**Regional community assembly drivers and microbial
environmental sources shaping bacterioplankton in an alpine
lacustrine district (Pyrenees, Spain)**

Rüdiger Ortiz-Álvarez¹, Joan Cáliz¹, Lluís Camarero¹, Emilio O.
Casamayor^{1*}

¹Integrative Freshwater Ecology Group, Center for Advanced Studies of
Blanes (CEAB - CSIC). C/Accés cala St Francesc n°14, E-17300 Blanes,
Catalonia, Spain

*e-mail corresponding author: casamayor@ceab.csic.es

Running title: Community assembly and bacterial sources in alpine lakes

Keywords: metacommunity, assembly, alpine lakes, bacteria, freshwater, 16S rRNA,
Illumina, NGS

Abstract

Microbial communities in natural ecosystems are subject to strong ecological rules. The study of local communities along a regional metacommunity can reveal patterns of community assembly, and disentangle the underlying ecological processes. In particular, we seek drivers of community assembly at the regional scale using a large lacustrine dataset (>300 lakes) along the geographical, limnological and physico-chemical gradients in the Pyrenees. By using high throughput amplicon sequencing of the 16S rRNA gene, and inferring environmental sources of bacterial immigrants, we showed that surface aquatic bacterial assemblages were strongly influenced by terrestrial populations from soil, biofilms or sediments, and primarily selected by a pH-alkalinity gradient. Indeed, source proportions explained 27% of the community variation, and chemistry 15% of the total variation, half of it shared with the sources. Major taxonomic groups such as Verrucomicrobia, Actinobacteria, and Bacteroidetes showed higher aquatic affinities than Parcubacteria, Gammaproteobacteria, Alphaproteobacteria or Betaproteobacteria, which may be recruited and selected through different hydrographic habitats. A regional fingerprint was observed with lower alpha diversity and higher beta diversity in the central Pyrenees than in both ends. We suggest an ecological succession process, likely influenced by complex interactions of environmental source dispersal and environmental filtering along the mountain range geography.

Originality-significance statement

Understanding the general rules of microbial community assembly is one of the current challenges in ecology. Biological communities assemble through dispersal from a global pool and environmental filtering processes to form local communities according to their environmental preferences. Within metacommunities, the study of whole regions with partially isolated local communities can disentangle the ecological processes leading to idiosyncratic spatial structures. In this study, we simultaneously analyzed >300 alpine lakes within the Pyrenean lacustrine district as an especially suitable living laboratory where local effects can be easily noticed and scaled to the regional level because of their remoteness, partial isolation, and strong environmental gradients. We showed that lacustrine plankton is strongly influenced by terrestrial populations that experience differential environmental pressures, and that these sources are the primary drivers of community structure. A spatial fingerprint was observed at the regional level, which could be related to the accumulation of these processes over time.

Introduction

Understanding the general rules of microbial community assembly is one of the current challenges in ecology. The ecological theory posits that communities assemble through dispersal from a global pool and environmental filtering processes [1] to form local communities [2] according to their environmental preferences [3, 4]. In general, local environmental filtering results from the combination of substrate (i.e., soil type), chemistry (i.e., water alkalinity), geologic history (i.e., rock formations), climatic factors (i.e., temperature gradients), and biotic-related factors (i.e., species competition). When evaluating local communities at the regional scale, assembled biological communities are expected to be the result of the complex integration of these processes.

In continental hydrological networks, the whole process of community assembly is complex, and is linked to a higher spatial heterogeneity than in the surface marine realm [5, 6]. In addition, hydrological networks have properties that combine the peculiarities of both isolated (i.e., lakes as water islands, surrounded by terrestrial environments) and highly connected flowing systems (i.e., terrestrial materials transfers subjected to temporal dynamics of streams or small flows at the catchment scale [7]). These fluxes are particularly relevant in alpine oligotrophic aquatic networks, where seasonal processes can exert a population seeding effect from the snowpack, littoral, or watershed soil water [8]. Niche specialization becomes evident in boreal hydrographic networks, where communities from rivers, sediments or soils are differentiated and directionally structured to colonize lake water [9]. The combination of water flows, terrestrial inputs and local filtering is however still not well understood at regional scales. Lake plankton aggregates organisms from different dispersal sources and integrate bacteria from processes acting at the catchment scale [10, 11], in transition between water, air and surrounding grounds [12].

Lakes are living laboratories where local effects can be easily noticed. Given the strong altitudinal gradient and the mostly highly diluted and oligotrophic nature of high-mountain waters, alpine microbial metacommunities are expected to display different ecological patterns than those of lowland freshwaters, which usually display higher organic matter contents [13, 14]. Environmental filters such as pH has been called to explain spatial gradients and temporal variations in alpine lakes [15, 16] and in boreal

hydrographic networks [9, 17]. Also, other environmental filters related to trophic state such as phosphorus or nitrates [18, 19], hydrographic factors such as water retention time [17, 20] or even species-area relationships [16, 21] have been reported to explain alpine metacommunity patterns. In highly connected hydrographic networks an interaction between hydrology and local pH has been called as the major force structuring communities [17] suggesting that a combined exploration of dispersal and environmental filtering is also needed in the more isolated lakes within the alpine biome. However, the strength of the whole set of environmental filters on immigrant species is unknown. Hypothetically, specific ranges and values of any environmental variable could selectively promote or prevent colonization, survival or growth of immigrants from different origins such as soils, sediments, rivers or ground waters from the watersheds.

The Pyrenees mountain range hold more than 1000 lakes with different local communities that according to zooplankton composition belong to the same eco-region [22]. So far, the extensive limnological tradition in this area has been mostly focused on a few lakes because of the remoteness, high isolation and difficulties to extensive sampling [11], although a few large spatial surveys have been carried out [18]. Because of its high reported microbial molecular diversity coupled to its strong environmental gradients [16, 23, 24], the Pyrenees lacustrine district region is an ideal natural laboratory to study microbial metacommunity processes. Reconstructions of ice regression chronology in the Pyrenees suggest an east-west gradient in ice extent during the last glacial maximum and a faster retreat of the east-Pyrenean ice sheet after temperature increments [26], which could have driven the structure of the contemporary hydrographic system, as observed in other ecosystems [27]. In addition, geological substrate is more alkaline in the west-Pyrenean region, and pH is known to structure alpine communities of the area at least within small spatial scales [16].

In the present investigation and given the opportunities provided by the Pyrenean lakes system, we aimed to uncover drivers of community assembly at the regional scale using a large Pyrenean lakes dataset (>300 lakes simultaneously studied) and looking for distribution patterns and source populations of local bacterial communities along geographical, limnological and physico-chemical gradients.

