https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026

Bacterial Community Structure and Predicted Functional Potential Across Andean Lakes with Contrasting Preservation Status Based on 16S rRNA Gene Profiles

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 play a key role in supporting agriculture and local livelihoods, but they are increasingly affected by anthropogenic pressures such as agricultural expansion and urbanization. These pressures can alter water quality and ecosystem functioning, yet their effects on microbial communities remain poorly understood. Here, we analyze bacterial community structure and predicted functional potential across four freshwater lakes in the Eastern Cordillera of Colombia, representing contrasting levels of environmental preservation. Using 16S rRNA gene sequencing, we characterized microbial diversity and community composition in two lakes subject to higher anthropogenic influence (Fúquene and Tota) and two relatively well-preserved systems (Calderona and Colorado). Lakes with lower preservation status exhibited higher microbial richness and diversity, likely associated with increased nutrient inputs and environmental heterogeneity. In contrast, more preserved lakes showed lower diversity but more stable community composition. Spatial heterogeneity also played a key role, with larger systems displaying greater horizontal variability linked to tributary inflows and extensive littoral zones, as well as vertical stratification driven by gradients in light, temperature, oxygen, and nutrients. Co-occurrence patterns among dominant genera suggest environmentally structured microbial associations across lakes. Inferred functional profiles indicate differences in predicted metabolic potential among preservation statuses, particularly in pathways related to carbohydrate metabolism and xenobiotic degradation. However, these results represent inferred functional capacity based on taxonomic composition rather than direct measurements of microbial activity. Overall, our findings indicate that environmental preservation status and spatial heterogeneity shape both the composition and predicted metabolic potential of bacterial communities in Andean freshwater ecosystems. These results provide a baseline for understanding microbial responses to environmental change and highlight the need for future studies integrating metagenomic and experimental approaches to validate ecosystem processes.

1. Introduction

Aquatic ecosystems, particularly highland lakes and lagoons, are of immense ecological and societal significance (Heino et al., 2021; Millennium Ecosystem Assessment, 2005). These water bodies provide essential resources for agricultural irrigation, sustain 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 shape the landscape, supporting local livelihoods and fostering biodiversity (Andrade-Sossa et al., 2023; Maasri et al., 2022; Prado et al., 2024). However, the sustainability of these ecosystems is increasingly threatened by rising anthropogenic pressures within their watersheds (Andrade-Sossa et al., 2023; Aranguren-Riaño et al., 2018; Prado et al., 2024; Torres-Barrera et al., 2017).

Human activities, including intensive agriculture and urbanization, have introduced a range of pollutants and disturbances into these aquatic environments (Aranguren-Riaño et al., 2018; Carpenter et al., 1998; Másmela-Mendoza et al., 2019; Torres-Barrera et al., 2017).

These disturbances disrupt ecosystem equilibrium and pose significant risks to water quality, aquatic life, and the integrity of these natural reservoirs (Barletta et al., 2010; Millennium Ecosystem Assessment, 2005; Prado et al., 2024). Such impacts align with well-established ecological principles of lake dynamics, where external pressures—such as nutrient enrichment (Carpenter et al., 1998), sedimentation (Smith and Schindler, 2009), and hydrological alterations (Jeppesen et al., 2009)—drive shifts in microbial communities, primary production, and overall ecosystem functioning (Andrade-Sossa et al., 2023; Forero-Pineda et al., 2021; Paerl and Huisman, 2009; Paver et al., 2020; Sessitsch et al., n.d.; Smith and Schindler, 2009). These processes, widely documented in Colombian lakes, illustrate how anthropogenic pressures exacerbate ecological imbalances and influence biodiversity and biogeochemical cycles (Aranguren-Riaño et al., 2018; Forero-Pineda et al., 2021; Másmela-Mendoza et al., 2019; Prado et al., 2024). Moreover, lakes play a significant role in global biogeochemical cycles and energy fluxes (Downing et al., 2006; Raymond et al., 2013; Tranvik et al., 2009).

