Exploratory characterization of bacterial communities and predicted functional profiles in six water samples from four Colombian Andean lakes using 16S rRNA gene amplicon sequencing

Andrés Gómez-Palacio[1*], Johana Marín-Suarez[2], Adriana Pedroza-Ramos[3], Nelson Aranguren-Riaño[3]

[1]Laboratorio de Investigación en Genética Evolutiva - LIGE. Universidad Pedagógica y Tecnológica de Colombia, Tunja, Boyacá, Colombia.

[2]Grupo de Estudios en Genética y Biología Molecular - GEBIMOL. Universidad Pedagógica y Tecnológica de Colombia, Tunja, Boyacá, Colombia.

[3]Unidad de Ecología en Sistemas Acuáticos- UDESA. Universidad Pedagógica y Tecnológica de Colombia, Tunja, Boyacá, Colombia.

**Correspondence to*: Andrés Gómez-Palacio (amgomezpa@gmail.com)

Abstract: Northern Andean highland lakes support agriculture, water provision, and local livelihoods, but their bacterial communities remain insufficiently characterized. We conducted an exploratory survey of bacterial community composition and predicted functional potential in six water samples from four freshwater lakes in the Eastern Cordillera of Colombia: Fúquene, Tota, Calderona, and Colorado. Five samples collected in 2019 comprised the primary dataset, while a previously published Tota sample collected in December 2018 was reprocessed as a historical reference. Bacterial communities were characterized using 16S rRNA gene amplicon sequencing, and functional profiles were inferred using PICRUSt2. After sequence processing and taxonomic filtering, 3,153 amplicon sequence variants were retained. Actinobacteria, Proteobacteria, Bacteroidetes, Cyanobacteria, and Verrucomicrobia were the predominant bacterial phyla. Coverage-standardized richness and diversity varied descriptively among

samples, with Colorado showing the lowest estimated ASV richness. Taxonomic composition and Bray–Curtis dissimilarities also varied among samples, reflecting differences in dominant bacterial groups. *Mycobacterium*-assigned ASVs were detected in Fúquene and Tota, *Leptospira*-assigned ASVs occurred in these same lakes at relative abundances of 0.01–0.36%, and *Legionella*-assigned ASVs occurred in all six samples at 0.11–4.60%. These genus-level assignments do not establish pathogenic species, viability, transmission, or risks to aquatic animals or humans. PICRUSt2 predicted variation in functional profiles, including pathways associated with naphthalene degradation, bacterial chemotaxis, biofilm formation, bacterial secretion systems, and pyrimidine and folate metabolism. Bray–Curtis ordination of predicted pathway profiles showed sample-level separation, and weighted NSTI values ranged from 0.160 to 0.181. These predictions represent inferred genomic potential rather than direct evidence of gene presence, expression, or metabolic activity. Because sampling was limited and unbalanced and lacked concurrent environmental measurements, the observed patterns cannot be attributed to specific environmental drivers. This study provides an initial descriptive baseline for bacterial communities in four Colombian Andean lakes and identifies patterns for evaluation through replicated sampling, physicochemical measurements, metagenomics, and targeted validation.

1. Introduction

Aquatic ecosystems, particularly highland lakes and lagoons, are of considerable ecological and societal importance (Heino et al., 2021; Millennium Ecosystem Assessment, 2005). These water bodies provide resources for agricultural irrigation, support dairy production systems, and supply freshwater for domestic use in many regions (Heino et al., 2021; Maasri et al., 2022; Millennium Ecosystem Assessment, 2005). In the Northern Andean highlands of east-central Colombia, a network of lakes and lagoons shapes the landscape, supports local livelihoods, and sustains biodiversity (Andrade-Sossa et al., 2023; Maasri et al., 2022; Prado et al., 2024). However, the ecological integrity of these ecosystems is increasingly threatened

by anthropogenic pressures within their watersheds (Andrade-Sossa et al., 2023; Aranguren-Riaño et al., 2018; Prado et al., 2024; Torres-Barrera et al., 2017; Yang et al., 2026).

Human activities, including intensive agriculture and urbanization, introduce pollutants, nutrients, sediments, and other disturbances into aquatic environments (Aranguren-Riaño et al., 2018; Carpenter et al., 1998; Másmela-Mendoza et al., 2019; Torres-Barrera et al., 2017). These pressures may affect water quality, aquatic biodiversity, and ecosystem functioning (Barletta et al., 2010; Millennium Ecosystem Assessment, 2005; Prado et al., 2024). Nutrient enrichment (Carpenter et al., 1998), sedimentation (Smith and Schindler, 2009), and hydrological alteration (Jeppesen et al., 2009) can influence microbial community composition, primary production, and biogeochemical processes in lakes (Andrade-Sossa et al., 2023; Forero-Pineda et al., 2021; Paerl and Huisman, 2009; Paver et al., 2020; Smith and Schindler, 2009). Similar environmental pressures have been documented in Colombian lakes, where they may be associated with changes in biodiversity and nutrient cycling (Aranguren-Riaño et al., 2018; Forero-Pineda et al., 2021; Másmela-Mendoza et al., 2019; Prado et al., 2024). More broadly, inland waters contribute substantially to global biogeochemical cycles and energy fluxes (Downing et al., 2006; Raymond et al., 2013; Tranvik et al., 2009). Although previous work, including Forero-Pineda et al. (2021), has begun to characterize microbial communities in Colombian highland lakes, comparable ASV-based metataxonomic surveys encompassing multiple Colombian Andean lentic systems remain scarce.

Microbial communities mediate key biogeochemical processes in aquatic ecosystems and influence nutrient cycling, organic-matter transformation, and water quality (Battin et al., 2016; Cai et al., 2026; Sessitsch et al., 2023; Shade et al., 2012). Characterizing bacterial diversity and potential functional attributes in lentic environments can therefore contribute

to understanding how microbial communities vary across aquatic systems and can provide baseline information for future ecological studies (Battin et al., 2016; Paver et al., 2020; Sessitsch et al., 2023; Shade et al., 2012). However, bacterial communities in Colombian Andean lakes remain insufficiently characterized, particularly across multiple lake systems with contrasting environmental contexts.

Beyond anthropogenic pressures, Colombian Andean lentic systems occur across heterogeneous geological and geomorphological settings that may influence their hydrology, sediment characteristics, water chemistry, and biological communities (Andrade-Sossa et al., 2023). The objective of this study was to provide an exploratory characterization of bacterial community composition and predicted functional potential in six water samples collected from four lakes in the Eastern Cordillera of Colombia: Fúquene, Tota, Calderona, and Colorado. The four water bodies represent contrasting geomorphological and hydrological settings. Fúquene occupies the Ubaté–Chiquinquirá intramontane basin, whose Pleistocene development reflects tectonic deformation, erosion, and sedimentation processes (Sarmiento et al., 2008). Tota is a large high-elevation lake with an extensive late-Quaternary sedimentary record and evidence of substantial historical lake-level fluctuations (Gibson et al., 2019). Calderona is a small high-elevation lake of glacial origin located in the Páramo El Bijagual region (Fundación Vive Ciénega, 2021), whereas Colorado—locally known as Laguna Las Coloradas—is a lower-elevation wetland in the municipality of Gachantivá and forms part of a landscape containing several lentic and wetland ecosystems (Corpoboyacá, 2022). These contrasting settings could influence basin morphology, sediment characteristics, hydrological connectivity, water residence time, and water chemistry, thereby contributing to differences in microbial habitats.

Specifically, we aimed to (i) describe sample-level patterns of bacterial richness, diversity, and taxonomic composition; (ii) explore variation in predicted metabolic pathways among samples; and (iii) conduct an exploratory phylogenetic placement of ASVs assigned to *Mycobacterium*. Because the sampling design was limited, unbalanced, and not fully contemporaneous, the study was not intended to test the effects of lake identity, environmental context, anthropogenic pressure, spatial position, or sampling period. Instead, the observed patterns are used to generate hypotheses for future studies based on replicated and contemporaneous sampling.

Because these lakes support human activities and provide habitats for native aquatic and terrestrial fauna, the detection of bacterial genera containing species of potential human, veterinary, or aquatic-animal health relevance warrants consideration from a One Health perspective. We applied a metataxonomic framework based on 16S rRNA gene amplicon sequencing to characterize bacterial diversity and community composition and used PICRUSt2 to infer potential functional profiles from taxonomic data. PICRUSt2 uses reference genomes and phylogenetic placement to predict gene-family abundances and metabolic pathways associated with detected taxa. Particular attention was given to ASVs assigned to *Mycobacterium* because nontuberculous mycobacteria (NTM) are widespread in natural and engineered aquatic environments and include species associated with infections in humans, cultured and wild fish, reptiles, and other aquatic organisms (Dowdell et al., 2019; Falkinham, 2021). Other detected genera, including *Leptospira* and *Legionella*, also contain species of potential human or animal-health relevance and are associated with aquatic or water-mediated ecological cycles (Bradley and Lockaby, 2023; Gattuso et al., 2022). The *Mycobacterium* analysis was therefore interpreted exclusively as exploratory phylogenetic placement rather than pathogen identification.

2. Materials and Methods

2.1. Study Area and Sampling

The study included six water samples from four Andean lentic systems in the Eastern Cordillera of Colombia: Fúquene Lagoon, Tota Lake, Calderona Lagoon, and Colorado Lagoon (Table 1). These systems differ in altitude, geological origin, surrounding land use, and degree of legal or environmental protection. The environmental context reported in Table 1 were assigned descriptively from available information on watershed land use, protection status, and previously reported environmental conditions. Because the sampling design was limited and unbalanced, these categories were not treated as statistically replicated explanatory factors.

