LETTER

Terrestrial origin of bacterial communities in complex boreal freshwater networks

Clara Ruiz-González,[1]*
Juan Pablo Niño-García[1,2] and
Paul A. del Giorgio[1]

Abstract

Bacteria inhabiting boreal freshwaters are part of metacommunities where local assemblages are often linked by the flow of water in the landscape, yet the resulting spatial structure and the boundaries of the network metacommunity have never been explored. Here, we reconstruct the spatial structure of the bacterial metacommunity in a complex boreal aquatic network by determining the taxonomic composition of bacterial communities along the entire terrestrial/aquatic continuum, including soil and soilwaters, headwater streams, large rivers and lakes. We show that the network metacommunity has a directional spatial structure driven by a common terrestrial origin of aquatic communities, which are numerically dominated by taxa recruited from soils. Local community assembly is driven by variations along the hydrological continuum in the balance between mass effects and species sorting of terrestrial taxa, and seems further influenced by priority effects related to the spatial sequence of entry of soil bacteria into the network.

Keywords

Aquatic bacterial communities, bacterial biogeography, boreal ecosystems, mass effects, network metacommunity, priority effects.

Ecology Letters (2015)

INTRODUCTION

Over two decades of molecular studies have revealed that aquatic bacterial communities are extremely diverse and highly dynamic in terms of taxonomic composition at different spatial and temporal scales. It is now well understood that freshwater communities are taxonomically distinct from estuarine and marine communities (Bouvier & del Giorgio 2002; Lozupone & Knight 2007; Tamames *et al.* 2010), but there is also a large degree of variation in bacterial community structure among and within different types of freshwater ecosystems (Crump *et al.* 2007; Lear *et al.* 2013, 2014; Barberán & Casamayor 2014; Liu *et al.* 2015; Read *et al.* 2015). Compared to the ocean, inland waters are much more spatially heterogeneous, ranging from isolated lakes fed essentially by rainfall, to complex aquatic networks where the different water bodies display varying degrees of connectivity among each other and with their surrounding terrestrial environment, and through which water and bacteria flow in the direction dictated by the landscape topography (Lindström *et al.* 2006; Crump *et al.* 2007, 2012; Nelson *et al.* 2009; Besemer *et al.* 2013). Features like the relative position in the network, the hydrologic connections and the local water residence time are known to influence the structure of aquatic microbial communities, as they regulate the balance between the immigration of bacteria from adjacent ecosystems, and the sorting of species by local environmental conditions, predation or competition (Lindström *et al.* 2005, 2006; Crump *et al.* 2007; Logue & Lindström 2010; Székely *et al.* 2013; Read *et al.* 2015).

Consequently, any attempt to understand community assembly within the different components of these complex aquatic networks must be placed in the context of the distribution and movement of microbes within the entire network.

In this regard, bacteria inhabiting these aquatic ecosystems can be considered as part of a broader metacommunity that is expressed differentially in response to local conditions and hydrology. Although this metacommunity concept provides a framework for approaching the study of spatial distribution of freshwater bacterial communities (e.g. Battin *et al.* 2007), its application in the context of complex aquatic networks is not straightforward. For example there may be multiple sources of bacteria that seed the different components of the network (i.e. aquatic sediments, rainfall, atmospheric deposition, metazoans) such that the metacommunity is the sum of disparate assemblages linked by dispersal but with little or no overall spatial structure or direction. Given the unidirectional flow of water in the landscape, however, it seems more plausible that there is an upstream source of bacteria that is dispersed downstream and expressed locally, thus generating a metacommunity that has a directional spatial structure. Establishing the spatial structure of the metacommunity within aquatic networks is thus key to understanding the mechanisms of local community assembly, but given the lack of cross-system field studies, our knowledge of the movement and interactions of microbes within complex aquatic networks remains largely limited.

Ultimately, most of the water feeding these aquatic networks comes from subsurface groundwater and direct surface

[1]*Groupe de Recherche Interuniversitaire en Limnologie et en Environnement Aquatique (GRIL), Département des Sciences Biologiques, Université du Québec à Montréal, Case Postale 8888, Succursale Centre-Ville, Montréal, QC H3C 3P8, Canada*

[2]*Escuela de Microbiología Universidad de Antioquia Ciudad Universitaria Calle 67 N° 53-108, Medellín, Colombia*
Correspondence: E-mail: clara.ruiz.glez@gmail.com

runoff, which inevitably carries bacteria from the surface and deeper layers of the surrounding soils. It is thus possible that the source of bacteria that generates the freshwater communities resides in the surrounding terrestrial environment. In this regard, several studies on river bacterial biogeography have suggested that headwater streams act as collectors and subsequent broadcasters of microbes from terrestrial sources (Crump *et al.* 2007; Nelson *et al.* 2009; Besemer *et al.* 2012, 2013; Savio *et al.* 2015), yet only one study has directly addressed this issue, reporting that the bacterial communities from a single arctic lake and its inlet were largely dominated by taxa from upstream soilwater (Crump *et al.* 2012). These results indicate that the community structure observed in aquatic ecosystems may be the local expression of a soil-derived inoculum, yet the generality of this pattern across landscapes, or the reach of this terrestrial influence within a given network, remain completely unknown. In particular, we do not know whether this terrestrial signature on local communities decreases along the hydrologic continuum or is maintained even in ecosystems with low connectivity with land and long water residence times. In addition, since soil and freshwater bacterial communities are consistently dominated by different bacterial taxa (Zwart *et al.* 2002; Lozupone & Knight 2007; Crump *et al.* 2012), a terrestrial origin of aquatic communities would imply major shifts in the expression of this common pool of microbes, but how and where these shifts occur along the terrestrial-aquatic continuum is still unclear.

These issues are particularly relevant in the vast boreal biome, which harbours not only the highest densities of freshwaters globally, but also the most complex and heterogeneous aquatic networks, and the highest degree of land–water connectivity. In this study, we reconstruct the spatial structure of the bacterial metacommunity within a complex aquatic network covering an area of ca. 13 000 km^2 in the Eastern boreal region of Québec (Canada). To explore the potential direc-

tionality in the spatial structure of the network metacommunity, we determined the composition of bacterial communities along the entire hydrologic continuum across 135 aquatic sites ranging from the smallest headwater streams to the largest rivers and lakes. In addition, we explicitly assessed whether aquatic communities originate in the terrestrial environment, by characterising the composition of 84 additional bacterial assemblages inhabiting the soils associated to our sampled waters, and exploring the changes in community composition and the recruitment of soil-derived taxa along the terrestrial-aquatic continuum.