Microbial communities are central to ecosystem resilience and functionality, as they mediate key biogeochemical processes and contribute to ecological stability (Battin et al., 2016; Sessitsch et al., n.d.; Shade et al., 2012). Bacteria and archaea interact with their environment in complex ways, influencing nutrient cycling, water quality, and ecosystem health (Battin et al., 2016; Sessitsch et al., n.d.; Shade et al., 2012). Therefore, characterizing microbial diversity and its potential functional attributes in lentic ecosystems is essential for understanding the processes underlying ecosystem functioning, stability, and resilience (Battin et al., 2016; Paver et al., 2020; Sessitsch et al., n.d.; Shade et al., 2012).

The main objective of this study is to determine how environmental preservation status influences the taxonomic structure and predicted functional potential of bacterial communities in Andean freshwater ecosystems. Specifically, we aim to (i) compare microbial diversity and community composition across lakes with contrasting levels of anthropogenic disturbance, and (ii) evaluate how these differences are reflected in inferred metabolic pathways relevant to ecosystem processes.

To achieve this, we applied a metataxonomic framework based on 16S rRNA gene sequencing to characterize bacterial diversity and community structure, combined with PICRUSt2 to infer potential functional profiles from taxonomic data. This approach leverages reference genome databases to predict gene-family abundances and metabolic pathways associated with detected taxa. However, these predictions rely on the assumption that closely related taxa share similar genomic content, which may not fully capture strain-level variation or functional diversity within microbial lineages. Therefore, the inferred functional profiles presented here should be interpreted as proxies of potential metabolic capacity rather than direct evidence of gene presence or activity, and are intended as a hypothesis-generating framework for future studies incorporating metagenomic and experimental validation.

2. Materials and Methods

2.1. Study Area and Sampling

On one hand, Fúquene Lagoon is located at 2568 meters above the sea level (m.a.s.l.) largest remaining freshwater system in the Northern Andean highland mountain range of Colombia, covering about 198 square kilometers of water surface. Fúquene is situated less than 100 kilometers away from Bogotá city and is surrounded by a diverse landscape, featuring extensive dairy farms and a variety of agricultural crops such as cereals, potatoes, and onions. Tota lake is located at 3015m in Boyacá, holds the title of Colombia's largest high mountain lake, covering around 60 square kilometers. Despite its popularity as a natural tourist destination, recent studies have highlighted an ongoing moderate eutrophication issue. This is primarily due to the unregulated use of fertilizers in green onion farming, various aquaculture practices, and the discharge of domestic wastewater. Unfortunately for both, these activities release various pollutants into the system (Aranguren-Riaño et al., 2018; Forero-Pineda et al., 2021; Másmela-Mendoza et al., 2019; Torres-Barrera et al., 2017).

On the other hand, Calderona is located at 3143 m.a.s.l., spanning approximately 0.4 square kilometers, is a glacial lagoon situated within the protected El Cañal Municipal Natural Park in Boyacá's Ciénega municipality. Calderona is surrounded by moor vegetation from the Páramo el Bijagual. Lastly, the Colorado lagoon, situated at 2430m, in the Boyacá's Gachantivá municipality, is a natural aquatic system pond surrounded by Andean forest, and nestled in the foothills of the Páramo de Iguaque. Only short-term tourism activities are permitted in this area. Further details about the origins and preservation statuses of these sites are shown in (Table 1).

Sampling was exclusively conducted in the limnetic zone of the aquatic systems, excluding the littoral zone, to focus specifically on planktonic communities. The limnetic zone allowing a better examination of planktonic composition, like bacterioplankton and their functional roles within the ecosystems. For Fúquene Lagoon we collected two samples in the first and in the second semester of 2019, whereas for Tota Lake, and Calderona and Colorado, only one sample was collected in October 2019 (Table 1). Additional raw dataset from the previous study of bacteriome structure of Tota Lake (BioProject: PRJNA720890), sampled in December 2018 was included as internal control (Forero-Pineda et al., 2021). The water collection procedure was consistent across all samples. We obtained a water sample, approximately 1.7 liters, from the middle part of the photic zone with a Schindler-Patalas

Bottle 1 sequi disk, so the sampling depth was closely related to the transparency of each system. The collected water underwent in situ prefiltration through 1.5 µm pore-size filter. Microorganisms were then concentrated onto two replicates of 0.22 µm Millipore Sterivex filters, following the recommendations of Paver et al. in 2020 (Paver et al., 2020). Subsequently, these filters were transferred into sterile plastic tubes containing 1.5 ml of Monarch® DNA/RNA protection reagent. They were transported to the laboratory within 2 to 4 hours, maintaining a cold chain, and stored at -20°C until DNA extraction could be performed, as recommended by the manufacturer (New England Biolabs).