First, we intend to identify the main sources providing bacterial populations to lakes and if a taxonomic signature was associated with these sources. Second, we assessed whether or not environmental factors constrained the different sources of diversity. We further quantified sources and environmental filters at the lake scale to sort primary drivers of community assembly at the regional scale. Finally, we aimed to unveil and understand spatial signals of bacterial diversity at the regional level.

Results

Environmental and geographic gradients

A multivariate analysis using PCA (Figure S2) with rotation ‘varimax’ reduced the geographic, chemical and limnological complexity of the lakes dataset to three main components that explained 47% of the total variance (Table 1). Lakes did not cluster in the dataset but arranged in a wide continuum of data points. The first rotated component (RC1) included TDP, TP, DOP, DON and DOC, representing the nutrient availability, which was higher but also more heterogeneous in the eastern Pyrenees. The second component (RC2) was indicative of the alkalinity-acidity axis along the latitudinal gradient with high loadings of pH, ALC, COND and DIC in the west. The third component (RC3) included lake morphometric parameters such as lake area, WRT, depth and lake volume.

Taxonomic diversity and regional community structure

Bacterial taxonomic diversity at the regional scale was represented by 53 identified phyla. Bacteroidetes, Betaproteobacteria, Actinobacteria, and Alphaproteobacteria were the most abundant taxa in the region, accounting for 89% of the relative abundance (Figure 1), with different distribution in the lakes (Figure 1, Figure S3). Interestingly, the relative abundance of Bacteroidetes was negatively correlated to Alphaproteobacteria ($r_s = -0.31$, $p < 0.001$), both showing differential abundance/richness relationships per sample (Figure 1a). While Betaproteobacteria and Actinobacteria were abundance-enriched, the rest of the phyla were richness-enriched. In particular, Deltaproteobacteria and Parcubacteria (Patescibacteria superphylum) showed the highest regional richness (Figure 1b), although they were among the taxa with lower relative proportions.

The resulting regional structure of the Pyrenean bacterioplankton showed a high dispersion without any clear clustered structure (Figure 2). Community dissimilarity matrix was correlated with two PCA rotated components (Table 1) (after transformation to equivalent dissimilarity matrices): the alkalinity-acidity axis RC2 ($r_M = 0.20$, $p < 0.001$) and the lake morphometric axis RC3 ($r_M = 0.06$, $p < 0.05$). Parameters that are usually reported as significant did not shown significant associations, such as the nutrient related axis RC1 ($r_M = 0.02$, $p > 0.05$), or the potential effect of the hydrological

framework, based on water renewal time (WRT). Alpha diversity distances, based on the Shannon indices, were related to the bacterioplankton structure ($r_M=0.38$, $p < 0.001$). Effects of altitude on community variability were weak ($r_M=0.098$, $p < 0.001$), although we observed a significant effect of altitude solely on lakes with extreme pH, both $pH > 8$ ($r_M=0.28$, $p=0.004$, $n=30$ lakes) and $pH < 6.5$ ($r_M=0.20$, $p=0.019$, $n=32$ lakes).

In addition, beta diversity was quantified according to the main variables related to the community structure, i.e., pH and alpha diversity. When dividing the region between lakes with high and low diversity based on the median H' value (4.29), there were no significant beta diversity differences. When estimating beta diversity with both H' and pH factors, the low H' group had significantly higher beta diversity than the high H' group (Kruskal-Wallis, $H=18.05$, $p<0.001$) (Figure 2). Still, pH levels did not show significant differences within each diversity group nor evaluating it as a single factor.

Associated environmental ontology (EnvO) of bacteria inhabiting the lakes

The predominant EnvO terms associated to bacterial counterparts in databases were intimately related to aquatic terms, mostly to 'lake' (37.6%), but also to 'epilimnion' (5.3%), 'river' (3.1%) or 'groundwater' (2.7%). Particle-associated and terrestrial sources terms such as 'soil' (2.8%), 'wetland' (2.0%), 'sediments' (1.5%) and 'biofilm' (1.9%) were also substantially detected. In further analyses we avoided the use of other abundant but redundant and correlated terms such as 'freshwater', 'freshwater lake' or 'surface water'. Bacterial populations that shared a common environmental annotation were grouped for each lake and showed significant differences on their ecological features ($p < 0.05$, ANOVA; Figure 3). *Post-hoc* Tukey showed different groups of significance depending on the diversity metric used, but in general, there was an equivalent contribution from aquatic and terrigenous sources to overall alpha diversity (Figure 3a and 3b). Betadispersion based on bray-curtis (Figure 3c) showed that aquatic sources ('Epilimnion', 'River' and 'Groundwater') had the lowest values, while betadispersion based with Raup-Crick dissimilarities (Figure 3d) showed few statistical differences among sources, again with 'River' and 'Groundwater' with the lowest values. Interestingly, we observed a taxonomic signature in the predominance of each environmental source (Figure 4). Phyla such as Cyanobacteria, Verrucomicrobia, Actinobacteria or Bacteroidetes were mostly related to aquatic habitats (lake and

epilimnion), while Deltaproteobacteria, Parcubacteria or Gammaproteobacteria predominated in the terrestrial origin (soil and sediment). It is worth to mention that those phyla with negative correlations mentioned above such as Alphaproteobacteria and Bacteroidetes (Figure S3) had markedly different associated EnvO term profiles (Figure 4).

Role of bacterial sources and environmental filtering on the regional community assembly and diversity

To compare the relative influence of the chemical variables and the different EnvO terms within the nMDS ordination space (Figure 2), we performed first a variation partitioning analysis in the whole Pyrenees region (Figure 5a). EnvO terms alone, explained 27% of the variation, while chemical variables explained 15% of the total variation, half of it shared between chemical variables and EnvO sources. Regarding the shared variation between chemistry and sources (8% in total, up to 21% in the west), we explored concrete relationships between proportion of EnvO terms per lake, and its chemical and limnological properties. Four EnvO terms showed relevant constraints ($r_s > 0.15$ and $p < 0.05$). ‘Epilimnion’ showed a significant positive relationship with pH and ALC, but also to nutrient content (negative to NO₃ and TDN, and positive to DOC). The ‘river’ term was significantly linked to geography (Latitude). ‘Sediment’ proportion had also a geographic signal, and was further correlated to the rotated components from the PCA (negative to RC1 and positive to RC2). The ‘Soil’ term only showed a negative relationship with Na.

