AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal

Pedro J. Cabello-Yeves,[a] Tamara I. Zemskaya,[b] Riccardo Rosselli,[a] Felipe H. Coutinho,[a] Alexandra S. Zakharenko,[b] Vadim V. Blinov,[b] Francisco Rodriguez-Valera[a]

[a]Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Alicante, Spain

[b]Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

ABSTRACT We present a metagenomic study of Lake Baikal (East Siberia). Two samples obtained from the water column under the ice cover (5 and 20 m deep) in March 2016 have been deep sequenced and the reads assembled to generate metagenome-assembled genomes (MAGs) that are representative of the microbes living in this special environment. Compared with freshwater bodies studied around the world, Lake Baikal had an unusually high fraction of *Verrucomicrobia*. Other groups, such as *Actinobacteria* and *Proteobacteria*, were in proportions similar to those found in other lakes. The genomes (and probably cells) tended to be small, presumably reflecting the extremely oligotrophic and cold prevalent conditions. Baikal microbes are novel lineages recruiting very little from other water bodies and are distantly related to other freshwater microbes. Despite their novelty, they showed the closest relationship to genomes discovered by similar approaches from other freshwater lakes and reservoirs. Some of them were particularly similar to MAGs from the Baltic Sea, which, although it is brackish, connected to the ocean, and much more eutrophic, has similar climatological conditions. Many of the microbes contained rhodopsin genes, indicating that, in spite of the decreased light penetration allowed by the thick ice/snow cover, photoheterotrophy could be widespread in the water column, either because enough light penetrates or because the microbes are already adapted to the summer ice-less conditions. We have found a freshwater SAR11 subtype I/II representative showing striking synteny with *Pelagibacter ubique* strains, as well as a phage infecting the widespread freshwater bacterium *Polynucleobacter*.

IMPORTANCE Despite the increasing number of metagenomic studies on different freshwater bodies, there is still a missing component in oligotrophic cold lakes suffering from long seasonal frozen cycles. Here, we describe microbial genomes from metagenomic assemblies that appear in the upper water column of Lake Baikal, the largest and deepest freshwater body on Earth. This lake is frozen from January to May, which generates conditions that include an inverted temperature gradient (colder up), decrease in light penetration due to ice, and, especially, snow cover, and oligotrophic conditions more similar to the open-ocean and high-altitude lakes than to other freshwater or brackish systems. As could be expected, most reconstructed genomes are novel lineages distantly related to others in cold environments, like the Baltic Sea and other freshwater lakes. Among them, there was a broad set of streamlined microbes with small genomes/intergenic spacers, including a new nonmarine *Pelagibacter*-like (subtype I/II) genome.

KEYWORDS Lake Baikal, metagenomics, metagenome-assembled genomes (MAGs), *Pelagibacter*, polynucleophage, Baltic Sea

Received 27 September 2017 **Accepted** 19 October 2017

Accepted manuscript posted online 27 October 2017

Citation Cabello-Yeves PJ, Zemskaya TI, Rosselli R, Coutinho FH, Zakharenko AS, Blinov VV, Rodriguez-Valera F. 2018. Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. Appl Environ Microbiol 84:e02132-17. https://doi.org/10.1128/AEM.02132-17.

Editor Harold L. Drake, University of Bayreuth

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Francisco Rodriguez-Valera, frvalera@umh.es.

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

Lake Baikal is the world's deepest (1,637 m), largest (by volume), and oldest (25 million years) lake (1). With a volume of 23,000 km^3, it represents close to 20% of the Earth's unfrozen freshwater (2). Due to its latitude (51 to 54°N) and continental climate, this water body is subjected to very low temperatures, and its surface is frozen for 4 to 4.5 months a year (3). The low surface temperatures create an inverted temperature profile, i.e., it is colder higher up, generating convective currents that keep the water column mixed for most of the year. This makes the Baikal Lake special among deep lakes by having high oxygen concentrations (9 to 14.5 mg/liter) and low temperatures (ca. 4°C in winter) throughout its depth (4, 5). Stratification takes place only in summer (July to September) and involves only the upper 100 m (the highest temperature reaches only 10 to 15°C in August for brief periods) (2, 6). The water of Lake Baikal is ultraoligotrophic, with high water transparency (1). Nutrient elements originating in the surface layers are recycled about four times before settling to deep waters (7). The oligotrophic level of the waters in the central basin of Lake Baikal was confirmed by chlorophyll, nutrient, and picoplankton concentrations; all of these parameters are similar in values to those found in the open ocean (8, 9). Only one peak of abundance and biomass of phytoplankton is characteristic of the lake, that being in spring when intense development of diatoms is recorded, whose mass concentrates in the upper 5- to 25-m-deep water layer (10). Primary production is higher in the summer due to mass development of picoplankton algae (11), and phytoplankton smaller than 10 μm are responsible for a significant proportion (60% to 100%) of total primary production in the epilimnion (10, 12). The lake remains frozen from January to May, and ice thickness reaches a maximum of 50 to 110 cm by late March. Still, approximately 65 to 80% of sunlight penetrates the ice to the underlying water layer, but as soon as snow accumulates, the amount of sunlight available in the sub-ice environment decreases approximately 10-fold (13). However, the sub-ice community remains active, and a dense layer of diatoms and other algae develops at the ice-water interphase (14, 15).

The microbiota of the water column has been studied by 16S rRNA sequencing, and they found common features with other freshwater bodies (16, 17), specifically the abundance of *Actinobacteria* or betaproteobacterial clones. For example, bacterial communities analyzed by pyrosequencing of 16S rRNA gene fragments in the sub-ice environment belonged to the phyla *Proteobacteria*, *Verrucomicrobia*, *Actinobacteria*, *Acidobacteria*, *Bacteroidetes*, and *Cyanobacteria* (13). Studies of the water column during the ice-free period indicated a decrease in *Actinobacteria* with depth, compensated by an increase of *Betaproteobacteria* (2). However, studies based on PCR amplification of 16S rRNA genes are PCR biased and do not provide insights about the community genomic information (18, 19).

Here, we present a study of Lake Baikal by deep metagenomic sequencing and genome assembly. We have studied samples taken at two depths (5 and 20 m) under the ice cover in March. We have assembled a large number of relatively complete metagenome-assembled genomes (MAGs). Most of them correspond to unique and novel species which so far have not been described, which helps characterize major components of the microbiota of the largest freshwater body on Earth. Among the most remarkable findings are the description of genomes of a novel group of *Pelagibacter*-related freshwater bacteria and a phage infecting the typical freshwater betaproteobacterium *Polynucleobacter*.

RESULTS AND DISCUSSION

16S rRNA community structure and general assembly features. The 16S rRNA gene fragments retrieved from the metagenomes provide a glimpse into the diversity of major phylogenetic groups. Since they are sequenced directly (that is, without PCR amplification, as is the case for amplicon sequencing), the results are not PCR biased. Similar samples from the sub-ice ultraoligotrophic lakes have not been obtained. Nonetheless, we show the classification of 16S rRNA fragments found in Lake Baikal and in other freshwater and brackish bodies selected at different latitudes and also retrieved by direct extraction of rRNA gene fragments from metagenomes (rather than

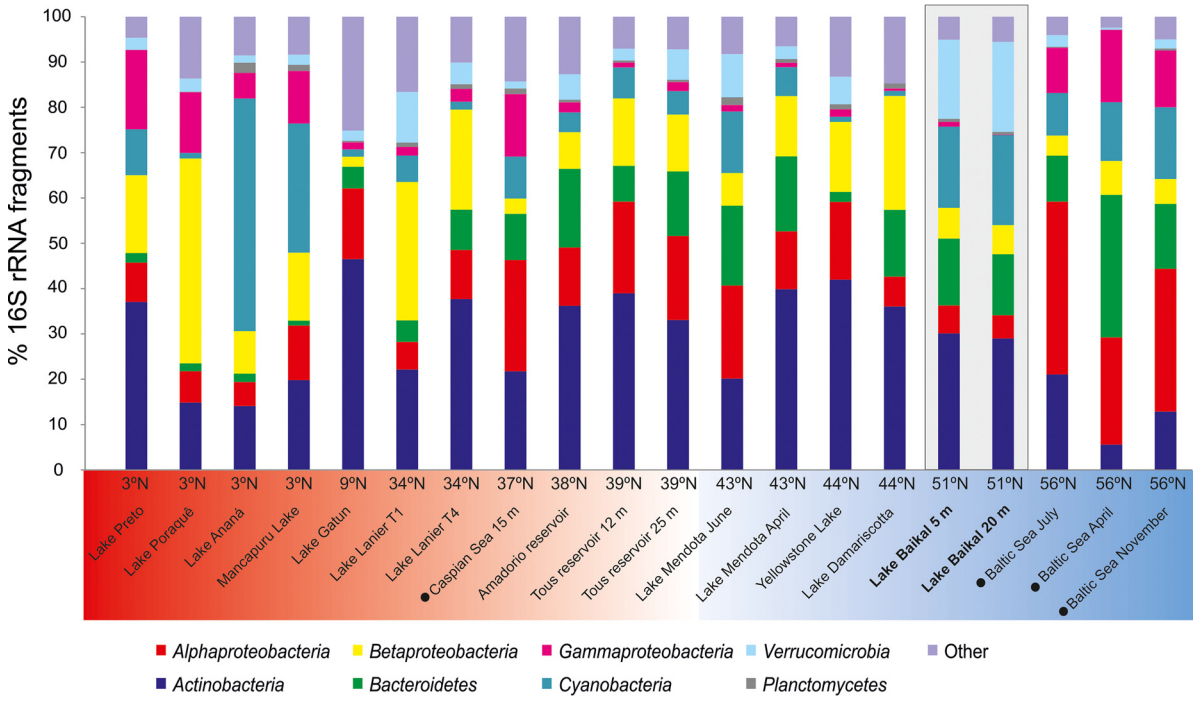


FIG 1 Prokaryotic community structure based on 16S rRNA gene fragments from unassembled Lake Baikal metagenomes from 5- and 20-m-deep samples compared to the 16S rRNA gene reads from different freshwater and brackish data sets coming from latitudes from 0 to 56°. All data are directly derived from metagenomes (no amplicon sequencing). The different represented lakes range from equatorial and temperate to cold. The red-to-blue gradient shows the increasing latitudes of the different freshwater and brackish data sets. Brackish data sets are indicated with a black dot. Lake Baikal data sets are highlighted in bold.

using amplicon sequencing) in order to compare data obtained by the same methodology (Fig. 1 and S1). The proportions of the predominant groups found in Lake Baikal and the other lakes available are similar, especially compared to high-latitude Swedish lakes, like Vattern (see Fig. S1 in the supplemental material). However, noticeable are the large fraction of *Verrucomicrobia* (20% of the total rRNA reads) and relatively small proportion of *Betaproteobacteria* compared with the other lakes, even those sampled at the winter season (Tous, Amadorio, Kalamas, or Lake Lanier in Fig. S1). The dominant groups in Lake Baikal using our approach are *Actinobacteria*, *Cyanobacteria*, *Verrucomicrobia*, and *Bacteroidetes*. In general, our data fit better with the amplicon sequencing carried out during the ice-covered period (13) when the community was dominated by *Actinobacteria*, *Acidobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*, followed by *Bacteroidetes* and *Verrucomicrobia*. During the period of water warming in the spring (2), for which we do not have data, surface and 25-m-deep layers showed the presence of *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Firmicutes*, and *Proteobacteria*, with *Proteobacteria* showing predominance of *Gammaproteobacteria* over *Betaproteobacteria* and *Alphaproteobacteria* (2). In our samples from the frozen season, we have observed very small proportions of *Gammaproteobacteria*, *Chloroflexi*, or *Firmicutes* compared to these results obtained by amplicon sequencing, while we observed higher proportions of *Cyanobacteria* and *Verrucomicrobia* (in both cases representing more than 15% of the rRNA reads). Still, by amplicon sequencing, sporadic blooms of *Verrucomicrobia* were also detected with proportions up to 40% (13).

To be able to compare and characterize the typical components of the microbiome of Lake Baikal, we have resorted to metagenome-assembled genome (MAG) analysis. We assembled the 5- and 20-m-depth genomes separately and also together (see Fig. S2 in the supplemental material). Although in general, the groups assembled best correspond with the most abundant groups (*Actinobacteria*, *Bacteroidetes*, and *Verrucomicrobia*) and are similar in the two samples, slight variations were observed when assembling the samples together. This allowed the completion of longer contigs for

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

most groups, and given the similarity of the two samples, they were considered henceforth together to generate MAGs by binning the contigs together. However, as an exception, the two *Cyanobacteria* described here were reconstructed from the separate assembly of the 5- and 20-m-depth metagenomes but could not be reconstructed with the combined assembly of both metagenomes, although the contigs were finally pooled. In order to simplify the work, we included in our analysis only bins which had more than 40% completeness obtained by CheckM estimations (20).

Major Lake Baikal MAGs. The genome features of the 35 reconstructed Lake Baikal MAGs are shown in Table 1. Their phylogenies based on protein-concatenated phylogenomic trees are shown in Fig. S3 to S13 in the supplemental material. We obtained eight actinobacterial genomes (Fig. S3), two of which affiliate close to Baltic Sea MAGs inside the *Acidimicrobidae* family, being clearly in different branches than the marine groups (21), freshwater acAcidi lineage (22), and brackish representatives from the Caspian Sea (23, 24). We were able to reconstruct a member of the *Thermoleophilia* family which has strong similarities with *Conexibacter* and *Gaiella* soil representatives. The remaining five *Actinobacteria* MAGs are relatives of the freshwater acI lineage. Actinobacterium-acI-Baikal-G5 has its closest relative in "*Candidatus* Planktophila versatilis" (25), while Actinobacterium-acI-Baikal-G4 affiliates with Baltic Sea MAGs inside the acIA lineage. Actinobacterium-acI-Baikal-G1, Actinobacterium-acI-Baikal-G2, and Actinobacterium-acI-Baikal-G3 are phylogenetically close to a Lake Soyang (South Korea) representative (26) and to the freshwater *Actinobacteria* acAMD-5 and acAMD-2 from the Amadorio reservoir (Spain) (22).