Table 1. Location and origin of lentic systems analyzed and raw data used in the core-metataxonomic description based on amplificon variant sequences (ASVs). *Indicates that the included data was retrieved from Forero-Pineda et al., (2021).

Locality	Sample	Coordinates		Altitude (m.a.s.l.)	Geological origin	Preserved status	Sampling (year.semester)	Number. of clean reads	Number of ASVs
		North	West						
Ciénega	Calderona	5° 23' 23.16"	73° 14' 40.99"	3,130	Glacial	High	2019.2	74690	514
Gachantivá	Colorado	5° 46' 29.43"	73° 33' 5.81"	2,450	Pluvial	High	2019.2	68737	761
Fúquene	Fúquene1	5° 27' 4.53"	73° 44' 47.82"	2,750	Tectonic	Low	2019.1	73484	756
	Fúquene2						2019.2	86752	848
Tota	Tota1	5° 32' 29.28"	72° 55' 27.75"	3,015	Tectonic-Glacial	Middle	2018.2*	171326	690
	Tota2						2019.2	81926	506

2.2. DNA Extraction, Sequencing, and Assembly

Genomic DNA was extracted from environmental filters using a modified cetyltrimethylammonium bromide (CTAB)-based chemical protocol, rather than a commercial extraction kit. The extraction buffer consisted of 1% (w/v) CTAB, 3% (w/v) sodium dodecyl sulfate (SDS), 100 mM Tris–HCl, 100 mM EDTA, and 1.5 M NaCl (pH 8.0), following previously described procedures (Forero-Pineda et al., 2021; Paver et al., 2020). Subsequently, we conducted amplicon sequencing of the 16S rRNA gene on the Miseq platform. We employed universal primers 341F (5′-CCT ACG GNGGC WGC AG-3′) and 805R (5′-GAC TAC HVGGT ATC TAA TCC -3′) to target the hypervariable V3 and V4 regions of the 16S rRNA gene, following established methods (Klindworth et al., 2013; Ren et al., 2017). The primers used in this study target conserved regions of the 16S rRNA gene present in both Bacteria and Archaea.

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 the DADA2 v1.20.0 package (Callahan et al., 2016). Forward reads were truncated to 200 bp and reverse reads to 150 bp (*truncLen* = c(200,150)), with the first 19 nt of forward and 20 nt of reverse reads removed (*trimLeft* = c(19,20)). Reads containing ambiguous bases were discarded (*maxN* = 0), and maximum expected errors were limited to two per read (*maxEE* = c(2,2)). Bases with quality scores below Q2 were trimmed (*truncQ* = 2), and phiX contaminants were removed (*rm.phix* = TRUE). After denoising, paired-end reads were merged, and chimeric sequences were removed using the “consensus” method implemented in *removeBimeraDenovo*. To reduce potential sequencing artefacts and low-frequency noise, we discarded amplicon sequence variants (ASVs) with fewer than 100 total reads across all samples. This threshold was selected to minimize spurious variants while retaining dominant and consistently detected taxa.

2.3.2. Taxonomic assignment and diversity analyses

Taxonomic classification was performed using the RDP training set (rdp_train_set_18) with the *assignTaxonomy* and *addSpecies* functions. The resulting ASV table was converted to relative abundances for diversity and ordination analyses using the function *transform*_sample_counts in the phyloseq v1.36.0 package (McMurdie and Holmes, 2013). Rarefaction curves were generated using the function *rarecurve* (step = 10) in vegan v2.6.4 (Oksanen et al., 2024). Diversity was further quantified using Hill numbers ($q = 0$, 1, and 2), calculated with the iNEXT package (Chao et al., 2014; Hsieh et al., 2024), corresponding to species richness, Shannon diversity, and Simpson diversity, respectively.