Fúquene Lagoon is located in the northern Andean highlands of Colombia, less than 100 km from Bogotá. Its watershed includes extensive dairy farming and agricultural crops such as cereals, potatoes, and onions. Tota Lake, located in Boyacá, is Colombia's largest high-mountain lake and is an important site for tourism, agriculture, and aquaculture. Previous studies have reported nutrient enrichment and moderate eutrophication associated with fertilizer use in green-onion cultivation, aquaculture activities, and domestic wastewater inputs in both lakes (Aranguren-Riaño et al., 2018; Forero-Pineda et al., 2021; Másmela-Mendoza et al., 2019; Torres-Barrera et al., 2017).

Calderona Lagoon is a small glacial water body located within El Cañal Municipal Natural Park in the municipality of Ciénega, Boyacá. It is surrounded by páramo vegetation associated with the Páramo El Bijagual. Colorado Lagoon is located in the municipality of Gachantivá, Boyacá, near the foothills of the Páramo de Iguaque. It is surrounded by Andean

forest, and human activity in the area is mainly limited to short-term tourism. Additional characteristics of the four systems are summarized in Table 1.

Sampling was restricted to the limnetic zone to characterize planktonic bacterial communities while avoiding direct representation of littoral habitats. The primary exploratory dataset comprised five samples collected in 2019: two from Fúquene in April and September, one each from Calderona and Colorado in October, and one from Tota in November (Table 1). In the Colombian Andean region, seasonal variation is more closely related to rainfall regime than to temperate climatic seasons. The sampled months fall within different portions of the regional bimodal rainfall cycle, but because sampling was not replicated across months or lakes, rainfall-related or seasonal patterns were not evaluated. Because these samples were collected in different months and were not obtained under a balanced, replicated design, each sample was treated as an individual observational unit. No sample was considered a biological, seasonal, or temporal replicate of another, and no seasonal trends were inferred from this sampling design.

A sixth sample from Tota, collected in December 2018 and previously reported under BioProject PRJNA720890, was retrieved from Forero-Pineda et al. (2021) and reprocessed using the same bioinformatic workflow. This sample was treated exclusively as a historical reference for Tota and was analyzed separately from the 2019 cross-lake dataset. It was not considered an internal control, a replicate of the 2019 Tota sample, or evidence of temporal change.

For each field sample, approximately 1.7 L of water was collected from a single depth within the photic zone using a custom-made, non-commercial Schindler–Patalas-type water sampler constructed from a thoroughly cleaned plastic bottle. Water transparency was measured *in situ* using a white, 20-cm-diameter Secchi disk lowered from the shaded side of

the boat, and the observed Secchi depth was used qualitatively to guide selection of the water-sampling depth. No predefined transparency range or fixed fraction of Secchi depth was applied. Only one depth was sampled during each collection event; therefore, vertical variation and stratification within the water column were not evaluated. Water was prefiltered *in situ* through a 40-mm-diameter filter with 1-mm openings to remove large suspended debris and organisms. Microorganisms passing through the prefilter were subsequently concentrated onto two sterile Sterivex-GP pressure filter units fitted with 0.22 µm hydrophilic polyethersulfone membranes (10 cm^2 filtration area; catalog no. SVGP01050, Merck Millipore, Darmstadt, Germany), following Paver et al. (2020). Each Sterivex unit was transferred to a sterile tube containing 1.5 mL of Monarch® DNA/RNA Protection Reagent (catalog no. T2011, New England Biolabs, Ipswich, MA, USA). Samples were transported to the laboratory under refrigerated conditions within 2–4 h and stored at −20 °C until DNA extraction. DNA was extracted separately from the two Sterivex filters obtained from each water sample, and the resulting extracts were pooled before library preparation, producing one sequencing library per water sample.

The unequal sampling effort reflected differences in field accessibility, logistical constraints, and available resources rather than a balanced comparative design. Access to Calderona and Colorado was especially restricted because of their protected or remote settings. Consequently, the dataset was treated as an exploratory collection of individual water samples rather than as a replicated lake-level comparison.

Table 1. Location, qualitative environmental context, sampling period, sequencing depth, and number of retained amplicon sequence variants for the six water samples included in the study. The five samples collected in 2019 constituted the primary exploratory dataset.

192 *Indicates data retrieved from Forero-Pineda et al. (2021) and reprocessed here using the
193 same bioinformatic workflow applied to the 2019 samples.
194

Locality	Sample	Coordinates		Altitude (m.a.s.l.)	Geological origin	Environmental context	Sampling period	Clean reads (n)	Retained ASVs (n)
		North	West						
Ciénega	Calderona	5° 23' 23.16"	73° 14' 40.99"	3,143	Glacial	Protected municipal natural park; páramo vegetation	Oct. 2019	74690	514
Gachantivá	Colorado	5° 46' 29.43"	73° 33' 5.81"	2,430	Rainfall-fed wetland	Andean forest; limited tourism	Oct. 2019	68737	761
Fúquene	Fúquene1	5° 27' 4.53"	73° 44' 47.82"	2,568	Tectonic	Dairy farming and intensive agriculture within the watershed	Apr. 2019	73484	756
	Fúquene2						Sept. 2019	86752	848
Tota	Tota1	5° 32' 29.28"	72° 55' 27.75"	3,015	Tectonic-Glacial	Agriculture, aquaculture, tourism, and domestic wastewater inputs	Dec. 2018*	171326	690
	Tota2						Nov. 2019	81926	506

195

196 **2.2. DNA extraction and 16S rRNA gene amplicon sequencing**

197 Genomic DNA was extracted from environmental filters using a modified
198 cetyltrimethylammonium bromide (CTAB)-based protocol to maintain methodological
199 consistency with previous analyses of Colombian highland lake samples, including the
200 historical Tota sample reprocessed in the present study. CTAB was used as a cationic
201 detergent to promote cell lysis and facilitate the removal of polysaccharides and other
202 potential inhibitors. The extraction buffer contained 1% (w/v) CTAB, 3% (w/v) SDS,
203 100 mM Tris–HCl, 100 mM EDTA, and 1.5 M NaCl (pH 8.0), following previously
204 described procedures (Forero-Pineda et al., 2021; Hassan et al., 2018; Paver et al., 2020).

205 The V3–V4 region of the bacterial 16S rRNA gene was amplified using primers 341F
206 (S-D-Bact-0341-b-S-17; 5′-CCTACGGGNGGCWGCAG-3′) and 785R, also commonly
207 referred to as 805R (S-D-Bact-0785-a-A-21; 5′-GACTACHVGGGTATCTAATCC-3′),
208 following Klindworth et al. (2013). These primers target conserved regions of the 16S
209 rRNA gene present in both Bacteria and Archaea.

210 PCR amplification was performed in 50 µL reactions containing 0.3 mg mL^{-1} bovine
211 serum albumin, 250 µM dNTPs, 0.5 µM of each primer, 1 U Phusion High-Fidelity DNA
212 Polymerase, and Phusion HF buffer containing 1.5 mM $MgCl_2$. The thermal-cycling
213 program consisted of an initial denaturation at 95 °C for 5 min, followed by 25 cycles of
214 denaturation at 95 °C for 40 s, annealing at 55 °C for 2 min, and extension at 72 °C for 1
215 min, with a final extension at 72 °C for 7 min. These conditions were adapted from
216 Klindworth et al. (2013). The resulting amplicons were submitted to Macrogen Inc.
217 (https://dna.macrogen.com/) for library preparation, indexing, quality control, and
218 paired-end sequencing on the Illumina MiSeq platform.

The historical Tota sample was generated using the same DNA extraction protocol, primer pair, target region, sequencing platform, and library-preparation procedure as the 2019 samples (Forero-Pineda et al., 2021; Hassan et al., 2018; Paver et al., 2020).

2.3. Taxonomic and Functional Assignment of the Microbiome

2.3.1. Sequence processing and ASV inference

Raw sequencing reads were processed in R using DADA2 v1.20.0 (Callahan et al., 2016). Forward and reverse reads were truncated at positions 200 and 150, respectively, and the first 19 nt of forward reads and 20 nt of reverse reads were removed (*truncLen* = c(200,150); *trimLeft* = c(19,20)), yielding retained read lengths of 181 and 130 nt. Reads containing ambiguous bases were discarded (*maxN* = 0), and the maximum number of expected errors was set to two for both read directions (*maxEE* = c(2,2)). Reads were truncated at the first base with a quality score ≤2 (*truncQ* = 2), and reads matching phiX were removed (*rm.phix* = TRUE) (Callahan et al., 2016).

After filtering, forward- and reverse-read error models were learned from the filtered data, reads were dereplicated, and ASVs were inferred independently for each read direction using DADA2 (Callahan et al., 2016). Because the retained forward and reverse reads did not provide sufficient overlap, denoised forward and reverse-complemented reverse reads were concatenated using *mergePairs* (*justConatenate* = TRUE) with a 10-nt N spacer representing only the unresolved intervening region and not introducing inferred nucleotide information (Callahan et al., 2016). A sequence table was then constructed, and chimeric sequences were assessed and removed using the consensus method implemented in r*emoveBimeraDenovo* (Callahan et al., 2016). No additional minimum-abundance threshold was applied to the resulting ASVs.

2.3.2. Taxonomic assignment and diversity analyses

Taxonomy was assigned to the concatenated ASVs using the *assignTaxonomy* function in DADA2 and the naïve Bayesian classifier against the RDP training set 18 (*rdp_train_set_18*) (Callahan et al., 2016). DADA2 supports taxonomic classification of sequences generated using *justConcatenate* = TRUE, allowing the classifier to use the combined information from the independently denoised forward and reverse read segments despite the intervening N spacer. Taxonomic interpretation was therefore kept conservative, and species-level assignments were not used for general community characterization. Species-level assignments were additionally generated using *addSpecies* but were used only for exploratory purposes. ASVs lacking a phylum-level assignment and sequences classified as chloroplasts or non-target eukaryotic lineages were excluded from bacterial-community summaries. ASVs assigned at the phylum level but unresolved at lower taxonomic ranks were retained and represented as unclassified family or genus categories. The resulting ASV table was retained as count data for rarefaction and alpha-diversity analyses and transformed to relative abundances for taxonomic composition and descriptive beta-diversity analyses using *transform_sample_counts* in phyloseq v1.36.0 (McMurdie and Holmes, 2013).