To fully uncover the complex ecological associations in the regional community structure, we applied a network analysis based on Spearman correlations between the axis of the nMDS ordination, chemistry, morphometric variables, alpha diversity, and the contribution of annotated EnvO terms to each lake (Figure 5b). We identified three distinct network modules of a total of 55 significant interactions (modularity 0.317). The main connector of the three modules and nMDS axis was ‘Epilimnion’, followed by the Shannon index. The first module (nMDS1) further showed strong negative relationships with pH ($r_s = -0.52$, $p < 0.001$) and related variables, but also a positive relationship to ‘Groundwater’ and ‘Soil’. The second module (nMDS2) was linked to alpha diversity (richness $r_s = -0.72$ and wtd.PD $r_s = -0.58$, $p < 0.001$), nutrient variables such as DOC or NO₂, and EnvO terms such as ‘River’, ‘Biofilm’, ‘Wetland’ and ‘Lake’, with an effect of Altitude. Interestingly the terms ‘Soil’, ‘Biofilm’ and ‘River’ correlated positively with

richness or wtd.PD, while both alpha diversity indices had a negative relation with ‘Lake’. Shannon index was not associated to any EnvO term. The third module (nMDS3) comprised two chemical variables: K and SRP (a measure of available phosphate in soils for plants), which were strongly associated to ‘Sediment’ ($r_s=0.52$, $p<0.001$) showing a trade-off with ‘Epilimnion’. The complex network of correlations results in three major combined forces structuring the community: (1) pH and ‘Epilimnion’, (2) nutrient chemistry and multiple sources, and (3) ‘Sediment’, with a geographic effect linked to the three forces.

Community assembly in the geographic context

Despite there was not a significant linear relationship between geographic distances and community distances ($r_M=0.04$, $p=0.06$), mantel correlograms over multiple distance classes, showed a small but significant effect of geographic distance on nearby lakes (~ 7.3 km.) (Figure S4). Indeed, nMDS axis 1 and 3 linearly correlated to geographic coordinates (Figure 5b), and in the variation partitioning analysis of the nMDS (Figure 5a), the proportion of variance explained by geography was very low (6.1%) but significant. The eastern and central Pyrenees reflect the overall Pyrenees trend, whereas the western Pyrenees showed a substantial increase in the proportion of variance explained by geography and associated chemical variables, reaching a total 75% of variation explained. Furthermore, we evaluated potential non-linear geographic relationships with alpha diversity considering geographic sub-regions (West, Central and East) and the distance to Pyrenees geographic center (Figure 6). We observed a relationship between the geographic distance to Pyrenees center and H' (Figure 6a), lacking with the two other alpha diversity metrics. This trend was significant towards the eastern region ($r_s=0.18$, $p=0.047$), but non-significant towards the more alkaline western region. Surprisingly, when evaluating the alpha diversity of zOTUs classified according to their EnvO terms, the trend towards the western region was significant in five EnvO terms (Table 2), while the trend in the western region was only significant for the term ‘Lake’. Regarding chemistry, although there was a trend towards alkaline lakes in the west and lakes with higher nutrients (DOC or TP, see Fig. S2) in the east, those factors were not correlated to alpha diversity, except for a weak positive relationship between DOC and PD (Figure 5b).

Regarding beta diversity in the sub-regions, there were significant differences (based on Kruskal-Wallis tests: Bray-Curtis: $H=7.26$, $p<0.05$; Raup-Crick: $H= 57.17$, $p<0.01$), with Central Pyrenees showing significantly higher dispersion than Western and Eastern Pyrenees (Figures 6b and 6c). Remarkably, betadispersion differences increased using the Raup-Crick metric, implying that alpha diversity differences are not the reason behind the betadispersion differences. When combining geography with EnvO terms into a two-way ANOVA, we observed significant differences between EnvO terms regarding beta diversity but the differences in beta diversity by geographic cluster within each EnvO term were lost. This indicates that it is required to consider the full pool of environmental sources to observe a difference between geographic clusters. The aquatic terms shared a significance group ('Epilimnion', 'Groundwater', 'River' and 'Lake') and those with solid particles involved shared a distinct significance group ('Sediment', 'Soil', 'Wetland' and 'Biofilm'), similarly to the beta diversity without considering geography (Figure 3c). Also, beta diversity per altitude ranges (below 2000, 2000-2500, above 2500) had no significant differences.

Discussion

Understanding biological community assembly is a complex issue in ecology that involves several processes, such as environmental selection and dispersal among others [1]. Inland water bodies offer good opportunities to approach this complexity for the microbial world [4, 9]. The assembly process is even worse understood at the regional scale and there is a lack of studies integrating multiple regional processes in the microbial assembly process, with few examples [28].

The role of hydrographic niches in lake communities

The results that we report here point to a critical relevance of dispersal from multiple hydrographic niches on the structure of the local bacterioplankton communities in the Pyrenees mountain range. Although water fluxes through the catchment are known to export solid particles passively, movement and origin of bacteria has traditionally been hard to infer [10]. Only more recently, there have been comprehensive attempts to uncover these fluxes, mainly regarding soils [9, 29–31]. Indeed, tracking dispersal processes based on the potential source populations [32] allowed us to identify bacteria related to soils and sediments, but also to rivers, groundwater and biofilms. Obviously, populations from lakes and particularly the epilimnion layer were among the most abundant EnvO terms of the Pyrenean surface bacterioplankton. Phyla containing zOTUs with higher aquatic affinities, such as Cyanobacteria, Verrucomicrobia, Actinobacteria, and Bacteroidetes may represent autochthonous bacteria [16, 33] whereas the remarkable proportion of soil-sediment affinities ascribed to Parcubacteria, Gammaproteobacteria, Alphaproteobacteria and Betaproteobacteria zOTUs may represent allochthonous ‘seed’ bacteria, some of them known to be recruited through the hydrographic network and later thriving in lakes [9, 12, 29]. As expected, habitat preference of bacterial phyla shows a certain degree of coherence [34], but also some habitat diversification. There are specific populations within the latter groups well adapted to the planktonic realm, e.g., the Alphaproteobacterial LD12 lineage [35], and the Betaproteobacterial genera *Polynucleobacter* and *Limnohabitans* [36, 37], among others.

Furthermore, bacteria related to rivers and groundwater could be indicative of the mass effects on water movement through the hydrographic surface network [17] and the sub-surface reservoirs [38], respectively. Biofilms as a source may be directly related to

particle-associated bacteria transported by the hydrological system [39] or to ‘floc’ formation that has been observed in the neustonic fraction of the lakes [40]. These may also be free-living stages of epilithic biofilms, known to have a large diversity and an active role in Pyrenean oligotrophic lakes [41]).