2.3.3. Co-occurrence network analysis

To investigate potential microbial associations, ASVs were aggregated to the genus level using tax*_glom* (phyloseq). Only genera present in at least 50 % of samples and with mean relative abundance greater than 0.1 % were retained. Pairwise Spearman correlations (ρ) were calculated based on genus-level relative abundances. Statistical significance was assessed by generating null distributions through 1 000 random permutations of sample labels, recalculating ρ for each iteration. Raw p-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) method. Associations were considered significant for $|\rho| > 0.6$ and FDR-corrected $p < 0.01$.

In parallel, probabilistic co-occurrence analysis (Veech, 2013) was conducted using the *site_spp* null model implemented in cooccur v1.3. Positive and negative associations were defined based on significant deviations from expected co-occurrence frequencies ($\alpha = 0.05$).

Network visualization was performed using visNetwork v2.1.2. These co-occurrence patterns represent statistical associations and should not be interpreted as direct ecological interactions.

2.3.4. Phylogenetic analysis of *Mycobacterium*

For non-tuberculous *Mycobacterium* (NTM) identification, 25 sequences corresponding to ASV29 were extracted from the dataset (Katoh et al., 2009) including 96 available sequences drawn from RefSeq database (Table S2). Maximum-likelihood (ML) phylogenetic tree was conducted in IQ-Tree 2 (Minh et al., 2020) with 1000 Ultrafast Bootstrap replicates under the GTR+F substitution model. Topology visualization and edition was performed in ggtree v.3.0.4 (Yu, 2020) package in R.

2.3.5. Functional prediction and differential analysis

Functional profiles were inferred from 16S rRNA gene data using the PICRUSt2 package (Douglas et al., 2020), which predicts gene content based on reference genomes. Using amplicon sequence variants (ASVs) as input, we inferred Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologs (KOs) and their associated metabolic pathway abundances through the complete PICRUSt2 pipeline. These predictions rely on phylogenetic inference and assume that closely related taxa share similar genomic content. However, this assumption may not fully account for strain-level variability and functional heterogeneity within microbial lineages. Consequently, inferred functional profiles should be interpreted as estimates of potential metabolic capacity rather than direct measurements of gene presence or activity.

A per-ASV nearest sequenced taxon index (NSTI) cutoff of < 2 was applied following PICRUSt2 guidelines. ASVs approaching this threshold were retained but interpreted with caution due to reduced prediction accuracy.

To evaluate differences in predicted functional pathways among lakes, we conducted differential abundance analyses using ALDEx2 v1.28.1 (Paulson et al., 2013), implemented through the ggpicrust2 package (Yang et al., 2023). Count data were transformed using a centered log-ratio (CLR) transformation after generating 128 Monte Carlo Dirichlet instances per feature to account for compositional variability.

Statistical significance was assessed using both Welch's t-test and the Wilcoxon rank-sum test, as implemented in the *aldex.ttest* function. Resulting p-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) correction. KEGG pathways were considered significant at FDR < 0.01, while MetaCyc pathways were evaluated at FDR < 0.05. These thresholds were selected to account for differences in pathway resolution and database structure. Effect sizes were calculated using *aldex.effect* but were not subjected to an additional cutoff.

This analytical framework provides a robust approach for identifying compositional differences in inferred functional profiles while controlling for false discovery, although results should be interpreted as hypothesis-generating rather than confirmatory of microbial activity.

3. Results

3.1. Description of microbial communities present in aquatic systems

After pre-processing, consisting of high-quality filtering, a total of 556,915 reads was obtained from the study population, with an average of 92,819 reads per sample, by sequencing the 16S rRNA gene fragment (Table 1). The number of assigned reads to ASVs varied from 42,171 in Colorado to 86,554 in Tota lake 1 sample (Fig. 1a). After excluding those ASVs harboring < 100 sequences, and those with no available taxa name, a total of

https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026

3,153 assigned to taxa ASVs were retained for metataxonomic description in the 6 samples analyzed. The number of ASVs identified in samples varied from 514 in Colorado to 802 (± 65) in the two sampling of Fúquene lake (Table 1), and significant differences ($p < 0.001$) in ASV abundance was calculated between all samples, except for Calderona and Tota2, Colorado and Fúquene, Colorado and Tota1, and Fúquene1 and Tota1 (Fig. 1b). Furthremore, significative differences ($p < 0.001$) were also observed for the temporal replicates conducted in Tota and Fúquene (Fig. 1b). Similar number of observed and rarified ASVs, and asymptotic rarefaction curves showed in all samples were indicating that further sequencing effort would not greatly affect the results (Fig. S1).

