

## Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat

Pedro J. Cabello-Yeves,<sup>1</sup> Tamara I. Zemskaya,<sup>2</sup> Alexandra S. Zakharenko,<sup>2</sup> Mariya V. Sakirko,<sup>2</sup>  
Vyacheslav G. Ivanov,<sup>2</sup> Rohit Ghai,<sup>3</sup> Francisco Rodriguez-Valera<sup>1,4\*</sup>

<sup>1</sup>Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel, Hernández, San Juan de Alicante, Alicante, Spain

<sup>2</sup>Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

<sup>3</sup>Institute of Hydrobiology, Department of Aquatic Microbial Ecology, Biology Center of the Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic

<sup>4</sup>Moscow Institute of Physics and Technology, Dolgoprudny, Russia

### Abstract

Lake Baikal is the deepest lake in the world. Its depth provides the only bathypelagic (> 1000 m deep) freshwater habitat on Earth and its oxic, ultra-oligotrophic features make it a freshwater counterpart of the deep ocean. Here we have analyzed metagenomes from 1250 and 1350 m deep samples and built 231 metagenome-assembled genomes (MAGs). We detected high fractions of Thaumarchaeota (ca. 20% of 16S rRNA reads) and members of the candidate phyla radiation (CPR) (3–4.5%). Among the MAGs, we obtained ammonia-oxidizing archaea (AOA, Nitrosopumilaceae) and bacteria (AOB, Nitrosomonadaceae), and nitrite-oxidizers (Nitrospirae) indicating very active nitrification. A new clade of freshwater SAR202 Chloroflexi and methanotrophs (*Methyloglobulus*) were also remarkably abundant, the latter reflecting a possible role of methane oxidation as well. Novel species of streamlined and cosmopolitan bacteria such as *Ca. Fonsibacter* or *aci* Actinobacteria were more abundant at the surface but also present in deep waters. Conversely, CPRs, Myxococcales, Chloroflexi, DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota) archaea, or Gammaproteobacteria were found only in bathypelagic samples. We noted various important taxonomic and metabolic differences between deep aphotic region of Lake Baikal and marine waters of similar depth: Betaproteobacteriales, CPR, and DPANN superphylum were only found in bathypelagic Baikal, while Deltaproteobacteria, Gammaproteobacteria, or Alphaproteobacteria prevailed in oceanic samples. The genes mediating ammonia and methane oxidation, aromatic compound degradation, or alkane/methanesulfonate monooxygenases were detected in higher numbers in deep Baikal compared to their oceanic counterparts or its own surface. Overall, depth seems to be less relevant than salinity in configuring the microbial community.

Lake Baikal, located in an intracontinental rift zone of southcentral Siberia, is the world's deepest lake (1637 m,

average of 758 m), with the largest volume of fresh liquid water (23,000 km<sup>3</sup>) (Hampton et al. 2008). The lake is morphologically divided into three deep basins that are detached from each other by underwater elevations. The southern and central basins are the deepest reaching more than 1400 m depth (Galazy 1993). The presence of oxygen down to the bottom, a trait shared with the ocean but unique among deep lakes (> 800 m), explains the presence of multicellular life and the evolution of an extensive, mostly endemic fauna in profound depths of the lake (Kozhov 1962).

Hydrological factors influence the distribution and phylogenetic diversity of pelagic microorganisms in Lake Baikal (Likhoshway et al. 1996; Parfenova et al. 2000; Kurilkina et al. 2016). The rapid lowering of waters from the trophogenic layer favors the entry of small-size prokaryotic and eukaryotic organisms into the near-bottom and deep layers of the water column

\*Correspondence: frvalera@umh.es

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Additional Supporting Information may be found in the online version of this article.

**Author Contribution Statement:** F.R.-V., T.I.Z., and P.J.C.-Y. conceived the project. T.I.Z., A.S.Z., M.V.S., and V.G.I. took the samples and analyzed the metadata. P.J.C.-Y., F.R.-V., and R.G. performed the metagenomic analyses. F.R.-V., P.J.C.-Y., and T.I.Z. wrote the manuscript.

**Data Availability Statement:** All metagenomic data sets and microbial genomes used in this work are publicly available in the NCBI/SRA databases.

(Likhoshway et al. 1996; Parfenova et al. 2000). Small-celled organisms can move from the surface layers to the deep ones and vice versa due to a combination of multiple mechanisms that allow vertical mixing of waters (Shimaraev and Granin 1991; Weiss et al. 1991; Shimaraev et al. 2011, 2015). In addition, microorganisms can come with gas-containing fluids from sediments during earthquakes that are characteristic of the Baikal depression (Granin et al. 2018). Evidently, these processes can lead to the presence of microorganisms from the photic zone or even the sediment at various depths, influencing the content and functioning of microbiocenoses in the pelagic Lake Baikal. The composition of the particle fluxes studied using sedimentation traps showed the presence of diatom shells and organic matter of autochthonous origin at all depths of the water column. At depths of more than 1000 m, a slight decrease in the content and flow rates of total organic carbon and nitrogen (Vologina and Sturm 2017) was observed. The study of microbial communities in Lake Baikal with molecular techniques and sequencing has been mostly focused on their diversity and structure in the trophogenic layer (Zakharova et al. 2013; Mikhailov et al. 2015) and surface subice waters (Bashenkhaeva et al. 2015; Cabello-Yeves et al. 2018). Other studies from the summer period showed that bacterial communities in deep-water samples (700 and 1400 m) were mainly represented by Proteobacteria, Actinobacteria, Acidobacteria, and Chloroflexi (Glöckner et al. 2000; Bel'kova et al. 2003; Parfenova et al. 2006; Zakharova et al. 2013; Kurilkina et al. 2016). The dynamics of physicochemical conditions throughout the water column and their relative stability in the deep layers appeared to be decisive factors that shape the pattern of bacterial communities in Lake Baikal (Kurilkina et al. 2016).

In the last years, some freshwater lakes have been studied with high-throughput sequencing metagenomics and metagenome assembled genome (MAG) analysis (Oh et al. 2011; Eiler et al. 2014; Ghai et al. 2014; Cabello-Yeves et al. 2017b; Garcia et al. 2018; Linz et al. 2018; Tran et al. 2018), but all of them are less than 150 m deep. The availability of sequencing data on bacterial communities from deep freshwater lakes has been limited to amplicon sequencing (Glöckner et al. 2000; Urbach et al. 2001; De Wever et al. 2005; Kurilkina et al. 2016). In this sense, only Crater Lake (ca. 600 m of maximum depth) could be compared to Lake Baikal in terms of depth and extreme oligotrophy (both showing 40 m of Secchi Disk depth). New 16S rRNA phylogenetic groups from uncultured microbes such as CL500-11 Chloroflexi, CL120-10 Verrucomicrobiales, ACK4 actinomycetes, group I marine crenarchaeota, or uncultured and rare OP10, ABY1, and CL0-1 were described from Crater Lake (Urbach et al. 2001). Other lakes are comparable to Lake Baikal in terms of depth, mostly African Rift Valley lakes such as Tanganyika (maximum depth of 1470 m) and Malawi (maximum depth of 706 m), but are anoxic in their hypolimnion (De Wever et al. 2005) due to thermal stratification. Other relatively deep oligotrophic lakes such as O'Higgins-San Martín (Chile-Argentina) or Great Slave

and Quesnel (Canada) have maximum depths of about 600 m but their average depth is much less and thus far, there are no metagenomes derived from them.

A metagenomic study from the surface (5–20 m) subice waters of Lake Baikal revealed 35 novel MAGs, including the first Pelagibacteraceae member from the marine clade I to be found in a nonmarine related environment, although similar genomes were found before in brackish regions of the Baltic Sea (Hugerth et al. 2015). That study provided a first glimpse into the genomes present within the photic layer and revealed a certain degree of endemism and similarities of MAGs with some freshwater and brackish Baltic Sea microbes (Cabello-Yeves et al. 2018). Here, we have done a new study of the same lake during the winter period (subice samples) dealing with samples from bathypelagic depths (> 1000 m). We provide a set of 231 high-quality MAGs assembled from 1250 and 1350 m samples, and a short overview of the metabolic characteristics of the main taxa that inhabit them. We have compared the microbiome of both deep layers to the photic zone of Lake Baikal. We also compared the data to bathypelagic and mesopelagic (> 800 m depth) marine systems which resemble Lake Baikal in most physicochemical parameters such as depth, temperature, oxygen concentration, oligotrophy, or pH. Lake Baikal microbiota, even at these depths, resembled much more other lake hypolimnia, underscoring again the main divide that salinity introduces (akin to animals or plants) in prokaryotic microbes.

## Methods

### Metadata from 1250 and 1350 m samples, CTD (Conductivity, temperature and depth) profiles, sampling procedure

The water samples were taken on 29 March 2018, using 4-liter bathometers, at the station of the ice camp, which was located at 7 km from the Listvyanka settlement, at coordinates 51°50,910'N and 104°47,234'E. Ice thickness in the studied period was 70 cm thick and depth of the water column at the sampling site was 1405 m. The water samples were taken from two horizons of 1250 m and 1350 m, respectively. Measurements of the temperature profile throughout the water column were made using SBE 25 Sealogger CTD (Sea-Bird Electronics) accurate within 0.002°C and with a resolution of 0.0003°C (Supporting Information Fig. S1). Within a few hours, the samples (26–30 L) at a temperature of approximately +4°C were delivered to the laboratory. Then, each 26–30 L of water were filtered through a net (size 27 µm) and then filtered through nitrocellulose filter with a pore size of 0.22 µm (Millipore, France), and the material from the filter was transferred to sterile flasks with 20 mL of lysis buffer (40 mmol L<sup>-1</sup> EDTA, 50 mmol L<sup>-1</sup> Tris/HCl, 0.75 mol L<sup>-1</sup> sucrose) and stored at -20°C. As previously described (Cabello-Yeves et al. 2018), DNA was extracted according to the modified method of phenol-chloroform-isoamyl alcohol extraction and stored at -70°C until further use. The DNA samples were placed in a 70% alcohol solution

and were forwarded to the laboratory for sequencing. The different physicochemical profiles, hydrological conditions, and the mineralization and concentration of chemical components in the water column of the studied area corresponded to the data that were previously recorded during the ice period in Lake Baikal (Supporting Information Fig. S1, Text, and Methods) (Votintsev et al. 1975; Khodzher et al. 2016).

### Total cell counts

The total cell counts found in our samples ( $1.9 \times 10^6$  to  $0.17 \times 10^6$  cells mL<sup>-1</sup>) were in the normal range reported for Lake Baikal previously (Maksimova and Maksimov 1989). Its highest values were detected in the photic layer of 0–50 m in the near-bottom layer (Supporting Information Fig. S1).