MATERIALS AND METHODS

Study sites, sampling and basic parameters

During July 2013, we collected 219 samples for the characterisation of bacterial communities, including soils ($n = 43$), soilwaters ($n = 41$), small streams (Strahler order ≤ 2, $n = 55$), rivers (Strahler order 2–8, $n = 32$) and lakes ($n = 48$). The different ecosystem types were randomly sampled over a 3-week period and thus there is no temporal bias in our results. All sites belong to La Côte-Nord region of Québec (44–56°N, 64–80°W, Fig. 1a), located in the northern margin of the Saint Laurence estuary. This is a boreal maritime ecozone, with an average annual temperature of 0.8 °C (−15 °C to 15 °C) and precipitation of 1150 mm. The vegetation is dominated by coniferous trees such as balsam fir (*Abies balsamea*) and spruce (*Picea* sp.). Deciduous trees including birch (*Betula* sp.), alder (*Alnus* sp.) and willow (*Salix* sp.) occur along the stream channels, and mosses dominate mires and riparian wetlands. The area has a high density of interconnected surface waters and is a runoff hotspot in the northern region of Québec.

Elevation of the sampled sites ranged from 7 to 600 m.a.s.l. Sites were sampled once during summer low flow conditions,
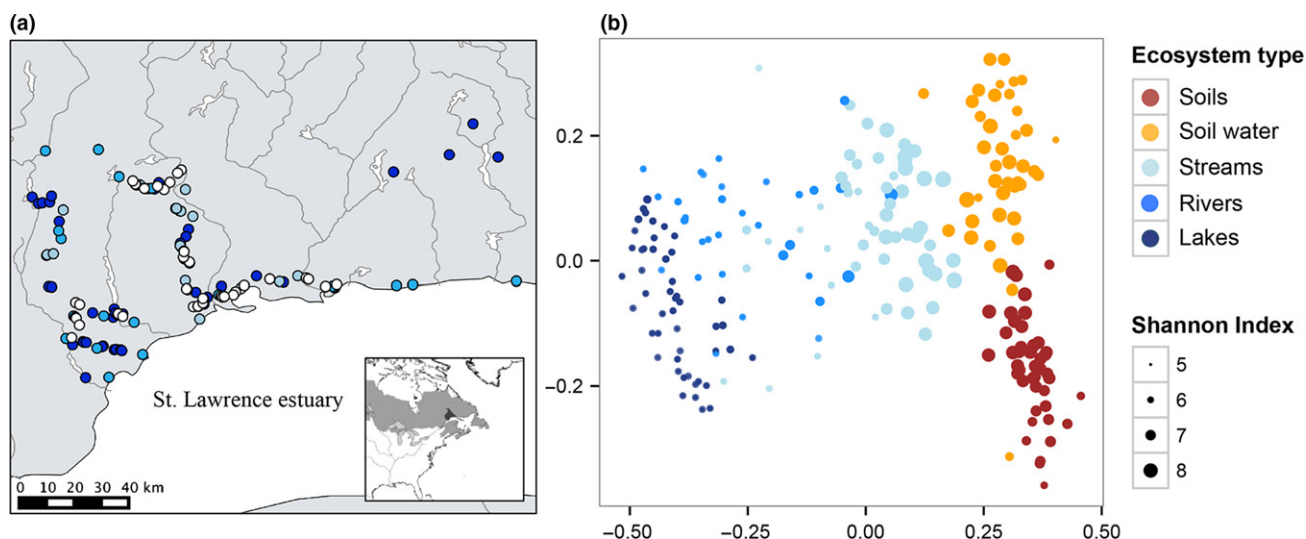

Figure 1 (a) Map showing the location and distribution of sites across the sampled region. The inset map indicates the position of La Côte-Nord region in North America, and the shaded area represents the boreal zone. Colour indicates lakes (dark blue), rivers (Strahler order >2, blue) and streams (Strahler order ≤ 2, light blue and white), white dots indicate streams where the adjacent soil and soil water samples were also taken. (b) Non-metric multidimensional scaling (NMDS) plots based on Bray–Curtis distances of taxonomic composition of bacterial communities. Colours indicate the different system types and dot size is proportional to the taxonomic diversity (Shannon index) of each community.

and cover a wide range in lake area (0.002–50 km^2) and river order (Strahler order 0–8). We defined as streams of order 0 those that did not appear in topographic maps. The surrounding soils ranged from peat-dominated to mineral and sandy soils (Table S1). Water samples were collected from lakes at a depth of 0.5 m at the deepest measured spot, and near the shore in streams and rivers. At 43 of these sites (corresponding to the smallest order streams), soil and soilwater samples were also taken (Fig. 1a). Soilwater samples were collected with piezometers installed at an average depth of 50 cm ± 20 cm, at 30–80 cm from the nearest stream. The initial volume was discarded, and fresh soilwater was collected after 1–2 h with a peristaltic pump from 3 to 5 randomly chosen piezometer locations. Water temperature, dissolved oxygen (DO), pH and conductivity were measured *in situ* or in the collected soilwater with a YSI probe. At each site, a water sub-sample was filtered *in situ* through 0.45 μm and stored in acid-washed glass vials for dissolved organic carbon (DOC) and optical analyses, and another sample was kept in acid-rinsed bottles for DNA processing in the laboratory. DOC concentration was measured on an OI1010 TOC analyser, and coloured dissolved organic matter (cDOM) was quantified as the absorbance at 440 nm using an Ultrospec3100 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Soil samples were collected from the top layer (0–20 cm) of the soil around the streams pooling samples from five randomly chosen locations. Soil pH was determined with a pH3310 pH-metre in 1:2.5 (w/v) soil:water extracts (Van Reeuwijk 2000).

Geographical analyses

Lake surface, river length and order, catchment areas and elevation of the sampled sites were obtained using the ArcMap 10 and ArcGIS V10 software (ESRI Inc., Redland, CA) applied on Digital Elevation Models (DEM) derived from (1:50 000) maps. For each catchment, we averaged slope based on DEM, and determined various landcover properties from Geobase (2009). Lake water residence time was estimated from lake volume (estimated as mean depth × lake area), catchment area and mean annual regional runoff. River residence time was calculated based on the measured water velocity (using a FlowTracker Handheld-ADV (Sontek, San Diego, CA, USA) when possible, or timing a small floating device) and the total upstream distance.