A total of 34 bacterial phyla were identified across all samples, along with members of the archaeal phylum Euryarchaeota, consistent with amplification of both bacterial and archaeal 16S rRNA genes (Fig. 1c–d; Table S1). The most abundant bacterial phyla included Actinobacteria (34.5%), Proteobacteria (29.7%), Bacteroidetes (12.7%), Cyanobacteria (8.5%), and Verrucomicrobia (7.9%), whereas approximately 80% of detected phyla exhibited low relative abundances (< 1%), including Euryarchaeota (Fig. 1c; Table S1).

Across preservation statuses, Proteobacteria and Actinobacteria were consistently dominant in all aquatic systems, whereas Bacteroidetes, Verrucomicrobia, and Cyanobacteria showed variable abundances among lakes (Fig. 1d). Additionally, although Planctomycetes were detected in all systems, they reached relatively high abundance in the Colorado lake (11.3%) (Fig. 1d).

Across all samples, a total of 366 genera belonging to 179 families were identified following taxonomic assignment (Table S1; Fig. 1e–f). Low-abundance (< 1%) archaeal genera, including *Methanobacterium*, *Methanothrix*, and *Methanocella*, were detected in the

Colorado lagoon, whereas *Methanosarcina* was identified in the second sampling of Fúquene lagoon (Table S1). The most abundant genera included *Polynucleobacter* (Proteobacteria; 29.5%) and *Flavobacterium* (Bacteroidetes; 20.3%), both dominant in Calderona, followed by *Phragmitibacter* (Verrucomicrobia; 24.2%) in Colorado (Fig. 1f).

