rRNA gene surveys are accessible from public databases, facilitating the integration of phylogeny and community ecology. Initial attempts have focused on the global distribution of prokaryotic communities in oceans (Pommier et al. 2007, Fuhrman et al. 2008), soils (Fierer & Jackson 2006), and a mixture of terrestrial and aquatic environments (Lozupone & Knight 2007, Auguet et al. 2010), and general ecological patterns have emerged from the study of 16S rRNA gene data

sets, e.g. taxa–area relationships (e.g. Horner-Devine et al. 2004, Reche et al. 2005). Statistical methods that compare phylogenetic tree topology of different communities and metrics that quantify the distribution of taxa in a single sample relative to a pool of taxa (Webb 2000) can be applied to microbes (Horner-Devine & Bohannan 2006, Newton et al. 2007).

Here we focused on whether or not phylogenetic community patterns exist for some of the most abundant planktonic bacterial groups (*Alpha*-, *Beta*- and *Gammaproteobacteria*, *Actinobacteria*, *Cyanobacteria* and *Bacteroidetes*) in a worldwide range of lakes and seas. Even though the bacterial groups selected have members with different ecophysiology, consistent ecological trends have been observed at such broad clustering levels (e.g. Glöckner et al. 1999, Bouvier & del Giorgio 2002, Demergasso et al. 2004). At each site, we assessed the community phylogenetic diversity with an index (Faith 1992) that takes into account the number of sequences (i.e. species richness) and the phylogenetic relationship among them ('branch length'). The different tree topologies (shape) were examined using 2 indices, i.e. the net relatedness index (NRI) and the nearest taxa index (NTI) (Webb 2000) to estimate whether or not the bacterial assemblages exhibited significant phylogenetic structure. Although different scenarios could lead to the same patterns observed in the data, it has been proposed that those communities primarily structured by competitive exclusion would be less closely related than expected by chance (overdispersed, i.e. evenly spread across the phylogenetic tree), whereas communities structured by habitat filtering would be more closely related (i.e. clustered). Finally, phylogenetic  $\beta$ -diversity patterns were explored (changes in species composition along environmental/spatial gradients and among communities), as well as the qualitative salinity effect by comparing thalassohaline (i.e. salt composition similar to seawater) and athalassohaline (salt composition different from seawater) water masses. Inland (i.e. isolated water bodies understood as 'pond as a patch') and marine sites (well-connected environments) offered a useful contrasting framework to explore the most plausible mechanisms responsible for the phylogenetic diversity, structure and  $\beta$ -diversity patterns observed for the different bacterial groups.

## MATERIALS AND METHODS

**Data set characteristics.** After a bibliographic search we found 34 different locations that carried out extensive clone libraries (>40 valid sequences of the bacterial 16S rRNA gene per site) of surface or epilimnetic waters, for a total of 7154 sequences. However, some data sets were already available in databases as non-

redundant 97 % identity-grouped sequences, a consensus value for delimiting bacterial species, but not without concerns (Rosselló-Mora & Amann 2001), and only 1 sequence for each 97 % cut-off group was reported (see Table S1 in the supplement available at [www.int-res.com/articles/suppl/a059p001\\_app.pdf](http://www.int-res.com/articles/suppl/a059p001_app.pdf)). Therefore, to minimize the bias introduced by the different sampling efforts for each clone library and to agree on a conservative phylogenetic criterion, we grouped the remaining data at 97 % identity using the percentage sequence identity (PSI) algorithm of the online program FastGroupII (<http://biome.sdsu.edu/fastgroup/>). Sequences shorter than 300 nucleotides were also discarded. Therefore, a non-redundant, consistent (all sequences obtained from the same cloning methodology), and balanced (equal number of marine sites and lakes, although with different sampling effort in each case that we corrected using randomized subsamples, see below) data set of 4495 sequences was retained for further analysis (see Table S1 for more detailed information).

**16S rRNA gene sequence data analysis.** The 16S rRNA gene pool was automatically aligned with the NAST alignment tool (<http://greengenes.lbl.gov/NAST>). Next, the aligned data set was imported into the ARB software package ([www.arb-home.de](http://www.arb-home.de)) and was added by parsimony to the optimized Greengenes tree provided by default (>130 000 sequences, May 2007 release). Each 16S rRNA gene sequence was assigned to a bacterial phylogenetic group (e.g. *Alpha*-, *Beta*-, *Gamma*-, and *Deltaproteobacteria*, *Actinobacteria*, *Cyanobacteria*, *Bacteroidetes*, *Firmicutes*, *Planctomycetes*, and *Verrucomicrobia*) in a hierarchy based on the Ribosomal Database Project (<http://rdp.cme.msu.edu>).

**Community phylogenetic analyses.** PD for each community was calculated as the sum of the branch lengths associated with the overall 16S rRNA sequences obtained from that sample. Thus, PD takes into account the number of sequences (i.e. species richness) and the phylogenetic relationship among them (Faith 1992). Richness and Shannon diversity index were calculated for each site at the taxonomic level selected. To correct for unequal sample sizes, we calculated the mean of the richness, the Shannon diversity, and PD of 1000 randomized subsamples for each community (Bryant et al. 2008). The subsample size was the number of sequences present in the smallest community.