Bacterial community composition

0.25 g of soil or 300–500 mL of water filtered onto 0.22 μm filters were used for genomic analyses. Bacterial DNA was extracted using the MoBio (Carlsbad, CA, USA) PowerWater (for lake and river samples) and PowerSoil (soil and soilwater samples) DNA extraction kits, after checking that processing a same sample with both kits yielded similar results. Tagged amplicons of the 16SrRNA gene (V3–V4 region) were obtained with the primers 515F and 806R using a two-step PCR and the Access Array barcode library (Fluidigm), and sequenced on an Illumina MiSeq2000 using a paired-end approach (Caporaso *et al*. 2012a). Paired-end reads were assembled with flash (Magoc & Salzberg 2011) and sequences

between 250 and 290 bp were used for downstream analyses in QIIME to remove primers, low-quality, archaeal and chloroplast reads (Caporaso *et al*. 2010). Quality sequences were binned into operational taxonomic units (OTUs, ≥97% similarity) using uclust v1.22q (Edgar 2010) and RDP classifier (Wang *et al*. 2007). Representative sequences were then aligned against the SILVAv108 reference alignment (Pruesse *et al*. 2007). To ensure that rare bacteria were not the result of sequencing errors, we discarded all OTUs present in <10 samples and/or showing <10 sequences, resulting in a mean of 77 000 sequences recovered per sample (range 36 423– 301 519). To enable comparisons between samples, the OTU table was randomly subsampled to ensure an equal number of sequences per sample, based on the sample with the least number of reads (36 423 sequences).

Statistical analyses

Differences in bacterial taxonomic composition between ecosystem types were tested with ANOSIM (Clarke 1993). The Shannon index was calculated as an estimate of bacterial taxonomic diversity. Bray–Curtis distance was used as an estimator of taxonomic dissimilarity between samples, and was visualised by non-metric multidimensional scaling (NMDS) analysis. We performed least squares linear regression to establish relationships between individual variables. All analyses were performed in jmp 9.0.1 (SAS Institute, NC, USA) or r 3.0.0 software (R Core Team, 2013, Vegan package).

RESULTS

The sampled sites spanned a large range of variation in physical and chemical conditions both within a given ecosystem type, and also between the different types of ecosystem (Table S1). We recovered 20 858 980 quality sequences, which clustered into 263 646 OTUs (97% sequence similarity). After subsampling our OTU table to 36 423 reads/sample, 8 741 520 sequences (236 904 OTUs) were retained. Altogether, 29 bacterial phyla were detected, yet the total pool of sequences was mainly dominated by the phyla Proteobacteria (48% of the total number of sequences), followed by Actinobacteria (14%), Acidobacteria (7%), Bacteroidetes (7%) and Verrumicomicrobia (4%). The OTU accumulation curve showed a clear plateau, suggesting that we had good coverage of the regional bacterial richness (Fig. S1).

Structure of bacterial communities within complex aquatic networks

We observed a striking segregation of communities based on the type of ecosystem, which were organised following a clear sequence from soils and soilwaters, through small streams to larger rivers, and finally lakes (Fig. 1b, ANOSIM$_{ECOSYSTEM}$ $R = 0.87$, $P < 0.0001$). In turn, pH explained the differences between communities within each type of ecosystem (Fig. S2). In terms of taxonomic composition, whereas soil communities showed the highest proportions of Acidobacterial classes, the typical freshwater classes Betaproteobacteria and Actinobacteria gradually increased along the aquatic continuum and

Table 1 Average operational taxonomic units (OTU) number, Shannon diversity index, % of unique OTUs and sequences per bacterial community within each type of ecosystem and total OTUs and sequences within each ecosystem type

System type	OTU number	Shannon	% unique OTUs	% unique sequences	Total OTUs per ecosystem	Total sequences per ecosystem
Soils	8980 (1405)[a]	7.4 (0.4)[a]	12.3 (3.9)[a]	4.3 (1.4)[a]	74 389	1 566 189
Soil water	9029 (1806)[a]	7.5 (0.5)[a]	9.0 (4.0)[b]	3.9 (2.3)[a]	88 342	1 493 343
Streams	10352 (2388)[b]	7.2 (1.0)[a]	8.1 (2.1)[b]	2.9 (1.2)[b]	146 071	2 003 265
Rivers	7330 (1797)[c]	6.0 (0.7)[b]	7.5 (3.8)[b]	1.7 (0.7)[c]	98 224	1 165 536
Lakes	5588 (902)[d]	5.7 (0.4)[b]	16.8 (3.9)[c]	3.1 (1.1)[b]	83 712	1 748 304

Values are means (±standard deviation) of the sites within each ecosystem type. Different letters represent means that are significantly different from each other (Tukey's *post hoc* test; $P < 0.001$). Unique OTUs were those that were exclusively found in a single type of ecosystem.

dominated river and lake communities (Fig. S3). Most OTUs were shared among different ecosystem types, and ecosystem-specific OTUs (Unique OTUs, Table 1) averaged only 11% of all OTUs within communities, and represented less than 3% of the total community sequences. Diversity and richness also varied along this terrestrial-aquatic continuum: mean OTU richness increased from soils, peaked in small streams and decreased again towards lakes (Table 1). The Shannon index of soil, soilwater and stream communities was significantly higher than that of river and lake assemblages (Fig. 1b, Table 1).

Terrestrial origin of the metacommunity in complex freshwater networks

The directional structure shown in Fig. 1b and the spatial patterns in richness and diversity point to a common terrestrial origin of the communities within the network. To verify this, we investigated the extent to which the changes in composition along the terrestrial-aquatic continuum were due to variations in the relative abundances of taxa originating in the terrestrial environment, as opposed to the appearance of new taxa along this continuum. To do so, we assigned each OTU to the farthest upstream environment where it was first detected (as in Crump *et al.* 2012) along the gradient depicted by the NMDS1 axis (soil/soilwater/stream/river/lake). This analysis shows that there were new OTUs appearing through-

out the terrestrial-aquatic continuum (Fig. 2a), and on average, 30–40% of the total number of OTUs within any given ecosystem type were detected for the first time in that ecosystem. The contribution of these new OTUs to the total number of sequences associated to a given ecosystem type was on average much lower, ranging from 19% in soilwaters to less than 3% in lakes (Fig. 2b). In contrast, although the proportion of soil-derived OTUs declined systematically towards lakes, where they represented only 17% of the total number of OTUs detected (Fig. 2a), these soil-derived taxa clearly dominated the sequence number in all aquatic ecosystems: together, soil- and soilwater-derived OTUs accounted for over 80% of all aquatic sequences (Fig. 2b).