We also reconstructed novel genomes inside the *Planctomyces-Verrucomicrobia-Chlamydia* (PVC) superphylum. Seven representatives were similar to the still poorly studied freshwater *Verrucomicrobia* (Fig. S4). Six of these genomes affiliate with *Opitutae* representatives and one genome affiliates with *Pedosphaera parvula*, with all of them having their closest relatives in temperate freshwater Spanish reservoirs (100). We also assembled three members of the *Planctomycetes* phylum (Fig. S5), with their nearest relatives coming from diverse environments, like water treatment plants (BioProject no. PRJNA301005), an algal reef in northern Florida (BioProject no. PRJNA281489), and soil (BioProject no. PRJNA311679).

Lake Baikal *Bacteroidetes* MAGs were classified inside the *Flavobacteriales* (5 MAGs) and *Chitinophagaceae* (1 MAG) families (Fig. S6). Two of them, Flavobacteriales-Baikal-G1 and Flavobacteriales-Baikal-G2 showed similarity with a *Bacteroidetes* bacterium, UKL13-3, obtained from Klamath Lake (27). On the other hand, Flavobacteriales-Baikal-G4 affiliates closely with a *Cryomorphaceae* representative from the Baltic Sea. The other two *Flavobacteriales* (Flavobacteriales-Baikal-G3 and Flavobacteriales-Baikal-G5) are phylogenetically close to *Flavobacteriales* bacterium BRH_c54 (obtained from a rock porewater metagenome, BioProject no. PRJNA257561) and *Flavobacterium terrae*, isolated from greenhouse soil (28), respectively. The only representative from the *Chitinophagaceae* family affiliates closely with a bacterium from a Kulunda Steppe salt lake (BioProject no. PRJNA286221).

Representatives of autotrophic picoplankton (genera *Synechococcus* and *Cyanobium*) were previously recorded in ice communities in this region (14). The two *Cyanobacteria* assembled here both affiliate with the 5.2 subcluster (Fig. S7), which comprises euryhaline, marine, brackish, and freshwater strains (29). The phylogenetically closest organism found to MAG *Synechococcus* sp. Baikal-G1 was *Synechococcus* sp. CB0101, isolated from the Chesapeake Bay (PRJNA46503), while the closest genome to MAG *Cyanobium* sp. Baikal-G2 was that of the Baltic Sea MAG cyanobacterium BACL30 MAG-120619-bin27 (23).

With regard to *Proteobacteria*, we were able to assemble one *Betaproteobacteria* genome (Fig. S8) and two *Alphaproteobacteria* genomes (Fig. S9). The *Betaproteobacteria* genome assembled was relatively large (3.4 Mb), with strikingly high gene similarities to *Bordetella* representatives, especially to *Bordetella petrii*, which has been described as a mosaic versatile microbe with features typical of environmental bacteria

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

TABLE 1 Summary statistics of the 35 Lake Baikal MAGs

MAG	Assembly size (Mb)	GC content (%)	Median intergenic spacer (bp)	Completeness (%) in CheckM (% of contamination)	Estimated genome size (Mb) (CheckM)	Closest MAG/SAG/cultured organism taxonomy	Origin of closest organism (source or reference)
Actinobacterium-acI-Baikal-G5	0.87	48.30	11	57.14 (5.93)	1.53	"*Ca.* Planktophila versatilis"	Lake Zurich (Neuenschwander et al. [25])
Acidimicrobium-Baikal-G1	0.98	51.36	17	67.06 (0.43)	1.47	*Acidimicrobium* sp. BACL27 MAG-120823-bin4	Baltic Sea (Hugerth et al. [23])
Actinobacterium-acI-Baikal-G4	0.89	49.16	9	64.59 (7.19)	1.38	*Actinobacteria* bacterium BACL15 MAG-120619-bin91	Baltic Sea (Hugerth et al. [23])
Actinobacterium-acI-Baikal-G1	1.31	52.48	11	70.59 (2.19)	1.85	*Actinobacteria* bacterium IMCC26077	Lake Soyang (Kang et al. [26])
Actinobacterium-acI-Baikal-G2	1.02	53.62	21	55.08 (1.28)	1.85	*Actinobacteria* bacterium IMCC26256	Lake Soyang (Kang et al. [26])
Actinobacterium-acI-Baikal-G3	1.26	56.93	20	55.47 (0)	2.27	*Actinobacterium* acAMD-5	Amadorio reservoir (Ghai et al. [22])
Thermoleophilia-Baikal-G1	1.52	64.06	21	69.89 (12.18)	2.17	*Gaiella* sp. SCGCAG-212-M14	Soil (BioProject no. PRJNA311679)
Acidimicrobium-Baikal-G2	1.46	51.16	12	45.31 (9.57)	3.22	*Acidimicrobium* sp. BACL19MAG-120924-bin39	Baltic Sea (Hugerth et al. [23])
Opitutae-Baikal-G1	1.55	63.80	23	67.32 (42)	2.34	Opitutae-AMD-G1	Tous reservoir (Cabello-Yeves et al. [100])
Opitutae-Baikal-G2	1.04	61.35	43	45.51 (0)	2.29	Opitutae-AMD-G3	Tous reservoir (Cabello-Yeves et al. [100])
Pedosphaera-Baikal-G1	2.25	64.27	71	62.24 (2.7)	3.67	Pedosphaera-Tous-C6FEB	Tous reservoir (Cabello-Yeves et al. [100])
Opitutae-Baikal-G3	2.64	62.82	77	70.41 (0.68)	3.77	Opitutae-Tous-C4FEB	Tous reservoir (Cabello-Yeves et al. [100])
Opitutae-Baikal-G4	2.42	62.68	85	58.84 (2.74)	4.11	Opitutae-Tous-C4FEB	Tous reservoir (Cabello-Yeves et al. [100])
Opitutae-Baikal-G5	0.99	62.95	34	46.76 (0)	2.11	Opitutae-AMD-G1	Tous reservoir (Cabello-Yeves et al. [100])
Opitutae-Baikal-G6	0.82	48.37	20	60.34 (0)	1.37	Opitutae-Tous-C2FEB	Tous reservoir (Cabello-Yeves et al. [100])
Nitrosoarchaeum-Baikal-G1	1.19	31.02	42	99.03 (1.94)	1.21	"*Ca.* Nitrosoarchaeum koreensis" MY1	Soil (BioProject no. PRJNA67913)
Thaumarchaeota-Baikal-G2	1.13	30.27	39	99.03 (0)	1.15	Casp-Thauma-1	Caspian Sea (Mehrshad et al. [24])
Nitrospirae-Baikal-G1	1.67	57.91	57	78.59 (0.91)	2.14	*Nitrospira* sp. SCGC AG-212-E16	Soil (BioProject no. PRJNA311679)
Gemmatimonadetes-Baikal-G1	2.05	65.52	39	55.72 (0)	3.67	*Gemmatimonas phototrophica*	Swan Lake, Gobi Desert (Zeng et al. [36])
Gemmatimonadetes-Baikal-G2	2.69	64.44	38	92.31 (8.79)	2.93	*Gemmatirosa kalamazoonesis*	Soil (BioProject no. PRJNA194094)
Acidobacterium-Baikal-G1	3.05	65.13	26	91.4 (5.13)	3.34	*Luteitalea pratensis*	Soil (Vieira S. et al. [32])
Pelagibacteraceae-Baikal-G1	1.14	28.68	5	90.97 (8.53)	1.26	*Pelagibacteraceae* bacterium BACL5 MAG-121015-bin10	Baltic Sea (Hugerth et al. [23])
Rhodospirillaceae-Baikal-G1	2.92	55.66	50	96.02 (1.39)	3.05	*Caenispirillum salinarum*	Coastal seawater (BioProject no. PRJNA176297)
Alcaligenaceae-Baikal-G1	3.37	53.02	40	98.36 (3.68)	3.45	*Bordetella* sp. FB-8	Former uranium-mining district, Ronneburg, Germany (BioProject no. PRJNA187096)
Synechococcus sp. Baikal G1	1.78	66.24	59	89.54 (5.57)	2.05	*Synechococcus* sp. CB0101	Chesapeake Bay (BioProject no. PRJNA46501)
Cyanobium sp. Baikal G2	1.16	63.64	30	59.24 (1.63)	2.08	*Cyanobacterium* BACL30 MAG-120619-bin27	Baltic Sea (Hugerth et al. [23])
Planctomycetaceae-Baikal-G1-1L	1.46	59.72	39	93.04 (0.91)	3.63	*Planctomycetaceae* bacterium BBD1991-11	Algae reef in northern Florida (BioProject no. PRJNA281489)
Planctomycetaceae-Baikal-G4R	2.44	44.44	51	55.13 (1.16)	4.46	*Planctomycetaceae* bacterium SCGCAG-212-F19	Soil (BioProject no. PRJNA311679)
Planctomycetaceae-Baikal-G1-2R	3.38	57.42	73	93.04 (13.22)	3.63	*Planctomycetaceae* bacterium Ga0077529	Drinking water treatment plant (BioProject no. PRJNA301005)
Flavobacteriales-Baikal-G1	1.46	42.67	31	73.59 (2.86)	1.98	*Bacteroidetes* bacterium UKL13-3	Klamath Lake (BioProject no. PRJNA290651)
Flavobacteriales-Baikal-G2	1.92	40.67	39	94.36 (5.48)	2.04	*Bacteroidetes* bacterium UKL13-3	Klamath Lake (BioProject no. PRJNA290651)
Flavobacteriales-Baikal-G3	1.65	35.85	56	70.77 (2.54)	2.34	*Flavobacteriales* bacterium BRH_c54	Rock porewater (BioProject no. PRJNA257561)
Flavobacteriales-Baikal-G4	1.36	55.00	14	80.77 (0.56)	1.69	*Cryomorphaceae* bacterium BACL18 MAG-120924-bin36	Baltic Sea (Hugerth et al. [23])
Flavobacteriales-Baikal-G5	0.95	31.84	60.5	52.92 (0.53)	1.79	*Flavobacterium terrae*	Greenhouse soil (BioProject no. PRJNA331506)
Chitinophagaceae-Baikal-G1	1.61	37.29	52	65.31 (3.2)	2.46	*Chitinophagaceae* bacterium T5-Br10_B2g1	Kulunda Steppe salt lake (BioProject no. PRJNA286221)

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

and virulence traits of pathogenic bordetellae (30). We also reconstructed an alphaproteobacterial genome with its closest affiliation to *Caenispirillum salinarum* (isolated from coastal seawater) and a novel SAR11 member which affiliates inside the subclasses I and II together with Baltic Sea *Pelagibacteraceae* MAGs (23) and marine *Pelagibacter ubique* strains, being a unique freshwater representative of the SAR11 subclasses I and II (see below).

One member each of the *Acidobacteria* and *Nitrospirae* phyla was also assembled (Fig. S10 and S11). The closest *Nitrospirae* representative to the Lake Baikal MAG is a soil single-cell amplified genome, that of *Nitrospira* sp. SCGCAG-212-E16 (BioProject no. PRJNA311679) (31), while one of the nearest isolates to Acidobacteria-Baikal-G1 was described as *Luteitalea pratensis*, a temperate grassland novel acidobacterium classified into the subdivision 6 *Acidobacteria* (32). *Gemmatimonadetes*, a poorly described phylum that has not received much attention so far, has been detected only in soil, making up 2% of the total soil bacteria (33), wastewater treatment plants, and freshwater and saline lakes (34), but it is absent in marine systems. Here, we were able to reconstruct two novel genomes (Fig. S12), one of them clearly related to *Gemmatimonadetes phototrophica*, a photosynthetic organism containing pheophytin-quinone-type photosynthetic reaction centers from phototrophic *Proteobacteria* (35), isolated from Swan Lake in the western Gobi Desert (36). The remaining *Gemmatimonadetes* appear to be more closely related to the soil strain *Gemmatirosa kalamazoonensis* (37). A growing number of *Gemmatimonadetes* isolates and MAGs are expected to be discovered from different freshwater environments in the next future.

The two archaeal representatives assembled here affiliate inside the *Thaumarchaeota* class (Fig. S13). One of these genomes is phylogenetically close to the soil representative "*Candidatus* Nitrosoarchaeum koreensis" MY1. Remarkably, the other archaeal MAG showed a 99% average nucleotide identity (ANI) with Casp-Thauma-1, a MAG reconstructed from the Caspian Sea (24), which is a case of a cosmopolitan thaumarchaeon that appears to be able to adapt from the extremely cold freshwater of the Lake Baikal to the permanently brackish temperate waters of the Caspian Sea.