Environmental filtering of specific environmental sources

Environmental sources explain better the community heterogeneity than chemical and geographic variables combined. The potential immigration effect from rivers, wetlands, and biofilms indicates watershed fluxes that play a relevant dynamic role in community structuration. In fact, a temporal study has shown that the resident community strongly determines future communities regardless of further dispersal sources, which may behave in a transient manner [31]. The highly diluted nature of the Pyrenean lakes, in turn, makes that river and soil sources significantly enhance richness, in detriment of true lake bacteria proportion. Furthermore, our results showed that metacommunity beta dispersion was lowest in bacteria from some aquatic sources (rivers and groundwaters) while Shannon diversity was similar between aquatic and terrestrial sources. The Raup-Crick dissimilarities were homogeneous among sources, indicating that their beta-diversities may be primarily driven by its contribution in richness [43], and revealing that all sources may be similarly important when contributing to lake community structures.

In addition, how these fluxes are related with chemical filtering was previously unclear. Potential immigrant bacteria are sorted by the local environmental filtering to finally colonize the lakes showing major effects of pH and of the NO₃-DOC relationship. Bacteria specialized to the epilimnion were favored by alkaline pH and the nutrient related axis (DOC, NO₃ and TP), in detriment of those from the sediment. The environmental filter by pH associated to catchment lithology has been recurrently reported in the area [16, 44], and in additional studies in the literature such as lake sediments [45] streams or rivers [9] and lakes [9, 17, 46]. Similarly, influence of nutrients has also been widely reported on community structures and alpine lake processes in the area [16, 18, 47], however the magnitude of these variables structuring the whole bacterial metacommunity was substantially lower than expected, given previous observations in eukaryotic communities [23] or at the catchment scale [16]. Since we could not explain a fraction of the metacommunity variation (particularly in

central and eastern Pyrenees), we assume that additional unmeasured variables may be affecting the high mountain lake surface communities, such as UV radiation [13], wind effects on flocc proliferation [40], the long-term recurrence of aerial depositions [48, 49], or biotic relationships with eukaryotic plankton [50, 51]. In boreal systems it has been shown the influence of hydrology [17] in community composition, although in the Pyrenees alpine system we did not observe effects by any metrics of water turnover such as water renewal time (WRT) or lake size.

Spatial fingerprints in the regional metacommunity

Local communities are subject to all the factors and processes already mentioned, but within the whole region, historical processes such as past environmental selection and initial species sorting may have long-term effects on community structure [52], which likely exhibits biogeographic patterns [3]. In general, we observed a negative relationship between Shannon diversity and beta dispersion, potentially indicating a process of community convergence, potentially derived from ecological succession [53–55]. Furthermore, when we combined in the geographic context alpha and beta diversity metrics, we observed the same association. In general, when communities were more similar (low beta diversity), they had higher alpha diversity; and along sub-regions, we observed lower alpha diversity and higher beta diversity in the central Pyrenees than in the eastern and western parts. The higher beta-diversity in Central Pyrenees indicates a greater strength of stochastic processes than in the Pyrenees extremes, with an enhanced determinism. Given that the pattern overlap between pH and geography solely happens in the western Pyrenees, and that beta diversity did not change neither across pH nor altitude ranges, solely these factors do not adequately represent all the historical processes that may have taken place, although it may do so at the catchment level [14, 56]. Instead, founder effects of zooplankton populations and a trend of their alpha diversity towards the eastern Pyrenees has been already detected [57].

Furthermore, since the lake surface is connected to surrounding soils, it makes sense to infer that at least a fraction of aquatic bacterial populations may also respond to soil community biogeography, where palaeoclimate can explain a unique proportion of community heterogeneity [27]. Therefore, the freshwater regional diversity pattern observed could respond to a combination of historical processes such as niche dispersal, priority effects [9], present and former environmental conditions [52] ultimately shaping the ecological succession process [53].

382

383 **Conclusions**

384 In summary, our results show that the processes driving community assembly in high
385 mountain lakes are primarily related to the selective habitat dispersal of environmental
386 sources and secondarily to the environmental filtering of a wide repertoire of
387 environmental sources of bacterial immigrants. We found chemical selection towards
388 specific epilimnetic bacteria and a significant impact of sediment, river, wetland,
389 groundwater and soil populations on local community structures. Despite a minor effect
390 of geography on community assembly, our analysis of alpha and beta diversity indicates
391 that the complex interactions of environmental source dispersal and environmental
392 filtering leaves a spatial signature in the bacterioplankton along the mountain range
393 geography.

394

395

396

Experimental methods

Sampling and environmental variables

The dataset (304 lakes analyzed) contained both alpine and montane Pyrenean lakes that were sampled in summer 2011 along a spatial gradient ranging from longitudes 0°42'21.6" W to 2°27'45.9" E and latitudes 42°25'10.0" N to 42°56'06.4" N (see Figure S1 for a detailed map and altitudinal distributions, and Table S1). The dataset covered heterogeneous lakes in their physico-chemical properties and of different sizes (range 0.1-56.9 ha) and altitudinal gradient distributions (range 1459-2990 meters above sea level) (Table S2). Higher lake alkalinity has been reported in the west part of the Pyrenees mountain range [44, 58]. The Pyrenean alpine catchments are composed of bare rock, scree, and alpine meadows, with 40% soil coverage on average. The predominant basin bedrock is plutonic rock (48%). Other lithologies present are metamorphic rocks (schist and slate, 27%), limestones (18%), and detrital rocks (18%). Pyrenean lakes are typically very poor in nutrients [18, 59]. Surface water samples were directly collected close to the outflow of the lakes, sampling c. 200 mL in 0.22-micron Sterivex pressure driven sterile filter units (Millipore R), pre-filtered through a pore size of 50 µm, avoiding disturbance of the littoral zone [24]. Sterivex cartridges were filled in situ with lysis buffer, and DNA was extracted in the laboratory as previously reported [60]. We measured 30 chemical and geomorphological variables: lake area, catchment area, lake depth, lake volume, water renewal time (WRT), altitude, longitude, latitude, pH, alkalinity (ALK), conductivity (COND), dissolved inorganic carbon (DIC), dissolved organic carbon (DOC), nitrites (NO₂), nitrates (NO₃), ammonium (NH₄), dissolved organic nitrogen (DON), total dissolved nitrogen (TDN), dissolved organic phosphorus (DOP), total dissolved phosphorus (TDP), total phosphorus (TP), soluble reactive phosphorus (SRP), particulate Phosphorus (PP), dissolved reactive silica (DRSi), chlorides (Cl), sulphate (SO₄), potassium (K), calcium (Ca), magnesium (Mg) and sodium (Na). Analytical methods were carried out as reported [59]).