Within the aquatic network, there was a gradual decline in the relative contribution of terrestrially derived OTUs to the total number of sequences towards systems with lower connectivity with land, from small (order 0–2) headwater streams, to the largest (order 6–8) rivers and lakes (Fig. 3a), and along a gradient of increasing water residence time (Fig. 3b). Regardless of their position along the aquatic continuum, however, communities were consistently dominated by terrestrially derived OTUs, which in all cases accounted for over 75% of all sequences. This numerical dominance of terrestrial OTUs in aquatic ecosystems was mostly due to the increase in relative abundance of a small subset of the taxa that were originally loaded from soils (Fig. S4A, B). Interestingly, of all
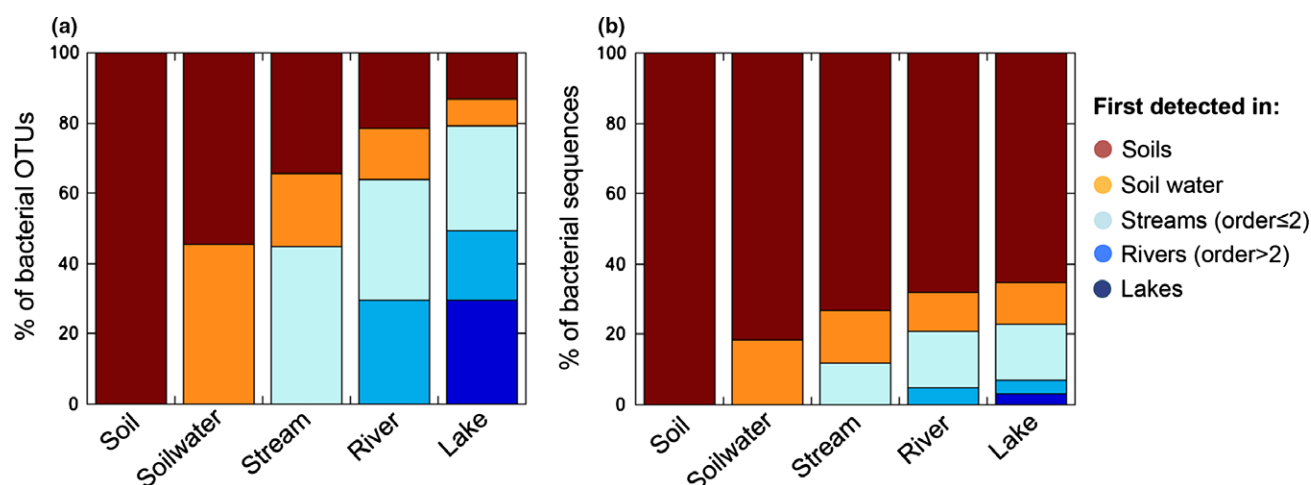

Figure 2 (a) Proportion of taxa (operational taxonomic units (OTUs)) categorised by the farthest upslope environment where they were first detected (see Methods), following the gradient depicted by the NMDS (Fig. 1b). (b) Proportion of sequences associated to those taxa. Values are expressed as a fraction of the total OTUs (a) or sequences (b) for each type of system.
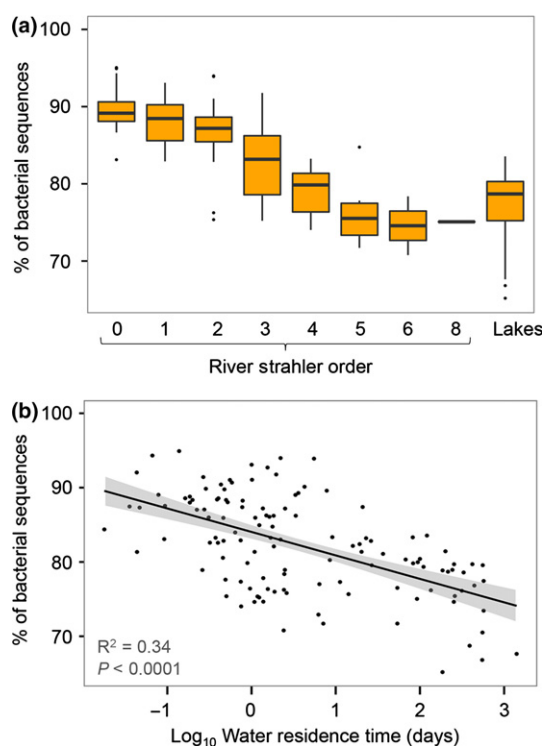
Figure 3 Contribution of terrestrially derived operational taxonomic units (OTUs) (i.e. those OTUs categorised as of soil or soilwater origin, see Fig. 2) to total bacterial sequences in open water communities as a function of river order (a), and estimated water retention time (b).

the terrestrial taxa, members the orders Burkholderiales (Cl. Betaproteobacteria) and Actinomycetales (Cl. Actinobacteria), which commonly dominate freshwaters, were the ones showing the steepest increases in relative abundance from soils to lakes, together comprising more than 45% of the sequences of lake and river communities (Fig. S4C). This evidence collectively suggests that dominant freshwater bacterial taxa are mainly recruited from the terrestrial environment.

Recruitment of rare bacteria along the aquatic continuum

To address whether there is recruitment of rare soil bacteria along the aquatic continuum, we identified the OTUs that showed the largest changes in relative abundances between communities at the network scale (i.e. between all pairs of sites). This change was calculated as the average Euclidean distance of the relative abundance of each OTU between all pairs of sites (see Appendix S1 for R codes), which was strongly positively related to the maximum change in relative abundance shown by any OTU between two sites ($R = 0.64$, $P < 0.0001$, $n = 233\,606$). We identified the OTUs showing the largest changes in relative abundance (termed 'shifting' OTUs) as those presenting a mean distance >10 (see Fig. S5), which roughly corresponded to an average maximum change of 704 sequences between two sites (range 60–10 440 sequences). This procedure recovered 603 'shifting' OTUs, and a detailed exploration of their spatial dynamics uncovered two clear groups that also had very different taxonomic composition (Fig. S6): OTUs that showed large increases in abundance from the ter-

restrial environment to rivers and lakes, which were dominated by Betaproteobacterial and Actinobacterial orders ($n = 264$, Fig. S6A–B), and OTUs that were dominant in soils and decreased progressively towards lakes, which comprised a higher number of bacterial orders, the most abundant belonging to Acidobacteria and Alphaproteobacteria ($n = 339$, Fig. S6C–D). Although both groups are of interest, for this study we focused on the former group, which represents rare soil 'seed' taxa that were recruited aquatic ecosystems and become numerically dominant. These 264 'seed' OTUs represented 0.1% of the regional pool of taxa, yet they accounted for 32% of all the sequences recovered. Furthermore, out of the 264 'seed' OTUs, 32 were first detected in streams, whereas the rest were found to be of terrestrial origin (as per categories shown in Fig. 2), and their contribution to community sequences increased from soils towards rivers and lakes (Fig. 4a). Interestingly, the contribution of these 'seed' OTUs to total number of sequences increased steeply from soils to streams of order 3, and then stabilised around 60–70% of community sequences in larger rivers and lakes (Fig. 4a).