Subsequent analyses were carried out selecting the most abundant groups (to make the phylogenetic inference reliable) present at all sites (to give consistency to the global patterns observed). Members of the *Bacteroidetes* were not separated in smaller phylogenetic units because *Sphingobacteria* and *Flavobacteria* were mostly absent from marine and inland waters, respectively. To assess driving processes in bacterio-

plankton community assemblages, we compared the different tree topologies using 2 indices, i.e. the NRI and the NTI (Webb 2000). These indices measured the degree of phylogenetic relatedness of the different taxa from a sample relative to the pool of taxa from all the samples. Relatedness information provides a different window into bacterial communities than does information concerning richness or taxonomic composition. NRI reveals wide patterns across the phylogeny, while NTI focuses on terminal taxa. High and positive values of these indices indicate clustering of taxa across the overall phylogeny, whereas low or negative values indicate overdispersion of taxa across the phylogeny. In other words, a positive value indicates that bacteria tended to co-occur with other bacteria that were more closely related than expected by chance (for more details see Webb 2000, Horner-Devine & Bohannan 2006). We calculated NRI and NTI using Phylocom 3.41 ([www.phylodiversity.net/phylocom](http://www.phylodiversity.net/phylocom)).

Distance matrices among environments were calculated using the UniFrac metric (<http://bmf.colorado.edu/unifrac>). The UniFrac distance is calculated as the percent of branch length leading to descendants from only 1 of the locations represented in the phylogenetic tree (Lozupone & Knight 2007). Hence, UniFrac is a  $\beta$ -diversity metric that quantifies community similarity based on phylogenetic relatedness.

**Statistical analyses.** We used standard and partial Mantel tests to determine the correlation between similarity in microbial communities (based on UniFrac matrices) and geographic (S) and environmental (E) matrices. Genetic databases usually lack environmental metadata associated with sequence information, and this limitation should be corrected in the future. Therefore, we tried to compile environmental data from the original publications and by requesting information from the authors. Unfortunately, it was not possible to obtain sufficient environmental variables simultaneously for all sites to more finely contextualize the information. Thus, we selected general variables traditionally used in limnology and oceanography. Lakes are usually classified by their geomorphology, as it encapsulates most of their variation and correlates with other fundamental variables. For marine sites, we used the physico-chemical parameters that define broad water masses. Data used are shown in Table S1. The spatial matrix (S) was calculated following detailed geographic coordinates and the Earth's curvature. The environmental matrix (E) was constructed with general geomorphological features for lakes (i.e. area, depth, elevation), whereas for oceanic waters we used physico-chemical data (i.e. temperature, salinity, dissolved oxygen) provided by the authors or estimated from the NOAA World Ocean Atlas ([www.nodc.noaa.gov/OC5/WOA05/pr\\_woa05.html](http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html)). We ran non-

metric multidimensional scaling (NMDS) to represent the ordering relationships obtained from the UniFrac distance matrices. To further explore environmental effects, each sample was classified into 2 different categories: one based on salt quality or composition (i.e. thalassohaline and athalassohaline) and the other on salt quantity or concentration (i.e. freshwater, saline, hypersaline). Thalassohaline habitats have salt composition similar to seawater, whereas athalassohaline environments are inland salt ponds with salt composition different from seawater. To assess the significance of the salinity categories, we performed analyses of similarities (ANOSIM) based on 1000 permutations. The ANOSIM R statistic is based on the difference of mean ranks between groups and within groups and ranges from 0 (no separation) to 1 (complete separation). Statistical analyses were carried out in the R environment ([www.r-project.org](http://www.r-project.org)) using the vegan package (<http://vegan.r-forge.r-project.org>).

## RESULTS

The number of 16S rRNA gene sequences analyzed for each site, as well as geographical coordinates, available environmental description, and reference in

the literature for each of the aquatic environments surveyed in the present study are presented in Table S1. Overall, 18 inland water bodies of different sizes, altitudes, and trophic status and 16 mainly coastal marine sites were analyzed using ca. 4500 gene sequences. The number of 16S rRNA gene sequences assigned to each bacterial group at each site is also shown in Table S2 in the supplement (available at [www.int-res.com/articles/suppl/a059p001\\_app.pdf](http://www.int-res.com/articles/suppl/a059p001_app.pdf)).

Clone libraries indicated that surface bacterioplankton was mainly dominated by 2 phyla, i.e. *Proteobacteria* (especially *Alpha*-, *Beta*-, and *Gammaproteobacteria*) and *Bacteroidetes* (especially *Flavobacteria* and *Sphingobacteria*). Inland waters harbored more bacterial groups ( $p = 0.017$ ,  $t$ -test), and these were more diverse ( $p < 0.001$ ,  $t$ -test) than in marine waters. Additionally, lakes presented greater (but not significant) PD than marine samples ( $p = 0.079$ ,  $t$ -test). Fig. 1 shows the respective fractions of 16S rRNA gene sequence types in freshwater and marine habitats, testing for significant (i.e. *Alpha*-, *Beta*-, and *Gammaproteobacteria*, *Actinobacteria*, *Firmicutes*, *Flavobacteria*, and *Sphingobacteria*;  $p < 0.05$ ,  $t$ -test) and not significant (e.g. *Cyanobacteria*, *Deltaproteobacteria*, *Planctomycetes*, *Verrucomicrobia*) differences between the pairs of frequencies. Glöckner et al. (1999) used