https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026

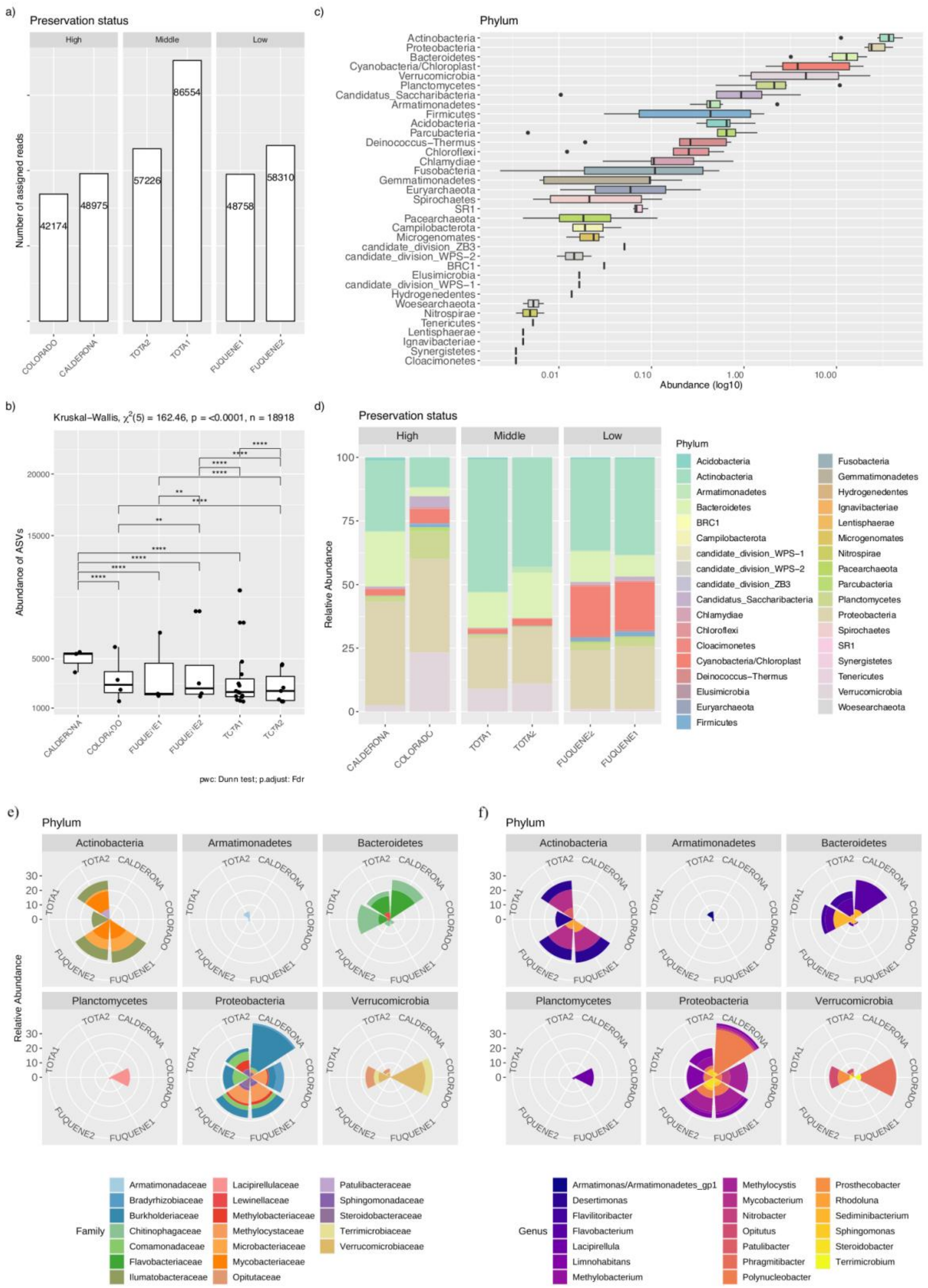


Figure 1. Sequencing output, diversity patterns, and taxonomic composition across aquatic systems with contrasting preservation status. (a) Number of reads assigned to amplicon sequence variants (ASVs) per sample. (b) Differences in ASV abundance among samples (Kruskal–Wallis test with Dunn's post hoc comparisons; FDR-adjusted p-values shown). (c) Relative abundance of bacterial phyla across all samples. (d) Phylum-level community composition by preservation status. Relative abundance of dominant taxa at the (e) family and (f) genus levels.

3.2. Microbiome diversity estimation in aquatic systems of Northern Andean Region of Colombia

Microbial diversity was assessed using Hill numbers, where q represents the order of diversity. Specifically, q0 corresponds to species richness (all taxa weighted equally), q1 corresponds to the exponential of Shannon diversity (sensitive to abundant taxa), and q2 corresponds to the inverse Simpson index (emphasizing dominant taxa). For q0, Colorado exhibited the highest species richness (128 taxa), followed by Fúquene 1, Fúquene 2, and Tota 2. However, diversity decreased substantially at q1 and q2 in Colorado, indicating that a relatively small number of taxa accounted for a large proportion of the total abundance. In contrast, Calderona, despite having lower richness (q0 = 73), showed a more gradual decline from q0 to q1 and q2, suggesting a more even distribution of taxa. Across systems, Tota exhibited noticeable variation in diversity patterns between sampling points, reflecting temporal or spatial heterogeneity in community composition (Fig. 2a).

3.3. Co-occurrence network analysis

The co-occurrence patterns of bacterial communities across aquatic systems were assessed using the most abundant genera. The resulting network was statistically significant ($p < 0.01$) and comprised 20 nodes and 226 edges, representing taxa from six major phyla: Proteobacteria, Actinobacteria, Bacteroidetes, Verrucomicrobia, Planctomycetes, and Armatimonadetes (Fig. 2b). Several genera from different phyla, including *Nitrobacter*, *Methylobacterium*, *Opitutus*, *Flavilitoribacter*, and *Sphingomonas*, exhibited strong positive correlations across samples. In contrast, *Mycobacterium* showed a more restricted pattern, co-occurring primarily with *Polynucleobacter* (Fig. 2b). These associations represent statistical co-occurrence patterns and should not be interpreted as direct ecological interactions.