Rarefaction curves were generated using the *rarecurve* function in vegan v2.6.4 with a step size of 10 reads (Oksanen et al., 2024). lpha diversity was quantified using Hill numbers of orders q=0, q=1, and q=2. These represent ASV richness, exponential Shannon diversity, and inverse Simpson diversity, respectively, and were estimated using the iNEXT package (Chao et al., 2014; Hsieh et al., 2024). Ordination analyses were used only to visualize sample-level compositional dissimilarities and were not subjected to inferential testing for lake identity or environmental context. Bray–Curtis dissimilarities were calculated

from sample-level relative-abundance profiles and visualized using principal coordinates analysis.

The Tota sample collected in December 2018 had been examined previously (Forero-Pineda et al., 2021) using a different analytical approach. In the present study, this historical sample was newly reprocessed together with the 2019 samples using the DADA2-based ASV workflow described above, so that taxonomic and diversity summaries were generated under a comparable bioinformatic framework. However, because this sample was non-contemporaneous with the 2019 dataset, it was presented separately as a historical reference and was excluded from ordination analyses and cross-lake summaries of the five-sample 2019 dataset.

Diversity, taxonomic composition, and ordination results were summarized descriptively at the individual-sample level. Owing to the limited, unbalanced, and non-contemporaneous sampling design, no hypothesis tests were conducted to evaluate the effects of lake identity, preservation status, sampling period, or water-column depth. Accordingly, no p-values or claims of statistically significant differences are reported.

2.3.3. Exploratory phylogenetic placement of ASVs assigned to *Mycobacterium*

ASVs assigned taxonomically to the genus *Mycobacterium* were aligned with 96 reference 16S rRNA gene sequences retrieved from RefSeq using MAFFT v7.525 (Katoh et al., 2009). Reference accession numbers and taxonomic information are provided in Table S2. A maximum-likelihood tree was inferred in IQ-TREE 2 (Minh et al., 2020) using the GTR+F substitution model and 1,000 ultrafast bootstrap replicates. The tree was visualized using ggtree v3.0.4 (Yu, 2020). Because the analysis was based on the V3–V4 fragment, phylogenetic proximity to named reference sequences was interpreted only as tentative placement and not as species-level identification.

2.3.4. Functional prediction and descriptive analysis

Functional profiles were computationally inferred from the 16S rRNA gene amplicon data using PICRUSt2 (Douglas et al., 2020). ASV sequences and their corresponding abundance table were used as input for phylogenetic placement and prediction of KEGG Ortholog (KO) family abundances based on the genomic content of available reference taxa. These outputs represent computational predictions rather than direct metagenomic or functional measurements.

A per-ASV Nearest Sequenced Taxon Index (NSTI) threshold of <2 was applied to exclude ASVs with particularly large phylogenetic distances from the reference genomes used for prediction. Sample-level weighted NSTI values generated by PICRUSt2 were retained as descriptive measures of the average phylogenetic proximity of the detected ASVs to available reference genomes.

Predicted KO abundances were subsequently mapped and aggregated into KEGG pathway categories using the *ko2kegg_abundance* function implemented in ggpicrust2 v2.5.17 (Douglas et al., 2020), based on the internal KO-to-KEGG pathway reference distributed with the package. For ordination, predicted pathway abundances from the complete pathway table were converted to within-sample relative abundances. For descriptive visualization, the 30 pathways with the highest total predicted abundances across the six samples were selected. Their predicted abundances were displayed by lake, while the two sampling events available for Fúquene and Tota were retained as separate stacked segments and were not treated as biological replicates. Bray–Curtis dissimilarities were then calculated among individual samples and visualized using principal coordinates analysis.

PICRUSt2 outputs were interpreted as predicted genomic potential associated with particular gene families and pathways and not as direct evidence of gene presence,

expression, metabolic activity, or ecosystem process rates. Predictions may be affected by the availability and representativeness of reference genomes and may not capture strain-level genomic variation or genes absent from the reference database. KEGG pathways with organism-, disease-, drug-resistance-, or eukaryote-associated labels were treated only as database annotation categories containing homologous genes and were not interpreted as evidence of the named organisms, pathogenicity, virulence, antimicrobial resistance, infection risk, or the corresponding eukaryotic processes in the sampled lakes.

3. Results

3.1. Taxonomic composition of microbial communities among samples

After quality filtering, denoising, paired-read concatenation, and chimera assessment, 556,915 reads were retained across the six samples, corresponding to an average of 92,819 reads per sample. Of these, 341,997 reads remained in ASVs retained after the taxonomic filtering described in the Methods. The number of reads represented in the retained ASVs ranged from 42,174 in Colorado in October 2019 to 86,554 in Tota in November 2019 (Fig. S1a). Following taxonomic filtering, 3,153 ASVs were retained for community characterization. ASVs assigned at the phylum level but unresolved at the family or genus level were retained and represented as unclassified categories.

A total of 34 bacterial phyla and the archaeal phylum Euryarchaeota were detected across the six samples (Fig. 1a–b; Table S1). Phylum-level relative abundances were strongly right-skewed, with a limited number of phyla accounting for most of the sequences and most detected phyla occurring at low abundance (Fig. 1a). Across all samples, the predominant bacterial phyla were Actinobacteria (34.5%), Proteobacteria (29.7%), Bacteroidetes (12.7%), Cyanobacteria (8.5%), and Verrucomicrobia (7.9%). Approximately 80% of the detected

phyla each contributed less than 1% of the total relative abundance, with Euryarchaeota among these low-abundance phyla (Fig. 1a; Table S1).

At the individual-sample level, Actinobacteria and Proteobacteria were major components of the microbial communities, whereas the relative contributions of Bacteroidetes, Verrucomicrobia, Cyanobacteria, Firmicutes, and other less-abundant phyla varied among samples (Fig. 1b). Planctomycetes were detected in all four lakes but reached their highest relative abundance in Colorado, where they represented 11.3% of the community (Fig. 1b; Table S1).

Taxonomic assignment recovered 179 families and 366 genera across the six samples (Table S1). Figure 1c summarizes the 20 most abundant families and the 20 most abundant genera within six selected phyla: Actinobacteria, Armatimonadetes, Bacteroidetes, Planctomycetes, Proteobacteria, and Verrucomicrobia. Less-abundant taxa were grouped as "Other families" or "Other genera." Family- and genus-level profiles varied among the six samples, including between the two samples available for Fúquene and Tota (Fig. 1c). Because each sampling date was represented by a single sample, these differences were not interpreted as evidence of temporal change or stability.

In Calderona, *Polynucleobacter* and *Flavobacterium* reached relative abundances of 29.5% and 20.3%, respectively, whereas *Phragmitibacter* reached 24.2% in Colorado (Fig. 1c; Table S1). Low-abundance archaeal genera were also detected: *Methanobacterium*, *Methanothrix*, and *Methanocella* occurred in Colorado, whereas *Methanosarcina* occurred in Fúquene in September 2019 (Table S1). Other bacterial genera, including *Leptospira* and *Legionella*, were also detected. *Leptospira* occurred in the Fúquene and Tota samples at 0.01–0.36%, whereas *Legionella* occurred in all six samples at 0.10–4.60%, reaching its

highest relative abundance in Calderona (Table S1). These assignments are reported only at the genus level and do not establish species identity, viability, pathogenicity, or health risk.

3.2. Alpha and beta diversity among samples

Observed ASV richness varied among samples, with the highest value recorded in Fúquene in September 2019, followed by Colorado and Fúquene in April 2019. Intermediate richness was observed in Tota in November 2019, whereas Calderona and the historical Tota sample from December 2018 had the lowest values (Fig. 1d). Observed and rarefied richness estimates were closely aligned (Fig. S1b), and the rarefaction curves approached asymptotes in all samples (Fig. S1c), suggesting that the sequencing depth captured most of the ASV diversity detectable under the analytical conditions used.

Alpha diversity was further evaluated using coverage-standardized Hill numbers at a common sample coverage of 0.999 (Fig. 1d). The diversity orders q=0, q=1, and q=2 represent ASV richness, exponential Shannon diversity, and inverse Simpson diversity, respectively. Increasing values of q assign progressively greater weight to abundant ASVs.

At q=1, the two Fúquene samples had the highest diversity, corresponding to comparatively large effective numbers of common ASVs. Colorado showed comparatively high richness at q=0 but lower diversity at q=1 and q=2, consistent with a less even distribution of ASV relative abundances. The historical Tota sample from December 2018 had comparatively low richness but the highest q=2 diversity, corresponding to a relatively large effective number of dominant ASVs. In contrast, the Tota sample from November 2019 had lower q=1 and q=2 diversity despite its intermediate richness.

Bray–Curtis principal coordinates analysis showed differences in community composition among samples in ordination space (Fig. 1e). The first two axes accounted for 67.6% of the variation in Bray–Curtis dissimilarities, with PCoA1 explaining 35.1% and

PCoA2 explaining 32.5%. Calderona and Colorado were positioned on the positive side of PCoA1, whereas the Fúquene and Tota samples were positioned on its negative side. The two Fúquene samples plotted relatively close to one another and were separated from the Tota samples primarily along PCoA2. The two Tota samples were also positioned relatively close to one another in the lower-left region of the ordination, including the historical sample reprocessed in the present study.