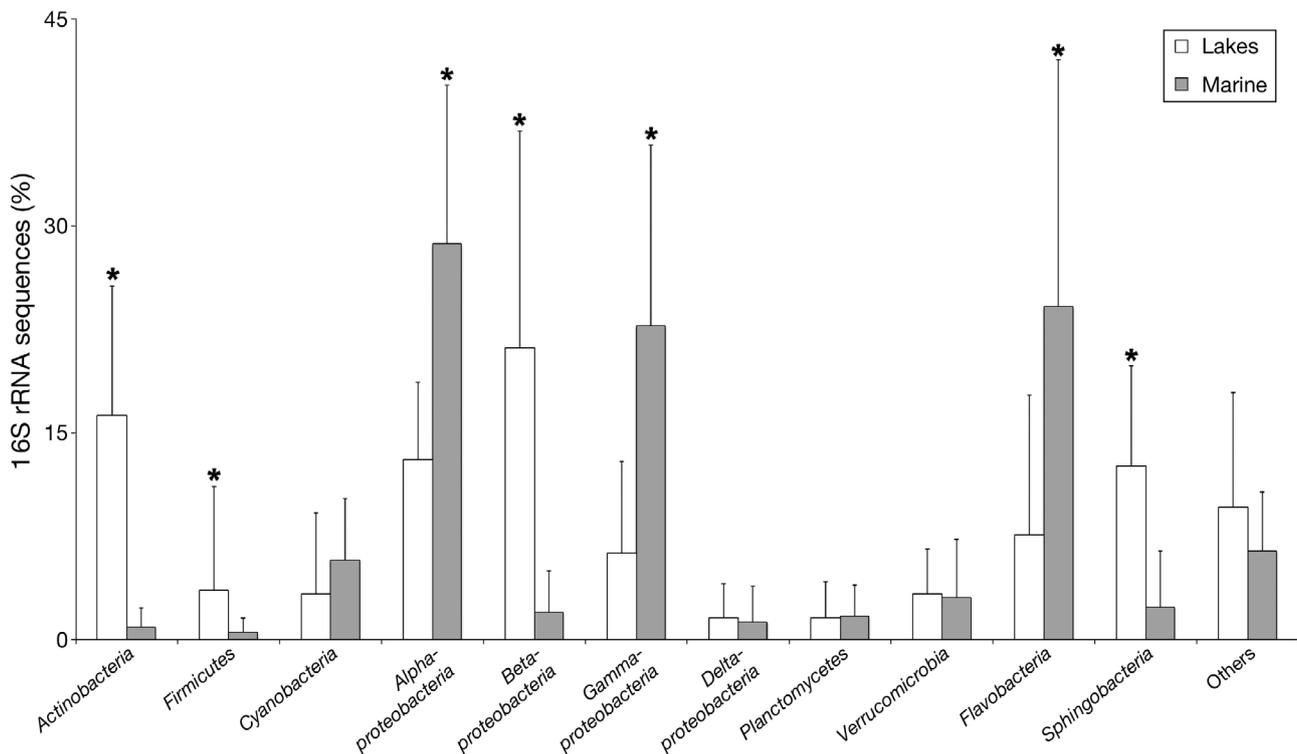


Fig. 1. Relative abundance of bacterial groups present in the different lakes and marine habitats examined using the 4495 16S rRNA gene sequence data set (grouped at 97% identity). See detailed information in Table S2 in the supplement. Error bars indicate 1 SD. Groups with significant differences between pairs of frequencies ( $p < 0.05$ ,  $t$ -test) are labeled with an asterisk

whole-cell fluorescence *in situ* hybridization (FISH) to indicate for the first time significant differences in the composition of marine and lake bacterioplankton with a limited number of samples and environments. Here, we used a larger data set obtained by cloning and sequencing that confirmed more detailed striking differences in global relative abundance and distribution for the most abundant groups in the 2 types of environments. In inland waters, *Betaproteobacteria*, *Actinobacteria*, *Sphingobacteria*, and *Firmicutes* were the most abundant, while in marine bacterioplankton *Alphaproteobacteria*, *Flavobacteria*, and *Gammaproteobacteria* dominated (Fig. 1).

For the analysis of processes driving community phylogenetic structure, both NRI and NTI indicated that bacterial assemblages had significantly higher phylogenetic clustering than expected by chance (Table 1; for further details, see Table S3 in the supplement, available at [www.int-res.com/articles/suppl/a059p001\\_app.pdf](http://www.int-res.com/articles/suppl/a059p001_app.pdf)), suggesting that communities were structured by habitat filtering. Although there was some degree of variation, such clustered structure was consistent for the different environments and different bacterial groups considered. Marine habitats showed a higher percentage of clustered sites (~55% for NRI and 77% for NTI) than lakes (~33% for NRI and 57% for NTI) ( $p = 0.031$  for NRI and  $p = 0.079$  for NTI, *t*-test).

Table 1. Summary of net relatedness index (NRI) and nearest taxa index (NTI) results. Numbers are the percentages of significant structured sites found for each group (for details see Table S3 in the supplement). Significance was assessed at values of 0.05 and 0.10, and was tested with a null model (999 random permutations across the entire phylogeny). A 2-tailed test evaluated the rank of observed values

Community	NRI		NTI	
	Overdispersed	Clustered	Overdispersed	Clustered
<b>All aquatic environments</b>				
All bacteria	3	53	0	91
<i>Bacteroidetes</i>	3	33	0	67
<i>Alphaproteobacteria</i>	3	31	3	59
<i>Gammaproteobacteria</i>	9	59	0	73
<b>Lakes</b>				
All bacteria	0	44	0	89
<i>Bacteroidetes</i>	0	6	0	53
<i>Alphaproteobacteria</i>	6	19	6	38
<i>Gammaproteobacteria</i>	0	50	0	50
<i>Actinobacteria</i>	7	33	0	47
<i>Betaproteobacteria</i>	0	43	7	64
<b>Marine habitats</b>				
All bacteria	6	63	0	94
<i>Bacteroidetes</i>	6	63	0	81
<i>Alphaproteobacteria</i>	0	44	0	81
<i>Gammaproteobacteria</i>	13	63	0	81
<i>Cyanobacteria</i>	0	42	0	50