Lake Baikal waters have low salinity, and the total concentrations of dissolved salts are approximately 100 mg/liter (10, 12, 38, 39); the salinity level is constant throughout the year in the pelagic area of the lake. However, salinity increases in certain biotopes, including the zones of ice community formation from the 4- to 32-m-depth layers. Still, in our case, the ionic content of the samples was determined and did not exceed the threshold of 100 mg/liter of ions in either of the two samples (Fig. 2). Accordingly, most of the Lake Baikal MAGs showed highest similarities with other freshwater microbes, such as acI *Actinobacteria* from Lake Zurich and Lake Soyang, *Verrucomicrobia* from the Tous and Amadorio reservoirs, and *Flavobacteriales* from Klamath Lake. However, it was highly noticeable that some members reconstructed here showed close similarities to marine or salt-adapted microbes. We noticed that certain of the phylogenetically closest microbes to those reconstructed here were obtained from a brackish environment (Baltic Sea) that in spite of being connected to the global ocean was the water body where some of the microbes most closely related to the Lake Baikal dwellers were found. Specifically, the Baltic Sea samples analyzed contained salinity between 5 and 7 ppt. The Baltic Sea is partially covered in ice for significant periods of the year, and temperature-wise, it is among the most similar water bodies that have been studied by similar approaches (MAG analysis) (23).

Incomplete Lake Baikal MAGs. In general, the reconstruction of genomes from the most abundant phyla *Actinobacteria*, *Bacteroidetes*, and *Verrucomicrobia* corresponds with the observations of their abundance by the 16S rRNA read analysis. However, we experienced more difficulties in getting relatively complete MAGs from some bacterial phyla, like *Proteobacteria*, which are abundant not only in Lake Baikal but in other freshwater metagenomes worldwide (see above). We found it particularly difficult to bin *Alphaproteobacteria* and *Betaproteobacteria* MAGs with >40% completeness. In the case of the *Betaproteobacteria*, we only got one MAG (Alcaligenaceae-Baikal-G1), but we

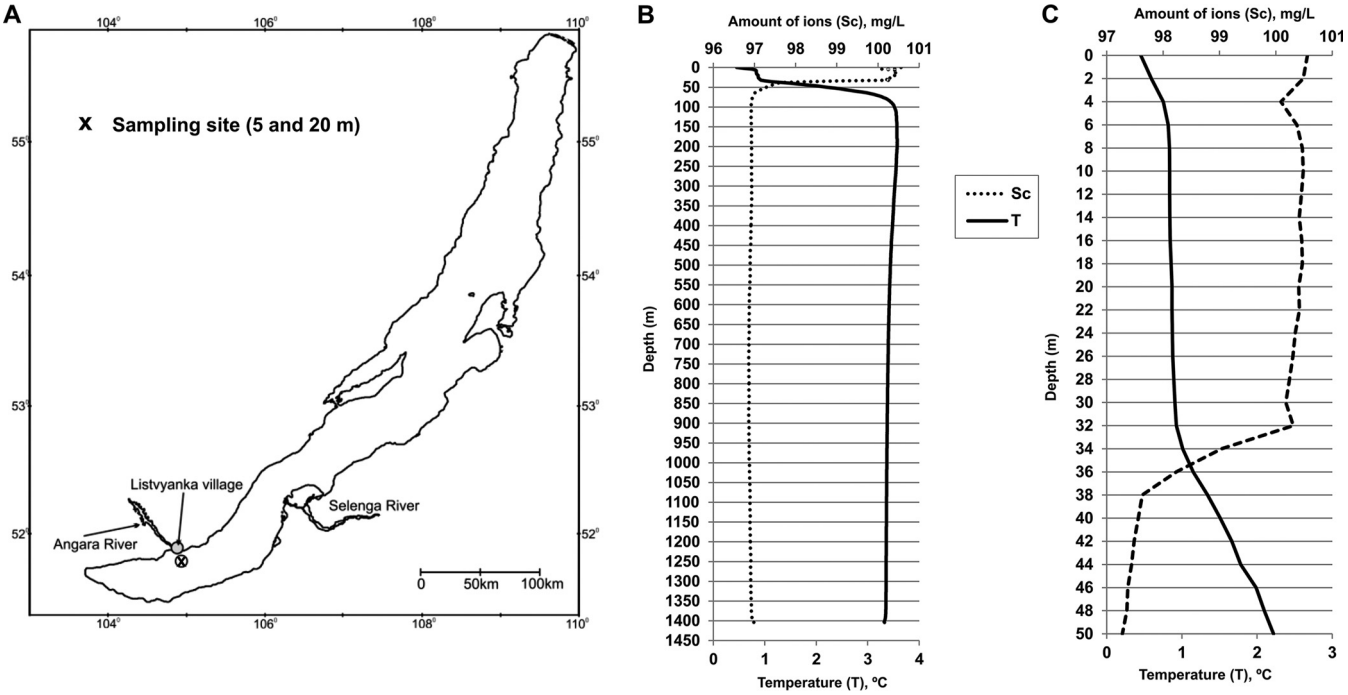


FIG 2 (A) Sampling point at the Lake Baikal station of the ice camp, located 7 km from the Listvyanka settlement (51°47.244′ 104°56.346′). (Modified from reference 3 with permission of the publisher.) (B and C) Measurements of temperature (°C) and total amount of ions (in milligrams per liter) (m) along the water column at the time and site of sampling at 0 to 1,450 m depth (B) and 0 to 50 m depth (C).

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

observed some contigs that although not reaching the 40% threshold were significant. For example, up to 0.7 Mb of contigs affiliating closely with *Methylopumilus planktonicus* (40) and some others affiliating with *Polynucleobacter* species (41–44) were detected. Not all community members can be reconstructed from metagenomes due to their complex population structure. This seems to be the case for the *Polynucleobacter* or *Methylopumilus* relatives, which are abundant and cosmopolitan freshwater microbes (40, 41). In the case of *Gammaproteobacteria*, we observed neither a high percentage of 16S reads nor assembled contigs in the Lake Baikal data sets.

Predominance of small genomes in an ultraoligotrophic environment. Considering that Lake Baikal is among the most oligotrophic lakes in the world, it is not surprising that most of the reconstructed genomes shown here are small (45), especially those of the phyla *Actinobacteria*, *Bacteroidetes*, *Thaumarchaeota*, *Nitrospirae*, *Cyanobacteria*, and *Verrucomicrobia*, which are also the most abundant microbes in this freshwater system based on our 16S rRNA classification and total number of assembled contigs. We can observe two clearly differentiated MAG groups, first, 24 genomes with estimated genome sizes of <2.7 Mb (mainly *Actinobacteria*, *Bacteroidetes*, *Thaumarchaeota*, *Cyanobacteria* and some of the *Verrucomicrobia*) and 11 genomes with estimated genome sizes above the standard average (3 Mb), belonging to *Planctomycetes*, *Verrucomicrobia*, *Betaproteobacteria*, *Acidobacteria*, and one *Alphaproteobacterium* (see Fig. S14 in the supplemental material).

The case of aquatic *Actinobacteria* having small genomes has long been known, being the acI lineage the most abundant group of small microbes in different freshwater ecosystems (46), like Soyang Lake (26), the Amadorio reservoir (22), and Lake Zurich (25). Here, we have reconstructed actinobacterial genomes inside the acI lineage comprising genome sizes 1.1 to 1.9 Mb and small median intergenic spacers (9 to 20 bp), the typical pattern observed for acI lineage freshwater genomes which show a high degree of streamlining (Fig. S15A in the supplemental material). On the other hand, we have reconstructed two *Acidimicrobium* genomes with estimated sizes of 1.2 and 2.3 Mb and the related *Thermoleophilia* organism that has an estimated genome size of

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

2 Mb. Here, we have also observed a pattern of small *Bacteroidetes* genomes (*Flavobacteriales*) with sizes of 1.9 to 2.4 Mb. It is also noticeable that the small *Bacteroidetes* genomes reconstructed have low GC content (31 to 42%). As shown in Fig. S15B, the smallest *Bacteroidetes* genomes inside the *Flavobacteriales* family have been assembled from the Baltic Sea and Lake Baikal.

Verrucomicrobia were described as being very diverse and abundant in many freshwater ecosystems and comprised different genome size ranges from small to large microbes (100). Here, we observed the same pattern found in temperate freshwater ecosystems, with the predominance of either small (4 MAGs) or large (3 MAGs) *Verrucomicrobia* genomes. *Thaumarchaeota* MAGs exhibit small estimated genome sizes, as is the case of most described members of this phylum, and were assembled at >95% completeness. Nitrospirae-Baikal-G1 was found to be (with its 2.13 Mb of estimated genome size) the smallest *Nitrospirae* of all soil and aquatic genomes analyzed thus far (Fig. S15C).

Novel truly freshwater representative of the marine SAR11 subtype I-II clade. Remarkably, we were able to assemble a novel member of the *Pelagibacteraceae* family which affiliates closely with *Pelagibacter* MAGs from the Baltic Sea (but ANI below 80%) and the marine *Pelagibacter ubique* HTCC strains. To our knowledge, this is a new truly freshwater *Pelagibacter* representative, since the Baltic Sea is connected to the global ocean, contradicting the classical view that the LD12 clade is the only SAR11 relative in freshwater bodies.

Since its discovery (47), *Pelagibacter ubique* has been extensively studied because of its unique features of genome streamlining (45) and abundance in the oceans worldwide, being probably one of the most abundant microbes on Earth (48, 49). Subtypes I and II of the SAR11 clade have been considered exclusive to offshore marine waters (50), while clade IIIa comprises representatives from brackish waters (SAR11-HIMB114) (51) and from the Arctic Ocean (SAR11-IMCC9063) (52). On the other hand, the LD12 clade has been described as the freshwater SAR11 subtype IIIb, being the most abundant *Alphaproteobacteria* group in freshwater bodies (53, 54). In this work, we have assembled a novel freshwater SAR11 member which affiliates inside the marine subclades I and II together with Baltic Sea MAGs (23). This discovery is even more surprising considering that it comes from Lake Baikal, distant many thousand kilometers from the closest marine waters.

The MAG of this novel Pelagibacteraceae-Baikal-G1 represents the 90% of a putative genome of 1.25 Mb (as estimated by CheckM; see Materials and Methods), with 1,193 predicted coding sequences (CDSs) and a median intergenic spacer of 5 bp, which confirms the same pattern of genome streamlining seen in marine *Pelagibacter ubique* (45). As presented in Fig. 3A, the closest affiliations with Pelagibacteraceae-Baikal-G1 were two Baltic Sea MAGs (23) inside clades I and II, which demonstrates that these contain marine, brackish, and now freshwater representatives. To increase the robustness of this placement, we also added LD12 single-cell amplified genomes (SAGs) (53) to the whole SAR11 phylogeny with the PhyloPhlAn tool (see Materials and Methods), obtaining a tree with topology practically identical to the protein concatenation-based tree made with 83 different COGs shown in Fig. S9.

The isoelectric point of proteins is generally associated with the salinity of the natural environment of microbes (55). The isoelectric point evaluated for all the predicted proteins for several representatives of SAR11 subtypes I, II, IIIa, and IIIb (Fig. 3B) clearly shows a distinction between the freshwater Pelagibacteraceae-Baikal-G1 and the marine *Pelagibacter*. Apart from the freshwater LD12 SAGs, only the brackish SAR11-HIMB114 (51), Qinghai Lake SAR11-QL1 (56), and the Arctic Ocean SAR11-IMCC9063 (52) representatives display similar isoelectric patterns with shifts toward basicity similar to the freshwater Pelagibacteraceae-Baikal-G1. The contrast between freshwater and marine microbes was also evident, being noteworthy the shift toward basicity in freshwater LD12 SAGs and the other freshwater and brackish genomes. The truly marine genomes (HTCC strains) and some brackish MAGs (coming from the Baltic Sea) show different patterns, with approximately 10% more acidic proteins (around 4.5