All alpha- and beta-diversity patterns were interpreted descriptively. No PERMANOVA or other lake-level or date-level inferential comparisons were performed because only six samples were available, sampling was uneven among lakes, and each sampling date was represented by a single sample. Consequently, the proximity of samples from the same lake in the ordination was not interpreted as evidence of temporal stability.


a)
Phylum
Actinobacteria
Proteobacteria
Bacteroidetes
Cyanobacteria/Chloroplast
Verrucomicrobia
Planctomycetes
Candidatus_Saccharibacteria
Armatimonadetes
Firmicutes
Acidobacteria
Parcubacteria
Deinococcus-Thermus
Chloroflexi
Chlamydiae
Fusobacteria
Gemmatimonadetes
Euryarchaeota
Spirochaetes
SR1
Pacearchaeota
Campilobacterota
Microgenomates
candidate_division_ZB3
candidate_division_WPS-2
BRC1
Elusimicrobia
candidate_division_WPS-1
Hydrogenedentes
Woesearchaeota
Nitrospirae
Tenericutes
Ignavibacteriae
Lentisphaerae
Cloacimonetes
Synergistetes
0.0
0.1
1.0
10.0
100.0
Abundance (log10)
b)
Tota Nov 2019
Tota Dec 2018
Fúquene Sep 2019
Fúquene Apr 2019
Colorado Oct 2019
Calderona Oct 2019
0%
20%
40%
60%
80%
100%
Relative abundance
Phylum
Actinobacteria
Bacteroidetes
Verrucomicrobia
Candidatus_Saccharibacteria
Firmicutes
Euryarchaeota
Proteobacteria
Cyanobacteria/Chloroplast
Planctomycetes
Armatimonadetes
Acidobacteria
Other phyla
c)
Actinobacteria
Armatimonadetes
Bacteroidetes
Planctomycetes
Proteobacteria
Verrucomicrobia
Family
Genus
Relative abundance (%)
Tota Nov-2019
Tota Dec-2018
Calderona Oct-2019
Colorado Oct-2019
Fúquene Sep-2019
Fúquene Apr-2019
Family
Ilumatobacteraceae
Mycobacteriaceae
Intrasporangiaceae
Flavobacteriaceae
Burkholderiaceae
Lacipirellulaceae
Comamonadaceae
Methylocystaceae
Candidatus_Pelagibacter
Methylophilaceae
Acetobacteraceae
Steroidobacteraceae
Caulobacteraceae
Sphingomonadaceae
Bradyrhizobiaceae
Verrucomicrobiaceae
Other families
Unclassified family [Actinobacteria]
Genus
Unclassified genus [Actinobacteria]
Unclassified genus [Ilumatobacteraceae]
Unclassified genus [Intrasporangiaceae]
Rhodoluna
Flavobacterium
Lacipirellula
Methylocystis
Unclassified genus [Methylophilaceae]
Steroidobacter
Unclassified genus [Comamonadaceae]
Other genera
Mycobacterium
Desertimonas
Unclassified genus [Microbacteriaceae]
Unclassified genus [Chitinophagaceae]
Sediminibacterium
Polynucleobacter
Unclassified genus [Candidatus_Pelagibacter]
Limnohabitans
Unclassified genus [Acetobacteraceae]
Phragmitibacter
d)
Common sample coverage = 0.999
q = 0
ASV richness
q = 1
Exponential Shannon
q = 2
Inverse Simpson
Coverage-standardized Hill diversity
Calderona Oct-2019
Colorado Oct-2019
Fúquene Apr-2019
Fúquene Sep-2019
Tota Dec-2018
Tota Nov-2019
e)
Fúquene Sep-2019
Fúquene Apr-2019
Colorado Oct-2019
Calderona Oct-2019
Tota Dec-2018
Tota Nov-2019
PCoA2 (32.5%)
PCoA1 (35.1%)
Dataset
Primary 2019
Historical reprocessed

398

399

Figure 1. Taxonomic composition and diversity of microbial communities in six water samples from four Colombian Andean lakes. (a) Distribution of phylum-level relative abundances on a log10 scale. (b) Relative abundance of dominant phyla in each sample. (c) Family- and genus-level composition within six selected phyla; radial scales were adjusted independently among panels. (d) Coverage-standardized Hill diversity at q=0, q=1, and q=2, with 95% confidence intervals. The confidence intervals associated with coverage-standardized Hill estimates represent uncertainty in the diversity estimator and were not used to test differences among lakes or samples. (e) Bray–Curtis PCoA of ASV relative abundances; PCoA1 and PCoA2 explained 35.1% and 32.5% of the variation, respectively.

3.3. Exploratory phylogenetic placement of *Mycobacterium* ASVs detected in Tota and Fúquene

ASVs assigned to the genus *Mycobacterium* were included in an exploratory phylogenetic analysis together with reference 16S rRNA gene sequences from named *Mycobacterium* species. Some ASVs were placed within well-supported clades containing reference sequences annotated as *M. celatum*, *M. noviomagense*, *M. szulgai*, *M. pseudokansasii*, and *M. cookii* (Fig. S2). Other ASVs occupied less-resolved positions within clades containing reference sequences from multiple species, including *M. attenuatum*, *M. innocens*, *M. arosiense*, *M. haemophilum*, *M. nebraskense*, *M. cookii*, and *M. pseudoshottsii*.

These placements indicate phylogenetic affinity to lineages represented by the selected reference sequences but do not establish species identity. Because the V3–V4 region of the 16S rRNA gene provides limited discrimination among closely related *Mycobacterium* species, the observed associations were interpreted only as exploratory phylogenetic

placements rather than definitive species-level assignments. No conclusions were drawn regarding organism viability, pathogenicity, virulence, host infection, or health risk.

3.4. Functional prediction of bacterial communities presents in aquatic systems of Northern Andean Region of Colombia

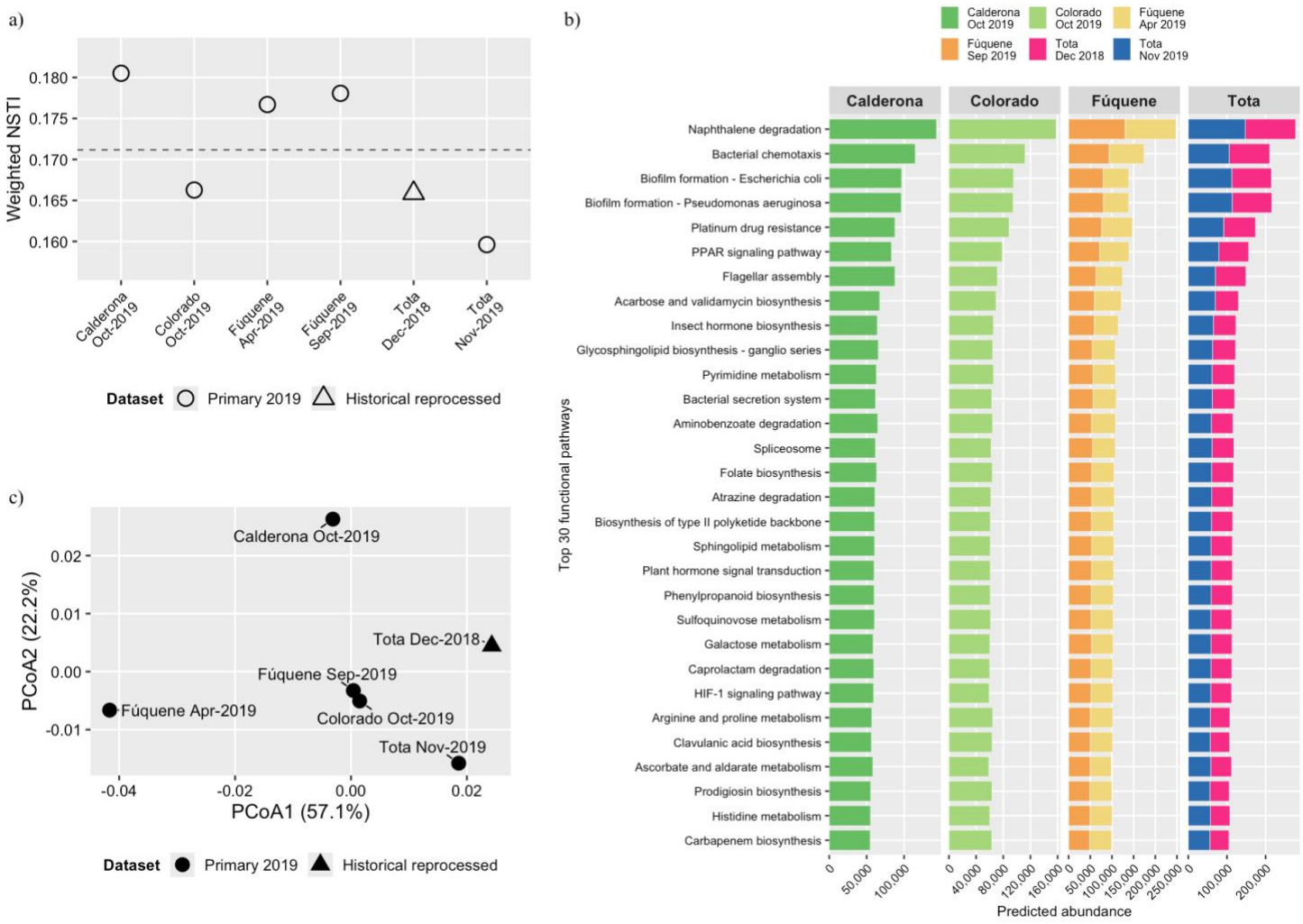
The functional potential of the microbial communities was inferred using PICRUSt2 across the six samples. Sample-level weighted Nearest Sequenced Taxon Index (NSTI) values ranged from 0.160 in Tota in November 2019 to 0.181 in Calderona, with a mean of 0.171 (Fig. 2a). At the individual-ASV level, ASVs with NSTI values ≥2 were excluded from functional prediction. Under this criterion, 569 of 3,153 ASVs (18.0%) were excluded, whereas 2,584 ASVs (82.0%) were retained. The 30 pathways with the highest total predicted abundances included database categories associated with naphthalene degradation, bacterial chemotaxis, biofilm formation, bacterial secretion, flagellar assembly, pyrimidine metabolism, folate biosynthesis, and sphingolipid metabolism (Fig. 2b). Their predicted abundances varied among individual samples.