### Sequencing, assembly, binning, and annotation

The two samples from 1250 and 1350 m depth were sequenced in an entire lane of Illumina HiSeq X Ten PE 2X150 bp (Novogene Company), which provided about 564 million reads for the 1250 m sample (average 53.64% of GC) and 480 million reads for the 1350 m sample (average 54.04% GC). A first cross-assembly of about 1 billion reads of both samples was done with IDBA-UD (Peng et al. 2012), which provided a total of 76,177 contigs > 5 Kb (average of 14,477 bp) and a total of 1102 Mb of output. Subassemblies of 20 million reads were done in both 1250 and 1350 to retrieve some of the most abundant microbes, such as *Ca. Fonsibacter*, *Methyloglobulus*, and Nitrospirae. These bacteria assembled poorly when combining both samples; hence, we reduced the total number of reads to obtain more complete bins of these representatives. Annotation of contigs and MAGs was assessed with BLAST (Nr), COG (Tatusov et al. 2001), TIGFRAM (Haft et al. 2001), tRNAscan (Lowe and Eddy 1997), ssu-align (Nawrocki and Eddy 2010), RAST (Overbeek et al. 2013), Kegg-KO (Kanehisa and Goto 2000), CDD-SPARCLE (Marchler-Bauer et al. 2016), and BLAST Koala (Kanehisa et al. 2016). A first manual inspection of bins was done assigning a hit (based on BLAST against Nr) to each CDS, which allowed us to classify each contig taxonomically into different phyla. Then, an initial binning step was applied for each set of contigs assigned to each phyla with METABAT2 (Kang et al. 2019). Afterward, further manual inspection of contigs was applied with GC, coverage, and tetranucleotide frequencies to refine the bins (Rice et al. 2000; Lê et al. 2008). Finally, we only used MAGs at < 5% contamination and > 50% of completeness estimated by CheckM package (Parks et al. 2015).

### Phylogenomic classification of MAGs

MAGs were classified according to the latest version of GTDB (Parks et al. 2018). Specific trees of Nitrospirae, Candidate Phyla Radiation (CPR), Latescibacteria, Omnitrophica, Thaumarchaeota, and Methylococcaceae were obtained with

this methodology. PhyloPhlan tool (Segata et al. 2013) was used to construct a tree of Chloroflexi MAGs.

### Comparison with other publicly available freshwater and marine metagenomes

All metagenomic data sets used in this work are referenced and are publicly available: Lake Baikal (Cabello-Yeves et al. 2018), Lake Zurich (Mehrshad et al. 2018b), Caspian Sea (Mehrshad et al. 2016), Mediterranean Sea (Haro-Moreno et al. 2018), and North Atlantic and Indian Ocean data sets (Sunagawa et al. 2015). The 16S rRNA gene reads from different metagenomes were obtained using ssu-align. Then, a BLASTN was performed against SILVA database (Quast et al. 2012) (SILVA\_132\_SSURef\_Nr99\_Tax\_silva from December 2018). Hierarchical cluster analysis (dendrograms) of different metagenomic samples with k-mer = 21 bp was assessed with SIMKA (Benoit 2015) and Bray-Curtis indexes of presence/absence were obtained. GC content distribution and whole-proteome isoelectric points (Rice et al. 2000) of all metagenomes were also obtained.

A subset of 10 million reads of each marine and freshwater compared metagenomes was analyzed with BLASTX against Nr database with Diamond (Buchfink et al. 2015), with options more-sensitive, max-target-seqs 1, e-value 0.00001, > 50 bp of alignment length and > 50% identity. The top hits were analyzed in search of specific genes. The ratio of recA/gene was assessed to estimate the abundance of different metabolic genes in each dataset and a Z-score was assessed to estimate the differences between data sets. RecA was used assuming that it was present as a single marker copy gene per genome.

### Recruitment and reads assignment of MAGs

Recruitment was assessed considering BLASTN hits of > 95% identity and 50 bp of alignment length as belonging to the same species. A microbe was considered present in a metagenome if it was detected at > 2 reads per Kb of genome per Gb of metagenome (RPKG). All the RPKG values for all MAGs retrieved from this work on Lake Baikal metagenomes are shown in Supporting Information Table S3. Recruitment plots that covered the core genome at > 95% of identity/50 bp alignment lengths were used as threshold for the species presence.

### Availability of metagenomes and MAGs

All objects derived from this work are publicly available under the NCBI Bioproject number PRJNA521725. The Lake Baikal 1250 m metagenome was under NCBI-SRA number SRR8561390 and the 1350 m sample under number SRR8561391. GenBank accession numbers for MAGs were assigned accordingly from SHUD00000000 to SICZ00000000.

## Results and discussion

### Metadata and general features of deep Lake Baikal

The concentration of chemical components in the water column of the studied area corresponded to the data that were

previously recorded during the ice period in Lake Baikal (Supporting Information Fig. S1 and Text) (Votintsev et al. 1975; Khodzher et al. 2016). Other deep lakes (like Tanganyika and Malawi) have an anoxic hypolimnion (De Wever et al. 2005), so in this sense, and with the obvious difference of salinity, deep Lake Baikal (dLB) is a waterbody that resembles more the ocean than other deep freshwater lakes. Actually, the changes of these parameters with depth follow the pattern of offshore oceanic waters (DeLong et al. 2006). However, aside from the salt concentrations, a significant difference was detected at the level of total cell counts in dLB that were higher (particularly at the deepest sample,  $0.4 \times 10^6$  cells mL<sup>-1</sup>) than is typical for other marine oligotrophic samples of equivalent depth at about  $1 \times 10^5$  cells mL<sup>-1</sup> (Liang et al. 2017; Haro-Moreno et al. 2018).

We have analyzed two samples (1250 and 1350 m depth), both taken during winter from the deep (subice) waters of Lake Baikal. These samples were still 57 and 157 m above the bottom. The GC content and the microbial population in both samples varied little, but there were a few differences that were already detected by the 16S rRNA fragments (see below). However, for MAG reconstruction, we used cross-assembly with a total of about 1 billion reads.

### Overall similarity to other aphotic freshwater and marine metagenomes

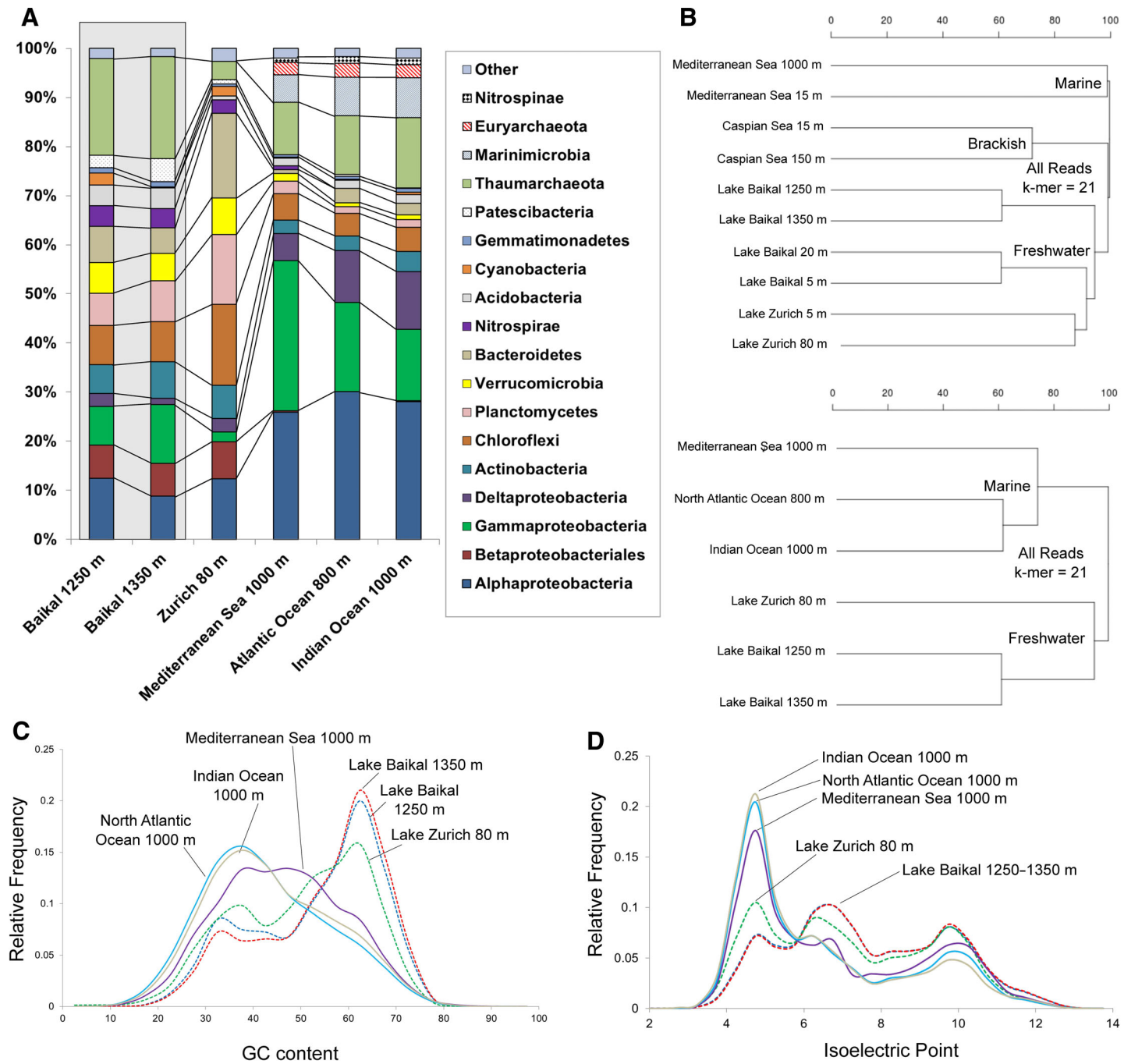
We first compared the 16S rRNA gene fragments (obtained from unassembled reads) from dLB, Lake Zurich and similar depth (ca. 1000 m) marine samples from different origins (Mediterranean Sea, North Atlantic, and Indian Ocean) to describe the large-scale taxonomic differences (Fig. 1A). The sample from the aphotic hypolimnion of Lake Zurich (aerobic but only 80 m deep) was used as a freshwater reference due to the scarcity of available metagenomes from deeper lakes. In addition, Lake Zurich has been extensively studied by both classical and metagenomic tools (Salcher et al. 2015; Neuenchwander et al. 2017; Mehrshad et al. 2018b; Andrei et al. 2019). As shown in Fig. 1A, classes Alphaproteobacteria and Deltaproteobacteria, and the phylum Gammaproteobacteria were at much higher proportions in deep ocean samples compared to dLB and Lake Zurich (at least two times higher in all cases). On the other hand, Bacteroidetes, Verrucomicrobia, Planctomycetes, Chloroflexi, and Actinobacteria were more abundant in both aphotic lake layers. Thaumarchaeota and Acidobacteria were present at higher percentages in dLB compared not only with oceanic data sets but also with Lake Zurich. Notably, some taxons appeared to be absent either from marine or freshwater samples. For instance, order Betaproteobacteriales, phylum Nitrospirae, and the recently described Patescibacteria (CPR) were only present significantly in dLB and Lake Zurich, while Euryarchaeota, Marinimicrobia, and Nitrospirinae, all present in deep marine ecosystems, appear to be absent from aphotic lake waters.