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

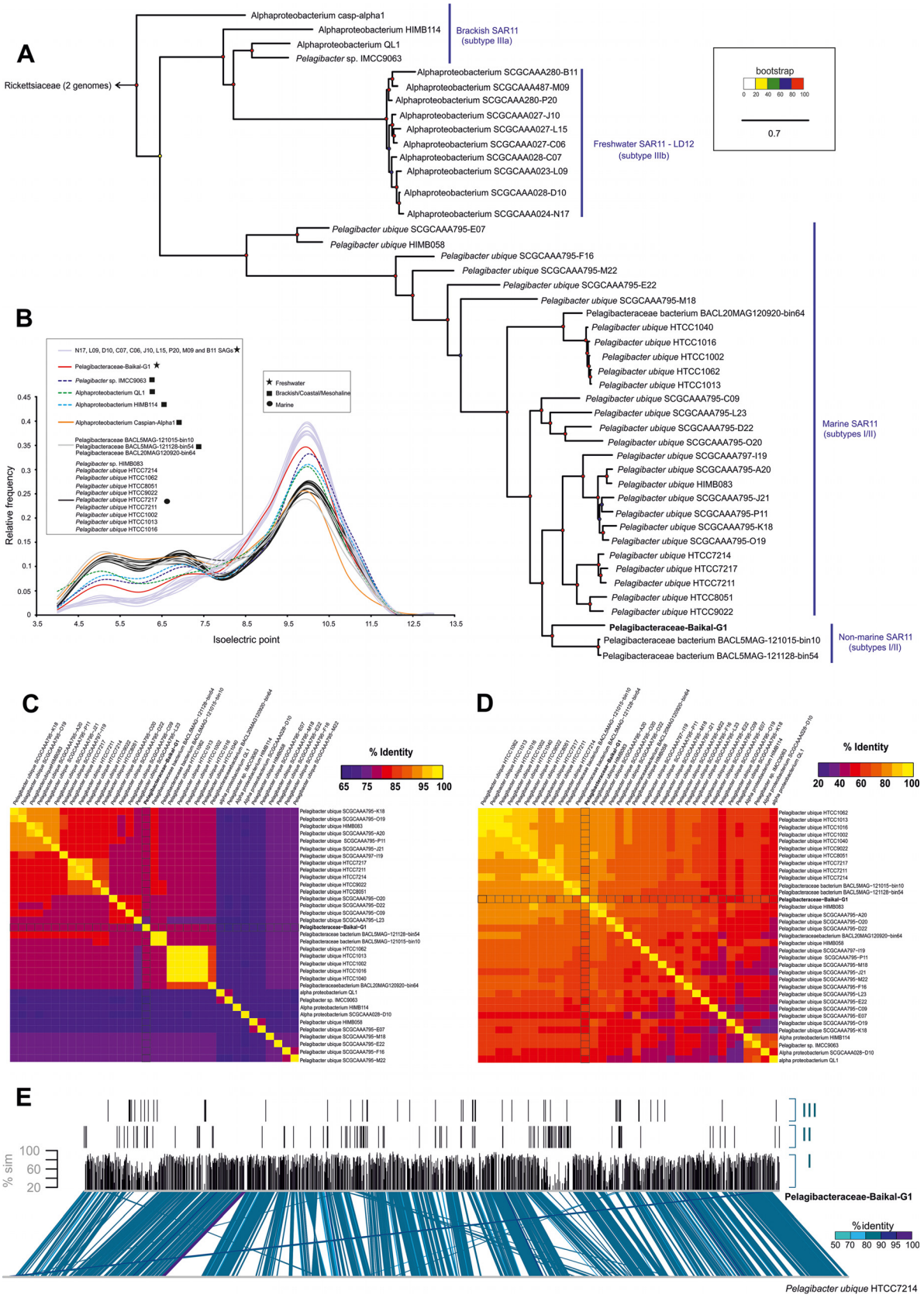


FIG 3 (A) Phylogenomic tree of Lake Baikal SAR11 reconstructed genome together with clade I/II, IIIa, and IIIb representatives. Two Rickettsiaceae genomes were used to root the tree. (B) Protein relative frequencies versus isoelectric point (IP) plot evaluated on a subset of marine, brackish, and freshwater SAR11 representatives. (C and D) Average nucleotide identity (ANI) (C) and average amino acid identity (AAI)

(Continued on next page)

to 5 isoelectric points) and 10% less basic proteins (around 9.5 to 10 isoelectric points). These data point to a key role of the protein charge in adaptation to different salinities in their environment, even at relatively small changes, like those between brackish and freshwater environments, as has been detected for other microbes (57).

Remarkably, we observed low average nucleotide identity (ANI) between this novel microbe and its closest representatives from the Baltic Sea or marine *P. ubique* HTCC strains, always being less than 77% (Fig. 3C). The ANIs between the Lake Baikal MAG and the LD12 SAGs were in all cases lower than 70%. Average amino acid identity (AAI) values were also relatively low compared to the SAR11 cluster (Fig. 3D). Despite the low ANI and AAI values compared to the known SAR11 representatives, the synteny of the 1,193 predicted CDSs is remarkably conserved with other *P. ubique* genomes, exemplified by the cultured strain HTCC7214 that we used as a reference. We performed BLASTN (>70% of identity threshold hits and >200 bp of alignment lengths) (see Fig. 3E) and TBLASTX comparisons (50% of similarity hits and >200 bp of alignment length) between the two microbes (Fig. 3E-I). The genome of Pelagibacteraceae-Baikal-G1 (Fig. 3E-II) contains only 120 genes not present in the reference HTCC7214, including several transporters and the entire carbon monoxide dehydrogenase cluster (which is present in some other *P. ubique* strains). The genome also encodes 51 proteins that have not been identified in any other sequenced *Pelagibacter* genome (Fig. 3E-III and Table S1). Among these, a complete *hisHAFI* operon shows that this freshwater strain may be prototrophic for histidine, in spite of being the third less abundant amino acid of its whole proteome. Also exclusive to this MAG is the presence of an *ureABDE* operon (urea transport) and other transporters that extend the ability of Pelagibacteraceae-Baikal-G1 to import organic compounds directly from the environment. In summary, in spite of the low ANI value and the differences of net charge of the proteome, the overall gene content and synteny are remarkably conserved in Pelagibacteraceae-Baikal-G1 compared to other *Pelagibacter* genomes.

We also compared the local synteny between the Lake Baikal representative and the more distant freshwater LD12 alphaproteobacterium SCGCAAA028-D10 and the closest Baltic *Pelagibacteraceae* bacterium BACL5MAG-121128-bin54 (Fig. S16). It is evident that Pelagibacteraceae-Baikal-G1 shows higher shared genomic content with its marine (e.g., *P. ubique* HTCC7214, as shown in Fig. 3D) and Baltic relatives (Fig. S16). The lower shared genomic content between the LD12 freshwater representative from Lake Mendota and the Lake Baikal representative (Fig. S16) also reflects the genetic distance between these two freshwater SAR11 lineages and confirms the taxonomic placement of the novel Lake Baikal SAR11 inside the traditional marine clades I and II. Furthermore, its origin as a freshwater-adapted microbe, showing the typical basicity shift pattern observed in the proteomes of freshwater microbes (although with a slight decrease in basicity compared to LD12), expands the diversity of marine subclass I-II inside the SAR11 lineage. Considering that Lake Baikal is among the most ancient lakes in the world, it is puzzling how this *Pelagibacter* kept the vast majority of genomic content of its closest marine representatives rather than acquired genetic material from freshwater relatives. The discovery of a truly freshwater SAR11 with closest synteny and core genome to marine and brackish SAR11 genomes opens new perspectives on the evolutionary models interconnecting marine and freshwater microbes.

MAG key metabolic pathways. A summary of the metabolic features of all reconstructed bins is shown in Fig. 4. It must be considered that all metabolic pathways displayed here have been robustly found in the different MAGs, although a note of caution must be added, considering that some metabolic potential could have been missed because of the incompleteness of the MAGs.

FIG 3 Legend (Continued)
(D) between Pelagibacteraceae-Baikal-G1 and a subset of SAR11 clade I/II, IIIa, and IIIb reference genomes. (E) Alignment of the Pelagibacteraceae-Baikal-G1 to *Pelagibacter ubique* HTCC7214 by BLASTN with >70% identity hits and >200 bp of alignment length. (E-I) Location and similarity of hits TBLASTX, >50% and >200 bp of alignment length. (E-II) Locations of 120 genes from Pelagibacteraceae-Baikal-G1 that do not match with the reference *Pelagibacter ubique* HTCC7214. (E-III) Location of Pelagibacteraceae-Baikal-G1 51 unique genes absent in other SAR11 clade I-II genomes. sim, similarity.

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

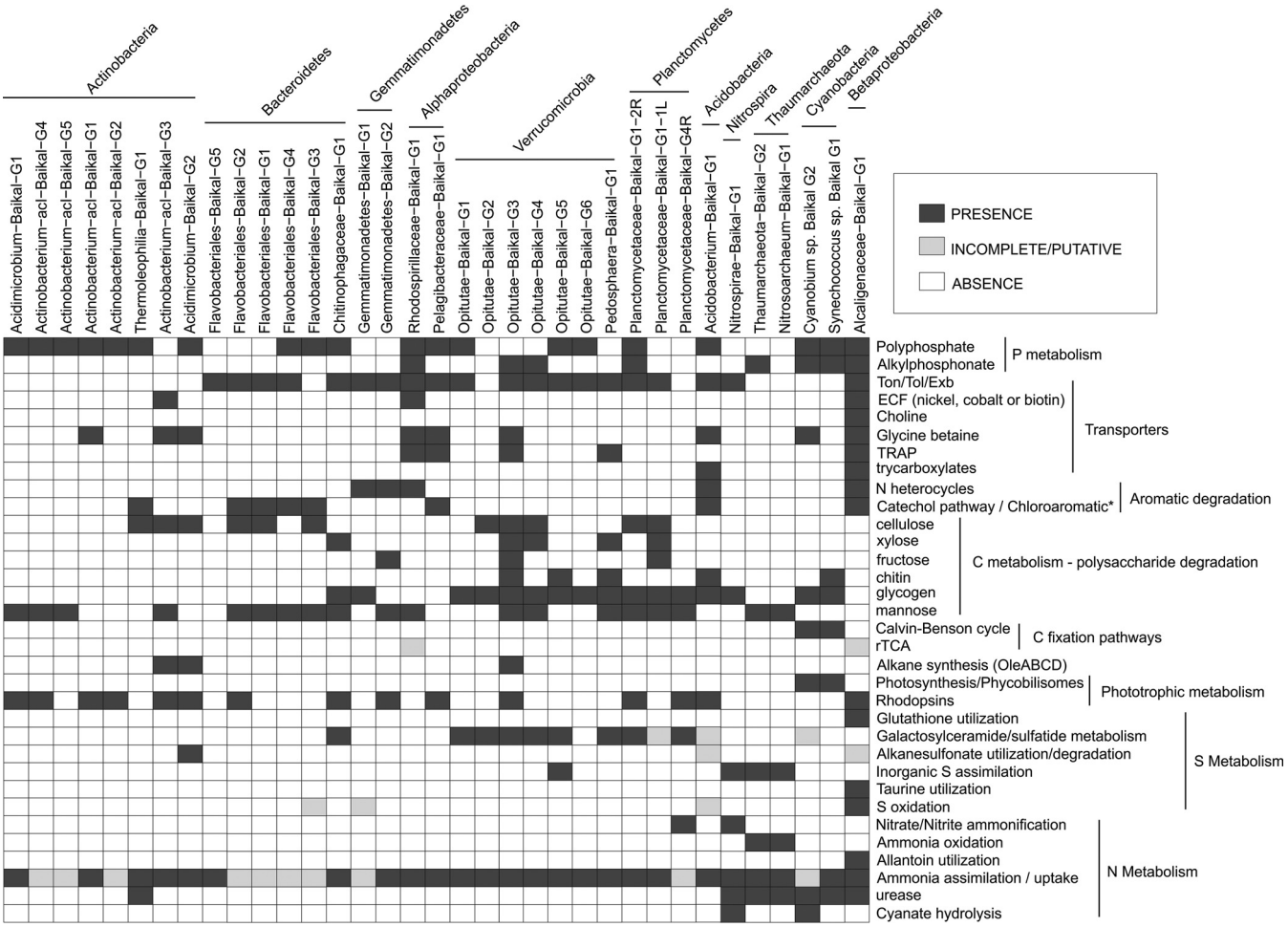


FIG 4 Summary of different metabolic pathways found in the 35 Lake Baikal MAGs. The presence of a pathway is denoted by black boxes. The absence of a pathway is denoted by gray boxes. Incomplete or putative pathways are denoted by white boxes. The incompleteness of MAGs has to be considered when assessing the absence/incompleteness of metabolic pathways.

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

Carbon fixation pathways. Among all the reconstructed MAGs, we were able to identify the Calvin-Benson cycle only in *Cyanobacteria*. However, some of the bins showed evidence for certain alternative carbon fixation pathways, like reverse tricarboxylic acid (rTCA) in Alcaligenaceae-Baikal-G1 and Rhodospirillaceae-Baikal-G1, which contains the rTCA key enzymes fumarate reductase, 2-oxoglutarate:ferredoxin oxidoreductase, ATP citrate lyase, and a citryl-coenzyme A (citryl-CoA) lyase (58). We did not observe any pheophytin-quinone-type photosynthetic reaction centers in the *Gemmatimonadetes phototrophica* genome relative reconstructed from Lake Baikal.

Organic matter degradation. Members of the PVC superphylum have been described as one of the major polysaccharide degraders (59). All *Verrucomicrobia* and *Planctomycetes* MAGs described here contain key enzymes and pathways for the effective degradation of at least two polysaccharides, disaccharides and amino sugars (cellulose, xylose, fructose, mannose, chitin, and glycogen). It is particularly interesting that Opitutae-Baikal-G3 contains putative pathways for the degradation of all five polymers, while Planctomycetaceae-Baikal-C1-1L has all of them except for chitin. It was remarkable the ability to degrade cellulose in some *Actinobacteria* and *Bacteroidetes* MAGs. However, from our data, the major polymer degraders seem to be *Planctomycetes* and *Verrucomicrobia*.

The degradation of aromatic compounds is very important in nature, since some N-heterocycles or chloroaromatic compounds are toxic for animals, plants, and humans and are hazardous contaminants to the environment (60). Here, we have found strong

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

evidence for N-heterocycle degradation in *Gemmatimonadetes*, Rhodospirillaceae-Baikal-G1, Acidobacterium-Baikal-G1, and the Betaproteobacterium Alcaligenaceae-Baikal-G1. The catechol degradation pathway was also detected in many *Bacteroidetes* MAGs, Pelagibacteraceae-Baikal-G1, three *Actinobacteria* MAGs, Acidobacterium-Baikal-G1, and Alcaligenaceae-Baikal-G1. Chlorophenols are toxic xenobiotics that certain microorganisms, like the betaproteobacterium *Alcaligenes xylosoxidans* and other *Alcaligenes* spp. can use as carbon input (61). Alacaligenaceae-Baikal-G1, the betaproteobacterium with closest affiliations with *Alcaligenes* and *Bordetella*, contains key enzymes for biphenyl and chlorophenol degradation, which highlights the importance of this bacterium as sentinel if it were to increase in numbers in Lake Baikal ultraoligotrophic clear waters used for human consumption or agriculture.