At a global scale and pooling together the bacterial groups as a whole (Fig. 2A), the ordination analysis of the phylogenetic  $\beta$ -diversity matrix grouped the aquatic environments into 3 salinity concentration categories ( $R = 0.92$ ,  $p < 0.001$ , ANOSIM). Salt quality or salinity composition was also important ( $R = 0.87$ ,  $p < 0.001$ ), although hypersaline environments tended to cluster together with either thalassohaline and athalassohaline environments. We observed the same pattern for *Bacteroidetes* ( $R = 0.92$ ,  $p < 0.001$  for salt concentration, and  $R = 0.80$ ,  $p < 0.001$  for salt composition) (Fig. 2B). However,  $\beta$ -diversity patterns for *Alphaproteobacteria* and *Gammaproteobacteria* (Fig. 2C,D) appeared less sensitive to salt concentration ( $R = 0.81$ ,  $p < 0.001$  and  $R = 0.73$ ,  $p < 0.001$ , respectively) than to salinity composition ( $R = 0.91$ ,  $p < 0.001$  and  $R = 0.86$ ,  $p < 0.001$ , respectively).

Overall, when all aquatic environments were assembled and their phylogenetic  $\beta$ -diversity contextualized, the correlation analysis showed the significance of the environmental component (after fixing the spatial component by partial Mantel tests) for the different groups considered (Table 2). After splitting environments into lakes and marine habitats, environmental determinism was still the best explanatory dynamic for most of the groups, but some relevant exceptions appeared for inland bacterioplankton. Inland *Actinobacteria* and *Betaproteobacteria* showed, in addition to the environmental component, a significant contribution from the spatial component (after fixing the environmental component by partial Mantel tests) (Table 2). The geographical distribution merely explained the correlations found in inland *Gammaproteobacteria*, but not in the marine counterparts (Table 2). In Table 2 we also tentatively assigned the most appropriate metacommunity model following Cottenie (2005), bearing in mind that the assignment of different metacommunity models is dependent on the *p*-value used to define statistical significance and would lead to non-identical beta-diversity patterns (in Table 2, see bold and italic values for significant [ $p < 0.05$ ] positive correlations and bold-faced print for the more relaxed  $p < 0.1$  positive correlations that we used for discerning patterns). Both all aquatic environments and marine habitats agreed with the SS (species-sorting) model in all cases. Freshwater *Actinobacteria* and *Betaproteobacteria* showed, in addition, a

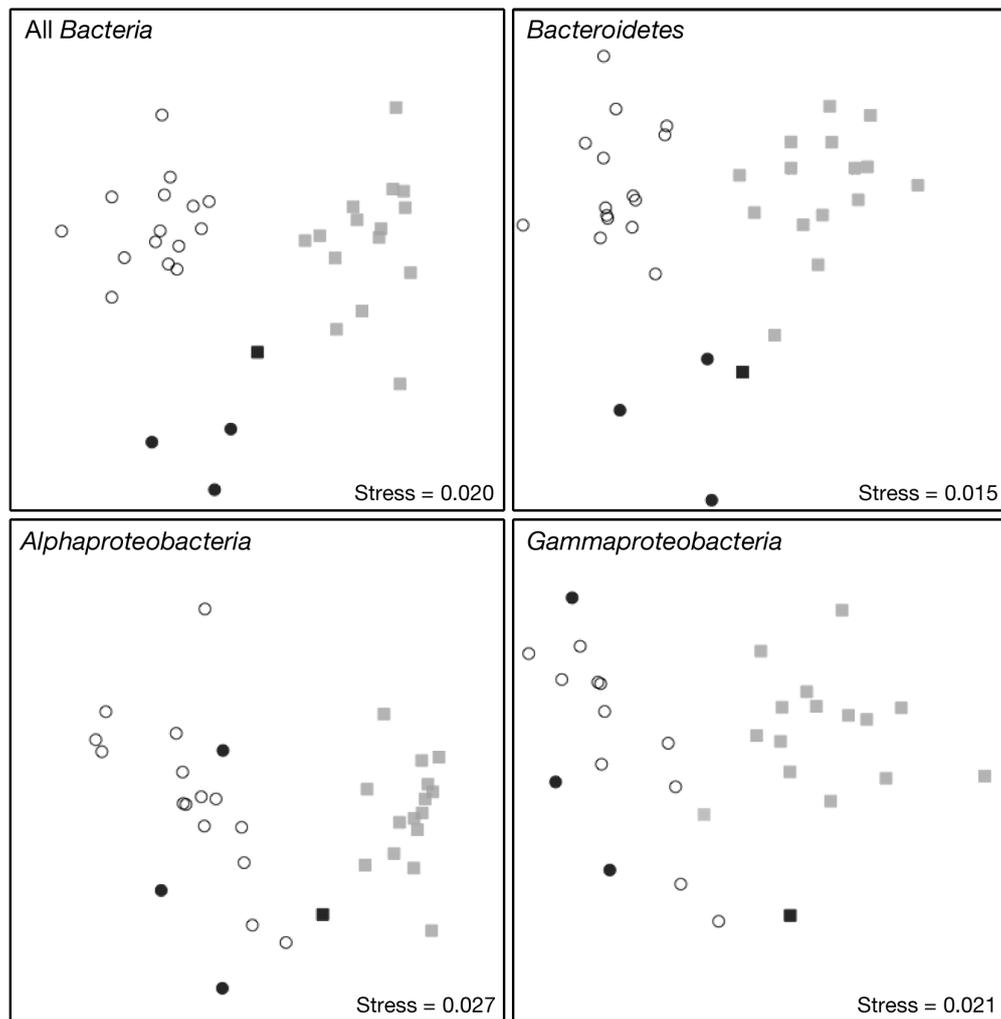


Fig. 2. Non-metric multidimensional scaling (NMDS) plots based on the UniFrac matrices for all *Bacteria*, *Bacteroidetes*, *Alphaproteobacteria*, and *Gammaproteobacteria* (i.e. those groups present at all sites with enough available sequences). Circles: inland salt composition (athalassohaline); squares: marine salt composition (thalassohaline); open symbols: freshwater; grey: saline water; black: hypersaline water. Normalized stress values are shown for each plot. Stress is a measure of the mismatch between original distance measures (the UniFrac distance) and distance in ordination space

signal consistent with the ME metacommunity model. Inland *Gammaproteobacteria*, in turn, formed a spatial pattern equally explained by neutral or patch dynamics (see more details in the 'Discussion' section).