The comparison of raw reads (Fig. 1B) shows that deep marine datasets group together among themselves while the aphotic lakes group in a different branch. Another comparison that we carried out was the overall similarity of individual reads in metagenomes of aquatic habitats of different salinities and depths to gauge the relative effect of both factors. The results shown in the dendrograms of Fig. 1B indicate that salinity is the most influential factor in the similarity of the reads (microbes), that is, deep samples cluster with their corresponding surface sample and not by depth. As mentioned above, a small but significant difference could be found between the two dLB samples. Specifically Alphaproteobacteria were less represented in the deepest sample while CPR were twice more abundant there. Another notable difference was the presence of Cyanobacterial and Chloroplast reads (1%) in the 1250 m sample, likely deriving from surface waters by sedimentation or vertical mixing (Weiss et al. 1991; Shimaraev et al. 2011).

Another significant metagenomic parameter is the overall GC content that in marine waters has been shown to increase with depth (Mende et al. 2017). In the freshwater samples (even in the Lake Zurich sample that is only 80 m deep), there was the expected marked shift to higher GC values (Fig. 1C). Actually, the shift was more extreme in dLB and Lake Zurich, with a peak between 60% and 70%, while marine data sets of similar depth mean GC content was about 40%. A shift toward higher GC content in deeper waters is generally interpreted to be due to the change between an N-limited environment near the surface and an energy-limited one at aphotic depths. However, the GC content of surface samples of Lake Baikal was also higher (maximum peak at 55% GC, average of 44.6%) (Cabello-Yeves et al. 2018) than similarly oligotrophic waters in the North Pacific Ocean (34%) (Mende et al. 2017) or the Mediterranean Sea (38.6%) (Haro-Moreno et al. 2018). Therefore, there might be also a taxonomically driven change or nitrogen might be overall less limiting in freshwaters. Previously, we have shown that the whole pI of the predicted metaproteomes in freshwater systems is less acidic and shifted toward neutral and basic values compared to marine ones (Cabello-Yeves and Rodriguez-Valera 2019). Along these lines, peaks at neutral (6–7) and basic (from 8 to 10) pI values were more accentuated in Lake Zurich and dLB samples (Fig. 1C).

### Recovery of novel genomes from dLB

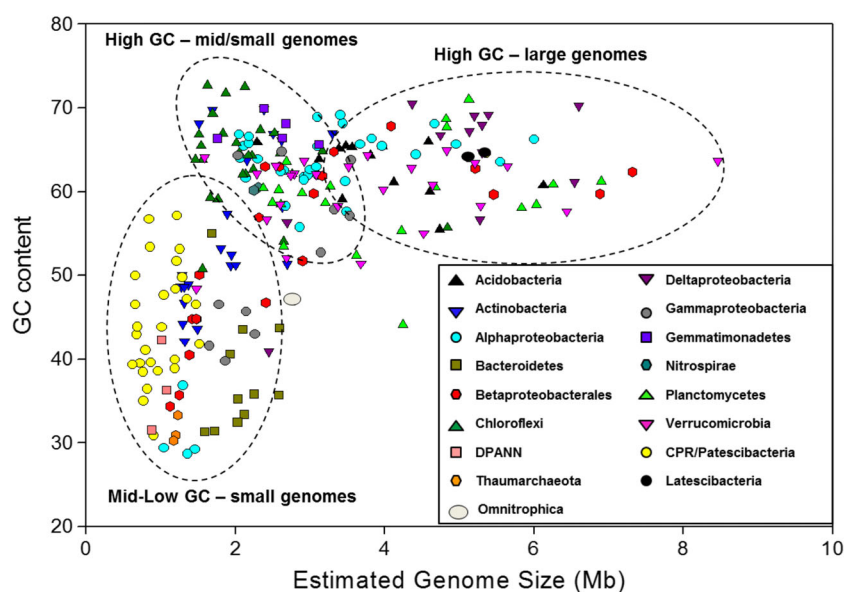
We assembled a total of 231 MAGs at > 50% of completeness and < 5% of contamination (Parks et al. 2015). A representation of the estimated genome size (Mb) vs. GC content of all of our MAGs is shown in Fig. 2. Their general characteristics and phylogenetic affiliations are shown in Table 1 and detailed individually for each MAG in Supporting Information Table S1, Figs. S2–S9, and Text. The overall representation of the MAGs was calculated by the amount of reads recruited by any of the genomes at the species threshold of > 95% identity (Goris et al. 2007). With this threshold, about 50% of the reads were recruited to the genomes described here, indicating



**Fig. 1.** Deep/aphotic freshwater (Lake Baikal, Lake Zurich) vs. marine (Indian Ocean, North Atlantic Ocean, Mediterranean Sea) data sets comparison. **(A)** 16S rRNA metagenomic read fragments classification in the different data sets. Each phylum is color-coded. **(B)** Dendrogram of metagenomic data sets (all reads) similarity based on 21 bp k-mers. Presence/absence Bray-Curtis index was used to assess similarity. **(C)** GC content vs. relative frequency of marine and freshwater aphotic/deep data sets. Dashed lines represent freshwater and continuous lines marine data sets. **(D)** Isoelectric point (pI) distribution of freshwater and marine deep/aphotic data sets. Dashed lines represent freshwater and marine data sets continuous lines.

that many (likely as much as half) of the microbial species present in the sample were recovered as high quality MAGs. In fact, when we used 85% and 80% as threshold (a range that might approximately define taxa at the level of genera), we matched 72% and 77% of reads, respectively. The reads that

did not map with any MAG were mostly associated with other species with similarities from 80% to 94.9% average nucleotide identity (ANI) to some of the 231 MAGs. These species consisted of small and streamlined microbes showing a high inter- and intraspecific diversity, such as Nitrosopumilaceae or



**Fig. 2.** Estimated genome size (Mb) vs. GC content of the 231 dLB MAGs.

*Ca. Fonsibacter* (Supporting Information Fig. S10). Other abundant MAGs (*Methyloglobulus* or Nitrospiraceae) showed lower inter and intraspecific diversity (Supporting Information Fig. S11). Other nonrecruited reads mapped to viruses (1647 contigs > 5 kb that were not included in this analysis) or other MAGs below the threshold of 50% of completeness. Among taxa that could not be assembled over this threshold were some Acidobacteria, Firmicutes, or Flavobacteriaceae.

Many of the novel taxa described here, such as CPR/Patescibacteria, DPANN superphylum Nitrospirae, Gemmatimonadetes, Acidobacteria, Deltaproteobacteria, or some unclassified members of Alpha/Betaproteobacteriales, have their highest identities with MAGs described from groundwater aquifers in the U.S.A. (Rifle) and deep subsurface sediments (Anantharaman et al. 2016; Hemsdorf et al. 2017), although the same taxa were never found at both environments, that is, those described in this work are novel representatives, at least at the level of genus. Nonetheless, having found these similarities with groundwater is not surprising since aquifers should be also dark, energy limited freshwater habitats. Still, from the environmental point of view, the closest habitat to the dLB is the deep ocean. Deep lakes tend to be anoxic due to the thermal stratification (De Wever et al. 2005). Even when oxic, the aphotic zone of other lakes (such as Lake Zurich), being much shallower, tends to be much more affected by the sedimentation of biomass from the upper layers, and this is mainly observed by the high presence of widespread surface microbes in deep layers such as *aci* Actinobacteria or *Limnohabitans* (Kasalický et al. 2013; Neuenchwander et al. 2017). The deep ocean, on the other hand, is oligotrophic, aerobic (particularly when there is no oxygen minimum zone) and has a temperature similar to dLB (close to

4°C). Therefore, the main difference is the nearly complete absence of marine salts in dLB. Furthermore, although as stated in the previous section, the major driver of the community structure seems to be salinity, there are groups of not very divergent microbes that are shared by marine/brackish waters and freshwaters. For instance, MAGs from surface Lake Baikal (Cabello-Yeves et al. 2018), especially Flavobacteriales, *aci* Actinobacteria, or SAR11 (Pelagibacteraceae) members (also retrieved in this work), have relatives (at ANI values < 85%) at the cold, eutrophic, and brackish Baltic Sea (Hugerth et al. 2015). Even more remarkable was the case of a *Nitrosopumilus* sp. MAG (G182) that was found in surface/deep Baikal, Lake Zurich and the Caspian (Mehrshad et al. 2016) and Baltic Seas (Hugerth et al. 2015). In fact, this MAG presented ANI values close to 99% to the MAG obtained from the Caspian Sea. In the case of the dLB, we have also found what appears to be the first members of a sister clade (with some representatives within) to the thus far typically marine deep dweller SAR202 (Mehrshad et al. 2018a) Chloroflexi cluster (Supporting Information Fig. S2). The existence of freshwater counterparts from these clades open up novel possibilities to find out the key elements that allow adaptation to live in freshwater for what are likely ancestral marine lineages and might contribute new perspectives to the evolutionary models of prokaryotic marine-freshwater transitions (Logares et al. 2009; Herlemann et al. 2011; Simon et al. 2017; Paver et al. 2018; Cabello-Yeves and Rodriguez-Valera 2019).

### Microbial dark matter in dLB

Members of the bacterial CPRs and DPANN archaeal superphylum are often considered “microbial dark matter” (Rinke et al. 2013) given the lack of information about their biology.