3.4. Temporal bacteriome composition in Tota and Fúquene aquatic systems

Bacterial community composition at the genus level (relative abundance > 0.1%) was broadly similar between the two sampling events in both Tota and Fúquene (Fig. 2c-d). In Tota, relatively consistent abundances were observed for *Desertimonas*, *Flavobacterium*, *Limnohabitans*, and *Sediminibacterium* across both sampling events. In contrast, noticeable variation was detected for *Mycobacterium* and *Patulibacter* (Fig. 2c). In Fúquene, the dominant genera showed relatively stable abundances between sampling events, with *Mycobacterium*, *Desertimonas*, *Methylocystis*, and *Rhodoluna* consistently prevalent (Fig. 2d). These patterns suggest temporal stability in dominant taxa, with localized variability in specific genera.

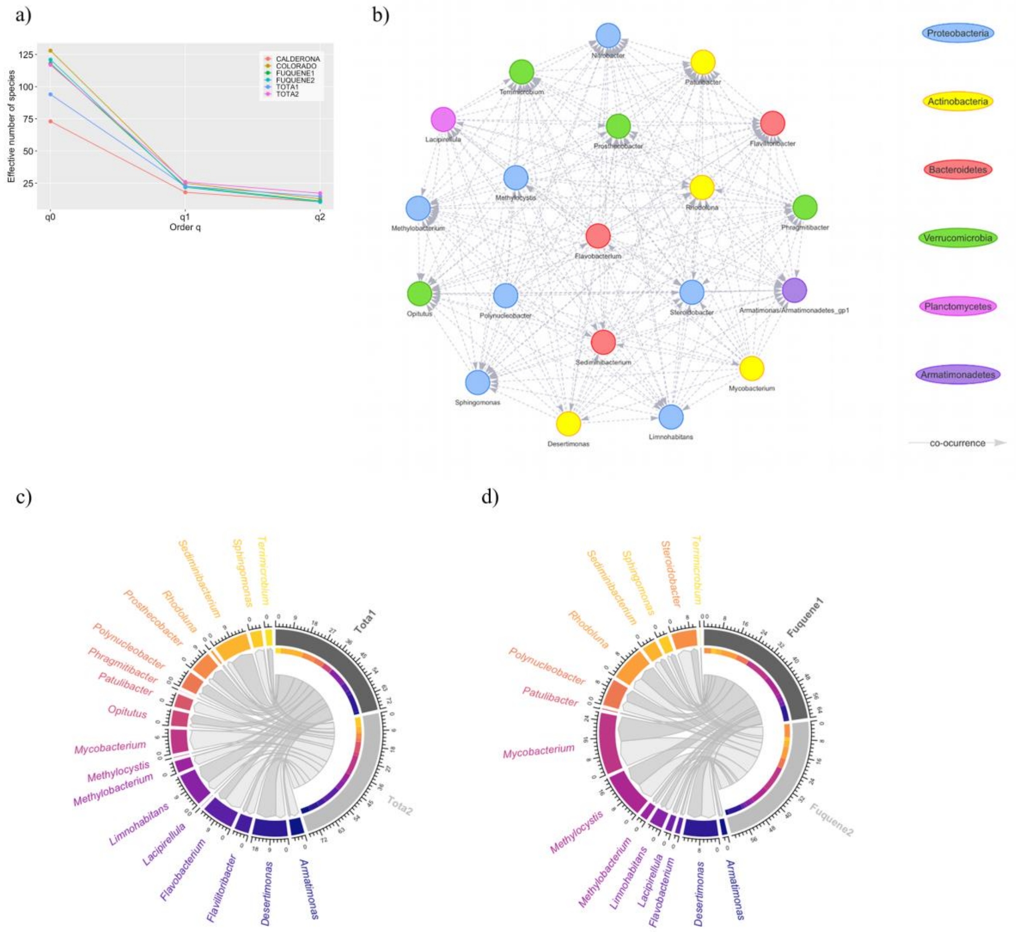


Figure 2. Diversity patterns, co-occurrence structure, and temporal variation of bacterial communities across aquatic systems. (a) Hill numbers (effective number of taxa) across sampling localities, where q represents the order of diversity (q0 = richness, q1 = Shannon diversity, q2 = Simpson diversity). (b) Co-occurrence network of the most abundant bacterial genera; nodes represent genera, edges indicate significant correlations, and node colors correspond to phylum-level classification (see inset). (c–d) Relative abundance of dominant bacterial genera (> 0.1%) across two sampling events in (c) Lake Tota and (d) Lake Fúquene.