## DISCUSSION

The correlation approach used in the present study showed community assembly based on environmental selection as a relevant potential mechanism for bacterioplankton biogeographical structure at a global scale. Recently, several studies have emphasized the importance of environment in shaping aquatic bacterial communities at a local scale (Langenheder & Ragnarsson 2007, Van Der Gucht et al. 2007, Sommaruga

& Casamayor 2009). Hence, environmental adaptations explain the abundance and distribution of bacterial assemblages at multiple scales of study.

Community assembly rules are intriguing for microbial ecologists, due to the lack of knowledge about the dispersal potential and effective niche space for bacteria. Some recent studies suggest that dispersal ability is probably not high for all bacteria (Papke & Ward 2004, Hervàs et al. 2009). Here, we consistently observed distinct bacterial composition in the plankton of marine and inland habitats even using broad taxonomic classification. These results match previous findings by whole-cell FISH (Glöckner et al. 1999), where a limited number of samples (4 lakes, 4 marine sites) were surveyed using group-specific probes (*Alpha*-, *Beta*- and *Gammaproteobacteria*, *Bactero*-

Table 2. Summary of the results obtained with the UniFrac matrices against the environmental ([E]) and the spatial ([S]) components, and tentative assignment to the most appropriate metacommunity model following Cottenie (2005). Mantel correlations ( $r_M$ ) and p-values obtained after 999 permutations. Bold and italic values: significant ( $p < 0.05$ ) positive correlations; bold; partially significant ( $p < 0.1$ ) positive correlations; not highlighted values: non-significant or negative correlations. [E] and [S]: standard Mantel tests; [E|S], [S|E]: partial Mantel tests with environment independent of space and space independent of environment, respectively; SS: species-sorting; NM: neutral model; PD: patch dynamics; ME: mass-effects models

Community	— [E] —		— [S] —		— [E S] —		— [S E] —		Metacommunity type
	$r_M$	p	$r_M$	p	$r_M$	p	$r_M$	p	
<b>All aquatic environments</b>									
All bacteria	<b>0.521</b>	<b>0.001</b>	0.040	0.241	<b>0.520</b>	<b>0.001</b>	-0.011	0.469	SS
<i>Bacteroidetes</i>	<b>0.475</b>	<b>0.001</b>	0.080	0.119	<b>0.471</b>	<b>0.001</b>	0.039	0.275	SS
<i>Alphaproteobacteria</i>	<b>0.493</b>	<b>0.001</b>	-0.047	0.206	<b>0.501</b>	<b>0.001</b>	-0.108	0.022	SS
<i>Gammaproteobacteria</i>	<b>0.563</b>	<b>0.001</b>	<b>0.108</b>	<b>0.063</b>	<b>0.558</b>	<b>0.001</b>	0.062	0.187	SS
<b>Lakes</b>									
All bacteria	<b>0.237</b>	<b>0.004</b>	0.157	0.153	<b>0.203</b>	<b>0.006</b>	0.095	0.249	SS
<i>Bacteroidetes</i>	<b>0.192</b>	<b>0.016</b>	0.051	0.310	<b>0.186</b>	<b>0.014</b>	-0.005	0.546	SS
<i>Alphaproteobacteria</i>	<b>0.110</b>	<b>0.090</b>	-0.087	0.270	<b>0.141</b>	<b>0.039</b>	-0.125	0.171	SS
<i>Gammaproteobacteria</i>	<b>0.228</b>	<b>0.018</b>	<b>0.576</b>	<b>0.001</b>	0.068	0.253	<b>0.546</b>	<b>0.002</b>	NM/PD
<i>Actinobacteria</i>	<b>0.168</b>	<b>0.047</b>	<b>0.207</b>	<b>0.071</b>	<b>0.155</b>	<b>0.062</b>	<b>0.197</b>	<b>0.083</b>	SS + ME
<i>Betaproteobacteria</i>	<b>0.253</b>	<b>0.007</b>	<b>0.326</b>	<b>0.026</b>	<b>0.234</b>	<b>0.008</b>	<b>0.312</b>	<b>0.034</b>	SS + ME
<b>Marine habitats</b>									
All bacteria	<b>0.388</b>	<b>0.012</b>	-0.236	0.033	<b>0.364</b>	<b>0.067</b>	-0.237	0.900	SS
<i>Bacteroidetes</i>	<b>0.500</b>	<b>0.009</b>	-0.109	0.228	<b>0.486</b>	<b>0.041</b>	-0.125	0.721	SS
<i>Alphaproteobacteria</i>	<b>0.430</b>	<b>0.010</b>	-0.239	0.030	<b>0.411</b>	<b>0.057</b>	-0.187	0.857	SS
<i>Gammaproteobacteria</i>	<b>0.392</b>	<b>0.007</b>	-0.192	0.073	<b>0.366</b>	<b>0.056</b>	-0.319	0.982	SS
<i>Cyanobacteria</i>	<b>0.374</b>	<b>0.089</b>	-0.076	0.294	0.358	0.111	-0.081	0.628	Not found