**Table 1.** General features of the bathypelagic Lake Baikal MAGs.

Phylum/class	No. of MAGs	Range estimated genome size (Mp)*	Range GC	Range median intergenic spacer (bp)	Average completeness	Average contamination	Taxonomic affiliation of MAGs (GTDB, referenced groups)
Alphaproteobacteria	33	1.04–6 (3.01)	28.7–69.2 (60.93)	4–64 (39.6)	74.82	1.23	1 <i>Ca. Fonsibacter</i> , 1 Pelagibacteraceae, 2 <i>Pseudolabrys</i> , 4 Rhodospirillaceae, 4 Rhodospirillales, 11 unclassified Alphaproteobacteria, 1 Rickettsiales, 4 Acetobacteraceae, 1 Xanthobacteraceae, 1 Elsteraceae, 3 Sneathielles
Chloroflexi	26	1.51–4.84 (2.15)	50.8–72.7 (64)	11–139 (76.67)	78.03	2.30	14 sister clade of SAR202, 1 Dehalococcoidia, 5 <i>Ca. Bathosphaeria</i> , 1 Umbricyclopiia, 1 Limnocylinidria, 4 unclassified Chloroflexi
Verrucomicrobia	26	1.48–8.46 (3.85)	48.4–64.9 (60.25)	26–111 (63.38)	71.41	2.08	14 Pedosphaerales, 4 <i>Opitutus</i> , 2 Opitutales, 3 Opitutaceae, 2 <i>Lacumisphaera</i> , 1 Verrucomicrobiae
CPR/Patescibacteria	25	0.66–1.52 (1)	30.9–57.1 (44.7)	12–93 (42.54)	68.47	1.52	1 Levybacteriales, 1 Nomurabacteraceae, 4 Taylorbacteraceae, 2 Paceibacteria, 1 Uhrbacterales, 2 Peribacteraceae, 2 Zambryskibacteraceae, 1 Ryanbacterales, 1 Margulisbacteria, 1 Saccharimonadia, 3 Kaiserbacteraceae, 1 Magasanikbacterales and 2 Staskawiczbacteriaceae, 2 unclassified Patescibacteria
Actinobacteria	22	1.27–3.3 (1.9)	42.1–69.8 (55.43)	8–33 (17.22)	79.42	1.87	5 <i>Ca. Planktophila</i> , 1 <i>Ca. Nanopelagicus</i> , 1 Nanopelagiales, 2 Nanopelagicaceae, 4 Ilumatobacteraceae, 3 Thermoleophila, 3 Acidimicrobia, 1 Microtrichales, 1 Solirubrobacterales
Planctomycetes	20	2.37–6.9 (4.06)	44.1–71 (59.62)	17–113 (54.225)	81.81	1.69	8 Nemodillakiaceae, 3 Nixiaceae, 3 Gemmataceae, 1 <i>Vodnikaceae</i> , 5 unclassified planctomycetes
Betaproteobacteriales	18	1.24–7.32 (3.23)	35.7–67.8 (55.34)	12–73 (36.44)	77.44	1.26	3 <i>Limnohabitans</i> , 1 <i>Ca. Methyloputimilus</i> , 2 <i>Methylotenera</i> , 1 <i>Rhodiferax</i> , 2 Nitrosomonadaceae, 9 unclassified Betaproteobacteriales
Acidobacteria	13	2.29–6.12 (3.94)	55.5–66 (62.92)	23–119 (43.88)	66.59	3.39	2 Acidobacteriae, 3 Solibacteraceae, 8 Luteitaleales
Gammaaproteobacteria	11	1.65–3.55 (2.53)	39.8–64.8 (52.45)	16–67 (39.9)	80.86	0.97	1 <i>Methyloglobulus</i> , 2 Methylomonadaceae, 1 Pseudohongiellaceae, 1 Xanthomonadales, 2 Steroidobacteraceae, 4 unclassified Gammaaproteobacteria

(Continues)

Table 1. Continued

Phylum/class	No. of MAGs	Range estimated genome size (Mp)*	Range GC	Range median intergenic spacer (bp)	Average completeness	Average contamination	Taxonomic affiliation of MAGs (GTDB, referenced groups)
Deltaproteobacteria	11	2.44–6.6 (4.88)	40.9–70.5 (63.28)	16.5–124 (50.41)	85.86	2.29	8 Myxococcota, 1 Bdellovibrionota, 2 unclassified Deltaproteobacteria
Bacteroidetes	10	1.59–2.59 (2.05)	31.3–55 (38.51)	10–72 (38.85)	84.20	1.44	2 Chitinophagaceae, 5 Flavobacteriaceae, 3 unclassified bacteroidetes (2 Ignavibacteria)
Gemmatimonadetes	5	1.65–3.12 (2.51)	65.7–70 (67.31)	13–30 (22.6)	97.25	0.69	5 unclassified Gemmatimonadetes
Thaumarchaeota	3	1.17–1.23 (1.2)	30.3–33.3 (31.48)	37–40 (38.66)	75.05	0.90	2 Nitrosoarchaeum, 1 Nitrosopumilus
DPANN	3	0.88–1.08 (0.98)	31.3–42.3 (36.72)	36–44 (39)	61.27	0.00	2 Pacearchaeales, 1 Woesearchaeota
Latescibacteria	2	5.12–5.36 (5.24)	64.1–64.5 (64.3)	19–68 (43.5)	82.76	2.59	2 Latescibacterota
Nitrospirae	2	2.24–2.31 (2.27)	60.2–60.5 (60.35)	51.5–60 (55.75)	67.44	1.03	2 unclassified Nitrospirae
Omnitrophica	1	2.78	46.97	48.00	91.38	0.00	1 Omnitrophales

\* The average of each field appears between brackets.

They have been thus far detected in either unusual (e.g., hyperthermophilic) or anaerobic environments (aquifer, sediments, human gut) (Anantharaman et al. 2016; Danczak et al. 2017; Castelle et al. 2018; McLean et al. 2018; Vavourakis et al. 2018). We assembled 28 CPR MAGs of very small assembled and estimated genome sizes (0.6–1.5 Mb), with very variable median intergenic spacers (12–93 bp) (Fig. 2), most of which were novel lineages from different classes, orders, and families based on the GTDB classification scheme (Supporting Information Fig. S8, Table 1 and Supporting Information Table S1). Their small genome sizes and simple metabolisms have led others to propose a probable symbiotic or parasitic association with hosts that still remains to be elucidated (Castelle et al. 2018). Their high abundance in dLB (from 3% in 1250 m to 4.5% in 1350 m) could be expected from the similarity to the conditions in groundwater aquifers that also have limited light and strong soil/sediment influence (Anantharaman et al. 2016; Danczak et al. 2017). As far as we know, this is the first time that significant numbers of such microbes have been detected in a clear-cut aerobic pelagic environment. Along similar lines, DPANN archaea of extremely small genomes (estimated from 0.8–1 Mb) were also found at dLB at < 1% of total archaeal 16S rRNA sequences, what corresponds with previous observations in oligotrophic alpine lakes by 16S rRNA sequencing (Ortiz-Alvarez and Casamayor 2016). We retrieved three MAGs of DPANN archaea which had as closest relatives two Pacearchaeales and one Woesearchaeia from the Rifle groundwater aquifer (see Supporting Information Table S1 and Fig. S9).

A first exploration of the ORFs from these small genomes provided mostly hypothetical proteins. However, among their limited metabolic capabilities detected (Supporting Information Table S2), we observed that our DPANN genomes lacked ATPases, while these were present in the majority of CPRs. Almost all of them contained all the subunits of FOF1 ATP synthase and presented various gene/subunit rearrangements comparable to other bacteria. It has been suggested that they have an anaerobic lifestyle (Castelle et al. 2018) and devoid of NADH oxidoreductases and cytochromes. Lactate dehydrogenases were recurrently found in 11 CPRs, confirming their putative fermentative lifestyle (Castelle et al. 2018). However, in our case, the majority presented superoxide dismutases (sodA), and a few of them harbored NADH dehydrogenases, CcdA Cytochrome c biogenesis proteins, NADPH FMN reductases, and other genes listed in Supporting Information Table S2, which are typically associated to aerobic microbes. A feature that was previously detected in CPR and DPANN was the unusual amount of functional introns and self-splicing mechanisms within tRNAs, 16S/23S rRNA, and other genes (Castelle et al. 2018). The majority of our genomes showed indeed this feature and in some cases, we observed from five to four introns in 23S and 16S genes, respectively (for instance in *Ca. Taylorbacteria* G199). Homing endonucleases (presenting LAGLIDADG motifs), which are typically encoded by

introns, were also detected in two CPRs (G199 and G212) and one DPANN (Woese archaeota G139), while a *rtcB* splicing ligase was observed in Pacearchaeota G141. Additionally, a RuBisCO subunit (form IIIB) was detected in one of our DPANN genomes (Pacearchaeota G140). These subunits were also observed previously within DPANN and CPRs and resemble those observed in methanogens (Castelle et al. 2018).

Five CPRs and one DPANN presented at least one subunit of CRISPR-Cas, although complete systems were not observed. Transposases were also seen in various genomes although usually not more than one per genome. Finally, we observed type II systems for pili formation (PufEF and gspEF) and specific proteins such as *virB4* in CPRs and *virB11* in DPANN, which were detected in the majority of these small genomes. These subunits have been associated with a type IV secretion system which could work as an injector (McLean et al. 2018), thus helping CPRs and DPANN translocating nutrients, proteins, and DNA to a putative host, as previously suggested for Nanoarchaeota (DPANN) (Hamm et al. 2019), Saccharibacteria, and other CPRs (McLean et al. 2018). The presence of these subunits would thus point toward a close interaction of these small-sized microbes to other cells. Although the reasons of their presence (and relatively high abundance) in an oxic bathypelagic environment such as dLB contributes to the mystery surrounding the way of life of these microbes rather than clarifying it.