Transporters. Specific choline and ECF class (nickel, cobalt, or biotin) transporters were exclusively present in Alcaligenaceae-Baikal-G1. TonB transporters for iron and biopolymers uptake were ubiquitous in *Bacteroidetes*, *Verrucomicrobia*, *Planctomycetes*, *Gemmatimonadetes*, *Proteobacteria*, *Acidobacteria*, and *Nitrospirae*, while transporters for glycine betaine involved in osmoregulation were found in *Actinobacteria*, *Betaproteobacteria*, *Acidobacteria*, and only one *Verrucomicrobia* MAG. Tricarboxylate transporters were only found in Alcaligenaceae-Baikal-G1 and Acidobacteria-Baikal-G1. Tripartite ATP-independent periplasmic (TRAP) transporters, used to incorporate organic acids or molecules with carboxylate or sulfonate groups, were present in two *Verrucomicrobia* representatives and both *Alphaproteobacteria* and *Betaproteobacteria* MAGs, but the presence of more than 40 genes related to these transporters in Alcaligenaceae-Baikal-G1 was remarkable, suggesting that this microbe tends to accumulate and incorporate organic acids with carboxylate and sulfonate groups inside the cell.

Sulfur metabolism. We have observed certain patterns involving different pathways to efficiently oxidize and reduce sulfur intermediates in some of the MAGs, particularly in Alcaligenaceae-Baikal-G1, the two *Thaumarchaeota*, Nitrospirae-Baikal-G1, and some of the *Verrucomicrobia* representatives. For instance, the inorganic sulfur assimilation or assimilatory sulfate reduction to transform it to hydrogen sulfide via 3′-phosphoadenosine-5′-phosphosulfate (PAPS), sulfate adenylyltransferase (SAT), adenylyl sulfate reductase (APSR), ferredoxin sulfite reductase, and the ABC sulfate transporters were detected only in the *Nitrospirae* and *Thaumarchaeota* representatives. Alkanesulfonates are degraded by some lineages of freshwater *Actinobacteria* (22); here, alkanesulfonate monooxygenase was detected only in one acidimicrobium, although some enzymes involved in alkanesulfonate utilization were also detected in the betaproteobacterial and acidobacterial representatives. It is noticeable that *Verrucomicrobia* appear to be involved in the metabolism of particularly two glycosphingolipids, galactosylceramides and sulfatides, a pathway that is typically found in eukaryotes, plants (62), and algae (63), presenting several enzymes, like aryl-sulfatases, sialidases, and beta-galactosidases, which could be used by the *Verrucomicrobia* to effectively degrade these abundant plant and algae sulfur-containing lipids, as suggested before (100).

Sulfur oxidation pathways in freshwater ecosystems have been previously described in chemolithotrophic and phototrophic microbes (64) and particularly within the betaproteobacterial genera *Polynucleobacter* (41, 42) and *Sulfuricella* (65). Alcaligenaceae-Baikal-G1 shows a cluster of *sox* genes together with cytochrome *c* and sulfite reductases, which have its highest resemblance to the genes found in cosmopolitan freshwater *Polynucleobacter* strains (see Fig. S17 in the supplemental material). This microbe also shows a proteorhodopsin proton pump together with some key enzymes of the rTCA cycle for carbon fixation (see below), which suggests a photo- and chemolitotrophic metabolism. The capability to utilize taurine or glutathione as other S sources, together with the previously mentioned sulfur oxidation pathways, indicates that this bacterium has a key role in the S cycle of Lake Baikal, being capable of degrading and utilizing different sulfur sources. The *Gemmatimonadetes* MAGs also present sulfite dehydrogenase (SoxD), cytochrome *c* biogenesis protein

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

(CcdA), and other *sox* genes (*soxC* and *soxH*), although the whole sulfur oxidation pathway was incomplete.

Nitrogen metabolism. Some of the pathways involved in nitrogen metabolism found in Lake Baikal are very ubiquitous in nature, like ammonia assimilation and uptake (found in all the MAGs) or ammonia oxidation by *Thaumarchaeota* (66). In contrast, we have found certain pathways unique to some MAGs, most of them directly involving *Nitrospirae* and *Thaumarchaeota* representatives, which could be the key microorganisms in the N cycle of Lake Baikal. For instance, *Nitrospirae* representatives are known for nitrification and commamox processes (67), although here, we found pathways involved in nitrate and nitrite ammonification through respiratory nitrate (NarGHIJC) and nitrite reductases in Nitrospirae-Baikal-G1. Assimilatory nitrate reductase and nitrite reductase were also found in the MAG Planctomycetaceae-Baikal-C4R. Other pathways, including that for urea degradation, were ubiquitous in *Cyanobacteria*, *Acidobacteria*, *Nitrospirae*, *Betaproteobacteria*, and *Thaumarchaeota* MAGs. Urea could be a particularly significant substrate for ammonia oxidizers when aquatic systems are ice covered, as already proposed for Arctic waters (68). On the other hand, cyanate hydrolysis giving rise to CO_2 and ammonia also occurs in Cyanobium-Baikal-G2 and Nitrospirae-Baikal-G1. It is clear that among all microbes reconstructed here, the *Nitrospirae* representative contains the widest set of pathways to utilize ammonia. Finally, we have found unique pathways involved in allantoin ($C_4H_6N_4O_3$) utilization as a N source in Alcaligenaceae-Baikal-G1.

Photoheterotrophy through rhodopsin pumps. Among the 35 reconstructed genomes presented here, we identified 15 MAGs containing rhodopsin pumps. This fact is remarkable because, although rhodopsins are extremely widespread in the photic zone of aquatic habitats, including lakes, the ice cover of Lake Baikal at the time of sampling would deprive microbes of a significant amount of light (up to 10 times less light when snow is accumulated to some extent, what is the usual situation) (13). Still, many of the genomes reconstructed here contained rhodopsins, indicating a possible photoheterorophic lifestyle. However, since the lake is not perennially covered with ice, during most months of the year, such rhodopsins may be able to harvest more intense solar radiation. Thus, as displayed in Fig. 5, the reconstructed genomes of Gemmatimonadetes-Baikal-G2, Opitutae-Baikal-G3 (*Verrucomicrobia*), Acidobacterium-Baikal-G1, Alcaligenaceae-Baikal-G1 (*Betaproteobacteria*), and Pelagibacteraceae-Baikal-G1 (*Alphaproteobacteria*) contained rhodopsins, all of which affiliate inside the proteobacterial proton pumps. Two of the *Bacteroidetes* (Flavobacteriales-Baikal-C2 and Chitinophagaceae-Baikal-C1) contain one rhodopsin, each which affiliates with *Bacteroidetes* proteorhodopsins. Six rhodopsins were found in *Actinobacteria* MAGs. Four of the reconstructed *Actinobacteria* inside the acI lineage contained rhodopsins which clearly affiliate with the actinobacterial xanthorhodopsins (69) (Actinobacterium-acI-Baikal-G1, Actinobacterium-acI-Baikal-G2, Actinobacterium-acI-Baikal-G3, and Actinobacterium-acI-Baikal-G4). The other two actinobacterial rhodopsins were found in *Acidimicrobidae* MAGs; Acidimicrobium-Baikal-G2 possesses a rhodopsin inside the actinobacterial acidirhodopsins (21), while Acidimicrobium-Baikal-G1 contained a novel acidirhodopsin which affiliates with other similar proteins from the Baltic Sea *Acidimicrobidae* MAGs (23). A novel branch of *Verrucomicrobia* rhodopsins which affiliate closely with the *Exiguobacterium* rhodopsins was discovered from Spanish freshwater reservoirs (100). Here, we also confirmed the presence of two *Planctomycetes* rhodopsins inside this novel proton pump affiliation, which could expand the branch for the general PVC superphylum rhodopsins, including *Planctomycetes* and *Verrucomicrobia* representatives.

The alignment of the rhodopsins described for these Lake Baikal MAGs (see Fig. S18 in the supplemental material) confirms that all of them are green-light-absorbing proton pumps based on the presence of the L105 or M105 residues in the retinal pocket (70). In this study, we observed that all rhodopsins retrieved from Lake Baikal assemblies are green-light variants, while no blue-light-absorbing types were found. As occurs with the *Exiguobacterium* rhodopsins, the two *Planctomycetes* rhodopsins differ from the rest of the green-light proton pumps because they present a K residue two

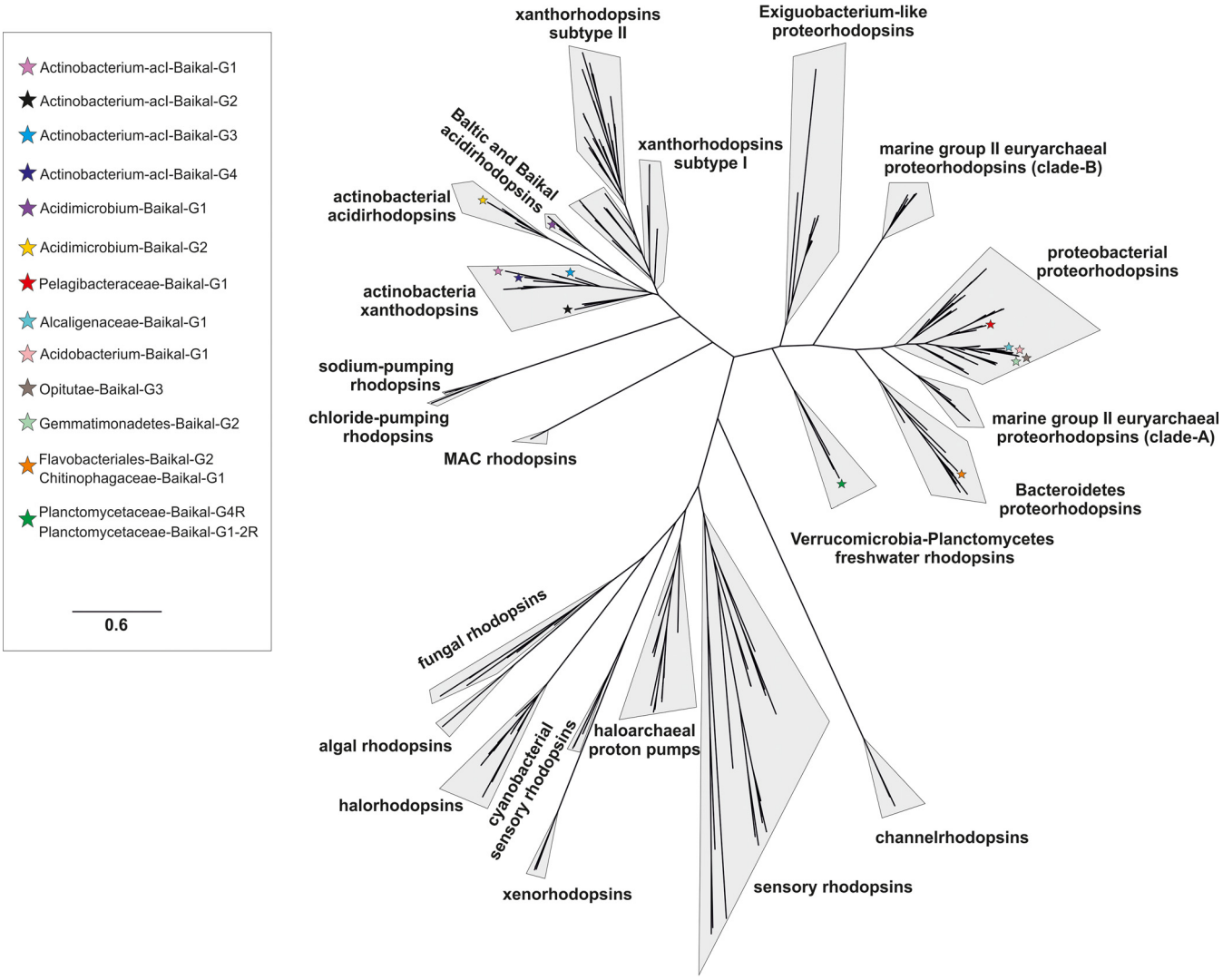


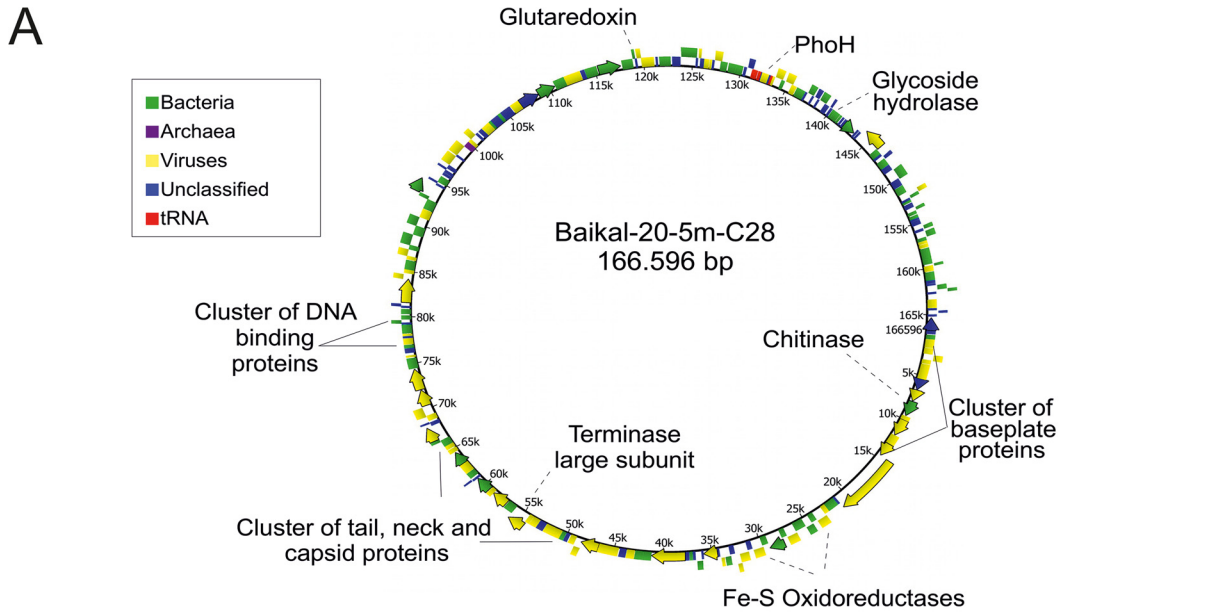
FIG 5 Rhodopsin phylogenetic tree made with >200 reference archaeal and bacterial rhodopsins. All known types of rhodopsin clades are included. Clade affiliations of the different rhodopsins from Lake Baikal MAGs are color coded and labeled with a star. MAC, marine actinobacterial clade.