16S rRNA gene amplification and sequence analyses

The V4 loop was amplified using the universal primer pair 515f and 806r [61] that is reported to cover most bacteria and archaea with a few known biases [24]. Amplicon libraries were prepared through dual-index sequencing strategy [62] to be sequenced through the MiSeq 2x250 technology. Samples were prepared and sequenced in the

genomics core of the RTSF-MSU (Michigan State University). Illumina paired-end raw data was processed using UPARSE v. 9.2.64 [63]. In brief, 9024683 bacterial 16S rRNA gene partial sequences were merged allowing a maximum of two ambiguities. Quality filtering step kept 6716745 merged sequences of 253 bp and with an expected error higher than 0.5. The UNOISE algorithm [64] recovered a total of 13385 zero-radius OTUs (zOTUs, i.e., 100% identity) after de-noising and chimera filtering under default parameters. This approach improves resolution of OTU clustering at 97% identity by accurately retrieving unique 16S rRNA gene clean sequences, which may hold higher ecological resolution. These zOTUs were taxonomically classified online using the SILVA-NGS pipeline [65] with the Silva 128 database using default parameters. We recovered 5563074 sequences, which clustered into 11804 zOTUs after removal of chloroplasts and mitochondria. Raw sequences are available in the NCBI SRA archive under the BioProject number PRJNA413654, and a detailed list of accessions with additional geographic information is shown in Table S1.

The descriptive environmental terms (environmental sources) were approached through the SeqEnv pipeline [66]. The SeqEnv pipeline selects up to one hundred similar sequences to zOTUs (we set the threshold to 98.5% identity) from public database repositories (NCBI Genbank), and retrieves environmental ontology (EnvO) terms [67] available for these sequences for each zOTU. SeqEnv results retrieve the probability of finding a zOTU in one or multiple environments given previously reported sequences. A total of 409 EnvO terms were associated to the most abundant zOTUs, corresponding on average to the 87.4% (standard deviation 0.06) of reads per sample. Weighted metrics of environmental sources were obtained after correcting for relative abundances (i.e., sum of relative abundances of all zOTUs with the ‘Soil’ annotation; weighted by the probability of finding that zOTU in ‘Soil’). A single zOTU can therefore contribute to multiple environmental sources. To further characterize each environmental source, we split zOTUs per environmental source annotation into smaller community matrices.

Statistical analyses

Statistical analyses were run in the R environment [68] with functions from packages ‘vegan’ [69], ‘agricolae’ [70], and ‘lfe’ [71]. Results were plotted with ‘ggplot2’ [72].

Alpha diversity metrics (Richness, Shannon, and Phylogenetic diversity) were estimated after rarefaction to 5500 sequences per sample, and to 2000 sequences per sample for the EnvO term communities. Shannon index (H') was calculated with the function *diversity*, and Phylogenetic diversity (Weighted Faith's PD index) with function *weighted_faith* after aligning sequences and reconstructing a phylogenetic tree using the FastTree maximum-likelihood algorithm[73]. Regional microbial communities were fitted under a non-metric multidimensional scaling (nMDS) based on Bray-Curtis dissimilarities calculated with function *vegdist*. To test whether the communities were deterministically or stochastically selected, we compared community subsets according to Raup-Crick dissimilarities, which compare sample differences towards a null model using presence/absence data using the *raupcrick* function (it also allows to measure strict beta-diversity, avoiding the influence of alpha diversity [43]). We quantified beta diversity per geographic region, altitude, pH ranges, H' groups, and by EnvO term communities using the function *betadisper*. Spearman correlations (r_s) and Mantel tests (r_M) were carried out to assess relationships between microbiological parameters and chemical, morphometric, geographical variables, and estimated EnvO terms (weighted per lake). We further quantified the impact of these 4 sets of variables on the community structure through a variance partitioning analysis (function *varpart*), with a prior forward selection of variables avoiding co-linearity. Finally, we carried out a comprehensive integration of variable correlations through network analysis in *gephi* [74]. We included 76 correlations with an $r_s > |0.15|$, at $p < 0.05$ threshold, to assess dispersal and filtering. For that purpose, we excluded interactions within variable categories (that is, nMDS axis, diversity metrics, environmental variables and sources). We tested for overall significant differences in categories using Kruskal Wallis test, and further post-hoc multiple comparison tests if data lacked normality. For normal distributions, ANOVA with a post-hoc Tukey HSD was used.

Acknowledgements. This research was funded by Grants BRIDGES CGL2015–69043-P from the Spanish Office of Science (MINECO) and European funding (ERDF) and DISPERSAL (Ref. 829/2013) from the Spanish National Parks research program (OAPN-MAGRAMA), to EOC. We thank technicians and students who participated in the field sampling expeditions. ROA was supported by the FPI fellowship programme

497 from the Spanish Government for project DARKNESS CGL2012-32747 to EOC. The
498 authors have not conflict of interest to declare.