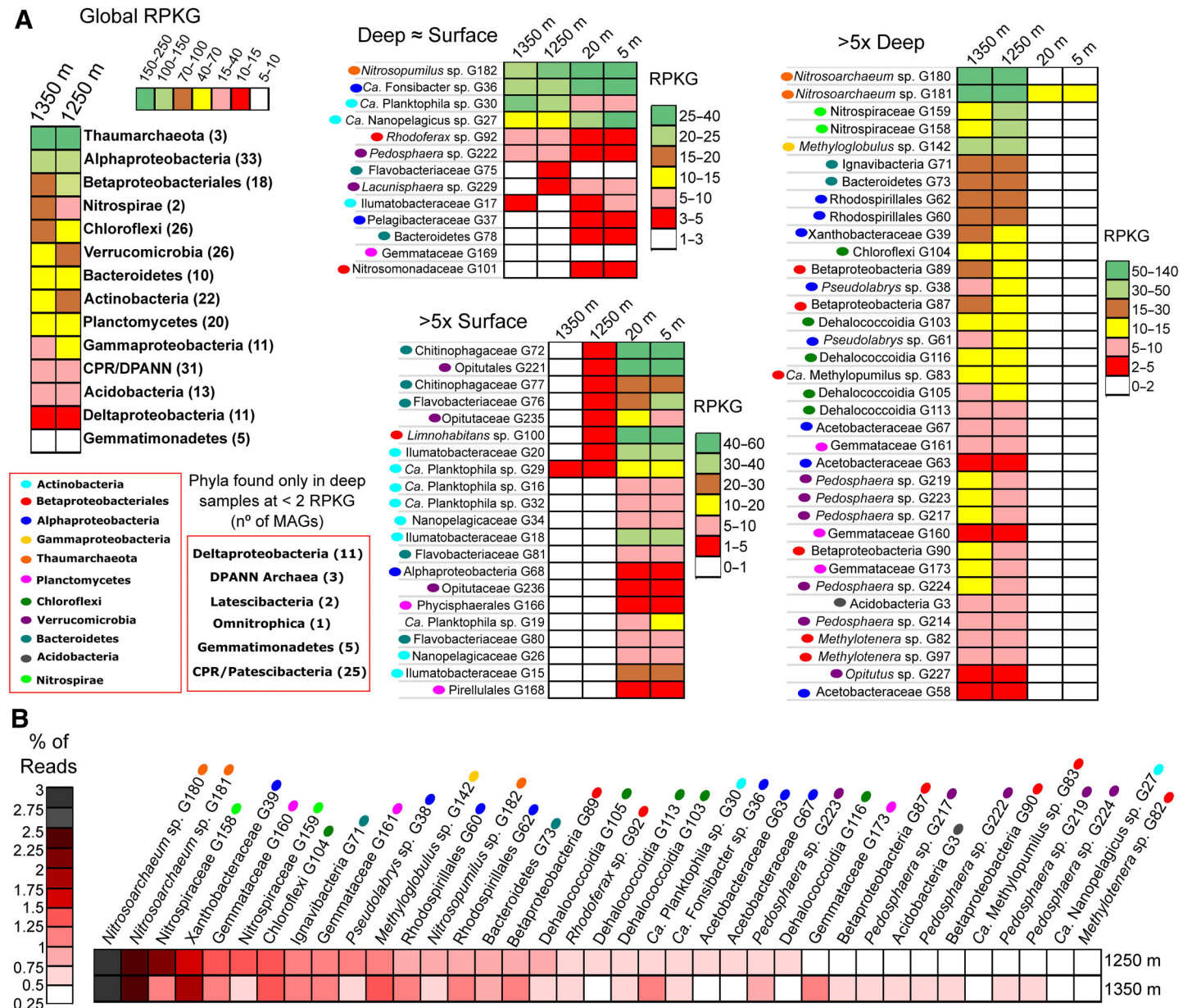
### Surface (epipelagic) vs. deep (bathypelagic) Lake Baikal

As expected from the values of the metagenomes, we observed an important difference in the GC content of MAGs, with a general trend of higher GC in bathypelagic waters (see Supporting Information Text and Fig. S12). We have made an in-depth MAG recruiting comparison between the raw reads of the subice surface and dLB. We ordered MAGs by their relative contribution to the recruitment of reads (Fig. 3A,B), being Thaumarchaeota (only three MAGs) the group that recruited the most reads from dLB metagenome (ca. 10% of the total), followed by Alphaproteobacteria, Betaproteobacteriales (both with many MAGs), Nitrospira (only two MAGs), Chloroflexi, and Verucomicrobia (both with 26 MAGs). Only 21 out of the 231 dLB MAGs were more abundant at the surface (Fig. 3A and Supporting Information Text). All of them are examples of cosmopolitan freshwater microbial taxa detected in other freshwater lakes (Kasalický et al. 2013; Neuenschwander et al. 2017). In addition, some microbes (13) were present at similar abundances (within five times RPKG numbers) in surface/deep waters. Among them, we observed the new *Ca. Fonsibacter* species (G36), two *act* actinobacteria (*Ca. Planktophila* and *Ca. Nanopelagicales*), one species of Thaumarchaeota (Nitrosopumilaceae) or the novel Pelagibacteraceae G37 (Cabello-Yeves et al. 2018). In their case, they could either be adapted to the deep waters or just transported by the deep convection currents (Shimaraev and Granin 1991; Weiss et al. 1991; Shimaraev et al. 2011, 2015).

In any case, the majority of our MAGs only recruited at dLB (absent from the surface). They included the novel *Nitrosoarchaeum* sp. G180, which was the most abundant MAG detected at dLB (3% of the reads, Fig. 3B), or *Nitrosoarchaeum* sp. G181, which was also retrieved from surface waters but was five times more abundant in dLB. Nitrospiraceae G158 and G159 were also found only in deep waters, being in a different phylogenetic branch (Supporting Information Fig. S4) from the Nitrospirae Baikal-G1 described from the surface (Cabello-Yeves et al. 2018). Among all retrieved methanotrophs (Supporting Information Fig. S3), a small genome (*Methyloglobulus* sp. G142) was particularly abundant (ca. 50 RPKG). Two Bacteroidetes (Ignavibacteria bacterium G71 and Bacteroidetes bacterium G73) and two Rhodospirillales were also at the top of most abundant microbes from deep waters (> 15 RPKG). Some other examples of specific dLB MAGs were the majority of Chloroflexi (including the freshwater sister clade to SAR202, see Supporting Information Fig. S2), some specific Alphaproteobacteria seen for the first time in a freshwater environment (Acetobacteraceae members or *Pseudolabrys* spp.) or methylotrophs such as *Methylotenera* (Kalyuzhnaya et al. 2012) and *Methylopumilus* (Salcher et al. 2015). In Fig. 3, we have displayed only MAGs recruiting > 2 RPKGs, but there were other deep-specific microorganisms recruiting at lower but significant values (see Supporting Information Table S3 for RPKG values of all dLB genomes). Examples of these dLB taxa that were present at lower abundances but still exclusive of dLB are Myxococcales (Deltaproteobacteria), unclassified members of Alphagammaproteobacteria and Betaproteobacteriales, Acidobacteria, CPR/DPANN, *Ca. Latescibacteria*, or *Omnitrophica* (Figs. 3, 4). The recruitment data corresponded well with the number of 16S rRNA fragments detected for each phyla and the families, genus, or species detected in each layer as summarized in Fig. 4A,B. From the recruitment analysis, we can conclude that most MAGs retrieved from dLB are characteristic and adapted to live there although some relatives could be found at the surface or even at other freshwater bodies around the world (see Supporting Information Text and Fig. S13). The large prevalence of AOA-AOB, nitrite oxidizers, and methylotrophs points to an ecosystem driven by chemolithotrophic metabolism (see below) what leads to a relatively large standing crop compared to marine waters of similar trophic status and depth.

### Metabolic differences between marine and freshwater aphotic systems

We have also studied the presence/absence and prevalence of several metabolic pathways of the carbon, nitrogen, and sulfur cycles (among others) in the data sets used in this work. We assigned a set of 10 million reads of each data set to key genes and estimated a ratio of metabolic pathway genes to *recA* (Fig. 5 and Supporting Information Table S4) to normalize per genome abundance (see "Methods" section). The surface data set of Lake Baikal was added as a control to establish differences between photic/aphotic systems. As

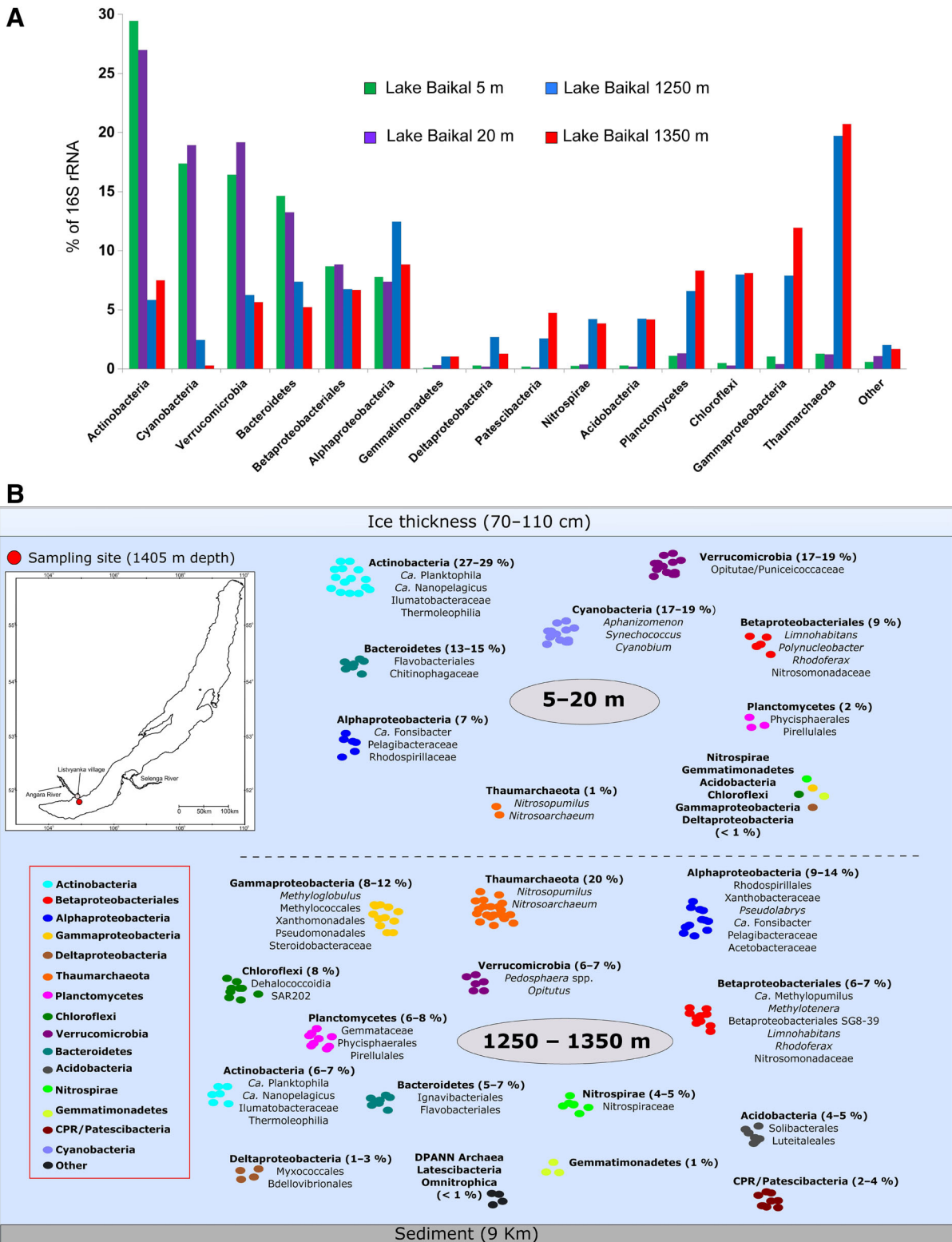


**Fig. 3.** Abundance of dLB MAGs in Lake Baikal. **(A)** Recruitment of dLB MAGs in epipelagic and bathypelagic Lake Baikal metagenomes. Global RPKG values are shown. **(B)** Percentage of read assignment to the most abundant dLB MAGs. MAGs are colored according to phylum classification.

expected, typically photic zone genes such as rhodopsins were much more abundant at the surface of Lake Baikal, which are corresponded to the high number of green-light variants observed in surface MAGs (Cabello-Yeves et al. 2018). There were specific genes that were much more present in all freshwater data sets (independently of their depth). Some of these examples included: carbonic anhydrases that might be required in low carbonate waters, cellulases/xylanases for degradation of recalcitrant matter, or degradation of some aromatic compounds such as biphenyls or chloroaromatic (3-oxoadipate enol-lactonase and CoA-transferase). On the other hand, there were genes that were

more prevalent in marine aphotic systems such as other different genes from aromatic compound degradation including the protocatechuate-3,4-dioxygenases or the 4-hydroxyphenylacetate 3-monooxygenases. The degradation of alkanes (a feature observed in practically all datasets) was more prevalent in surface Lake Baikal and oceanic samples, but was also detected in dLB at non-negligible values (see Supporting Information Table 4). These differences presumably reflect the different nature of resilient organic matter present at the different locations/habitats.