We compared the mean abundances of each 'seed' OTU between pairs of ecosystem types to identify the environmental interfaces along this gradient that promote changes in abundance of the 'seed' taxa (Fig. 4b). The largest shifts in the mean abundance of these OTUs occurred between the terrestrial (soil and soilwater) and the aquatic environments, with increases of 450- and 480-fold in mean abundance from soils to rivers and lakes respectively (Fig. 4b). Changes within either the terrestrial or the aquatic environment were much smaller (<20-fold increases, Fig. 4b).

We further explored whether there is an abundance threshold below which rare soil bacterial taxa cannot act as a seed and become dominant within the constraint of water residence time in the aquatic network. To do so, we selected all OTUs originating in terrestrial systems (soil- and soilwater-derived OTUs, $n = 114\,519$) and compared their mean abundance in terrestrial ecosystems to either i) the proportion of downstream sites (rivers, lakes) where they were abundant (>500 reads/site, Fig. 5a), and ii) to their mean downstream abundances (Fig. 5b). We found that those OTUs that were most often dominant in downstream sites (Fig. 5a) and which had higher average downstream abundances (Fig. 5b), tended to have low mean abundances (<10 sequences per sample) in terrestrial ecosystems, but at our sequencing depth we could not detect a minimum abundance threshold below which there is no further recruitment. Remarkably, all these OTUs had been previously categorised as 'seeds' based on our distance metric (Fig. 5). Analysis of the individual spatial dynamics of these OTUs revealed large increases in their abundances from soils to surface waters, although the peak abundances of different OTUs were attained at different points of the aquatic continuum, in streams (Fig. 5c), rivers (Fig. 5d–e) or lakes (Fig. 5f).

DISCUSSION

Our results demonstrate a clear directional structuring of the bacterial communities in these complex aquatic networks, revealing a sequential organisation of bacterial assemblages from soils and soilwater, through small streams towards rivers
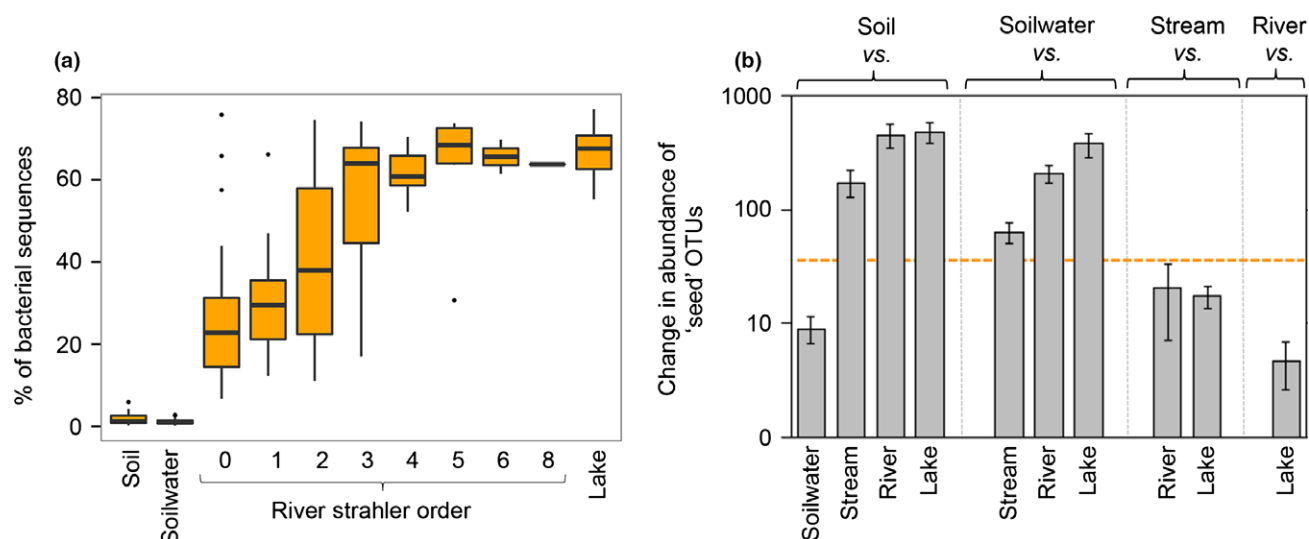
Figure 4 (a) Contribution of 'seed' operational taxonomic units (OTUs) to total bacterial sequences in the different types of ecosystems. Seed OTUs are defined as those that showed the largest changes in relative abundance at the network scale and increased from the soils to the lakes (see Results for further details). (b) Change in the mean relative abundance of the 'seed' OTUs between pairs of ecosystem types. Values are means (±standard error) of the ratio between the mean relative abundance of a given OTU in an upstream ecosystem relative to a downstream ecosystem. Note that the largest changes occurred when moving from terrestrial to surface water environments (above the orange line), while abundance changes between terrestrial or between aquatic ecosystems were much lower (below the orange line).

and lakes (Fig. 1b). The fact that very few taxa were unique to a given type of ecosystem (Table 1), together with the clear ecosystem-specific segregation of communities (Fig. 1b), points to a strong local selection of species from a shared pool of microbes, further supported by the clear pH-driven structuring of communities within each ecosystem type (Fig. S2). In addition, the lack of an obvious discontinuity between terrestrial and aquatic communities suggests that soils are clearly part of the network metacommunity.

In agreement with previous studies proposing that headwater streams collect communities from various upstream terrestrial environments (Nelson *et al.* 2009; Crump *et al.* 2012; Besemer *et al.* 2013), we found that the OTU richness peaked in the smallest streams and decreased gradually towards lakes (Table 1), a trend that was accompanied by gradual increases in the relative abundance of typical freshwater taxa (Fig. S3). This, together with the strong directional structure in bacterial community composition, points to a common terrestrial origin of the communities within the network. Accordingly, we found that all communities were numerically dominated by terrestrially derived taxa, which on average accounted for >75% of the sequences in aquatic communities (Figs 2b and 3a).