499 **References**

- 500 1. Vellend M. Conceptual synthesis in community ecology. *Q Rev Biol* 2010; **85**: 183–206.
- 501 2. Barberán A, Casamayor EO, Fierer N. The microbial contribution to macroecology.
- 502 *Front Microbiol* 2014; **5**: 1–8.
- 503 3. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al.
- 504 Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 2006; **4**:
- 505 102–112.
- 506 4. Lindström ES, Langenheder S. Local and regional factors influencing bacterial
- 507 community assembly. *Environ Microbiol Rep* . 2012. , **4**: 1–9
- 508 5. Tranvik LJ, Downing JA, Cotner JB, Loiselle SA, Striegl RG, Ballatore TJ, et al. Lakes
- 509 and reservoirs as regulators of carbon cycling and climate. *Limnol Oceanogr* 2009; **54**:
- 510 2298–2314.
- 511 6. Barberán A, Casamayor EO. Global phylogenetic community structure and β -diversity
- 512 patterns in surface bacterioplankton metacommunities. *Aquat Microb Ecol* 2010; **59**: 1–10.
- 513 7. Pace ML, Cole JJ, Carpenter SR, Kitchell JF, Hodgson JR, Van de Bogert MC, et al.
- 514 Whole-lake carbon-13 additions reveal terrestrial support of aquatic food webs. *Nature*
- 515 2004; **427**: 240–243.
- 516 8. Felip M, Wille A, Sattler B, Psenner R. Microbial communities in the winter cover and
- 517 the water column of an alpine lake: system connectivity and uncoupling. *Aquat Microb*
- 518 *Ecol* 2002; **29**: 123–134.
- 519 9. Ruiz-González C, Niño-García JP, del Giorgio PA. Terrestrial origin of bacterial
- 520 communities in complex boreal freshwater networks. *Ecol Lett* 2015; **18**: 1198–1206.
- 521 10. Felip M, Bartumeus F, Halac S, Catalan J. Microbial plankton assemblages, composition
- 522 and biomass, during two ice-free periods in a deep high mountain lake (Estany Redó,
- 523 Pyrenees). *J Limnol* 1999; **58**: 193–202.
- 524 11. Catalan J, Camarero L, Felip M, Pla S, Ventura M, Buchaca T, et al. High mountain
- 525 lakes: extreme habitats and witnesses of environmental changes. *Limnetica* 2006; **64**: 123–
- 526 145.
- 527 12. Monard C, Gantner S, Bertilsson S, Hallin S, Stenlid J, Torsvik V, et al. Habitat
- 528 generalists and specialists in microbial communities across a terrestrial-freshwater
- 529 gradient. *Sci Rep* 2016; **6**: 37719.
- 530 13. Laurion I, Ventura M, Catalan J, Psenner R, Sommaruga R. Attenuation of ultraviolet
- 531 radiation in mountain lakes: Factors controlling the among- and within-lake variability.
- 532 *Limnol Oceanogr* 2000; **45**: 1274–1288.
- 533 14. Ciccazzo S, Esposito A, Borruso L, Brusetti L. Microbial communities and primary
- 534 succession in high altitude mountain environments. *Ann Microbiol* 2016; **66**: 43–60.
- 535 15. Felip M, Camarero L, Catalan J. Temporal changes of microbial assemblages in the ice
- 536 and snow cover of a high mountain lake. *Limnol Oceanogr* 1999; **44**: 973–987.
- 537 16. Barberán A, Casamayor EO. A phylogenetic perspective on species diversity, β -diversity
- 538 and biogeography for the microbial world. *Mol Ecol* 2014; **23**: 5868–76.
- 539 17. Niño-García JP, Ruiz-González C, del Giorgio PA. Interactions between hydrology and
- 540 water chemistry shape bacterioplankton biogeography across boreal freshwater networks.
- 541 *ISME J* 2016; 1–12.
- 542 18. Camarero L, Catalan J. Atmospheric phosphorus deposition may cause lakes to revert
- 543 from phosphorus limitation back to nitrogen limitation. *Nat Commun* 2012; **3**: 1118.
- 544 19. Grossmann L, Jensen M, Pandey RV, Jost S, Bass D, Psenner R, et al. Molecular
- 545 investigation of protistan diversity along an elevation transect of alpine lakes. *Aquat Microb*

- 546 *Ecol* 2016; **78**: 25–37.
- 547 20. Sommaruga R, Casamayor EO. Bacterial ‘cosmopolitanism’ and importance of local
548 environmental factors for community composition in remote high-altitude lakes. *Freshw*
549 *Biol* 2009; **54**: 994–1005.
- 550 21. Reche I, Pulido-Villena E, Morales-Baquero R, Casamayor EO. Does ecosystem size
551 determine aquatic bacterial richness? *Ecology* 2005; **86**: 1715–1722.
- 552 22. Catalan J, Barbieri MG, Bartumeus F, Bitušík P, Botev I, Brancelj A, et al. Ecological
553 thresholds in European alpine lakes. *Freshw Biol* 2009; **54**: 2494–2517.
- 554 23. Triadó-Margarit X, Casamayor EO. Genetic diversity of planktonic eukaryotes in high
555 mountain lakes (Central Pyrenees, Spain). *Environ Microbiol* 2012; **14**: 2445–56.
- 556 24. Ortiz-Alvarez R, Casamayor EO. High occurrence of Paccarchaeota and
557 Woese archaea (Archaea superphylum DPANN) in the surface waters of oligotrophic
558 high-altitude lakes. *Environ Microbiol Rep* 2016; **8**: 210–217.
- 559 25. Ortiz-Álvarez R, Triadó-Margarit X, Camarero L, Casamayor EO, Catalan J. High
560 planktonic diversity in mountain lakes contains similar contributions of autotrophic,
561 heterotrophic and parasitic eukaryotic life forms. *Sci Rep* 2018; **8**: 4457.
- 562 26. Delmas M, Gunnell Y, Braucher R, Calvet M, Bourlès D. Exposure age chronology of
563 the last glaciation in the eastern Pyrenees. *Quat Res* 2008; **69**: 231–241.
- 564 27. Delgado-Baquerizo M, Bissett A, Eldridge DJ, Maestre FT, He JZ, Wang JT, et al.
565 Palaeoclimate explains a unique proportion of the global variation in soil bacterial
566 communities. *Nat Ecol Evol* 2017; **1**: 1339–1347.
- 567 28. Souffreau C, Van der Gucht K, van Gremberghe I, Kosten S, Lacerot G, Lobão LM, et
568 al. Environmental rather than spatial factors structure bacterioplankton communities in
569 shallow lakes along a >6000 km latitudinal gradient in South America. *Environ Microbiol*
570 2015; **17**: 2336–2351.
- 571 29. Crump BC, Amaral-Zettler LA, Kling GW. Microbial diversity in arctic freshwaters is
572 structured by inoculation of microbes from soils. *ISME J* 2012; **6**: 1629–1639.
- 573 30. Hauptmann AL, Markussen TN, Stibal M, Olsen NS, Elberling B, Børum J, et al.
574 Upstream freshwater and terrestrial sources are differentially reflected in the bacterial
575 community structure along a small Arctic river and its estuary. *Front Microbiol* 2016; **7**: 1–
576 16.
- 577 31. Comte J, Langenheder S, Berga M, Lindström ES. Contribution of different dispersal
578 sources to the metabolic response of lake bacterioplankton following a salinity change.
579 *Environ Microbiol* 2016; **19**: 1–28.
- 580 32. Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, et al. Bacterial diversity
581 along a 2600 km river continuum. *Environ Microbiol* 2015; **17**: 4994–5007.
- 582 33. Niño-García JP, Ruiz-González C, del Giorgio PA. Landscape-scale spatial abundance
583 distributions discriminate core from random components of boreal lake
584 bacterioplankton. *Ecol Lett* 2016; **19**: 1506–1515.
- 585 34. Philippot L, Andersson SGE, Battin TJ, Prosser JM, Schimel JP, Whitman WB, et al. The
586 ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol* 2010; **8**: 523–
587 529.
- 588 35. Henson MW, Lanclos VC, Faircloth BC, Thrash JC. Cultivation and genomics of the
589 first freshwater SAR11 (LD12) isolate. *ISME J* 2018; **12**: 1846–1860.
- 590 36. Hahn MW. Isolation of strains belonging to the cosmopolitan Polynucleobacter
591 necessarius cluster from freshwater habitats located in three climatic zones. *Appl Environ*
592 *Microbiol* 2003; **69**: 5248–5254.
- 593 37. Hahn MW, Kasalický V, Jezbera J, Brandt U, Jezberová J, Šimek K. Limnhabitans