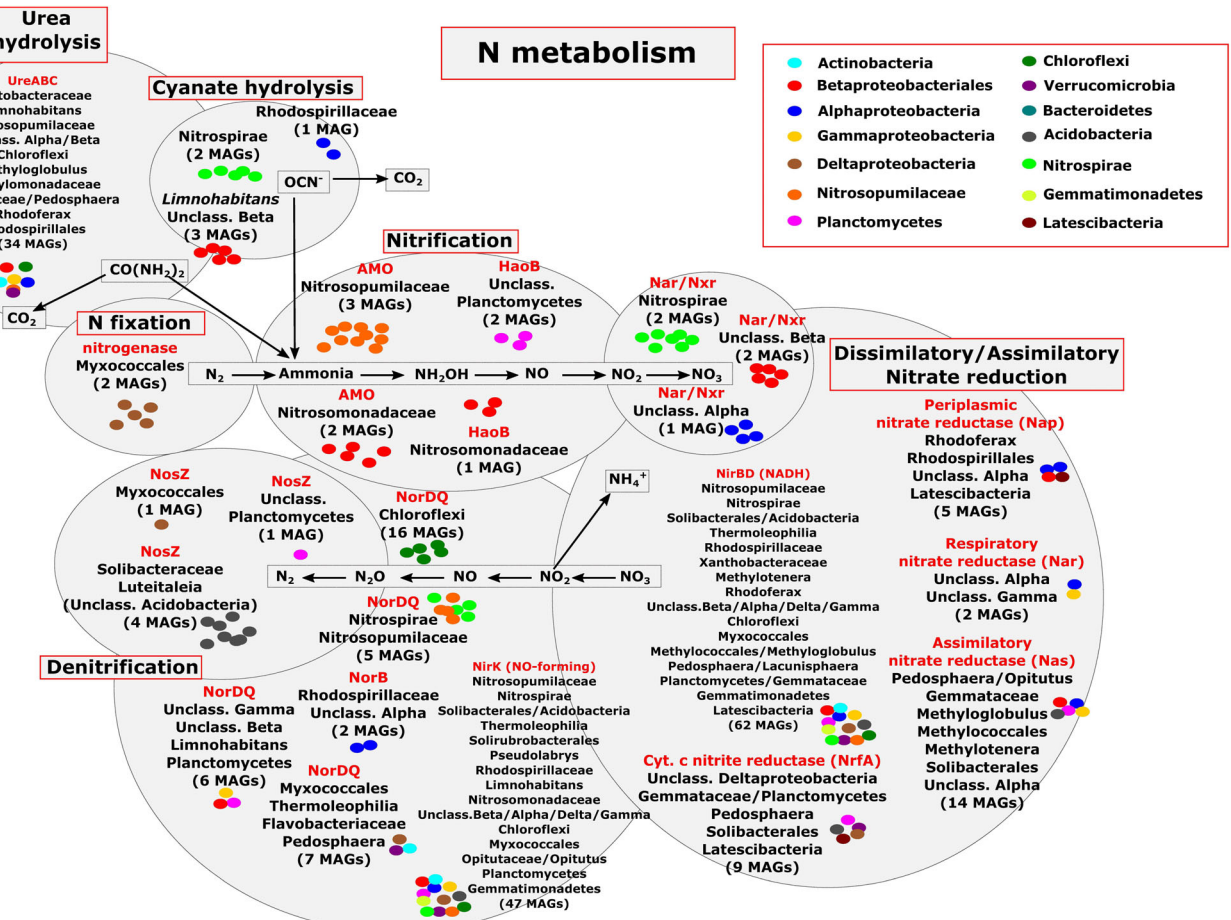
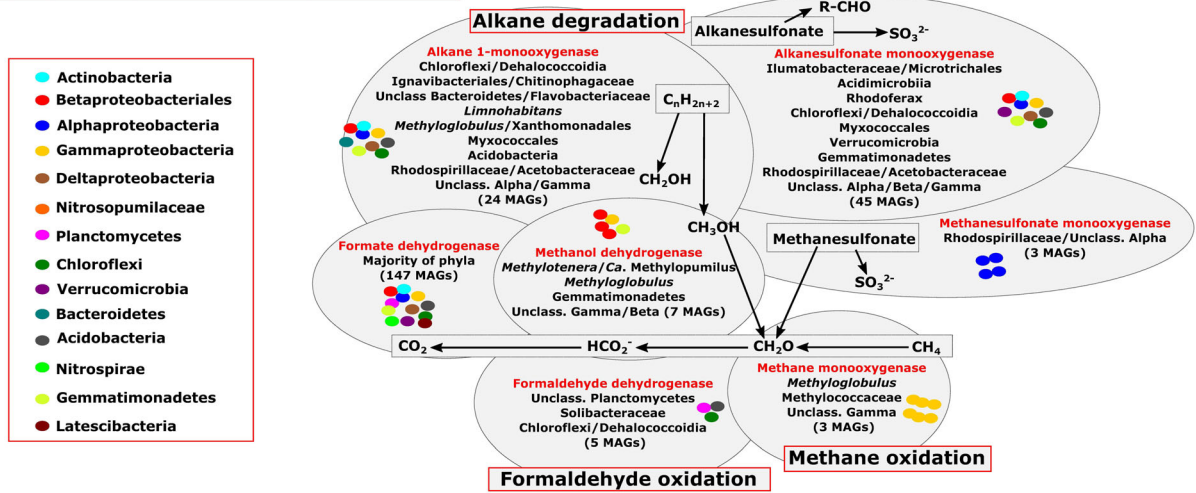
Pathways that were more prevalent specifically at the dLB were also observed. The alternative C fixation pathway (rTCA)



**Fig. 4.** Surface vs. dLB metagenomic data sets. **(A)** 16S rRNA gene reads classification to the phylum level of surface (5, 20 m) and deep (1250, 1350 m) data sets. **(B)** Summary of the presence/absence of the microbiome at epipelagic and bathypelagic horizons of Lake Baikal. Numbers in brackets represent the percentage of assigned 16S rRNA reads for each phylum. Phyla are color-coded.



**Methylotrophy/Alkane metabolism**



**Fig. 6.** C metabolism (methylotrophy and alkane pathways) and N metabolism (nitrification, dissimilatory/assimilatory nitrate reduction, N fixation, denitrification, cyanate hydrolysis, and urea hydrolysis) reconstructed pathways in dLB MAGs. Phyla are color-coded. Each pathway is identified by an enzyme and the number of MAGs presenting each pathway is shown in parenthesis.

19395390, 2020, 7, Downloaded from https://aspubs.onlinelibrary.wiley.com/doi/10.1002/lno.11401 by Universidad Miguel Hernandez De Elche, Wiley Online Library on [05/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

was detected in higher numbers, probably due to the abundance of Nitrospirae which behave as nitrite-oxidizing chemoautotrophs by using it (Lücker et al. 2010) and have been correlated well with the presence of Thaumarchaeota in other lakes (Alfreider et al. 2018). However, Wood-Ljungdhal or DC4HB pathways that have been described from microbes inhabiting marine systems (Hügler and Sievert 2010) were not found at high ratios in dLB. Noticeable were also enzymes for methylotrophy and methane metabolism such as methane monooxygenase, which were much more abundant in dLB, probably due to the higher supply of methane, which comes from the sediment and Selenga and adjacent rivers (Schmid et al. 2007). The large amounts of methane hydrates found in Lake Baikal sediment (Khlystov et al. 2013) attests to the high methane concentrations contained there. High concentrations of methane and oxidation rates were previously recorded in the deep layers of the water column of Lake Baikal (Zakharenko et al. 2015). Along the same lines, the methanotroph *Methyloglobulus* was one of the most abundant microbes in dLB recruiting >1% of the reads. On the other hand, *Methyloglobulus* was not detected in any other freshwater publicly available metagenome, suggesting that methane is a much more important resource in Lake Baikal. Actually, the abundance of oxygen throughout the water column makes dLB different from other lakes of lower latitudes, making the aerobic oxidation of methane a viable energy source. Hydrogen oxidation via different hydrogenases was another energy generating metabolic pathway that was more abundant in dLB. Methanesulfonate monooxygenases, the aromatic pathway of catechol degradation or branched-chain aa transporters (*liv* genes), that are actually transporters of hydrophobic molecules in general, were also significantly overrepresented in dLB. A relevant feature that we observed in dLB was the higher number of CRISPR-Cas systems, some of which were associated to some of the most abundant microbes in the sample (such as Nitrospirae). This phenomenon could be connected to the fact that Nitrospirae MAGs showed high population densities (recruitment rates) with little diversity (Supporting Information Fig. S11), both factors that facilitate phage predation.

Many oligotrophic freshwater systems (and Lake Baikal is no exception) are limited by sulfur (S), while marine systems are rich in it (Norici et al. 2005). In general, sulfate concentrations along the vertical profiles of Lake Baikal were very low (ca. 5 mg L<sup>-1</sup>), several orders of magnitude lower than in the ocean, hence one could expect that sulfur metabolism was not very widespread in dLB microbes. Nitrogen, on the other hand, is present in different compounds (such as cyanate, urea, nitrate, nitrite, or ammonia) in freshwater, and some of them can be limited in marine ecosystems (Bristow et al. 2017). We have explored the number of N and S related metabolic pathways comparing bathypelagic marine and dLB to detect their relative importance (Fig. 5). Our results showed that indeed the N-cycle related genes are more prevalent in

dLB. With regard to the nitrification pathway, we observed ammonia monooxygenase ratio at higher values in bathypelagic Baikal and oceanic data sets, which clearly corresponded to the abundance of Thaumarchaeota in both habitats, as observed from the 16S rRNA reads. The nitrite oxidation pathway, despite being in similar ratios in dLB and bathypelagic marine data sets, presented a significant taxonomic difference (both from reads and MAGs), *Nitrospirae* were the major nitrite-oxidizers in freshwater systems while *Nitrospinae* (Ngugi et al. 2016) were found in marine. On the other hand, except for the *Sox* oxidation pathway (more abundant in dLB samples), marine data sets presented a higher ratio in sulfite reductases, oxidases, and dehydrogenases, what reflects a larger contribution of the sulfur cycle in energizing the deep-ocean microbiota (Mehrshad et al. 2018a).