Dispersal followed by species sorting has been shown to shape the composition of downstream freshwater assemblages, and the relative contribution of both mechanisms is known to vary as a function of water residence time, among other factors (Lindström & Bergström 2004; Lindström *et al.* 2006; Crump *et al.* 2007). In this regard, if the numerical dominance of terrestrially derived taxa in aquatic communities was due to mass effects (i.e. massive advection of soil bacteria), we could expect that with increasing local residence time within the network and decreasing contact with the terrestrial environment, this terrestrial signature would be progressively lost. Interestingly, although the contribution of terrestrially derived

OTUs declined gradually as ecosystems become increasingly decoupled from land and water residence time increases (Fig. 3), the proportion of terrestrially derived sequences was seldom lower than 70%, and these OTUs numerically dominated the communities even in the systems with the longest water residence times (>1000 days). This implies that, even though the sorting of species is likely the dominant mechanism for community assembly in aquatic systems with long retention times, this selection operates on taxa that have been transported from soils. This dominance of terrestrially derived sequences in aquatic communities had been previously reported for individual ecosystems (Crump *et al.* 2012), yet our results extend this observation to entire regional aquatic networks in the boreal landscape. Moreover, our study was performed during an extremely dry summer, and thus the observed patterns represent a scenario of base flow and therefore of minimum connectivity between soils and surface waters, implying that the terrestrial influence on aquatic community assembly might even be higher under more average hydrological conditions.

In spite of the dominance of terrestrially derived OTUs in all communities, we observed the appearance of many new OTUs along the aquatic continuum (Fig. 2a). These OTUs that appeared during transit may be intrinsically aquatic and recruited within the aquatic network, or they may originate in soils but not be well represented in our 43 soils. Indeed, the species accumulation curve using only the soil samples did not reach a plateau, suggesting that our soil samples did not fully capture the regional soil diversity. In any case, it is remarkable that these new OTUs consistently accounted for a small proportion of the total aquatic sequences regardless of their origin (Fig. 2b). For example in lakes, individual OTUs of upstream soil origin had an average abundance 23-fold higher than that of OTUs appearing during transit in rivers and lakes (details not shown). The question then is, why do these
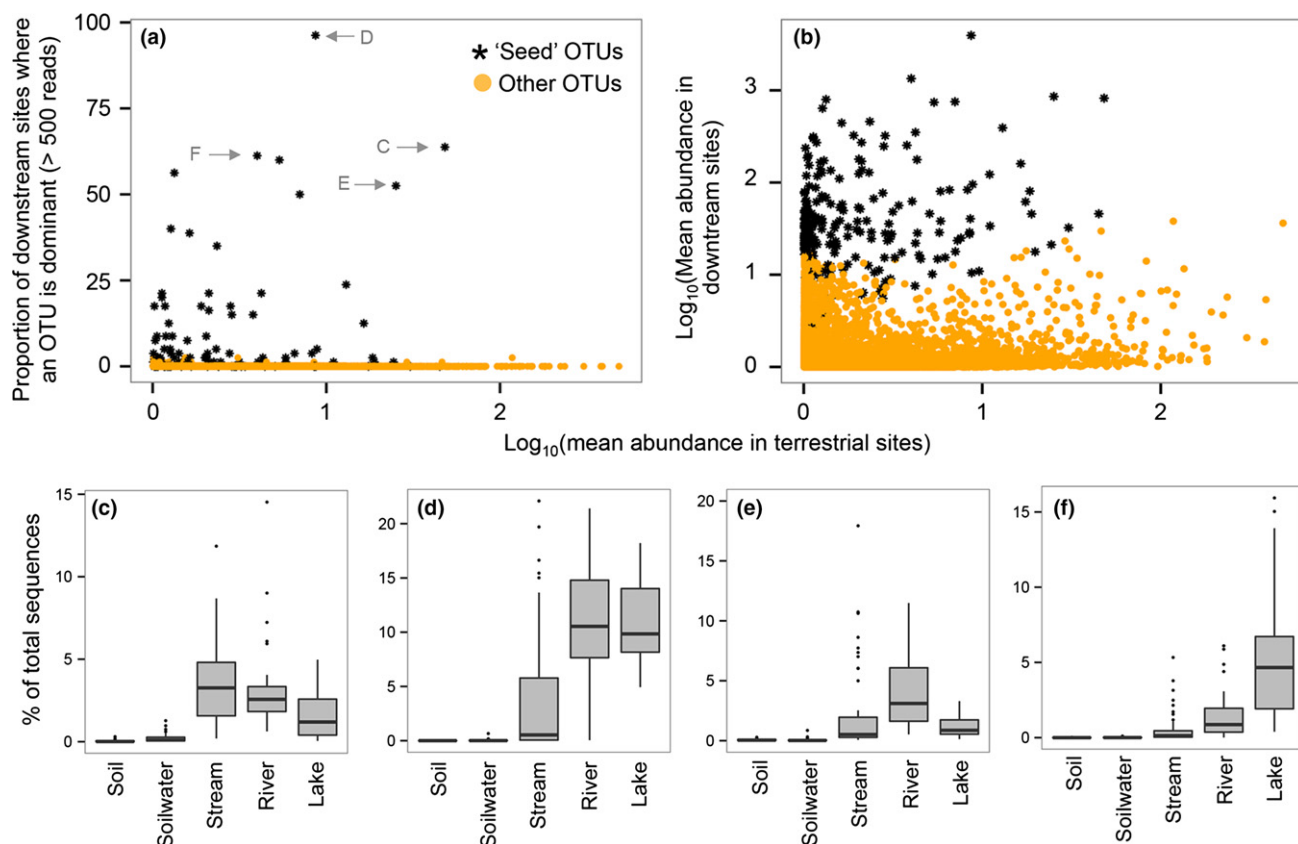
Figure 5 (a) Comparison between the proportion of downstream sites (rivers, lakes) where a given operational taxonomic unit (OTU) was abundant (>500 reads) and its mean relative abundance in upstream terrestrial sites (soil, soilwater). (b) Comparison between the mean abundance of OTUs in downstream sites and their mean relative abundance in upstream terrestrial sites. Only those OTUs originating in soil or soilwater from the rarefied OTU table where considered for this analysis (*n* = 114 519) since they were the ones that could be compared between upstream and downstream sites. Asterisks (*) represent which of those OTUs were classified as seed taxa based on our distance criterion (see Results). (c–f) Four examples of different spatial patterns of abundance shown of individual OTUs that were most often found to dominate in downstream waters (indicated by arrows in a), belonging to (c) Betaproteobacteria/Burkholderiales/Comamonadaceae; (d) Betaproteobacteria/Burkholderiales/Burkholderiaceae/*Polynucleobacter*; (e) Alphaproteobacteria/Sphingomonadales/Sphingomonadaceae; (f) Actinobacteria/Actinomycetales.

newly appearing OTUs never become numerically dominant, as opposed to the soil-derived bacteria that entered the network in the headwaters?