- 594 curvus gen. nov., sp. nov., a planktonic bacterium isolated from a freshwater lake. *Int J*
595 *Syst Evol Microbiol* 2010; **60**: 1358–1365.
- 596 38. Winter TC. Relation of streams, lakes, and wetlands to groundwater flow systems.
597 *Hydrogeol J* 1999; **7**: 28–45.
- 598 39. Zwart G, Van Hannen EJ, Kamst-van Agterveld MP, Van der Gucht K, Lindström ES,
599 Van Wichelen J, et al. Rapid Screening for Freshwater Bacterial Groups by Using
600 Reverse Line Blot Hybridization? *Appl Environ Microbiol* 2003; **69**: 5875–5883.
- 601 40. Drudge CN, Warren LA. Diurnal floc generation from neuston biofilms in two
602 contrasting freshwater lakes. *Environ Sci Technol* 2014; **48**: 10107–10115.
- 603 41. Bartrons M, Catalan J, Casamayor EO. High Bacterial Diversity in Epilithic Biofilms of
604 Oligotrophic Mountain Lakes. *Microb Ecol* 2012; **64**: 860–869.
- 605 42. Vila-Costa M, Bartrons M, Catalan J, Casamayor EO. Nitrogen-Cycling Genes in
606 Epilithic Biofilms of Oligotrophic High-Altitude Lakes (Central Pyrenees, Spain). *Microb*
607 *Ecol* 2014; **68**: 60–69.
- 608 43. Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, et al.
609 Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist.
610 *Ecol Lett* 2011; **14**: 19–28.
- 611 44. Catalan J, Ballesteros E, Gacia E, Palau A, Camarero L. Chemical composition of
612 disturbed and undisturbed high-mountain lakes in the Pyrenees - a reference for acidified
613 sites. *Water Res* 1993; **27**: 133–141.
- 614 45. Zeng J, Zhao D, Li H, Huang R, Wang J, Wu QL. A monotonically declining
615 elevational pattern of bacterial diversity in freshwater lake sediments. *Environ Microbiol*
616 2016; **18**: 5175–5186.
- 617 46. Li H, Zeng J, Ren L, Wang J, Xing P, Wu QL. Contrasting patterns of diversity of
618 abundant and rare bacterioplankton in freshwater lakes along an elevation gradient.
619 *Limnol Oceanogr* 2017.
- 620 47. Buchaca T, Felip M, Catalan J. A comparison of HPLC pigment analyses and
621 biovolume estimates of phytoplankton groups in an oligotrophic lake. *J Plankton Res* 2005;
622 **27**: 91–101.
- 623 48. Cáliz J, Triadó-Margarit X, Camarero L, Casamayor EO. A long-term survey unveils
624 strong seasonal patterns in the airborne microbiome coupled to general and regional
625 atmospheric circulations. *Proc Natl Acad Sci* 2018; **115**: 12229–12234.
- 626 49. Barberán A, Henley J, Fierer N, Casamayor EO. Structure, inter-annual recurrence, and
627 global-scale connectivity of airborne microbial communities. *Sci Total Environ* 2014; **487**:
628 187–95.
- 629 50. Šimek K, Vrba J, Pernthaler J, Posch T, Hartman P, Nedoma J, et al. Morphological
630 and compositional shifts in an experimental bacterial community influenced by protists
631 with contrasting feeding modes. *Appl Environ Microbiol* 1997; **63**: 587–595.
- 632 51. Šimek K, Nedoma J, Znachor P, Kasalický V, Jezbera J, Hornňák K, et al. A finely
633 tuned symphony of factors modulates the microbial food web of a freshwater reservoir in
634 spring. *Limnol Oceanogr* 2014; **59**: 1477–1492.
- 635 52. Andersson MGI, Berga M, Lindstrom ES, Langenheder S. The spatial structure of
636 bacterial communities is influenced by historical environmental conditions. *Ecology* 2014;
637 **95**: 1134–1140.
- 638 53. Ortiz-Álvarez R, Fierer N, De Los Ríos A, Casamayor EO, Barberán A. Consistent
639 changes in the taxonomic structure and functional attributes of bacterial communities
640 during primary succession. *ISME J* 2018; **12**: 1658–1667.
- 641 54. Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. Disentangling mechanisms that

642 mediate the balance between stochastic and deterministic processes in microbial
643 succession. *Proc Natl Acad Sci* 2015; **112**: E1326–E1332.

644 55. Ortiz-Álvarez R, Fierer N, de los Ríos A, Casamayor EO, Barberán A. Consistent
645 changes in the taxonomic structure and functional attributes of bacterial communities
646 during primary succession. *ISME J* 2018; **12**: 1658–1667.

647 56. Wang J, Meier S, Soininen J, Casamayor EO, Pan F, Tang X, et al. Regional and global
648 elevational patterns of microbial species richness and evenness. *Ecography (Cop)* 2016;
649 393–402.

650 57. Ventura M, Petrusek A, Miró A, Hamrová E, Buñay D, De Meester L, et al. Local and
651 regional founder effects in lake zooplankton persist after thousands of years despite high
652 dispersal potential. *Mol Ecol* 2014; **23**: 1014–1027.

653 58. Camarero L, Catalan J, Boggero A, Marchetto A, Mosello R, Psenner R. Acidification in
654 high mountain lakes in central, southwest, and southeast Europe (Alps, Pyrennees, Pirin).
655 *Limnologica* 1995; **25**.

656 59. Camarero L, Rogora M, Mosello R, Anderson NJ, Barbieri A, Botev I, et al.
657 Regionalisation of chemical variability in European mountain lakes. *Freshw Biol* 2009;
658 **54**: 2452–2469.

659 60. Hervas A, Casamayor EO. High similarity between bacterioneuston and airborne
660 bacterial community compositions in a high mountain lake area. *FEMS Microbiol Ecol*
661 2009; **67**: 219–228.

662 61. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et
663 al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.
664 *Proc Natl Acad Sci* 2011; **108**: 4516–4522.

665 62. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a
666 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence
667 data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; **79**: 5112–
668 20.

669 63. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads.
670 *Nat Methods* 2013; **10**: 996–8.

671 64. Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
672 sequencing. *bioRxiv* 2016.