*idetes*, and *Planctomycetes*). Essentially, the same results were obtained here with a larger data set and with a different approach (DNA extraction, PCR amplification, and gene cloning and sequencing). Thus, although pure quantitative composition cannot be inferred from clone libraries, these molecular techniques provide semi-quantitative information that can be used to assess such general trends as those observed in our meta-analytical study. In particular, we consistently observed that *Actinobacteria* and *Sphingobacteria*, in addition to the previously known *Betaproteobacteria*, were mostly absent from marine habitats and hypersaline lakes, while they were abundant in freshwater lakes, indicating that salt concentration may drive their potential distribution. Although a large number of environmental parameters covary with salinity, consistent differences have been shown in microbial assemblages along the quantitative salinity gradient from seawater to NaCl saturation (Casamayor et al. 2000) that were higher than along qualitative gradients (i.e. thalassohaline vs. athalassohaline sites; Demergasso et al. 2004). In addition to the previously reported fact that salinity was the major filtering factor for bacterial and archaeal community assemblage (Lozupone & Knight 2007, Auguet et al. 2010), our analysis also detected different responses to salt composition among the dominant bacterioplankton groups. We observed that, while  $\beta$ -diversity patterns for *Bacteroidetes* were mostly shaped by salinity con-

centration, patterns among *Alphaproteobacteria* and *Gammaproteobacteria* were controlled by salt composition.

Besides the difference in community composition between inland and marine environments, contrasting geographical  $\beta$ -diversity patterns were observed. Such disparity could be related to the connectivity among habitats in the marine continuum versus lakes as islands within a sea of land (Reche et al. 2005 and references therein). Accordingly, marine habitats showed  $\beta$ -diversity patterns not explained by the spatial component. This is in agreement with the concept of the ocean surface as a contiguous and relatively homogeneous environment for bacteria that facilitates dispersion, and may lead to the lack of geographical patterns in marine bacteria (Pommier et al. 2007). Nonetheless, we found that local environmental factors (not related to spatial distance) were still very powerful constraints in structuring marine bacterioplankton assemblages. In contrast, inland water bodies had a patchy distribution more likely to exhibit dispersal constraints in addition to environmental factors. Other studies have found a significant spatial component operating at the regional scale for lakes and ponds (e.g. Reche et al. 2005, 2007, Langenheder & Ragnarsson 2007). The disconnected and relatively heterogeneous nature of lakes (Papke & Ward 2004) may have promoted the higher bacterial richness and diversity observed in our study. Recently, freshwater environments have also

been unveiled as one of the largest reservoirs of archaeal diversity (Auguet et al. 2010). Hence, inland water bodies appear as promising environments for biological novelty in terms of phylogenetic diversity and metabolic potential (e.g. Demergasso et al. 2008).

The analysis of the phylogenetic  $\beta$ -diversity for all aquatic environments together and marine habitats agreed with the SS model for all the groups examined (Table 2), as interpreted by Cottenie (2005) emphasizing environmental adaptations to explain the distribution of species. For lakes, in addition to an environmental signal, a relationship between community similarity and distance was detected for inland *Actinobacteria* and *Betaproteobacteria*, consistent with the ME metacommunity model, which mainly focuses on the effect of high immigration on local dynamics, but also may reflect dispersal restrictions (or the influence of unmeasured environmental parameters; Table 2). In fact, SS + ME dynamics may indicate a metacommunity with species sorting and either high dispersal (which are true mass effects) or low dispersal between at least some local communities (species sorting and dispersal limitation), although it is difficult to find out which of those is the case (Ng et al. 2009). MEs emphasize the role of spatial dynamics on community structure in a system where species can be rescued from local competitive exclusion in communities where they are poor competitors by immigration. This will result in spatial patterns occurring independently of environmental patterns (Cottenie 2005). For this study and the spatial scales that are considered, species sorting coupled with dispersal limitation is the more likely scenario. True mass effects would mean that dispersal rates between the different sites are higher than the internal growth rates of local populations, and this is very unlikely given that most locations are not part of the same catchment area. However, both *Betaproteobacteria* and *Actinobacteria* tend to co-occur and are very abundant in nearly all lakes, suggesting continuous feeding with allochthonous bacteria from the catchment area, but atmospheric dust depositions should also be considered in remote freshwater ecosystems where these 2 groups have been detected with the potential to grow (Hervàs et al. 2009). *Gammaproteobacteria*, in turn, formed a spatial pattern equally explained by neutral or patch dynamics. Many members of this phylogenetic group are typical copiotrophs, adapted to relatively high nutrient concentrations such as those in culturing media, phytoplankton blooms, or coastal areas (e.g. Agogué et al. 2005, Fuchs et al. 2007) that show sporadic peaks of abundance along river–estuary transects (e.g. Bouvier & del Giorgio 2002) or in lakes (e.g. Casamayor et al. 2002). Thus, if they have a feast and famine strategy and frequently peak in abundance in response to nutri-

ent pulses, this would be rather an indication of species sorting than neutral or patch dynamics. However, some of the occurrences of suitable conditions may have an unpredictable and sporadic nature, and the pattern observed might be related to such local sporadic events that produce suitable conditions to proliferate or, alternatively, to integrate other non-considered variables in the analysis, such as the temporal variation hidden in the spatial signal (see below).