One plausible explanation is that these apparently aquatic OTUs become dominant during other times of the year, but our sampling does not allow addressing this possibility. A more intriguing explanation may be a direct consequence of the directional flow of bacteria throughout the network. When bacteria are dispersed into new habitats, they must not only face the prevailing environmental conditions, but also compete with the established local populations, which are likely better adapted to these conditions. As a result, taxa that are already well established have a competitive advantage compared to late-comers, in what has been termed 'priority effects' (e.g. Vannette & Fukami 2014; Fukami 2015). Transposed to the context of the directional network metacommunity structure, this would imply that inoculation of soil bacteria into the smallest headwater streams, where true aquatic communities are not yet developed, could facilitate their establishment and progressive growth as local parcels of water move downstream in the landscape. Although there will be sorting of species along the network, it will naturally favour those taxa firmly established in

the community. Consequently, late-arriving immigrants such as taxa directly inoculated from soils or rain into more established lake and river communities, will have lower establishment success than taxa that have entered the aquatic environment through headwater streams and attained higher abundances while travelling throughout the network. Our observation that taxa that are recruited downstream the aquatic continuum seldom attain high abundances (Fig. 2) agrees with the above scenario, pointing to the existence of strong priority effects likely via niche pre-emption (i.e. limitation of the local abundance that can be reached by late-arriving species due to space reduction by early-comers, Fukami 2015) in these aquatic bacterial metacommunities. This suggests that the network history, and in particular, the spatial sequence of arrival to local habitats, determines the fate of immigrant bacteria, and that together with hydrologically driven dispersal and environmental species sorting, priority effects may significantly modulate community assembly in these aquatic ecosystems.

The fact that many of the species initially found in soils disappeared towards the lakes (or were diluted by the growth of other taxa) further suggests that only a subset of these terrestrially derived taxa have the potential to grow and dominate

aquatic ecosystems (Fig. S4). To verify whether there is aquatic recruitment of rare soil bacteria, we identified the OTUs that showed the greatest shifts in relative abundance within the network. Interestingly, we found two clear spatial behaviours among those OTUs: OTUs abundant in soils that disappeared gradually towards the lakes, and OTUs rare in soils that showed steep increases in their abundances once in the aquatic environment (Fig. S6), which represent the terrestrial OTUs that seed the aquatic communities (i.e. 'seed' taxa). Although a handful of studies have shown that the recruitment of rare bacteria may explain seasonal or spatial patterns in community composition within a given ecosystem (Campbell *et al.* 2011; Caporaso *et al.* 2012b; Shade *et al.* 2014; Lindh *et al.* 2015) or among systems under transplant conditions (Sjöstedt *et al.* 2012; Comte *et al.* 2014), to our knowledge ours is the first study explicitly investigating recruitment patterns of bacteria at the landscape scale within complex aquatic networks. The behaviour of our 'seed' taxa is analogous to that of the conditionally rare taxa described before (Shade *et al.* 2014; Lynch & Neufeld 2015; Shade & Gilbert 2015), yet on a spatial rather than a temporal basis. Whether the pool of rare OTUs explaining seasonal changes in these boreal bacterioplankton communities also comprises terrestrial taxa deserves further exploration, but seems plausible based on our results.

Interestingly, these 'seed' OTUs were mostly of terrestrial origin, and showed the largest relative abundance shifts when transiting from the terrestrial to the aquatic environment, where they clearly dominated the communities in terms of sequence abundance (Fig. 4). This suggests that the land–water interface represents an important ecotone that triggers large shifts in the expression of the network metacommunity (Fig. 4) but that does not act as a dispersal barrier, since most terrestrial OTUs are still found in the smallest headwater streams (Fig. S4A). Once in flowing waters, the relative abundance of 'seed' OTUs increased gradually until rivers of order 3, beyond which the 'seed' taxa consistently comprised over 60% of the community sequences (Fig. 4a). This inflexion corresponded approximately to a local water residence time of 10 days (details not shown), suggesting that longer retention times do not result in further selection or growth of these 'seed' bacteria. Hydrological conditions determine the time that bacteria spend in a given ecosystem and therefore their capacity to grow in response to the local environment (e.g. Lindström *et al.* 2006; Crump *et al.* 2007; Nelson *et al.* 2009), and it is thus possible that there is an abundance threshold below which it is physiologically impossible for a bacterial taxon to become dominant within the time frame dictated by the movement of water in the landscape. Interestingly, we observed that most of the OTUs largely dominating river and lake communities were concentrated amongst those presenting the lowest abundances in soils (<10 reads, 0.03% of community sequences), and many fell right on the limit of detection. This suggests that we were probably not recovering all the potential 'seed' taxa within the network (Fig. 5), and that there may be effective recruitment of even rarer soil taxa.

This reservoir of rare bacteria that can grow under suitable conditions has been also referred to as the bacterial 'seed bank' (Lennon & Jones 2011). The existence of such a microbial seed bank was recently proposed for the ocean (Caporaso *et al.* 2012b; Gibbons *et al.* 2013), and our results suggest that freshwater ecosystems may also have a seed bank, which originates in the terrestrial environment and is integrated and channelled mostly through small streams feeding the diversity of boreal bacterioplankton communities. It seems, however, that at the current sequencing depth we are missing a fraction of the potential seed bank of freshwater communities. This has implications for our understanding of microbial biogeography as shallower sequencing efforts would lead to the erroneous conclusion that freshwater communities do not derive from a shared terrestrial pool of microbes, but rather from the immigration of bacteria from other sources, and raises the question of which sequencing depth is needed to capture the potential seed bank of freshwater ecosystems.

In summary, we show that the metacommunity of complex boreal freshwater networks has a directional spatial structure driven by a common terrestrial origin, which results in local aquatic communities numerically dominated by terrestrially derived taxa. We demonstrate that some rare terrestrial taxa have the potential to grow and become abundant in open water environments, although their dominance in aquatic ecosystems is dependent on both the sequence of their arrival and the time frame dictated by the hydrological conditions in the landscape. Our results further suggest that the study of the biogeographic patterns within discrete components of these aquatic networks (lakes, rivers) will only provide an incomplete perspective of the actual mechanisms shaping local community composition, and thus determining the spatial structure of bacterial metacommunities is essential for understanding microbial biogeography at the landscape scale.