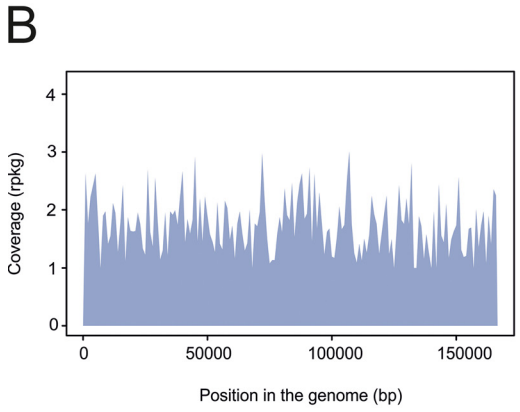
Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

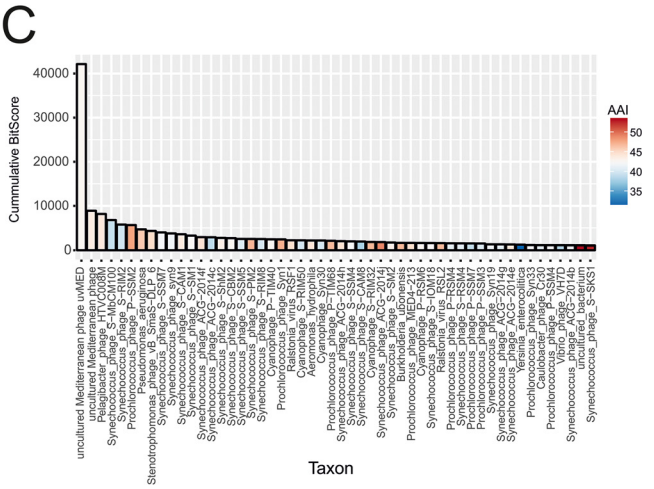
positions after the L/M105, which determines the green-light absorption. The meaning of this residue or whether it contributes to structural or biological features in these proteorhodopsins remains unknown.

Description of the Baikal-20-5m-C28 polynucleophage in Lake Baikal metagenomes. Among the viral scaffolds obtained from metagenome assembly, one (Baikal-20-5m-C28) is of particular relevance. It is a fragment of approximately 166 kbp, encoding 235 viral proteins and 6 tRNAs. The annotation of the proteins attests to the viral origin of this genome, which contains hallmark viral genes organized in the typical modular architecture of phage genomes, including DNA polymerase, terminase capsid, tail, and neck proteins (Fig. 6A). Read coverage across Baikal-20-5m-C28 was stable, and no spikes in coverage that are typical of chimeric contigs were detected (Fig. 6B). Querying the proteins from Baikal-20-5m-C28 against the NCBI NR database revealed strong similarity between its genome and other aquatic phage genomes (Fig. 6C). Finally, phylogenomic reconstruction based on the Dice method (71) placed Baikal-20-5m-C28 as a close relative of phage genomes discovered through metagenomics in both marine and freshwater habitats (Fig. 6D). Together, these results provide compelling evidence that Baikal-20-5m-C28 is a bona fide viral genome and not an artifact of sequence assembly. The overlap of the 5′ and 3′ ends of this sequence suggests that Baikal-20-5m-C28 is a complete circular phage genome.

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.







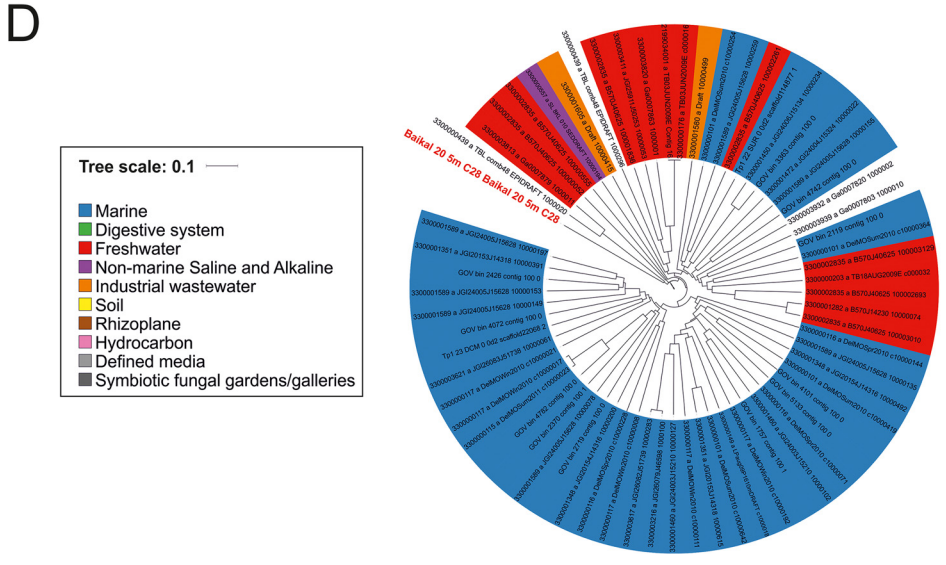


FIG 6 Baikal-20-5m-C28 polynucleophage genome from Lake Baikal. (A) Circular map of Baikal-20-5m-C28 phage genome displaying putative auxiliary metabolic genes and genes involved in phage replication and virion assembly. Genes are represented by arrows and color-coded

(Continued on next page)

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

One serine tRNA present in the Baikal-20-5m-C28 sequence matches the genomes of *Polynucleobacter necessarius* and *Polynucleobacter asymbioticus* with 100% identity and complete coverage, making these organisms the putative hosts of this phage. *Polynucleobacter* is a genus of chemoheterotrophic *Betaproteobacteria* that is abundant and widespread across global freshwater habitats (40, 41, 66). Despite this, no *Polynucleobacter* phages have been described to date. The genome of polynucleophage Baikal-20-5m-C28 includes several auxiliary metabolic genes, suggesting that it is capable of modulating host heterotrophic metabolism during infection (Fig. 6A), much like how cyanophages redirect host autotrophic metabolism toward pathways that favor viral replication (67). Among the annotated proteins encoded in Baikal-20-5m-C28 were a chitinase (CDS12) and a glycoside hydrolase (CDS 189). These proteins have polysaccharide-degrading activities; thus, they could provide *Polynucleobacter* with additional energy sources under the ultraoligotrophic conditions of Lake Baikal (7, 35) during the phage lytic cycle. A gene encoding phosphorus starvation inducible protein (PhoH) was also detected (CDS 170). This protein is involved in the process of scavenging of phosphorus (68, 69), an element that is essential for phage nucleic acid synthesis under low nutrient conditions. In addition, enzymes involved in redox reactions were also identified, namely, two Fe-S oxidoreductases (CDS 21 and 29) and glutaredoxin (CDS 143). Together, these observations suggest that during infection, the polynucleophage Baikal-20-5m-C28 uses auxiliary metabolic genes to enhance host nutrient uptake and utilization capabilities. The discovery of this phage and of its potential to modulate host metabolism during infection shed light onto the still poorly characterized repertoire of auxiliary metabolic genes present in freshwater phages and their potential to modulate host heterotrophic metabolism.

Abundance and cold adaptation of the novel microbes. In order to estimate the presence of the reconstructed genomes in different freshwater and brackish bodies, we performed fragment recruitment with >95% of identity values (i.e., above the species level). We used a wide variety of data sets (>150 different), ranging from tropical, to temperate, to cold (see Materials and Methods). We also assessed the presence of our MAGs in brackish systems, like the Baltic (23) and Caspian (24) Seas. Although we noticed a large number of contigs above these identities in the Baltic Sea and other freshwater data sets, these corresponded only to the 16S rRNA and other conserved genes. With these exceptions and above the 95% identity, we did not observe a significant presence of the Lake Baikal microbes in other environments (exceptions explained below). Hence, from what we know, the majority of the reconstructed genomes could be endemic to Lake Baikal, thus being microbes adapted to the cold and special hydrological and hydrochemical conditions existing in this lake.

Figure 7 shows the distribution pattern of each MAG in Lake Baikal 5- and 20-m-depth samples. Some of the acI lineage reconstructed MAGs were more abundant at 5 m depth than at 20 m depth (except for Actinobacterium-acI-G2, which is more abundant at 20 m). Opitutae-Baikal-G4 and Thermoleophilia-Baikal-G1 were found at higher reads per kilobase of genome per gigabase of metagenome (RPKG) also in the 5-m-deep layer. On the other hand, Gemmatimonadetes-Baikal-G1, Planctomycetes-Baikal-C1-2R, and Opitutae-Baikal-G3 appeared to be more abundant in the 20-m-depth waters. So far, the verrucomicrobial representative Opitutae-Baikal-G5 has been detected as the most abundant microbe in the 5- and 20-m-depth samples (between 60 and 210 RPKG), which also correlates with the high percentage of verrucomicrobial 16S rRNA fragments retrieved. The MAGs Opitutae-Baikal-G1 and Flavobacteriales-Baikal-C2

FIG 6 Legend (Continued)
according to the taxonomic affiliation of their best hits in the NCBI-NR protein database. Modules of proteins involved in baseplate, DNA binding, and tail, neck, and capsid proteins are highlighted. (B) Coverage plot along the Baikal-20-5m-C28 genome. (C) Bar plot displaying the cumulative BitScore and average amino acid identity (AAI) between Baikal-20-5m-C28 and genomes of phages and bacteria in NCBI-NR database. (D) Subset of the Dice phylogenomic tree displaying the placement of Baikal-20-5m-C28 and the closest relatives of these genomes. Branches are colored according to ecosystem source of the phage genomes.

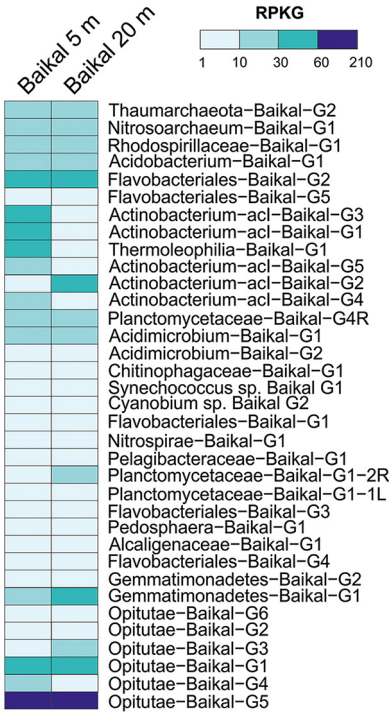


FIG 7 Metagenomic fragment recruitment of the 35 Lake Baikal MAGs (expressed as reads per kilobase of genome per gigabase of metagenome [RPKG]) in Lake Baikal 5- and 20-m-depth samples.

are the next most abundant microbes from both Lake Baikal samples (30 to 60 RPKG). Despite these findings, we could identify some microbes that were well adapted to different habitats, like the case of Thaumarchaeota-Baikal-G2, which was previously assembled in the Caspian Sea (24) and Opitutae-Baikal-G5, which shows similarities of 93% ANI with its relative in the temperate Spanish reservoir of Tous (100).

Many of the freshwater metagenomic data sets available thus far comprise temperate (North American and European), tropical (Amazon Lakes and Lake Gatun), and cold (North America, Sweden, and Finland) lakes. Nevertheless, an increase in high-latitude data sets is expected over the next few years. Future sampling on high-latitude lakes, especially during winter and sub-ice seasons, is crucial to establish relationships with the novel Lake Baikal microbes described in this paper. More Lake Baikal studies from both winter and summer seasons are under way.

MATERIALS AND METHODS

Sampling and metadata. Water samples were taken on 14 March 2016, using 4-liter bathometers (instrument similar to Niskin bottles), at the station of the ice camp, which was located 7 km from the Listvyanka settlement (51°47.244′N and 104°56.346′E). The ice thickness in the studied period was 72 cm, and the depth of the water column at the sampling site was 1,405 m. The water samples were taken from two depths, 5 and 20 m. 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. Within a few hours, the samples (30 liters) at a temperature of approximately 4°C were delivered to the laboratory. Then, each 30 liters of water was filtered through the net (size, 27 μm) and then filtered through nitrocellulose filters 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 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) and stored at −20°C (72). DNA was extracted according to a modified method of phenol-chloroform-isoamyl alcohol extraction (73), as was done with other freshwater samples (22, 57). The extracted DNA was stored at −70°C until further use. The DNA samples were placed in 70% alcohol solution and were forwarded to the laboratory.