A major focus in the field of biogeography is the relative influence of current environmental versus historical factors on present day distribution patterns. In the case of microorganisms, large population size and high dispersal may blur the potential effect of historical legacy on present community patterns (Fuhrman et al. 2008). For that reason, community phylogenetic structure is more likely to be generated by ecological rather than historical processes (i.e. diversification within a region causing many taxa to be, on average, more related to each other rather than taxa outside the region) (Webb et al. 2002). Presuming phylogenetic niche conservatism (i.e. related taxa are more ecologically similar), phylogenetic clustering indicates environmental selection, while overdispersion can be considered the result of competition between closely related species or facilitation between distantly related taxa. However, phylogenetic resolution and sampling scale may influence the observed patterns (for a discussion see Horner-Devine & Bohannan 2006). In the present study, we found that the closest bacterioplankton relatives tend to be in the same habitat (habitat filtering, owing to the sharing of adaptations needed for a particular habitat) in agreement with recently published reports for bacterial communities (Horner-Devine & Bohannan 2006, Newton et al. 2007, Bryant et al. 2008). The broad scales of environmental gradients, geography, and taxonomy explored in our study may have concealed overdispersed patterns (Horner-Devine & Bohannan 2006 and references therein). Despite this limitation, analytical exploration on such broad scales is still scarce for microorganisms (but see Fierer & Jackson 2006, Pommier et al. 2007) and relevant information is still to be unveiled. Some other methodological limitations inherent to the shortcomings in databases are also obvious in the work presented here, such as the rationale for the environmental variables chosen or the strong sampling bias toward northern temperate regions. Although we tried to homogenize the data sets by selecting clone library studies with high numbers of 16S rRNA gene sequences per site (>40) and by selecting superficial waters, other factors could have escaped our analysis, such as the temporal component that may lead to some predictable patterns (Fuhrman et al. 2006) or the influence of biotic factors on bacterial community composi-

tion (Kent et al. 2007). Other problems are intrinsic to the molecular ecology approach; for example, (1) PCR cloning focuses mostly on abundant taxa and may cause bias against rare taxa, resulting in semi-quantitative information, or (2) dereplicating 16S rRNA libraries at the species level (97% identity) implicitly loses the microdiversity inherent in most bacterial lineages. A minor sequence change in such a highly conserved gene may conceal relevant differences in whole genome content and, thus, in ecological capability, and there are some cases of different bacterial species with identical 16S rRNA genes (see a review in Rosselló-Mora & Amann 2001). In the worst case, methodological limitations will not affect the general patterns observed here because we focused on conservative criteria, and we dealt with the most abundant bacterial groups detected. The abundant members cover the part of the microbial biodiversity spectrum that includes the active units of the community (so-called 'core species') that drive most ecosystem functions, whereas those below numerical detection thresholds of molecular techniques form a seed bank of rare species that grow slowly or not at all. The rare species on the whole cannot be completely unveiled with traditional PCR methods, although biogeographical approaches would be greatly enhanced by increasing the genetic information deposited in databases and by improving the coverage of sequencing techniques (Galand et al. 2009, 2010). In addition, there is ample room for community phylogenetic approaches involving more specific bacterial clades and narrower scales (e.g. Newton et al. 2007).

Only very recently has it been observed that microorganisms may have similar ecological patterns to their macroorganism counterparts in terms of the taxa–area relationship (Horner-Devine et al. 2004, Reche et al. 2005, 2007), the latitudinal and elevational gradients of richness (Pommier et al. 2007, Bryant et al. 2008, Fuhrman et al. 2008), or the community assembly rules (Horner-Devine et al. 2007). One of the most puzzling aspects of ecology is the wide range of spatial, temporal, and taxonomical scales at which patterns may arise. As a result of their interrelationship, some patterns emerge as processes on a lower scale level (Levin 1992). Moreover, it has been shown that microbial groups have ecological idiosyncrasies and habitat preferences that are stable over time (Von Mering et al. 2007). Therefore, microbial biogeography, although traditionally disregarded, is a particularly motivating field of study, due to the wide range of scales involved (Dolan 2006).

Our empirical meta-analysis, which used a hierarchical phylogenetic approach, reinforces the niche explanation for the global distribution of surface bacterioplankton assemblages. The results indicate differences

among phylogenetic groups that deserve further investigations, and the patterns we observed might help in the future as a basis upon which to design studies and experiments to test various aspects related to community assembly theory. With current molecular techniques, we cannot definitely state whether everything is everywhere or not (no dispersal and colonization barriers for any bacterial cell), but certainly there is a strong aspect of environmental determinism shaping aquatic microbial assemblages on a global scale, and environmental forces are acting singularly on different bacterial groups.

*Acknowledgements.* We are thankful to authors who provided valuable data for the analysis, to Micah Hamady and Campbell O. Webb for technical assistance with UniFrac and Phylocom, respectively, to Antonio Fernández-Guerra for computing assistance, and to anonymous reviewers for helpful comments and improvements of earlier versions of this manuscript. This research was supported by Grant AERBAC 079/2007 to E.O.C. from the Spanish Ministerio de Medio Ambiente (MARM), and CONSOLIDER-INGENIO 2010 Grant GRACCIE CSD2007-00004 from the Spanish Ministerio de Ciencia e Innovación (MICINN). A.B. was supported by a FPU-MICINN predoctoral scholarship.