Several predicted pathways carried organism- or disease-associated KEGG labels, including biofilm-formation pathways named after *Escherichia coli* and *Pseudomonas aeruginosa*, platinum drug resistance, HIF-1 signaling, PPAR signaling, and plant- or insect-associated pathways (Fig. 2b). These labels reflect sequence homology and pathway nomenclature in the KEGG reference database and do not demonstrate the occurrence of the named organisms, pathogenic activity, drug resistance, or the corresponding eukaryotic biological processes in the sampled lakes.

Principal coordinates analysis based on Bray–Curtis dissimilarities showed variation among sample-level predicted functional profiles in ordination space (Fig. 2c). The first two axes accounted for 79.3% of the variation, with PCoA1 explaining 57.1% and PCoA2

447 explaining 22.2%. The ordination was interpreted descriptively and was not subjected to
448 inferential testing.

449 PICRUSt2 outputs represent predicted genomic potential based on phylogenetic
450 placement and available reference genomes rather than direct measurements of gene
451 presence, expression, metabolic activity, or ecosystem function. Because only six samples
452 were analyzed and the sampling design lacked balanced environmental and spatial
453 replication, variation in predicted functional profiles was not attributed to anthropogenic
454 pressure, lake size, spatial heterogeneity, geological origin, or sampling period.



455

456 **Figure 2. Predicted functional profiles of microbial communities across the six lake-**
457 **water samples.** (a) Sample-level weighted Nearest Sequenced Taxon Index (NSTI) values.
458 The dashed horizontal line indicates the mean weighted NSTI across the six samples (0.171).
459 (b) Predicted abundances of the 30 most abundant KEGG pathways inferred using
460 PICRUSt2. Results are displayed by lake; bars for Fúquene and Tota are partitioned
461 according to their two sampling events. (c) Principal coordinates analysis of Bray–Curtis
462 dissimilarities calculated from sample-level predicted functional profiles. PCoA1 and PCoA2
463 accounted for 57.1% and 22.2% of the variation, respectively. In panels (a) and (c), circles

represent the primary 2019 samples, whereas the triangle represents the historical Tota sample from December 2018 that was reprocessed in the present study. All functional profiles are predictions based on phylogenetic placement and reference genomes and should not be interpreted as direct evidence of gene presence, expression, metabolic activity, or ecosystem function. The ordination and NSTI patterns were interpreted descriptively and were not subjected to inferential testing.

4. Discussion

This exploratory survey provides a descriptive overview of bacterial community composition in six water samples from four Colombian Andean lakes. Actinobacteria, Proteobacteria, Bacteroidetes, Cyanobacteria, and Verrucomicrobia accounted for most of the classified sequences, with Actinobacteria and Proteobacteria being the most abundant phyla overall (Fig. 1a, b). Low-abundance ASVs assigned to archaeal taxa, including the methanogenic genera *Methanobacterium*, *Methanothrix*, *Methanocella*, and *Methanosarcina*, were also detected (Table S1). However, because the primers used primarily targeted the bacterial 16S rRNA gene V3–V4 region, archaeal diversity was probably incompletely represented. Previous studies have documented methanogenic archaea in lake sediments and have shown that methanogenesis can occur under locally anoxic physicochemical conditions (Ward and Frea, 1980). In eutrophic lakes, microbial decomposition of organic matter can contribute to oxygen depletion and create conditions favorable for methanogenic activity (Kersti Kangro et al., 2007; Smith and Schindler, 2009), while methane production constitutes an important component of freshwater carbon cycling and greenhouse-gas emissions (Raymond et al., 2013; Tranvik et al., 2009; Ward and Frea, 1980). Nevertheless, the detection of methanogen-related 16S rRNA gene sequences in the present water samples does not demonstrate active

488 methanogenesis, anaerobic conditions, or methane production. Because dissolved oxygen,
489 organic matter, nutrient concentrations, and methane fluxes were not measured concurrently,
490 the ecological activity of these archaeal lineages cannot be determined from the available
491 data.

492 Community composition and coverage-standardized diversity differed descriptively
493 among the six samples (Fig. 1b–e). Calderona showed comparatively low richness and
494 diversity, whereas Colorado showed relatively high richness at q = 0 but lower diversity at q
495 = 1 and q = 2 than the Fúquene and Tota samples (Fig. 1c, d). Both Calderona and Colorado
496 were characterized by high relative abundances of a limited number of taxa (Fig. 1c). In
497 particular, *Polynucleobacter* and *Flavobacterium* were abundant in Calderona, whereas
498 *Phragmitibacter* was abundant in Colorado (Fig. 1c; Table S1). Fúquene showed
499 comparatively higher richness and diversity, together with a different distribution of
500 dominant taxa (Fig. 1c, d). Similar variation among freshwater microbial communities has
501 been associated in other studies with differences in local environmental conditions, habitat
502 structure, and resource availability (Diao et al., 2017; Hoetzinger et al., 2019; Nuy et al.,
503 2020; Paver et al., 2020). Natural processes may also influence microbial composition
504 independently of direct anthropogenic disturbance (Sessitsch et al., 2023). For example,
505 interactions among sediments, littoral habitats, and limnetic waters can affect freshwater
506 microbial assemblages through sediment resuspension and littoral–pelagic coupling (Jones
507 and Lennon, 2010; Søndergaard et al., 2003). In Andean systems, surrounding vegetation and
508 catchment characteristics may additionally influence the quantity and quality of organic
509 matter entering lakes, while previous studies of Tota have documented hydroclimatic
510 variability that can modify both natural and anthropogenic pressures (Aranguren-Riaño et
511 al., 2018; Torres-Barrera et al., 2017). Nevertheless, these mechanisms were not evaluated

512 in the present study because no concurrent physicochemical, hydrological, or land-use
513 variables were measured. Therefore, the observed differences cannot be attributed causally
514 to anthropogenic pressure, preservation status, lake size, spatial heterogeneity, sampling
515 period, water-column depth, or any specific environmental driver. At a broader geographic
516 scale, the bacterial composition observed in these Colombian Andean lakes shares several
517 features with that reported from high-altitude and polar lake systems. Antarctic lake waters
518 examined by Cui et al. (2025) were dominated largely by Proteobacteria, Actinobacteriota,
519 and Bacteroidota, while Cai et al. (2026) similarly identified Proteobacteria,
520 Actinobacteriota, Bacteroidota, Verrucomicrobiota, and Cyanobacteria among the
521 predominant groups across ice-covered lakes of the Qing-Tibetan Plateau, Arctic, and
522 Antarctica. These groups also constituted important components of the Colombian samples,
523 although their relative abundances differed among lakes (Fig. 1a, b). Both studies further
524 documented substantial variation in microbial diversity and community composition among
525 lakes or geographic regions, supporting the broader observation that high-altitude and cold-
526 lake bacterial assemblages can be highly heterogeneous rather than exhibiting a uniform
527 community structure (Cai et al., 2026; Cui et al., 2025).

528 No physicochemical variables were measured concurrently with microbiological
529 sampling, including nutrient concentrations, dissolved oxygen, chlorophyll-a, pH,
530 temperature, conductivity, or indicators of organic matter. Consequently, relationships
531 between microbial community composition and environmental conditions could not be
532 evaluated directly. Previous studies have shown that eutrophication, nutrient enrichment,
533 thermal and oxygen gradients, hydrological variability, and biological interactions can
534 influence freshwater microbial assemblages (Andrade-Sossa et al., 2023; Giongo et al., 2023;
535 Hoang et al., 2023). Natural disturbances, such as flooding and seasonal mixing, and

anthropogenic disturbances, such as runoff and pollution, may modify resource availability and community assembly (Hoang et al., 2023). Likewise, changes in temperature, light, organic matter, rainfall, evaporation, phytoplankton abundance, and zooplankton grazing can affect bacterial communities by altering nutrient dynamics and substrate availability (Andrade-Sossa et al., 2023; Giongo et al., 2023; Hoang et al., 2023; Másmela-Mendoza et al., 2019). These mechanisms provide relevant context from the literature but were not measured or tested in the present study and therefore cannot be invoked as explanations for the observed differences among samples.

Differences between the two Tota samples likewise cannot be attributed to seasonal change. They represent only two sampling dates, one of which corresponds to a historical December 2018 sample reprocessed alongside the primary 2019 dataset, and they do not constitute replicated temporal sampling. Evaluation of microbial community stability, resistance, or resilience requires a defined disturbance and an adequately replicated temporal design that captures both community change and recovery (Shade et al., 2012). Accordingly, the observed variation in bacterial abundance and diversity (Fig. 1b–d) is interpreted only as sample-level variation and not as evidence of seasonal dynamics, temporal instability, or responses to nutrient, temperature, oxygen, hydrological, or biological gradients.