Sequencing, assembly, and annotation pipeline. Sequencing was performed using Illumina HiSeq 3000/4000 (Oklahoma Medical Research Foundation, USA). A Kapa DNA library was used for the library preparation. A total of 210 and 236 million sequence reads (PE 2 × 150 bp) representing 23 and 26 Gb of sequence data were produced for Lake Baikal (0.22-μm fraction) 5- and 20-m-depth samples,

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

respectively. The assembly pipeline was conducted using two different approaches: first, each data set was assembled independently using the IDBA-UD assembler (74), with the following parameters: mink, 70; maxk, 100; step, 10; and precorrection. With this first approach, we obtained 4,028 and 4,572 contigs larger than 10 kb, with an average contig size of 19 kb. Second, we assembled the two samples together using the same parameters described above, obtaining a total of 7,863 contigs larger than 10 kb and an average contig length of 20.9 kb. Gene predictions on the assembled contigs were done using Prodigal in metagenomic mode (75), tRNAs were predicted using tRNAscan-SE (76), and ribosomal rRNA genes were identified using ssu-align (77, 78) and meta-rna (79). Comparisons of predicted protein sequences against NCBI NR, COG (80), and TIGRFAM (81) databases were performed for taxonomic binning and functional annotation. In order to bin the different microbial groups described here, we first grouped the annotated contigs using taxonomy, principal-component analysis of tetranucleotide frequencies, GC content, and coverage values in Lake Baikal metagenomes. Tetranucleotide frequencies were computed using the compseq program in the EMBOSS package (82). Principal-component analysis was performed using the FactoMineR package in R (83). BLASTN, BLASTP, and TBLASTX (84) searches were performed when necessary. The redundancy and duplicity of the different contigs from each reconstructed genome were eliminated by assembling them together using the Geneious package (85), with default *de novo* assembly parameters.

16S rRNA read classification. In order to compare the 16S rRNA read classifications among different freshwater and brackish bodies, we first made a nonredundant version of the RDP database prepared by clustering all available 16S rRNA coding sequences (approximately 2.3 million) into approximately 800,000 sequences at 90% identity level using UCLUST (86). This database was used to identify candidate 16S rRNA fragments among the Illumina reads (unassembled). If a sequence matched this database at an E value of $<$1e-5, it was considered a potential 16S rRNA fragment. These candidate fragments were aligned to archaeal, bacterial, and eukaryal 16S/18S rRNA HMM models using ssu-align to identify true 16S/18S sequences (77). The 16S rRNA fragments retrieved were compared to the entire RDP database and classified into a high-level taxon if the sequence identity was $\geq$80% (BLASTN) and the alignment length was $\geq$90 bp. Fragments failing these thresholds were discarded.

Genome size estimation, completeness, and phylogenomics of the reconstructed genomes. Estimation of genome size, contamination, and completeness of the reconstructed genomes was assessed using CheckM (20). Phylogenomic trees were made for each MAG using the taxonomically closest microbes. We have used all genomes from NCBI available as of July 2017. First, MAGs were annotated using BLASTN and BLASTP searches for each CDS and protein against NCBI NR, and a top hit was assigned for each of them. This allowed us to determine the organisms closest to each CDS of each MAG. The closest selected bacterial genomes to ours were accordingly downloaded from NCBI, and each phylogenomic tree was done separately for each corresponding phylum, class, or genus. To create these whole-genome phylogenies, conserved proteins in the reconstructed genomes and the reference genomes were identified using the COG database (80) and were subsequently concatenated, aligned using Kalign (87), and trimmed using trimAl (88), with default parameters. Maximum likelihood trees were constructed using FastTree2 (89), a JTT+CAT model, a gamma approximation, and 100 bootstrap replicates. We also confirmed the robustness of SAR11 phylogeny with a new tree based on the PhyloPhlAn tool used for phylogenomic analysis (90).

Polynucleophage phylogenomic analysis. Protein sequences from Baikal-20-5m-C28 were queried against a database of proteins derived from viral genomes from NCBI RefSeq and from studies that described phage genomes through metagenomics (71, 91–93). A total of 1,253 phage genomes that shared at least five proteins with Baikal-20-5m-C28 (minimum identity 30% and minimum alignment length 30 amino acids) were selected for further analysis. Dice distances were calculated between phage genomes, as previously described (71), but replacing TBLASTX with Diamond (94) for querying proteins. The obtained distance matrix was used as input for phylogenomic reconstruction using the neighbor-joining algorithm (95) implemented in the Phangorn package of R.

Metagenomic data sets used for fragment recruitment and 16S rRNA fragment analysis. Metagenomics data sets are publicly available for the Amadorio (22) and Tous (57) reservoirs, Lake Lanier (96), the Dexter reservoir (BioProject no. PRJNA312985), the Klamath Iron Gate Dam (BioProject no. PRJNA312830), the Kalamas River (BioProject no. PRJNA304352), Lake Houston (BioProject no. PRJNA312986), Yellowstone (BioProject no. PRJNA60433), Lake Ontario and Lake Erie (BioProject no. PRJNA288501), Lake Michigan (BioProject no. PRJNA248239), Amazon Lakes (97), Lake Mendota (BioProject numbers PRJNA330170, PRJNA330171, and PRJNA330042), Swedish lakes and Trout Bog (98), Finnish lakes (99), and the Baltic (23) and Caspian (24) Seas.

Accession number(s). Lake Baikal 5- and 20-m-deep sample metagenomic data sets have been deposited in the NCBI SRA database with BioProject number PRJNA396997 (SRR5896115 and SRR5896114 for 5- and 20-m-deep samples, respectively). The 35 Lake Baikal MAGs have been deposited in the NCBI under Biosample identifiers SAMN07460786 to SAMN07460820. The viral sequence of Baikal-5-20m-C28 polynucleophage has been deposited in the NCBI under Biosample identifier SAMN07460823.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02132-17.

SUPPLEMENTAL FILE 1, PDF file, 2.9 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

ACKNOWLEDGMENTS

F.R.-V. was supported by grants "VIREVO" CGL2016-76273-P (AEI/FEDER, EU) (cofunded with FEDER funds), Acciones de dinamización "Redes de Excelencia" CONSOLIDER-CGL2015-71523-REDC from the Spanish Ministerio de Economía, Industria y Competitividad, and Prometeo II/2014/012 "Aquamet" from Generalitat Valenciana. T.I.Z. was supported by the Integration Project ISC SB RAS no. 4.1.2, the State Task no. 0345–2016–0007 "Geobiochemical studies of the methane cycles. . . ."

F.R.-V., T.I.Z., and P.J.C.-Y. conceived this work. T.I.Z., A.S.Z., and V.V.B. performed the sample collection, filtration, and DNA extraction. Analysis was carried out by P.J.C.-Y., F.H.C., and R.R. The manuscript was written by P.J.C.-Y., T.I.Z., and F.R.-V. All authors read and approved the final manuscript.

We declare no conflicts of interest.

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

REFERENCES

1. Hampton SE, Izmest'eva LR, Moore MV, Katz SL, Dennis B, Silow EA. 2008. Sixty years of environmental change in the world's largest freshwater lake–Lake Baikal, Siberia. Glob Change Biol 14:1947–1958. https://doi.org/10.1111/j.1365-2486.2008.01616.x.
2. Kurilkina MI, Zakharova YR, Galachyants YP, Petrova DP, Bukin YS, Domysheva VM, Blinov VV, Likhoshway YV. 2016. Bacterial community composition in the water column of the deepest freshwater Lake Baikal as determined by next-generation sequencing. FEMS Microbiol Ecol 92:fiw094. https://doi.org/10.1093/femsec/fiw094.
3. Galazy G. 1993. Atlas of Lake Baikal. GUGK, Moscow, Russia. (In Russian.)
4. Shimaraev M, Granin N, Zhdanov A. 1993. Deep ventilation of Lake Baikal waters due to spring thermal bars. Limnol Oceanogr 38:1068–1072. https://doi.org/10.4319/lo.1993.38.5.1068.
5. Shimaraev M, Verbolov V, Granin N, Sherstayankin P. 1994. Physical limnology of Lake Baikal: a review. Baikal Intl Cent Ecol Res.
6. Nagata T, Takai K, Kawanobe K, Kim D-S, Nakazato R, Guselnikova N, Bondarenko N, Mologawaya O, Kostrnova T, Drucker V, Satoh Y, Watanabe Y. 1994. Autotrophic picoplankton in southern Lake Baikal: abundance, growth and grazing mortality during summer. J Plankton Res 16:945–959. https://doi.org/10.1093/plankt/16.8.945.
7. Weiss R, Carmack E, Koropalov V. 1991. Deep-water renewal and biological production in Lake Baikal. Nature 349:665. https://doi.org/10.1038/349665a0.
8. Katano T, Nakano S-i, Ueno H, Mitamura O, Anbutsu K, Kihira M, Satoh Y, Drucker V, Sugiyama M. 2005. Abundance, growth and grazing loss rates of picophytoplankton in Barguzin Bay, Lake Baikal. Aquat Ecol 39:431–438. https://doi.org/10.1007/s10452-005-9000-8.
9. Nakano S-i, Mitamura O, Sugiyama M, Maslennikov A, Nishibe Y, Watanabe Y, Drucker V. 2003. Vertical planktonic structure in the central basin of Lake Baikal in summer 1999, with special reference to the microbial food web. Limnology 4:155–160. https://doi.org/10.1007/s10201-003-0100-7.
10. Votintsev K, Popovskaya G. 1979. The peculiarity of the biotic cycle in Lake Baikal. Dokladu Akademii Nauk SSSR 216:666–669.
11. Kozhova O. 1987. Phytoplankton of Lake Baikal: structural and functional characteristics. Arch Hydrobiol Beih 25:19–37.
12. Votintsev K, Mescheryakova A, Popovskaya G. 1975. Cycle of organic matter in Lake Baikal. Nauka, Novosibirsk.
13. Bashenkhaeva MV, Zakharova YR, Petrova DP, Khanaev IV, Galachyants YP, Likhoshway YV. 2015. Sub-ice microalgal and bacterial communities in freshwater Lake Baikal, Russia. Microb Ecol 70:751. https://doi.org/10.1007/s00248-015-0619-2.
14. Bondarenko NA, Belykh OI, Golobokova LP, Artemyeva OV, Logacheva NF, Tikhonova IV, Lipko IA, Kostornova TY, Parfenova VV, Khodzher TV, Ahn TS, Zo YG. 2012. Stratified distribution of nutrients and extremophile biota within freshwater ice covering the surface of Lake Baikal. J Microbiol 50:8–16. https://doi.org/10.1007/s12275-012-1251-1.
15. Bondarenko N, Belykh O, Golobokova L, Artemyeva O, Logacheva N, Tikhonova I, Lipko I, Kostornova TY, Parfenova V, Khodzher T. 2013. Chemical composition, bacteria and algae communities of the ice of Lake Baikal. Hydrobiol J 49:12–26. https://doi.org/10.1615/HydrobJ.v49.i3.20.
16. Bel'kova N, Parfenova V, Kostornova TY, Denisova LY, Zaichikov E. 2003. Microbial biodiversity in the water of Lake Baikal. Microbiology 72:203–213. https://doi.org/10.1023/A:1023224215929.
17. Mikhailov I, Zakharova YR, Galachyants YP, Usoltseva M, Petrova D, Sakirko M, Likhoshway YV, Grachev MA. 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 465:413–419. https://doi.org/10.1134/S1607672915060198.
18. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol 71:8966–8969. https://doi.org/10.1128/AEM.71.12.8966-8969.2005.
19. Kennedy K, Hall MW, Lynch MD, Moreno-Hagelsieb G, Neufeld JD. 2014. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. Appl Environ Microbiol 80:5717–5722. https://doi.org/10.1128/AEM.01451-14.
20. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
21. Mizuno CM, Rodriguez-Valera F, Ghai R. 2015. Genomes of planktonic *Acidimicrobiales*: widening horizons for marine actinobacteria by metagenomics. mBio 6:e02083-14. https://doi.org/10.1128/mBio.02083-14.
22. Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. 2014. Key roles for freshwater actinobacteria revealed by deep metagenomic sequencing. Mol Ecol 23:6073–6090. https://doi.org/10.1111/mec.12985.
23. Hugerth LW, Larsson J, Alneberg J, Lindh MV, Legrand C, Pinhassi J, Andersson AF. 2015. Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biol 16:1–18. https://doi.org/10.1186/s13059-014-0572-2.
24. Mehrshad M, Amoozegar MA, Ghai R, Fazeli SAS, Rodriguez-Valera F. 2016. Genome reconstruction from metagenomic datasets reveals novel microbes in the brackish waters of the Caspian Sea. Appl Environ Microbiol 82:1599–1612. https://doi.org/10.1128/AEM.03381-15.
25. Neuenschwander SM, Ghai R, Pernthaler J, Salcher MM. 13 October 2017. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. ISME J https://doi.org/10.1038/ismej.2017.156.
26. Kang I, Kim S, Islam MR, Cho J-C. 2017. The first complete genome sequences of the acI lineage, the most abundant freshwater Actinobacteria, obtained by whole-genome-amplification of dilution-to-extinction cultures. Sci Rep 7:42252. https://doi.org/10.1038/srep42252.
27. Driscoll CB, Otten TG, Brown NM, Dreher TW. 2017. Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. Stand Genomic Sci 12:9. https://doi.org/10.1186/s40793-017-0224-8.
28. Weon H-Y, Song M-H, Son J-A, Kim B-Y, Kwon S-W, Go S-J, Stackebrandt E. 2007. Flavobacterium terrae sp. nov. and Flavobacterium cucumis sp. nov., isolated from greenhouse soil. Int J Syst Evol Microbiol 57:1594–1598. https://doi.org/10.1099/ijs.0.64935-0.
29. Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F. 2009. Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev 73:249–299. https://doi.org/10.1128/MMBR.00035-08.
30. Gross R, Guzman CA, Sebaihia M, dos Santos VAM, Pieper DH, Koebnik R, Lechner M, Bartels D, Buhrmester J, Choudhuri JV, Ebensen T,

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

Gaigalat L, Herrmann S, Khachane AN, Larisch C, Link S, Linke B, Meyer F, Mormann S, Nakunst D, Ruckert C, Schneiker-Bekel S, Schulze K, Vorholter F-J, Yevsa T, Engle JT, Goldman WE, Puhler A, Gobel UB, Goesmann A, Blocker H, Kaiser O, Martinez-Arias R. 2008. The missing link: Bordetella petrii is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic bordetellae. BMC Genomics 9:449. https://doi.org/10.1186/1471-2164-9-449.