#### LITERATURE CITED

- Agogué H, Casamayor EO, Bourrain M, Obernosterer I, Joux F, Herndl GJ, Lebaron P (2005) A survey on bacteria inhabiting the sea surface microlayer of coastal ecosystems. *FEMS Microbiol Ecol* 54:269–280
- Auguet JC, Barberán A, Casamayor EO (2010) Global ecological patterns in uncultured Archaea. *ISME J* 4:182–190
- Baas-Becking LGM (1934) *Geobiologie, of inleiding tot de milieukunde*, Serie 18/19. Van Stockum's Gravenhange, The Hague
- Bouvier TC, del Giorgio PA (2002) Compositional changes in free-living bacterial communities along a salinity gradient in two temperate studies. *Limnol Oceanogr* 47:453–470
- Bryant JA, Lamanna C, Morlon HLN, Kerkhoff AJ, Enquist BJ, Green JL (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proc Natl Acad Sci USA* 105:11505–11511
- Casamayor EO, Calderón-Paz JI, Pedrós-Alió C (2000) 5S rRNA fingerprints of marine bacteria, halophilic archaea and natural prokaryotic assemblages along a salinity gradient. *FEMS Microbiol Ecol* 34:113–119
- Casamayor EO, Pedrós-Alió C, Muyzer G, Amann R (2002) Microheterogeneity in 16S ribosomal DNA-defined bacterial populations from a stratified planktonic environment is related to temporal changes and to ecological adaptations. *Appl Environ Microbiol* 68:1706–1714
- Cottenie K (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecol Lett* 8: 1175–1182
- Cottenie K, De Meester L (2004) Metacommunity structure: synergy of biotic interactions as selective agents and dispersal as fuel. *Ecology* 85:114–119
- Demergasso C, Casamayor EO, Chong G, Galleguillos P, Escudero L, Pedrós-Alió C (2004) Distribution of prokaryotic genetic diversity in athalassohaline lakes of the Atacama Desert, northern Chile. *FEMS Microbiol Ecol* 48:57–69

- Demergasso C, Escudero L, Casamayor EO, Chong G, Balagué V, Pedrós-Alió C (2008) Novelty and spatio-temporal heterogeneity in the bacterial diversity of hypersaline Lake Tebenquiche (Salar de Atacama). *Extremophiles* 12: 491–504
- Dolan JR (2006) Microbial biogeography? *J Biogeogr* 33: 199–200
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Fenchel T (2003) Biogeography for bacteria. *Science* 301: 925–926
- Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103: 626–631
- Fuchs BM, Spring S, Teeling H, Quast C and others (2007) Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proc Natl Acad Sci USA* 104:2891–2896
- Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* 103:13104–13109
- Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, Brown JH (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* 105:7774–7778
- Galand PE, Casamayor EO, Kirchman DL, Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *PNAS* 106:22427–22432
- Galand PE, Potvin M, Casamayor EO, Lovejoy C (2010) Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *ISME J*, doi:10.1038/ismej.2009.134
- Glöckner FO, Fuchs BM, Amann R (1999) Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl Environ Microbiol* 65:3721–3726
- Hervàs A, Camarero L, Reche I, Casamayor EO (2009) Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. *Environ Microbiol* 11:1612–1623
- Horner-Devine MC, Bohannan BJM (2006) Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* 87:100–108
- Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM (2004) A taxa–area relationship for bacteria. *Nature* 432:750–753
- Horner-Devine MC, Silver JM, Leibold MA, Bohannan BJM and others (2007) A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* 88: 1345–1353
- Hubbell SP (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ
- Hutchinson GE (1961) The paradox of the plankton. *Am Nat* 95:137–145
- Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD (2007) Synchrony in aquatic microbial community dynamics. *ISME J* 1:38–47
- Langenheder S, Ragnarsson H (2007) The role of environmental and spatial factors for the composition of aquatic bacterial communities. *Ecology* 88:2154–2161
- Leibold MA, Holyoak M, Mouquet N, Amarasekare P and others (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* 7:601–613
- Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* 73:1943–1967
- Lindström E, Logue JB (2008) Biogeography of bacterioplankton in inland waters. *Freshw Rev* 1:99–114
- Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* 104:11436–11440
- Newton RJ, Jones SE, Helmus MR, McMahon KD (2007) Phylogenetic ecology of the freshwater *Actinobacteria* acI lineage. *Appl Environ Microbiol* 73:7169–7176
- Ng ISY, Carr CM, Cottenie K (2009) Hierarchical zooplankton metacommunities: distinguishing between high and limiting dispersal mechanisms. *Hydrobiologia* 619:133–143
- Papke RT, Ward DM (2004) The importance of physical isolation to microbial diversification. *FEMS Microbiol Ecol* 48:293–303
- Pommier T, Canback B, Riemann L, Bostrom KH and others (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* 16:867–880
- Reche I, Pulido-Villena E, Morales-Baquero R, Casamayor EO (2005) Does ecosystem size determine aquatic bacterial richness? *Ecology* 86:1715–1722
- Reche I, Pulido-Villena E, Morales-Baquero R, Casamayor EO (2007) Does ecosystem size determine aquatic bacterial richness? Reply. *Ecology* 88:253–255
- Rosselló-Mora R, Amann R (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* 25:39–67
- Sommaruga R, Casamayor EO (2009) Bacterial ‘cosmopolitanism’ and importance of local environmental factors for community composition in remote high-altitude lakes. *Freshw Biol* 54:994–1005
- Van Der Gucht K, Cottenie K, Vloemans N and others (2007) The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proc Natl Acad Sci USA* 104:20404–20409
- Von Mering C, Hugenholtz P, Raes J, Tringe SG and others (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* 315: 1126–1130
- Webb CO (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am Nat* 156:145–155
- Webb CO, Ackerly DD, Mcpeck MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annu Rev Ecol Syst* 33:475–505
- Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science* 301:976–978

*Editorial responsibility: Paul del Giorgio, Montreal, Quebec, Canada*

*Submitted: August 18, 2009; Accepted: November 16, 2009  
Proofs received from author(s): March 8, 2010*