The exploratory phylogenetic placement of *Mycobacterium*-assigned ASVs from the Fúquene and Tota samples indicated affinities with several reference lineages included in the analysis (Fig. S2). However, the short V3–V4 region of the 16S rRNA gene provides insufficient resolution for reliable species-level identification within *Mycobacterium*, and proximity to a named reference sequence should not be interpreted as confirmation of species identity. The results therefore support only the occurrence of multiple *Mycobacterium*-related lineages in these samples. This observation is consistent with the broad distribution

of nontuberculous mycobacteria in freshwater and marine environments reported in previous studies (Barletta et al., 2010; Gcebe et al., 2018; Nuy et al., 2020). Cultured nontuberculous mycobacteria have also been associated with mycobacteriosis in fish and other aquatic organisms, with potential consequences for animal health and aquaculture (Gcebe et al., 2018; Hashish et al., 2018). Nevertheless, those findings cannot be extrapolated to the environmental ASVs detected here because the present analysis did not assess species identity, viability, abundance by targeted quantification, or pathogenic potential. Definitive identification would require recovery of cultured isolates followed by whole-genome sequencing or analysis using validated higher-resolution molecular markers.

ASVs assigned to *Leptospira* were detected in the Fúquene and Tota samples at low relative abundances of 0.01–0.36%, whereas *Legionella*-assigned ASVs occurred in all six samples at 0.11–4.60%, with the highest relative abundance observed in Calderona (Table S1). Previous studies have shown that some *Leptospira* species can persist in freshwater and cause leptospirosis following exposure to contaminated water, whereas some *Legionella* species are associated with aquatic systems and may proliferate under favorable conditions (Bradley and Lockaby, 2023; Gattuso et al., 2022). However, both genera include environmentally distributed lineages that cannot be distinguished from pathogenic species using the short 16S rRNA gene region analyzed here. Their detection therefore does not demonstrate the occurrence of disease-causing species, viable cells, active transmission, pathogenic activity, or public-health risk in the sampled lakes. These observations should instead be regarded as preliminary genus-level records that may guide future investigation using targeted species-specific assays, quantitative methods, culture-based analyses, and viability assessments.

PICRUSt2 indicated descriptive variation in predicted functional profiles among the six samples (Fig. 2b, c). Among the 30 pathways with the highest total predicted abundances were categories associated with naphthalene degradation, bacterial chemotaxis, biofilm formation, bacterial secretion systems, flagellar assembly, pyrimidine metabolism, folate biosynthesis, and sphingolipid metabolism (Fig. 2b). Bray–Curtis ordination also separated the sample-level predicted profiles, with the first two PCoA axes accounting for 79.3% of the variation (Fig. 2c). Because PICRUSt2 infers gene-family abundances from the phylogenetic placement of ASVs and the genomic content of reference taxa, this variation is expected to reflect, at least partly, the taxonomic differences observed among samples. Similar links between microbial community composition and functional potential have been described in freshwater ecosystems, where shifts in community structure may alter the representation of genes involved in carbon processing, nutrient cycling, and other metabolic functions (Battin et al., 2003, 2016; Paver et al., 2020). Comparable geographic heterogeneity has also been reported for predicted microbial metabolic functions in high-altitude and polar lakes. Cai et al. (2024) found that prokaryotic communities in ice-covered Hoh Xil lakes were comparatively enriched in functions associated with sulfur metabolism, whereas those from Arctic and Antarctic lakes showed greater representation of carbon- and nitrogen-related metabolic processes. Thus, the variation in predicted functional profiles among the Colombian lake samples is consistent, at a general level, with evidence that the potential metabolic repertoire of lake microbial communities varies substantially among geographically and environmentally distinct high-altitude systems (Cai et al., 2024). Nevertheless, the present predictions do not demonstrate that the corresponding genes were present, expressed, or metabolically active, and no inferential association with anthropogenic-pressure context or preservation status was tested.

607 The representation of pathways related to naphthalene degradation and biofilm formation
608 (Fig. 2b) suggests that genes homologous to those included in these KEGG pathways may
609 occur in the reference genomes most closely related to the detected ASVs. Microorganisms
610 capable of degrading naphthalene and other aromatic compounds employ diverse catabolic
611 pathways, while biofilm formation can facilitate surface colonization and microbial
612 persistence in aquatic environments (Mohapatra and Phale, 2021). However, the predicted
613 occurrence of these pathway categories does not demonstrate the presence of naphthalene or
614 other pollutants, active contaminant degradation, biofilm development, or adaptation to
615 measured environmental conditions in the sampled lakes. Such interpretations would require
616 direct chemical measurements, metagenomic confirmation, transcriptomic evidence, or
617 experimental activity assays.

618 Nutrient enrichment and eutrophication are widely recognized as drivers of changes
619 in microbial diversity and metabolic processes in freshwater ecosystems, with potential
620 consequences for primary production, organic-matter decomposition, oxygen availability,
621 and nutrient cycling (Carpenter et al., 1998; Cotner and Biddanda, 2002; Paerl and Huisman,
622 2009; Smith and Schindler, 2009). These studies provide relevant ecological context for
623 interpreting functional variation among freshwater microbial communities. However,
624 nutrient concentrations, chlorophyll-a, dissolved oxygen, organic matter, and other
625 physicochemical variables were not measured concurrently in the present study. Therefore,
626 the predicted functional patterns cannot be attributed to eutrophication, nutrient availability,
627 or other environmental gradients.

628 Several predicted pathways also carried KEGG labels referring to particular organisms,
629 infectious diseases, drug resistance, or eukaryotic processes. These labels reflect the structure
630 and nomenclature of the KEGG database and the presence of homologous genes within

reference pathways. They do not demonstrate the occurrence of the named organisms, pathogenicity, virulence, antimicrobial resistance, infection, or the corresponding eukaryotic biological processes in the lakes. Accordingly, the PICRUSt2 results should be interpreted as a descriptive approximation of potential genomic capabilities rather than as direct measurements of gene presence, ecological function, or metabolic activity.

5. Conclusions

This exploratory 16S rRNA gene survey provides a baseline description of microbial community composition in six water samples from four Colombian Andean lakes. Actinobacteria, Proteobacteria, Bacteroidetes, Cyanobacteria, and Verrucomicrobia were the predominant bacterial phyla, and the samples differed descriptively in taxonomic composition and coverage-standardized diversity. Exploratory phylogenetic placement supported the occurrence of multiple *Mycobacterium*-related ASVs in Fúquene and Tota, although the V3–V4 region did not provide sufficient resolution for species-level identification. ASVs assigned to *Leptospira* and *Legionella* were also detected, but these genus-level assignments do not establish pathogenicity, viability, or public-health risk.

PICRUSt2 predicted variation in functional profiles among samples, but these outputs represent inferred genomic potential rather than direct evidence of gene presence, expression, or metabolic activity. Because the study included only six samples and lacked concurrent physicochemical measurements and balanced spatial or temporal replication, the observed patterns cannot be attributed to anthropogenic pressure, environmental gradients, or temporal change. Future studies combining replicated sampling, concurrent environmental measurements, metagenomic sequencing, and targeted taxonomic validation

654 will be necessary to determine the ecological processes underlying the observed microbial
655 patterns.

Author contributions

AGP and NAR conceptualized and designed the study. APR and NAR conducted fieldwork. JMS performed laboratory procedures. AGP conducted data analysis. AGP, JMS, APR, and NAR contributed to writing and revising the manuscript. All authors reviewed and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Data Availability Statement

The raw sequencing datasets used in the present study are available in the NCBI Sequence Read Archive repository (the Bioproject; PRJNA720890).

Ethical approval

This study did not involve human participants, vertebrate animals, or any endangered or protected species. Water sampling was conducted in public lakes following standard environmental research protocols, without any intervention that could harm the ecosystem. Therefore, ethical approval was not required. All necessary permits for sample collection were obtained from the relevant local authorities.

Acknowledgments

Authors would also like to acknowledge the field-work assistance of the local community members; whose support and knowledge were invaluable to our sampling efforts.

Funding

This research was supported by Dirección de Investigaciones de la Universidad Pedagógica y Tecnológica de Colombia, Tunja, Boyacá.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT / OpenAI in order to improve drafting and readability of this manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Supplementary information

Available on https://doi.org/10.5281/zenodo.22695690

Figure S1. Sequencing depth and rarefaction assessment of bacterial ASVs across the six lake-water samples. (a) Number of reads assigned to ASVs retained after the taxonomic filtering described in the Methods. Values above bars indicate the number of retained reads

per sample. (b) Relationship between observed ASV richness and richness after rarefaction. The dashed diagonal line represents equality between observed and rarefied richness. (c) Sample-based rarefaction curves showing the accumulation of rarefied ASVs with increasing sequencing depth. Curves approaching an asymptote indicate that additional sequencing would be expected to recover relatively few additional ASVs under the analytical conditions used. Sample labels identify the lake and sampling event.

Figure S2. Exploratory maximum-likelihood phylogenetic placement of *Mycobacterium* ASVs detected in samples from Tota and Fúquene. The analysis included ASV sequences assigned to *Mycobacterium* and reference 16S rRNA gene sequences from named species. Branch-support values are shown at the corresponding nodes. Sequences identified in this study are highlighted in blue. Species names associated with reference sequences indicate phylogenetic affinity only and should not be interpreted as definitive species-level assignments because the analysis was based on the V3–V4 region of the 16S rRNA gene.

Supplementary table 1. Complete metataxonomic classification and relative abundance of ASVs detected in the six-freshwater lake-water samples analyzed. Taxonomic assignments are reported at the available ranks, with unresolved classifications indicated as unclassified or NA. Relative abundances are presented for each individual sample.

Supplementary table 2. Reference *Mycobacterium* sequences included in the exploratory phylogenetic analysis. The table lists the taxonomic identification of each reference sequence and its corresponding GenBank accession number.

6. References

Andrade-Sossa, C., Alvarez-Silva, J. P., Aranguren-Riaño, N., Cupitra-Gómez, O. S., Villabona-González, S. L., Torres-Bejarano, A. M., and López, C.: Zooplankton studies in Colombian fresh and brackish water ecosystems: A review and future perspectives, Limnologica, 100, 126081, https://doi.org/10.1016/j.limno.2023.126081, 2023.