Finally, we have also analyzed at the level of MAGs the evidence of some of these metabolic pathways by reconstructing metabolic maps of nitrogen, sulfur, or carbon pathways from dLB MAGs (Fig. 6, Supporting Information Fig. S14 and Text). Based on the evidence from reads, MAGs abundance, and gene annotation, we can conclude that nitrification and methylotrophy are the key energy production pathways at the bathypelagic Lake Baikal (Fig. 6).

## Conclusions

It is well known that convective currents throughout the whole lake can redistribute several microbes along the vertical and horizontal water profile. However, only a few microbes (in all cases examples of streamlined, cosmopolitan, abundant and with low surface/ratio volume) have been found in similar abundance in epilimnetic and hypolimnetic/bathypelagic waters of the southern basin of the lake. The taxonomic differences between surface and dLB show the contrast between the energy-limiting bathypelagic layers and the trophogenic epilimnetic one. Methylotrophy and nitrification are the key energy production pathways in dLB. Methane abundance throughout the lake has been observed in deep waters (probably coming from sediment-methanogenesis and rivers) and also in different parts/basins of the lake.

There is a certain degree of endemism and novelty in several microbes retrieved from dLB. Actually, most of the microbes described here such as representatives of CPRs, DPANN, Latescibacteria, Omnitrphica, Acidobacteria, and several unclassified members of Alpha-Gamma-Delta and Betaproteobacteriales are very different from their relatives found elsewhere. We cannot exclude the possibility that some of them could inhabit other similar freshwater systems, although in this sense Lake Baikal is rather unique.

Deep-ocean waters are considered extreme environments due to their lack of energy sources (light), low temperature, and high pressure. dLB is submitted to similar conditions plus a deficit of inorganic salts that are abundant in the sea. Our analysis of the microbiota inhabiting dLB indicates that, as other works have

suggested (Cabello-Yeves and Rodriguez-Valera 2019), salinity is the key factor influencing the community structure (with presence/absence of certain taxa) even at bathypelagic depths. The dLB presents much more similarity to its own surface or to aphotic layers of Lake Zurich than to similarly deep and cold marine habitats. The overall similarities that we have found between freshwaters in their 16S rRNA classification and reads identity gives us a clear evidence that they share similar abundant taxa such as LD12 *Ca. Fonsibacter* (Henson et al. 2018), *Ca. Planktophila* and *Ca. Nanopelagales* (Neuenschwander et al. 2017), Planctomycetes (Andrei et al. 2019), Chloroflexi (Mehrshad et al. 2018b), Verrucomicrobia (Cabello-Yeves et al. 2017a), and Betaproteobacteriales members such as *Ca. Methylopusillus* or *Limnohabitans* (Kasalický et al. 2013; Salcher et al. 2015). On the other hand, marine data sets group together due to the presence of characteristic deep ocean microbes like SAR202 Chloroflexi (Mehrshad et al. 2018a), marine Euryarchaeota, Rhodospirillaceae members, most Pelagibacteraceae, Marinimicrobia, or Gammaproteobacteria from different groups (such as SAR86 or Alteromonadaceae). However, there are some commonalities between marine bathypelagic depths and dLB, such as some Pelagibacteraceae, Chloroflexi of the SAR202 clade, and Nitrosopumilaceae (Thaumarchaeota). Some representatives of these groups must have properties that favor their success in flourishing at such depths.

### Compliance with ethical standards

Ethics approval was not required for the study.

### References

- Alfreider, A., V. Grimus, M. Luger, A. Ekblad, M. M. Salcher, and M. Summerer. 2018. Autotrophic carbon fixation strategies used by nitrifying prokaryotes in freshwater lakes. *FEMS Microbiol. Ecol.* **94**: fty163.
- Anantharaman, K., and others. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**: 13219.
- Andrei, A.-Ş., M. M. Salcher, M. Mehrshad, P. Rychtecký, P. Znachor, and R. Ghai. 2019. Niche-directed evolution modulates genome architecture in freshwater Planctomycetes. *ISME J.* **13**: 1056–1071.
- Bashenkhaeva, M. V., Y. R. Zakharova, D. P. Petrova, I. V. Khanaev, Y. P. Galachyants, and Y. V. Likhoshway. 2015. Sub-ice microalgal and bacterial communities in freshwater Lake Baikal, Russia. *Microb. Ecol.* **70**: 751–765.
- Bel'kova, N. L., V. V. Parfenova, T. Y. Kostornova, L. Y. Denisova, and E. F. Zaichikov. 2003. Microbial biodiversity in the water of Lake Baikal. *Microbiology* **72**: 203–213.
- Benoit, G. 2015. Simka: Fast kmer-based method for estimating the similarity between numerous metagenomic datasets. *RCAM*.
- Bristow, L. A., W. Mohr, S. Ahmerkamp, and M. M. M. Kuypers. 2017. Nutrients that limit growth in the ocean. *Curr. Biol.* **27**: R474–R478.
- Buchfink, B., C. Xie, and D. H. Huson. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**: 59–60.
- Cabello-Yeves, P. J., R. Ghai, M. Mehrshad, A. Picazo, A. Camacho, and F. Rodriguez-Valera. 2017a. Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. *Front. Microbiol.* **8**: 2131. doi:10.3389/fmicb.2017.02131
- Cabello-Yeves, P. J., J. M. Haro-Moreno, A.-B. Martin-Cuadrado, R. Ghai, A. Picazo, A. Camacho, and F. Rodriguez-Valera. 2017b. Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. *Front. Microbiol.* **8**: 1151.
- Cabello-Yeves, P. J., T. I. Zemskaya, R. Rosselli, F. H. Coutinho, A. S. Zakharenko, V. V. Blinov, and F. Rodriguez-Valera. 2018. Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. *Appl. Environ. Microbiol.* **84**: e02132–e02117.
- Cabello-Yeves, P. J., and F. Rodriguez-Valera. 2019. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome* **7**: 1–12.
- Castelle, C. J., C. T. Brown, K. Anantharaman, A. J. Probst, R. H. Huang, and J. F. Banfield. 2018. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* **16**: 629–645.
- Danczak, R. E., M. D. Johnston, C. Kenah, M. Slattery, K. C. Wrighton, and M. J. Wilkins. 2017. Members of the Candidate Phyla Radiation are functionally differentiated by carbon and nitrogen-cycling capabilities. *Microbiome* **5**: 112.
- De Wever, A., K. Muylaert, K. Van der Gucht, S. Pirlot, C. Cocquyt, J.-P. Descy, P.-D. Plisnier, and W. Vyverman. 2005. Bacterial community composition in Lake Tanganyika: Vertical and horizontal heterogeneity. *Appl. Environ. Microbiol.* **71**: 5029–5037.
- DeLong, E. F., and others. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Eiler, A., K. Zaremba-Niedzwiedzka, M. Martínez-García, K. D. McMahon, R. Stepanauskas, S. G. E. Andersson, and S. Bertilsson. 2014. Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics. *Environ. Microbiol.* **16**: 2682–2698.
- Galazy, G. I. 1993. Atlas of Lake Baikal. GUGK, [in Russ].
- Garcia, S. L., M. Buck, J. J. Hamilton, C. Wurzbacher, H.-P. Grossart, K. D. McMahon, and A. Eiler. 2018. Model communities hint at promiscuous metabolic linkages between ubiquitous free-living freshwater bacteria. *mSphere* **3**: e00202–e00218.
- Ghai, R., C. M. Mizuno, A. Picazo, A. Camacho, and F. Rodriguez-Valera. 2014. Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Mol. Ecol.* **23**: 6073–6090.