AUTHORSHIP

PdG, CRG designed the sampling, CRG, JPNG, PdG collected the data, CRG, JPNG analysed the data, CRG, JPNG, PdG, discussed and interpreted the results and CRG, PdG wrote the manuscript.

REFERENCES

Barberán, A. & Casamayor, E.O. (2014). A phylogenetic perspective on species diversity, β-diversity and biogeography for the microbial world. *Mol. Ecol.*, 23, 5868–5876.

Battin, T.J., Sloan, W.T., Kjelleberg, S., Daims, H., Head, I.M., Curtis, T.P. *et al.* (2007). Microbial landscapes: new paths to biofilm research. *Nature*, 5, 76–81.

Besemer, K., Peter, H., Logue, J.B., Langenheder, S., Lindstrom, E.S., Tranvik, L.J. *et al.* (2012). Unraveling assembly of stream biofilm communities. *ISME J.*, 6, 1459–1468.

Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W. & Battin, T.J. (2013). Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Proc. R. Soc. B*, 280, 20131760.

Bouvier, T. & del Giorgio, P.A. (2002). Compositional changesin free-livingbacterial communities along a salinity gradientin two temperate estuaries. *Limnol. Oceanogr.*, 47, 453–470.

Campbell, B.J., Yua, L., Heidelberg, J.F. & Kirchman, D.L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc. Natl Acad. Sci. USA*, 108, 12776–12781.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, 7, 335–336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N. *et al.* (2012a). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.*, 6, 1621–1624.

Caporaso, J.G., Paszkiewicz, K., Field, D., Knight, R. & Gilbert, J.A. (2012b). The Western English Channel contains a persistent microbial seed bank. *ISME J.*, 6, 1089–1093.

Clarke, K.R. (1993). Non-parametric multivariate analysis of changes in community structure. *Aust. J. Ecol.*, 18, 117–143.

Comte, J., Lindström, E.S., Eiler, A. & Langenheder, S. (2014). Can marine bacteria be recruited from freshwater sources and the air? *ISME J.*, 8, 2423–2430.

Crump, B.C., Adams, H.E., Hobbie, J.E. & Kling, G.W. (2007). Biogeography of bacterioplankton in lakes and streams of an Arctic tundra catchment. *Ecology*, 88, 1365–1378.

Crump, B.C., Amaral-Zettler, L.A. & Kling, G.W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J.*, 6, 1629–1639.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461.

Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.*, 46, 1–23.

Geobase (2009) Land Cover, Circa 2000-Vector.

Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R. & Gilbert, J.A. (2013). Evidence for a persistent microbial seed bank throughout the global ocean. *Proc. Natl Acad. Sci. USA*, 110, 4651–4655.

Lear, G., Washington, V., Neale, M., Case, B., Buckley, H. & Lewis, G. (2013). The biogeography of stream bacteria. *Global Ecol. Biogeogr.*, 22, 544–554.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E. & Buckley, H.B. (2014). Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. *ISME J.*, 8, 1715–1726.

Lennon, J.T. & Jones, S.E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.*, 119, 119–130.

Lindh, M.V., Sjöstedt, J., Andersson, A.F., Baltar, F., Hugerth, L.W., Lundin, D. *et al.* (2015). Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environ. Microbiol.*, 17, 2459–2476. doi:10.1111/1462-2920.12720.

Lindström, E.S. & Bergström, A.K. (2004). Influence of inlet bacteria on bacterioplankton assemblage composition in lakes of different hydraulic retention time. *Limnol. Oceanogr.*, 49, 125–136.

Lindström, E.S., Kamst-Van Agterveld, M.P. & Zwart, G. (2005). Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Appl. Environ. Microbiol.*, 71, 8201–8206.

Lindström, E.S., Forslund, M., Algesten, G. & Bergström, A. (2006). External control of bacterial community structure in lakes. *Limnol. Oceanogr.*, 51, 339–342.

Liu, L., Yang, J., Yu, Z. & Wilkinson, D.M. (2015). The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J.*, doi:10.1038/ismej.2015.29.

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

Lozupone, C.A. & Knight, R. (2007). Global pattens in bacterial diversity. *Proc. Natl Acad. Sci. USA*, 104, 11436–11440.

Lynch, M.D.J. & Neufeld, J.D. (2015). Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.*, 13, 217–229.

Magoc, T. & Salzberg, S. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27, 2957–2963.

Nelson, C.E., Sadro, S. & Melack, J.M. (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.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. *et al.* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.*, 35, 7188–7196.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Read, D.S., Gweon, H.S., Bowes, M.J., Newbold, L.K., Field, D., Bailey, M.J. *et al.* (2015). Catchment-scale biogeography of riverine bacterioplankton. *ISME J.*, 9, 516–526.

Savio, D., Sinclair, L., Ijaz, U.Z., Parajka, J., Reischer, G.H., Stadler, P. *et al.* (2015). Bacterial diversity along a 2600 km river continuum. *Environ. Microbiol.*, (DOI: 10.1111/1462-2920.12886).

Shade, A. & Gilbert, J.A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends Microbiol.*, 23, 335–340.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N. *et al.* (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio*, 5, e01371–e01314.

Sjöstedt, J., Koch-Schmidt, P., Pontarp, M., Canbäck, B., Tunlid, A., Lundberg, P. *et al.* (2012). Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Appl. Environ. Microbiol.*, 78, 1361–1369.

Székely, A.J., Berga, M. & Langenheder, S. (2013). Mechanisms determining the fate of dispersed bacterial communities in new environments. *ISME J.*, 7, 61–71.

Tamames, J., Abellán, J., Pignatelli, M., Camacho, A. & Moya, A. (2010). Environmental distribution of prokaryotic taxa. *BMC Microbiol.*, 10, 85.

Van Reeuwijk, L.P. (2000). Procedures for soil analysis. In: Technical Paper 9. (ed. Van Reeuwijk, L.P.). ISRIC, Wageningen, the Netherlands.

Vannette, R.L. & Fukami, T. (2014). Historical contingency in species interactions: towards niche-based predictions. *Ecol. Lett.*, 17, 115–124.

Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. (2007). Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*, 73, 5261–5267.

Zwart, G., Crump, B.C., Kamst-Van Agterveld, M.P., Hagen, F. & Han, S. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microb. Ecol.*, 28, 141–155.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).