Aranguren-Riaño, N. J., Shurin, J. B., Pedroza-Ramos, A., Muñoz-López, C. L., López, R., and Cely, O.: Sources of nutrients behind recent eutrophication of Lago de Tota, a high mountain Andean lake, Aquat. Sci., 80, 39, 2018.

Barletta, M., Jaureguizar, A. J., Baigun, C., Fontoura, N. F., Agostinho, A. A., Almeida-Val, V. M. F., Val, A. L., Torres, R. A., Jimenes-Segura, L. F., Giarrizzo, T., Fabré, N. N., Batista, V. S., Lasso, C., Taphorn, D. C., Costa, M. F., Chaves, P. T., Vieira, J. P., and Corrêa, M. F. M.: Fish and aquatic habitat conservation in South America: a continental overview with emphasis on neotropical systems, J. Fish Biol., 76, 2118–2176, https://doi.org/10.1111/j.1095-8649.2010.02684.x, 2010.

Battin, T. J., Kaplan, L. A., Newbold, J. D., and Hansen, C. M. E.: Contributions of microbial biofilms to ecosystem processes in stream mesocosms, Nature, 426, 439–442, 2003.

Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M., and Packmann, A. I.: The ecology and biogeochemistry of stream biofilms, Nat. Rev. Microbiol., 14, 251–263, https://doi.org/10.1038/nrmicro.2016.15, 2016.

Bradley, E. A. and Lockaby, G.: Leptospirosis and the Environment: A Review and Future Directions, Pathogens, 12, 1167, https://doi.org/10.3390/pathogens12091167, 2023.

Cai, M., Wang, B., Han, J., Yang, J., Zhang, X., Guan, X., and Jiang, H.: Microbial difference and its influencing factors in ice-covered lakes on the three poles, Environ. Res., 252, 118753, https://doi.org/10.1016/j.envres.2024.118753, 2024.

Cai, M., Zhang, P., Yang, J., Han, M., Huang, J., Xiao, H., Han, J., Zhang, X., and Jiang, H.: Salinity and nutrient availability dictate methane production and its temperature sensitivity in lakes, Water Res., 289, 124836, https://doi.org/10.1016/j.watres.2025.124836, 2026.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P.: DADA2: High-resolution sample inference from Illumina amplicon data, Nat. Methods, 13, 581–583, https://doi.org/10.1038/nmeth.3869, 2016.

9. Carpenter, S. R. *et al.* Nonpoint pollution of surface waters with Phophorus and Nitrogen. **8**, (1998).

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., and Ellison, A. M.: Rarefaction and extrapolation with Hill numbers: a framework for sampling and

estimation in species diversity studies, Ecol. Monogr., 84, 45–67, https://doi.org/10.1890/13-0133.1, 2014.

Cotner, J. B. and Biddanda, B. A.: Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems, Ecosystems, 5, 105–121, https://doi.org/10.1007/s10021-001-0059-3, 2002.

Cui, H., Han, J., Li, B., Sun, Y., Gong, D., Fan, X., Pavel, T., Zhang, D., Gao, L., and Jiang, H.: Bacterial Distribution in the Glacier Borehole Meltwater on the Eastern Broknes Peninsula of the Larsemann Hills and Adjacent Lake Water, East Antarctica, Microorganisms, 13, 679, https://doi.org/10.3390/microorganisms13030679, 2025.

Diao, M., Sinnige, R., Kalbitz, K., Huisman, J., and Muyzer, G.: Succession of Bacterial Communities in a Seasonally Stratified Lake with an Anoxic and Sulfidic Hypolimnion, Front. Microbiol., 8, https://doi.org/10.3389/fmicb.2017.02511, 2017.

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., and Langille, M. G. I.: PICRUSt2 for prediction of metagenome functions, Nat. Biotechnol., 38, 685–688, https://doi.org/10.1038/s41587-020-0548-6, 2020.

Dowdell, K., Haig, S.-J., Caverly, L. J., Shen, Y., LiPuma, J. J., and Raskin, L.: Nontuberculous mycobacteria in drinking water systems – the challenges of characterization and risk mitigation, Curr. Opin. Biotechnol., 57, 127–136, https://doi.org/10.1016/j.copbio.2019.03.010, 2019.

Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., McDowell, W. H., Kortelainen, P., Caraco, N. F., Melack, J. M., and Middelburg, J. J.: The global abundance and size distribution of lakes, ponds, and impoundments, Limnol. Oceanogr., 51, 2388–2397, https://doi.org/10.4319/lo.2006.51.5.2388, 2006.

Falkinham, J. O.: Ecology of Nontuberculous Mycobacteria, Microorganisms, 9, 2262, https://doi.org/10.3390/microorganisms9112262, 2021.

Forero-Pineda, N., Pedroza-Ramos, A., Marin-Suarez, J., Aranguren-Riaño, N., and Gómez-Palacio, A.: Bacteriome depiction and the trophic status of the largest Northern highland lake from Andes system: Lago de Tota, Boyacá, Colombia, Arch. Microbiol., 203, 3695–3705, https://doi.org/10.1007/s00203-021-02341-3, 2021.

Gattuso, G., Rizzo, R., Lavoro, A., Spoto, V., Porciello, G., Montagnese, C., Cinà, D., Cosentino, A., Lombardo, C., Mezzatesta, M. L., and Salmeri, M.: Overview of the Clinical and Molecular Features of Legionella Pneumophila: Focus on Novel Surveillance and Diagnostic Strategies, Antibiotics, 11, 370, https://doi.org/10.3390/antibiotics11030370, 2022.

Gcebe, N., Michel, A. L., and Hlokwe, T. M.: Non-tuberculous Mycobacterium species causing mycobacteriosis in farmed aquatic animals of South Africa, BMC Microbiol., 18, 32, https://doi.org/10.1186/s12866-018-1177-9, 2018.

Gibson, D. K., Bird, B. W., Wattrus, N. J., Escobar, J., Ahmed, M., Fonseca, H., Velasco, F., Fernandez, A., and Polissar, P. J.: Characterizing late Quaternary lake-level variability in Lago de Tota, Colombian Andes, with CHIRP seismic stratigraphy, J. Paleolimnol., 62, 319–335, https://doi.org/10.1007/s10933-019-00089-x, 2019.

Giongo, A., dos Anjos Borges, L. G., Simão, T. L. L., Eizirik, E., and Utz, L. R. P.: Structure and Dynamics of Periphyton in a Neotropical Freshwater Lake, with Emphasis on Ciliates and Their Relationships with Bacterial Taxa, Microb. Ecol., 86, 187–199, https://doi.org/10.1007/s00248-022-02101-w, 2023.

Hashish, E., Merwad, A., Elgaml, S., Amer, A., Kamal, H., Elsadek, A., Marei, A., and Sitohy, M.: Mycobacterium marinum infection in fish and man: epidemiology, pathophysiology and management; a review, Vet. Q., 38, 35–46, https://doi.org/10.1080/01652176.2018.1447171, 2018.

Hassan, M., Essam, T., and Megahed, S.: Illumina sequencing and assessment of new cost-efficient protocol for metagenomic-DNA extraction from environmental water samples, Braz. J. Microbiol., 49, 1–8, https://doi.org/10.1016/j.bjm.2018.03.002, 2018.

Heino, J., Alahuhta, J., Bini, L. M., Cai, Y., Heiskanen, A.-S., Hellsten, S., Kortelainen, P., Kotamäki, N., Tolonen, K. T., Vihervaara, P., Vilmi, A., and Angeler, D. G.: Lakes in the era of global change: moving beyond single-lake thinking in maintaining biodiversity and ecosystem services, Biol. Rev., 96, 89–106, https://doi.org/10.1111/brv.12647, 2021.

Hoang, D. Q., Wilson, L. R., Scheftgen, A. J., Suen, G., and Currie, C. R.: Disturbance-Diversity Relationships of Microbial Communities Change Based on Growth Substrate, bioRxiv, 2023.08.25.554838, https://doi.org/10.1101/2023.08.25.554838, 2023.

Hoetzinger, M., Schmidt, J., Pitt, A., Koll, U., Lang, E., and Hahn, M. W.: Polynucleobacter paneuropaeus sp. nov., characterized by six strains isolated from freshwater lakes located along a 3000 km North-South gradient across Europe, Int. J. Syst. Evol. Microbiol., 69, 203–213, https://doi.org/10.1099/ijsem.0.003130, 2019.

Hsieh, T. C., Ma, K. H., and Chao, A.: iNEXT: Interpolation and Extrapolation for Species Diversity, 2024.

Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K. M., Andersen, H. E., Lauridsen, T. L., Liboriussen, L., Beklioglu, M., Ozen, A., and Olesen, J. E.: Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations., J. Environ. Qual., 38, 1930–1941, https://doi.org/10.2134/jeq2008.0113, 2009.

Jones, S. E. and Lennon, J. T.: Dormancy contributes to the maintenance of microbial diversity, Proc. Natl. Acad. Sci., 107, 5881–5886, https://doi.org/10.1073/pnas.0912765107, 2010.

Katoh, K., Asimenos, G., and Toh, H.: Multiple alignment of DNA sequences with MAFFT, , Bioinforma. DNA Seq. Anal.39–64, 2009.

Kersti Kangro, Kalle Olli, Timo Tamminen, and Risto Lignell: Species-specific responses of a cyanobacteria-dominated phytoplankton community to artificial nutrient limitation in the Baltic Sea, Mar. Ecol. Prog. Ser., 336, 15–27, 2007.