- Glöckner, F. O., E. Zaichikov, N. Belkova, L. Denissova, J. Pernthaler, A. Pernthaler, and R. Amann. 2000. Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. *Appl. Environ. Microbiol.* **66**: 5053–5065.
- Goris, J., K. T. Konstantinidis, J. A. Klappenbach, T. Coenye, P. Vandamme, and J. M. Tiedje. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**: 81–91.
- Granin, N. G., and others. 2018. Lake Baikal's response to remote earthquakes: Lake-level fluctuations and near-bottom water layer temperature change. *Mar. Petrol. Geol.* **89**: 604–614.
- Haft, D. H., B. J. Loftus, D. L. Richardson, F. Yang, J. A. Eisen, I. T. Paulsen, and O. White. 2001. TIGRFAMs: A protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**: 41–43.
- Hamm, J. N., and others. 2019. Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl. Acad. Sci. USA* **116**: 14661–14670.
- Hampton, S. E., E. V. A. Izmet, R. Lyubov, M. V. Moore, S. L. Katz, B. Dennis, and E. A. Silow. 2008. Sixty years of environmental change in the world's largest freshwater lake—Lake Baikal, Siberia. *Glob. Chang. Biol.* **14**: 1947–1958.
- Haro-Moreno, J. M., M. López-Pérez, R. José, A. Picazo, A. Camacho, and F. Rodríguez-Valera. 2018. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* **6**: 128.
- Henson, M. W., V. C. Lanclos, B. C. Faircloth, and J. C. Thrash. 2018. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J.* **12**: 1846–1860.
- Herlemann, D. P. R., M. Labrenz, K. Jürgens, S. Bertilsson, J. J. Waniek, and A. F. Andersson. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**: 1571–1579.
- Hernsdorf, A. W., and others. 2017. Potential for microbial H<sub>2</sub> and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J.* **11**: 1915–1929.
- Hugerth, L. W., J. Larsson, J. Alneberg, M. V. Lindh, C. Legrand, J. Pinhassi, and A. F. Andersson. 2015. Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16**: 1–18.
- Hügler, M., and S. M. Sievert. 2010. Beyond the Calvin cycle: Autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci.* **3**: 261–289.
- Kalyuzhnaya, M. G., D. A. C. Beck, A. Vorobev, N. Smalley, D. D. Kunkel, M. E. Lidstrom, and L. Chistoserdova. 2012. Novel methylotrophic isolates from lake sediment, description of *Methylotenera versatilis* sp. nov. and emended description of the genus *Methylotenera*. *Int. J. Syst. Evol. Microbiol.* **62**: 106–111.
- Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**: 27–30.
- Kanehisa, M., Y. Sato, and K. Morishima. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**: 726–731.
- Kang, D., F. Li, E. S. Kirton, A. Thomas, R. S. Egan, H. An, and Z. Wang. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**: e7359.
- Kasalický, V., J. Jezbera, M. W. Hahn, and K. Šimek. 2013. The diversity of the Limnohabitans genus, an important group of freshwater bacterioplankton, by characterization of 35 isolated strains. *PLoS One* **8**: e58209.
- Khlystov, O., and others. 2013. Gas hydrate of Lake Baikal: Discovery and varieties. *J. Asian Earth Sci.* **62**: 162–166.
- Khodzher, T., V. M. Domysheva, L. M. Sorokovikova, and L. P. Golobokova. 2016. Methods for monitoring the chemical composition of Lake Baikal water, p. 113–132. *In* Mueller L., Sheudshen A., Eulenstein F. (eds) Novel methods for monitoring and managing land and water resources in Siberia. Springer.
- Kozhov, M. M. 1962. Biology of Lake Baikal. Publishing House of the Academy of Sciences of the USSR.
- Kurilkina, M. I., Y. R. Zakharova, Y. P. Galachyants, D. P. Petrova, Y. S. Bukin, V. M. Domysheva, V. V. Blinov, and Y. V. Likhoshway. 2016. Bacterial community composition in the water column of the deepest freshwater Lake Baikal as determined by next-generation sequencing. *FEMS Microbiol. Ecol.* **92**: f1w094.
- Lê, S., J. Josse, and F. Husson. 2008. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**: 1–18.
- Liang, Y., Y. Zhang, Y. Zhang, T. Luo, R. B. Rivkin, and N. Jiao. 2017. Distributions and relationships of virio- and picoplankton in the epi-, meso- and bathypelagic zones of the Western Pacific Ocean. *FEMS Microbiol. Ecol.* **93**: f1w238.
- Likhoshway, Y. V., A. Y. Kuzmina, T. G. Potyemkina, V. L. Potyemkin, and M. N. Shimaraev. 1996. The distribution of diatoms near a thermal bar in Lake Baikal. *J. Great Lakes Res.* **22**: 5–14.
- Linz, A. M., S. He, S. L. R. Stevens, K. Anantharaman, R. R. Rohwer, R. R. Malmstrom, S. Bertilsson, and K. D. McMahon. 2018. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ* **6**: e6075.
- Logares, R., J. Bråte, F. Heinrich, K. Shalchian-Tabrizi, and S. Bertilsson. 2009. Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). *Mol. Biol. Evol.* **27**: 347–357.
- Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.

- Lücker, S., and others. 2010. A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc. Natl. Acad. Sci. USA* **107**: 13479–13484.
- Maksimova, E. A., and V. N. Maksimov. 1989. Microbiology of Baikal water. *Izd. Irkutsk. Univ. Irkutsk*: 1–168.
- Marchler-Bauer, A., and others. 2016. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**: D200–D203.
- McLean, J. S., and others. 2018. Evidence of independent acquisition and adaptation of ultra-small bacteria to human hosts across the highly diverse yet reduced genomes of the phylum Saccharibacteria. *BioRxiv* 258137.
- Mehrshad, M., M. A. Amoozegar, R. Ghai, S. A. S. Fazeli, and F. Rodriguez-Valera. 2016. Genome reconstruction from metagenomic datasets reveals novel microbes in the brackish waters of the Caspian Sea. *Appl. Environ. Microbiol.* **82**: 1599–1612.
- Mehrshad, M., F. Rodriguez-Valera, M. A. Amoozegar, P. López-García, and R. Ghai. 2018a. The enigmatic SAR202 cluster up close: Shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. *ISME J.* **12**: 655–668.
- Mehrshad, M., M. M. Salcher, Y. Okazaki, S. Nakano, K. Šimek, A.-S. Andrei, and R. Ghai. 2018b. Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* **6**: 176.
- Mende, D. R., J. A. Bryant, F. O. Aylward, J. M. Eppley, T. Nielsen, D. M. Karl, and E. F. DeLong. 2017. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* **2**: 1367–1373.
- Mikhailov, I. S., Y. R. Zakharova, Y. P. Galachyants, M. V. Usoltseva, D. P. Petrova, M. V. Sakirko, Y. V. Likhoshway, and M. A. Grachev. 2015. Similarity of structure of taxonomic bacterial communities in the photic layer of Lake Baikal's three basins differing in spring phytoplankton composition and abundance. *Dokl. Biochem. Biophys.* **465**: 413.
- Nawrocki, E. P., and S. R. Eddy. 2010. ssu-align: A tool for structural alignment of SSU rRNA sequences.
- Neuenschwander, S. M., R. Ghai, J. Pernthaler, and M. M. Salcher. 2017. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* **12**, 185–198.
- Ngugi, D. K., J. Blom, R. Stepanauskas, and U. Stingl. 2016. Diversification and niche adaptations of Nitrospina-like bacteria in the polyextreme interfaces of Red Sea brines. *ISME J.* **10**: 1383–1399.
- Norici, A., R. Hell, and M. Giordano. 2005. Sulfur and primary production in aquatic environments: An ecological perspective. *Photosynth. Res.* **86**: 409–417.
- Oh, S., A. Caro-Quintero, D. Tsementzi, N. DeLeon-Rodriguez, C. Luo, R. Poretsky, and K. T. Konstantinidis. 2011. Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl. Environ. Microbiol.* **77**: 6000–6011.
- Ortiz-Alvarez, R., and E. O. Casamayor. 2016. High occurrence of Pacearchaeota and Woesearchaeota (Archaea superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. *Environ. Microbiol. Rep.* **8**: 210–217.
- Overbeek, R., and others. 2013. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**: D206–D214.
- Parfenova, V. V., and others. 2000. On the vertical distribution of microorganisms in Lake Baikal during spring deep-water renewal. *Microbiology* **69**: 357–363.
- Parfenova, V. V., N. L. Belkova, and L. Y. Denisova. 2006. Studies on the species composition of cultured heterotrophic organisms from Lake Baikal. *Biol. Vnutr. Vod.* **1**: 8–15.
- Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz, and G. W. Tyson. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**: 1043–1055.
- Parks, D. H., M. Chuvochina, D. W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil, and P. Hugenholtz. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**: 996–1004.
- Paver, S. F., D. Muratore, R. J. Newton, and M. L. Coleman. 2018. Reevaluating the salty divide: Phylogenetic specificity of transitions between marine and freshwater systems. *mSystems* **3**: e00232–e00218.
- Peng, Y., H. C. M. Leung, S.-M. Yiu, and F. Y. L. Chin. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.
- Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2012. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**: D590–D596.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16**: 276–277.
- Rinke, C., and others. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.
- Salcher, M. M., S. M. Neuenschwander, T. Posch, and J. Pernthaler. 2015. The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J.* **9**: 2442–2453.
- Schmid, M., M. De Batist, N. G. Granin, V. A. Kapitanov, D. F. McGinnis, I. B. Mizandrontsev, A. I. Obzhairov, and A. Wüest. 2007. Sources and sinks of methane in Lake Baikal: A synthesis of measurements and modeling. *Limnol. Oceanogr.* **52**: 1824–1837.
- Segata, N., D. Börnigen, X. C. Morgan, and C. Huttenhower. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**: 2304.
- Shimaraev, M. N., and N. G. Granin. 1991. On stratification and convection mechanism in Baikal. *Doklady Akad. Nauk.* **321**: 381–385.

- Shimaraev, M. N., R. Y. Gnatovskii, V. V. Blinov, and V. G. Ivanov. 2011. Renewal of deep waters of Lake Baikal revisited. *Dokl. Earth Sci.* **438**: 652.
- Shimaraev, M. N., N. G. Granin, R. J. Gnatovskij, and V. V. Blinov. 2015. The mechanism of oxygen aeration of bottom waters of Lake Baikal. *Dokl. Earth Sci.* **461**: 379–383.
- Simon, M., and others. 2017. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* **11**: 1483–1499.
- Sunagawa, S., and others. 2015. Structure and function of the global ocean microbiome. *Science* **348**: 1261359.
- Tatusov, R. L., and others. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**: 22–28.
- Tran, P., A. Ramachandran, O. Khawasek, B. E. Beisner, M. Rautio, Y. Huot, and D. A. Walsh. 2018. Microbial life under ice: Metagenome diversity and in situ activity of Verucomicrobia in seasonally ice-covered lakes. *Environ. Microbiol.* **20**: 2568–2584.
- Urbach, E., K. L. Vergin, L. Young, A. Morse, G. L. Larson, and S. J. Giovannoni. 2001. Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. *Limnol. Oceanogr.* **46**: 557–572.
- Vavourakis, C. D., A.-S. Andrei, M. Mehrshad, R. Ghai, D. Y. Sorokin, and G. Muyzer. 2018. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* **6**: 168.
- Vologina, E. G., and M. Sturm. 2017. Particulate fluxes in South Baikal: Evidence from sediment trap experiments. *Russ. Geol. Geophys.* **58**: 1045–1052.
- Votintsev, K. K., A. I. Mescheryakova, and G. I. Popovskaya. 1975. Cycle of organic matter in Lake Baikal. Nauka, Novosibirsk, 1975, 189 pp.
- Weiss, R. F., E. C. Carmack, and V. M. Koropalov. 1991. Deep-water renewal and biological production in Lake Baikal. *Nature* **349**: 665–669.
- Zakharenko, A. S., N. V. Pimenov, V. G. Ivanova, and T. I. Zemskaya. 2015. Detection of methane in the water column at gas and oil seep sites in central and southern Lake Baikal. *Microbiology* **84**: 90–97.
- Zakharova, Y. R., and others. 2013. The structure of microbial community and degradation of diatoms in the deep near-bottom layer of Lake Baikal. *PLoS One* **8**: e59977.

### Acknowledgments

This work was supported by grant “VIREVO” CGL2016-76273-P (AEI/FEDER, EU), (cofounded with FEDER funds), PROMETEO/2019/009 “HIDRAS3” from Generalitat Valenciana and by the State Task for Limnological Institute SB RAS 0345–2019–0007 and the Russian Foundation for Basic Research Grant ofi-m 17–29–05040. R.G. was supported by the research grant 17-04828S (Grant Agency of the Czech Republic). F.R.-V. was beneficiary of the Stop100-program of the Ministry for science and education of Russia. P.J.C.-Y. was supported by APOSTD/2019/009 Post-Doctoral fellowship from Generalitat Valenciana.

### Conflict of Interest

None declared.

Submitted 09 August 2019

Revised 08 October 2019

Accepted 23 November 2019

Associate editor: Hans-Peter Grossart