673 65. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of
674 general 16S ribosomal RNA gene PCR primers for classical and next-generation
675 sequencing-based diversity studies. *Nucleic Acids Res* 2013; **41**: e1.

676 66. Sinclair L, Ijaz UZ, Jensen L, Coolen MJ, Gubry-Rangin C, Chroňáková A, et al.
677 Seqenv: linking sequences to environments through text mining. 2016; **4**: e2690.

678 67. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE. The environment ontology:
679 contextualising biological and biomedical entities. *J Biomed Semantics* 2013; **4**: 43.

680 68. R Core Team. R: A Language and Environment for Statistical Computing. *R Found Stat*
681 *Comput Vienna, Austria* . 2017. , **0**: {ISBN} 3-900051-07-0

682 69. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan:
683 Community Ecology Package. *R Packag version 2.4-5* 2017.

684 70. De Mendiburu F. Agricolae: Statistical procedures for agricultural research. R package
685 version 1.2-0. <http://CRAN.R-project.org/package=agricolae> 2014.

686 71. Swenson NG. Functional and Phylogenetic Ecology in R. *Use R!* . 2014.

687 72. Wickham H. ggplot 2: Elagant graphics for data analysis. *Media* . 2016.

688 73. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees

689 for large alignments. *PLoS One* 2010.
690 74. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and
691 Manipulating Networks. *Int AAAI Conf Weblogs Soc Media; Third Int AAAI Conf Weblogs Soc*
692 *Media* 2009.
693

Legends

Table 1. Summary of principal components analysis with rotation ‘varimax’. Report of the weights of variables associated to the rotated components and % of variation explained.

Table 2. Relationships of alpha diversity (H') measured in each environmental source and the geographic context (measured as distance to the Pyrenees center and east/west direction). Significance levels: ** ($p < 0.01$), * ($p < 0.05$).

Figure 1. Richness (number of zOTUs) and relative abundance (proportion of reads) relationships of bacterial taxa detected in the Pyrenean dataset. (a) Mean and SD of phyla among the different samples. (b) Richness and abundance of phyla aggregating the total sum of samples of the metacommunity.

Figure 2: (a) Non-metric multidimensional scaling (nMDS) ordination of lake bacterial communities based on Bray-Curtis dissimilarities colored by the strongest correlated variable (pH) and dot-shapes split considering two Shannon index groups (above and below the median value) (b) Comparison of the beta dispersion between communities with high and low H' diversity, and classified in five groups according to pH (0.5 intervals).

Figure 3. Means and standard deviations of (a) Richness and (b) Shannon index for the main EnvO terms after the intensive SeqEnv database search; (c) community betadispersion of subsampled EnvO communities using Bray-Curtis dissimilarities; (d) community betadispersion of subsampled EnvO communities using Raup-Crick dissimilarities. All metrics showed significant differences with ANOVA, and significance groups based on *post-hoc* HSD test are indicated.

Figure 4. Environmental term affiliation of zOTUs within most abundant bacterial taxa, sorted by the ‘Lake’ EnvO term.

Figure 5: The upper panel (a) shows the percentages of variation in the nMDS ordinations (full Pyrenees and separated western, central and eastern Pyrenees areas) explained by chemistry, EnvO sources and geography through variation partitioning. Not explained (residuals) and shared variation between the metadata types is also shown. (b) Network of correlations between nMDS axis, chemistry, geography, EnvO sources, diversity metrics, and RC components from the environmental PCA (gephi layout: Forceatlas2). Node colors indicate three different modules connected by links with different thickness (Spearman’s r) and color (red: negative, grey: positive). Node size represents the number of links of that node.

Figure 6: Relationships between Pyrenean geography and bacterial diversity: (a) Sliding means of alpha diversity (H') starting from the geographic centroid separating the gradient to the east and to the west. Linear models and loess functions were used to draw the geographic trends observed; (b) Community betadispersion (Bray-Curtis) by Pyrenean sub-regions; (c) Community betadispersion (Raup-Crick) by Pyrenean sub-regions.

Figure S1. (a) Location of study lakes across the Pyrenees. Lakes colored by its altitudinal location, (b) longitudinal profile of altitudes and pH.

Figure S2. Principal components analysis (PCA) performed with the 30 environmental variables (log-transformed and standardized) and the geographic coordinates of 320 lakes. The diagram describes the lineal combination of variables showing the main environmental gradients (components PC1 and PC2 represented 37% of the total environmental variance. With the third component, PC3, the analysis accumulates 49% of the environmental variance). The different colors indicate the geographic position.

Figure S3. The proportion of the dominant (a) and more rare (b) phyla in the Pyrenees. Lakes are sorted by the most abundant in each case (Bacteroidetes and Gammaproteobacteria respectively).

Figure S4. Mantel correlogram showing spatial correlation (Mantel r) for the bacterial assemblages using: (a) the spearman method and (b) the Raup-Crick method. Colored circles (blue positive, red negative) indicate significant spatial autocorrelation after progressive Bonferroni corrections ($p = 0.05$, 999 permutations).

Table S1. Sample codes for BioProject PRJNA413654, geographic position and name of lakes.

Table S2. Environmental descriptors of the Pyrenean dataset. Mean values, standard deviation (SD), maximum and minimum values for 30 chemical and morphometric variables.

Table. 1

Variable	RC1 (18%)	RC2 (16%)	RC3 (12%)
TDP	<i>0,852</i>	N.S	-0,205
TP	<i>0,84</i>	0,102	-0,215
DOP	<i>0,826</i>	0,142	-0,211
DOC	<i>0,729</i>	0,154	-0,18
DON	0,7	N.S	N.S
Altitude	-0,313	-0,245	N.S
ALC	N.S	<i>0,954</i>	N.S
DIC	N.S	<i>0,946</i>	N.S
Ca	N.S	<i>0,937</i>	0,106
COND	0,142	<i>0,892</i>	0,199
pH	N.S	<i>0,755</i>	-0,213
Mg	0,232	<i>0,517</i>	0,23
Longitude	0,324	-0,511	0,106
Fond	-0,144	N.S	<i>0,813</i>
Lake area	-0,144	N.S	<i>0,812</i>
Volume	N.S	N.S	<i>0,794</i>
RWT	-0,154	N.S	<i>0,718</i>

Table 2.

EnvO term	Central-East	Central-West
Lake	0.2**	0.17*
Epilimnion	0,13	0.28**
River	0,12	0.23**
Wetland	0,11	0.22**
Soil	-0,01	0.18*
Groundwater	0,04	0,16
Sediment	-0,06	0,16
Biofilm	0	-0,02

a. Lake variability

b. Regional pool

Latitude

Altitude (m)

a **b**