Maasri, A., Jähnig, S. C., Adamescu, M. C., Adrian, R., Baigun, C., Baird, D. J., Batista-Morales, A., Bonada, N., Brown, L. E., Cai, Q., Campos-Silva, J. V., Clausnitzer, V., Contreras-MacBeath, T., Cooke, S. J., Datry, T., Delacámara, G., De Meester, L., Dijkstra, K.-D. B., Do, V. T., Domisch, S., Dudgeon, D., Erös, T., Freitag, H., Freyhof, J., Friedrich, J., Friedrichs-Manthey, M., Geist, J., Gessner, M. O., Goethals, P., Gollock, M., Gordon, C., Grossart, H.-P., Gulemvuga, G., Gutiérrez-Fonseca, P. E., Haase, P., Hering, D., Hahn, H. J., Hawkins, C. P., He, F., Heino, J., Hermoso, V., Hogan, Z., Hölker, F., Jeschke, J. M., Jiang, M., Johnson, R. K., Kalinkat, G., Karimov, B. K., Kasangaki, A., Kimirei, I. A., Kohlmann, B., Kuemmerlen, M., Kuiper, J. J., Kupilas, B., Langhans, S. D., Lansdown, R., Leese, F., Magbanua, F. S., Matsuzaki, S.-I. S., Monaghan, M. T., Mumladze, L., Muzon, J., Mvogo Ndongo, P. A., Nejstgaard, J. C., Nikitina, O., Ochs, C., Odume, O. N., Opperman, J. J., Patricio, H., Pauls, S. U., Raghavan, R., Ramírez, A., Rashni, B., Ross-Gillespie, V., Samways, M. J., Schäfer, R. B., Schmidt-Kloiber, A., Seehausen, O., Shah, D. N., Sharma, S., Soininen, J., Sommerwerk, N., Stockwell, J. D., Suhling, F., Tachamo Shah, R. D., Tharme, R. E., Thorp, J. H., Tickner, D., Tockner, K., Tonkin, J. D., Valle, M., Vitule, J., Volk, M., Wang, D., Wolter, C., and Worischka, S.: A global agenda for advancing freshwater biodiversity research., Ecol. Lett., 25, 255–263, https://doi.org/10.1111/ele.13931, 2022.

Másmela-Mendoza, J. E., Lizarazo, L. M., and Riaño, N. J. A.: Bacterias nitrificantes cultivables de la zona limnética del lago de Tota, Boyacá, Rev. UDCA Actual. Divulg. Científica, 22, 2019.

McMurdie, P. J. and Holmes, S.: phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data, PLOS ONE, 8, e61217, https://doi.org/10.1371/journal.pone.0061217, 2013.

Millennium Ecosystem Assessment (Ed.): Ecosystems and human well-being: synthesis, Island Press, Washington, DC, 1 p., 2005.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R.: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era, Mol. Biol. Evol., 37, 1530–1534, https://doi.org/10.1093/molbev/msaa015, 2020.

Mohapatra, B. and Phale, P. S.: Microbial Degradation of Naphthalene and Substituted Naphthalenes: Metabolic Diversity and Genomic Insight for Bioremediation, Front. Bioeng. Biotechnol., 9, https://doi.org/10.3389/fbioe.2021.602445, 2021.

Nuy, J. K., Hoetzinger, M., Hahn, M. W., Beisser, D., and Boenigk, J.: Ecological Differentiation in Two Major Freshwater Bacterial Taxa Along Environmental Gradients, Front. Microbiol., 11, https://doi.org/10.3389/fmicb.2020.00154, 2020.

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., Evangelista, H. B. A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M. O., Lahti, L., McGlinn, D., Ouellette, M.-H., Cunha, E. R., Smith, T., Stier, A., Braak, C. J. F. T., and Weedon, J.: vegan: Community Ecology Package (2.6-6.1), https://cran.r-project.org/web/packages/vegan/index.html (last access: 5 August 2024), 2024.

Paerl, H. W. and Huisman, J.: Climate change: a catalyst for global expansion of harmful cyanobacterial blooms., Environ. Microbiol. Rep., 1, 27–37, https://doi.org/10.1111/j.1758-2229.2008.00004.x, 2009.

Paver, S. F., Newton, R. J., and Coleman, M. L.: Microbial communities of the Laurentian Great Lakes reflect connectivity and local biogeochemistry, Environ. Microbiol., 22, 433–446, https://doi.org/10.1111/1462-2920.14862, 2020.

Prado, P. E., Modenutti, B., Aranguren-Riaño, N., Balseiro, E., Samanez, I., Campero, M., Fernández, C. E., Rivera-Rondon, C. A., Carvajal-Vallejos, F. M., López-París, D., Herrera-Martínez, Y., Ríos-Escalante, P. de los, and Steinitz-Kannan, M.: Andean Lakes, a proposal of lake districts, Inland Waters, 2024.

Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H., Meybeck, M., Ciais, P., and Guth, P.: Global carbon dioxide emissions from inland waters, Nature, 503, 355–359, https://doi.org/10.1038/nature12760, 2013.

Sarmiento, G., Gaviria, S., Hooghiemstra, H., Berrio, J. C., and Van der Hammen, T.: Landscape evolution and origin of Lake Fúquene (Colombia): Tectonics, erosion and sedimentation processes during the Pleistocene, Geomorphology, 100, 563–575, https://doi.org/10.1016/j.geomorph.2008.02.006, 2008.

Sessitsch, A., Wakelin, S., Schloter, M., Maguin, E., Cernava, T., Champomier-Verges, M.-C., Charles, T. C., Cotter, P. D., Ferrocino, I., Kriaa, A., Lebre, P., Cowan, D., Lange, L., Kiran, S., Markiewicz, L., Meisner, A., Olivares, M., Sarand, I., Schelkle, B., Selvin, J., Smidt, H., Van Overbeek, L., Berg, G., Cocolin, L., Sanz, Y., Fernandes, W. L., Liu, S. J., Ryan, M., Singh, B., and Kostic, T.: Microbiome Interconnectedness throughout Environments with Major Consequences for Healthy People and a Healthy Planet, Microbiol. Mol. Biol. Rev., 87, e00212-22, https://doi.org/10.1128/mmbr.00212-22, 2023.

Sessitsch, A., Wakelin, S., Schloter, M., Maguin, E., Cernava, T., Champomier-Verges, M.-C., Charles, T. C., Cotter, P. D., Ferrocino, I., Kriaa, A., Lebre, P., Cowan, D., Lange, L., Kiran, S., Markiewicz, L., Meisner, A., Olivares, M., Sarand, I., Schelkle, B., Selvin, J., Smidt, H., van Overbeek, L., Berg, G., Cocolin, L., Sanz, Y., Fernandes, W. L., Liu, S. J., Ryan, M., Singh, B., and Kostic, T.: Microbiome Interconnectedness throughout Environments with Major Consequences for Healthy People and a Healthy Planet, Microbiol. Mol. Biol. Rev. MMBR, 87, e00212-22, https://doi.org/10.1128/mmbr.00212-22, n.d.

Shade, A., Peter, H., Allison, S. D., Baho, D., Berga, M., Buergmann, H., Huber, D. H., Langenheder, S., Lennon, J. T., Martiny, J. B., Matulich, K. L., Schmidt, T. M., and Handelsman, J.: Fundamentals of Microbial Community Resistance and Resilience, Front. Microbiol., 3, https://doi.org/10.3389/fmicb.2012.00417, 2012.

Smith, V. H. and Schindler, D. W.: Eutrophication science: where do we go from here?, Trends Ecol. Evol., 24, 201–207, https://doi.org/10.1016/j.tree.2008.11.009, 2009.

Søndergaard, M., Jensen, J. P., and Jeppesen, E.: Role of sediment and internal loading of phosphorus in shallow lakes, Hydrobiologia, 506, 135–145, https://doi.org/10.1023/B:HYDR.0000008611.12704.dd, 2003.

Torres-Barrera, N. H., Grandas-Rincón, I. A., Torres-Barrera, N. H., and Grandas-Rincón, I. A.: Estimación de los desperdicios generados por la producción de trucha arcoíris en el lago de Tota, Colombia, Cienc. Tecnol. Agropecu., 18, 247–255, https://doi.org/10.21930/rcta.vol18_num2_art:631, 2017.

Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., Dillon, P., Finlay, K., Fortino, K., Knoll, L. B., Kortelainen, P. L., Kutser, T., Larsen, Soren., Laurion, I., Leech, D. M., McCallister, S. L., McKnight, D. M., Melack, J. M., Overholt, E., Porter, J. A., Prairie, Y., Renwick, W. H., Roland, F., Sherman, B. S., Schindler, D. W., Sobek, S., Tremblay, A., Vanni, M. J., Verschoor, A. M., von Wachenfeldt, E., and Weyhenmeyer, G. A.: Lakes and reservoirs as regulators of carbon cycling and climate, Limnol. Oceanogr., 54, 2298–2314, https://doi.org/10.4319/lo.2009.54.6_part_2.2298, 2009.

Ward, T. E. and Frea, J. I.: Sediment Distribution of Methanogenic Bacteria in Lake Erie and Cleveland Harbor, Appl. Environ. Microbiol., 39, 597–603, 1980.

Yang, J., Yao, B., Cai, M., Han, M., Wu, Z., Zhang, P., Xiao, H., Han, J., Zhang, X., and Jiang, H.: Salinity Change Overrides Nitrogen Increase in Affecting Microbial Abundance, Diversity, Community Composition and Organic Carbon Mineralization in Saline Lakes, J. Earth Sci., 37, 1709–1721, https://doi.org/10.1007/s12583-024-0139-4, 2026.

Yu, G.: Using ggtree to Visualize Data on Tree-Like Structures, Curr. Protoc. Bioinforma., 69, e96, https://doi.org/10.1002/cpbi.96, 2020.