31. Lonhienne TG, Sagulenko E, Webb RI, Lee K-C, Franke J, Devos DP, Nouwens A, Carroll BJ, Fuerst JA. 2010. Endocytosis-like protein uptake in the bacterium Gemmata obscuriglobus. Proc Natl Acad Sci U S A 107:12883–12888. https://doi.org/10.1073/pnas.1001085107.
32. Vieira S, Luckner M, Wanner G, Overmann J. 2017. Luteitalea pratensis gen. nov., sp. nov. a new member of subdivision 6 Acidobacteria isolated from temperate grassland soil. Int J Syst Evol Microbiol 67:1408–1414. https://doi.org/10.1099/ijsem.0.001827.
33. DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M. 2011. Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. Appl Environ Microbiol 77:6295–6300. https://doi.org/10.1128/AEM.05005-11.
34. Zeng Y, Baumbach J, Barbosa EGV, Azevedo V, Zhang C, Koblížek M. 2016. Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. Environ Microbiol Rep 8:139–149. https://doi.org/10.1111/1758-2229.12363.
35. Zeng Y, Feng F, Medová H, Dean J, Koblížek M. 2014. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. Proc Natl Acad Sci U S A 111:7795–7800. https://doi.org/10.1073/pnas.1400295111.
36. Zeng Y, Selyanin V, Lukeš M, Dean J, Kaftan D, Feng F, Koblížek M. 2015. Characterization of the microaerophilic, bacteriochlorophyll *a*-containing bacterium *Gemmatimonas phototrophica* sp. nov., and emended descriptions of the genus *Gemmatimonas* and *Gemmatimonas aurantiaca*. Int J Syst Evol Microbiol 65:2410–2419. https://doi.org/10.1099/ijs.0.000272.
37. DeBruyn JM, Fawaz MN, Peacock AD, Dunlap JR, Nixon LT, Cooper KE, Radosevich M. 2013. Gemmatirosa kalamazoonesis gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum Gemmatimonadetes. J Gen Appl Microbiol 59:305–312. https://doi.org/10.2323/jgam.59.305.
38. Falkner KK, Measures CI, Herbelin SE, Edmond JM, Weiss RF. 1991. The major and minor element geochemistry of Lake Baikal. Limnol Oceanogr 36:413–423. https://doi.org/10.4319/lo.1991.36.3.0413.
39. Grachev M. 2002. On the present state of the ecological system of Lake Baikal. SB RAS, Novosibirsk, Russia.
40. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. 2015. The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. ISME J 9:2442. https://doi.org/10.1038/ismej.2015.55.
41. Hoetzinger M, Schmidt J, Jezberová J, Koll U, Hahn MW. 2017. Microdiversification of a pelagic Polynucleobacter species is mainly driven by acquisition of genomic islands from a partially interspecific gene pool. Appl Environ Microbiol 83:e02266–16. https://doi.org/10.1128/AEM.02266-16.
42. Hahn MW, Scheuerl T, Jezberová J, Koll U, Jezbera J, Šimek K, Vannini C, Petroni G, Wu QL. 2012. The passive yet successful way of planktonic life: genomic and experimental analysis of the ecology of a free-living Polynucleobacter population. PLoS One 7:e32772. https://doi.org/10.1371/journal.pone.0032772.
43. Jezbera J, Jezberová J, Brandt U, Hahn MW. 2011. Ubiquity of Polynucleobacter necessarius subspecies asymbioticus results from ecological diversification. Environ Microbiol 13:922–931. https://doi.org/10.1111/j.1462-2920.2010.02396.x.
44. Jezberová J, Jezbera J, Brandt U, Lindström ES, Langenheder S, Hahn MW. 2010. Ubiquity of Polynucleobacter necessarius ssp. asymbioticus in lentic freshwater habitats of a heterogenous 2000 km^2 area. Environ Microbiol 12:658–669. https://doi.org/10.1111/j.1462-2920.2009.02106.x.
45. Giovannoni SJ, Thrash JC, Temperton B. 2014. Implications of streamlining theory for microbial ecology. ISME J 8:1553. https://doi.org/10.1038/ismej.2014.60.
46. Ghai R, McMahon KD, Rodriguez-Valera F. 2012. Breaking a paradigm: cosmopolitan and abundant freshwater actinobacteria are low GC. Environ Microbiol Rep 4:29–35. https://doi.org/10.1111/j.1758-2229.2011.00274.x.
47. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappe MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245.
48. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63. https://doi.org/10.1038/345060a0.
49. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420:806. https://doi.org/10.1038/nature01240.
50. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS. 2012. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. mBio 3:e00252-12. https://doi.org/10.1128/mBio.00252-12.
51. Herlemann DP, Woelk J, Labrenz M, Jürgens K. 2014. Diversity and abundance of "Pelagibacterales" (SAR11) in the Baltic Sea salinity gradient. Syst Appl Microbiol 37:601–604. https://doi.org/10.1016/j.syapm.2014.09.002.
52. Oh H-M, Kang I, Lee K, Jang Y, Lim S-I, Cho J-C. 2011. Complete genome sequence of strain IMCC9063, belonging to SAR11 subgroup 3, isolated from the Arctic Ocean. J Bacteriol 193:3379–3380. https://doi.org/10.1128/JB.05033-11.
53. Zaremba-Niedzwiedzka K, Viklund J, Zhao W, Ast J, Sczyrba A, Woyke T, McMahon K, Bertilsson S, Stepanauskas R, Andersson SG. 2013. Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade. Genome Biol 14:R130. https://doi.org/10.1186/gb-2013-14-11-r130.
54. Salcher MM, Pernthaler J, Posch T. 2011. Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria 'that rule the waves' (LD12). ISME J 5:1242. https://doi.org/10.1038/ismej.2011.8.
55. Elevi Bardavid R, Oren A. 2012. Acid-shifted isoelectric point profiles of the proteins in a hypersaline microbial mat: an adaptation to life at high salt concentrations? Extremophiles 16:787–792. https://doi.org/10.1007/s00792-012-0476-6.
56. Oh S, Zhang R, Wu QL, Liu W-T. 2014. Draft genome sequence of a novel SAR11 clade species abundant in a Tibetan lake. Genome Announc 2(6):e01137-14. https://doi.org/10.1128/genomeA.01137-14.
57. Cabello-Yeves PJ, Haro-Moreno JM, Martin-Cuadrado A-B, Ghai R, Picazo A, Camacho A, Rodriguez-Valera F. 2017. Novel Synechococcus genomes reconstructed from freshwater reservoirs. Front Microbiol 8:1151. https://doi.org/10.3389/fmicb.2017.01151.
58. Hügler M, Sievert SM. 2010. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Ann Rev Mar Sci 3:261–289.
59. Martinez-Garcia M, Swan BK, Poulton NJ, Gomez ML, Masland D, Sieracki ME, Stepanauskas R. 2012. High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. ISME J 6:113–123. https://doi.org/10.1038/ismej.2011.84.
60. Seo J-S, Keum Y-S, Li QX. 2009. Bacterial degradation of aromatic compounds. Int J Environ Res Public Health 6:278–309. https://doi.org/10.3390/ijerph6010278.
61. Arora PK, Bae H. 2014. Bacterial degradation of chlorophenols and their derivatives. Microb Cell Fact 13:31. https://doi.org/10.1186/1475-2859-13-31.
62. Michaelson LV, Napier JA, Molino D, Faure J-D. 2016. Plant sphingolipids: their importance in cellular organization and adaption. Biochim Biophys Acta 1861:1329–1335. https://doi.org/10.1016/j.bbalip.2016.04.003.
63. Gronnier J, Germain V, Gouguet P, Cacas J-L, Mongrand S. 2016. GIPC: glycosyl inositol phospho-ceramides, the major sphingolipids on earth. Plant Signal Behav 11:e1152438. https://doi.org/10.1080/15592324.2016.1152438.
64. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J. 2005. Prokaryotic sulfur oxidation. Curr Opin Microbiol 8:253–259. https://doi.org/10.1016/j.mib.2005.04.005.
65. Watanabe T, Kojima H, Fukui M. 2014. Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans skB26* and *Sulfuritalea hydrogenivorans sk43H*: genetic insights into the sulfur oxidation pathway of betaproteobacteria. Syst Appl Microbiol 37:387–395. https://doi.org/10.1016/j.syapm.2014.05.010.
66. Könneke M, Bernhard AE, de La Torre JR, Walker CB, Waterbury JB, Stahl DA. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437:543. https://doi.org/10.1038/nature03911.
67. Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, Jehmlich N, Palatinszky M, Vierheilig J, Bulaev A, Kirkegaard RH, von Bergen M, Rattei

Downloaded from https://journals.asm.org/journal/aem on 05 May 2026 by 193.147.143.24.

T, Bendinger B, Nielsen PH, Wagner M. 2015. Complete nitrification by Nitrospira bacteria. Nature 528:504. https://doi.org/10.1038/nature16461.
68. Alonso-Sáez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL, Lovejoy C, Tremblay J-É Potvin M, Heinrich F, Estrada M, Riemann L, Bork P, Pedros-Alio C, Bertillson S. 2012. Role for urea in nitrification by polar marine Archaea. Proc Natl Acad Sci U S A 109:17989–17994. https://doi.org/10.1073/pnas.1201914109.
69. Sharma AK, Sommerfeld K, Bullerjahn GS, Matteson AR, Wilhelm SW, Jezbera J, Brandt U, Doolittle WF, Hahn MW. 2009. Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. ISME J 3:726. https://doi.org/10.1038/ismej.2009.13.
70. Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, Spudich EN, Spudich JL, Béjà O. 2003. Diversification and spectral tuning in marine proteorhodopsins. EMBO J 22:1725–1731. https://doi.org/10.1093/emboj/cdg183.
71. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. 2013. Expanding the marine virosphere using metagenomics. PLoS Genet 9:e1003987. https://doi.org/10.1371/journal.pgen.1003987.
72. Martín-Cuadrado A-B, López-García P, Alba J-C, Moreira D, Monticelli L, Strittmatter A, Gottschalk G, Rodríguez-Valera F. 2007. Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. PLoS One 2:e914. https://doi.org/10.1371/journal.pone.0000914.
73. Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
74. Peng Y, Leung HC, Yiu S-M, Chin FY. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.
75. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:1. https://doi.org/10.1186/1471-2105-11-1.
76. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.0955.
77. Nawrocki E. 2009. Structural RNA homology search and alignment using covariance models. PhD thesis. Washington University in St. Louis, St. Louis, MO.
78. Nawrocki EP, Eddy SR. 2010. ssu-align: a tool for structural alignment of SSU rRNA sequences. http://eddylab.org/software/ssu-align/.
79. Huang Y, Gilna P, Li W. 2009. Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics 25:1338–1340. https://doi.org/10.1093/bioinformatics/btp161.
80. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22–28. https://doi.org/10.1093/nar/29.1.22.
81. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O. 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res 29:41–43. https://doi.org/10.1093/nar/29.1.41.
82. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.
83. Lê S, Josse J, Husson F. 2008. FactoMineR: an R package for multivariate analysis. J Stat Softw 25:1–18.
84. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
85. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meinties P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.
86. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.
87. Lassmann T, Sonnhammer EL. 2005. Kalign–an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics 6:1. https://doi.org/10.1186/1471-2105-6-1.
88. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
89. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
90. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304. https://doi.org/10.1038/ncomms3304.
91. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. Nature 536:425–430. https://doi.org/10.1038/nature19094.
92. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CP, Dutilh BE, Thompson FL. 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. Nat Commun 8:15955. https://doi.org/10.1038/ncomms15955.
93. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vague D, Tara Oceans Coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and biogeochemical impacts of uncultivated globally abundant ocean viruses. Nature 537:689–693. https://doi.org/10.1038/nature19366.
94. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
95. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.
96. Oh S, Caro-Quintero A, Tsementzi D, DeLeon-Rodriguez N, Luo C, Poretsky R, Konstantinidis KT. 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. https://doi.org/10.1128/AEM.00107-11.
97. Toyama D, Kishi LT, Santos-Júnior CD, Soares-Costa A, de Oliveira TCS, de Miranda FP, Henrique-Silva F. 2016. Metagenomics analysis of microorganisms in freshwater lakes of the Amazon Basin. Genome Announc 4:e01440-16. https://doi.org/10.1128/genomeA.01440-16.
98. Eiler A, Zaremba-Niedzwiedzka K, Martínez-García M, McMahon KD, Stepanauskas R, Andersson SG, Bertilsson S. 2014. Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics. Environ Microbiol 16:2682–2698. https://doi.org/10.1111/1462-2920.12301.
99. Peura S, Sinclair L, Bertilsson S, Eiler A. 2015. Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. Sci Rep 5:srep12102. https://doi.org/10.1038/srep12102.
100. Cabello-Yeves PJ, Ghai R, Mehrshad M, Picazo A, Camacho A, Rodriguez-Valera F. 2017. Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. Front Microbiol 8(November):1–17. https://doi.org/10.3389/fmicb.2017.02131.