3.5. Phylogenetic 16s rRNA-based identification of *Mycobacterium* sp. present in Tota lake and Fúquene lagoon

Phylogenetic analysis of 16S rRNA gene sequences revealed that, among the 25 sequences assigned to *Mycobacterium* sp. (ASV29), 10 clustered with reference sequences of known species with strong support (bootstrap > 0.9), including *M. celatum*, *M. noviomagense, M. szulgai, M. pseudokansasii*, and *M. cookii* (variant 2). These sequences formed well-supported clades with their respective reference taxa. The remaining sequences grouped into less-resolved clades, showing associations with multiple species, including *M. attenuatum* and *M. innocens, M. arosiense, M. haemophilum* and *M. nebraskense* (variant 2), as well as *M. cookii* and *M. pseudoshottsii* (Fig. 3). However, due to the limited resolution of the V3–V4 region of the 16S rRNA gene, species-level assignments should be considered provisional.

Sequences affiliated with other genera of potential relevance, such as *Leptospira* and *Legionella*, were also detected. These results indicate the presence of phylogenetically diverse bacterial lineages in the aquatic systems of Tota and Fúquene, including taxa commonly reported in environmental and water-associated habitats.

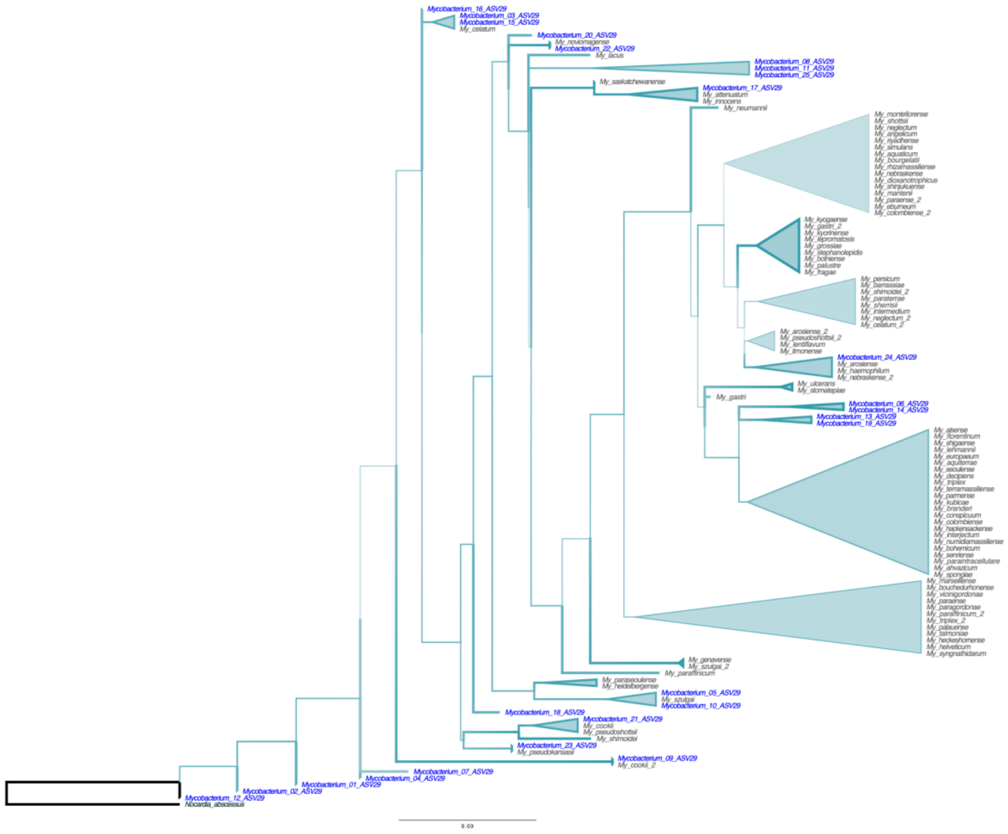


Figure 3. Maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences of non-tuberculous *Mycobacterium* (NTMs) detected in Tota and Fúquene aquatic systems. Sequences corresponding to ASV29 identified in this study are highlighted in blue. Branch thickness indicates bootstrap support, and clades with bootstrap values > 0.9 are collapsed.

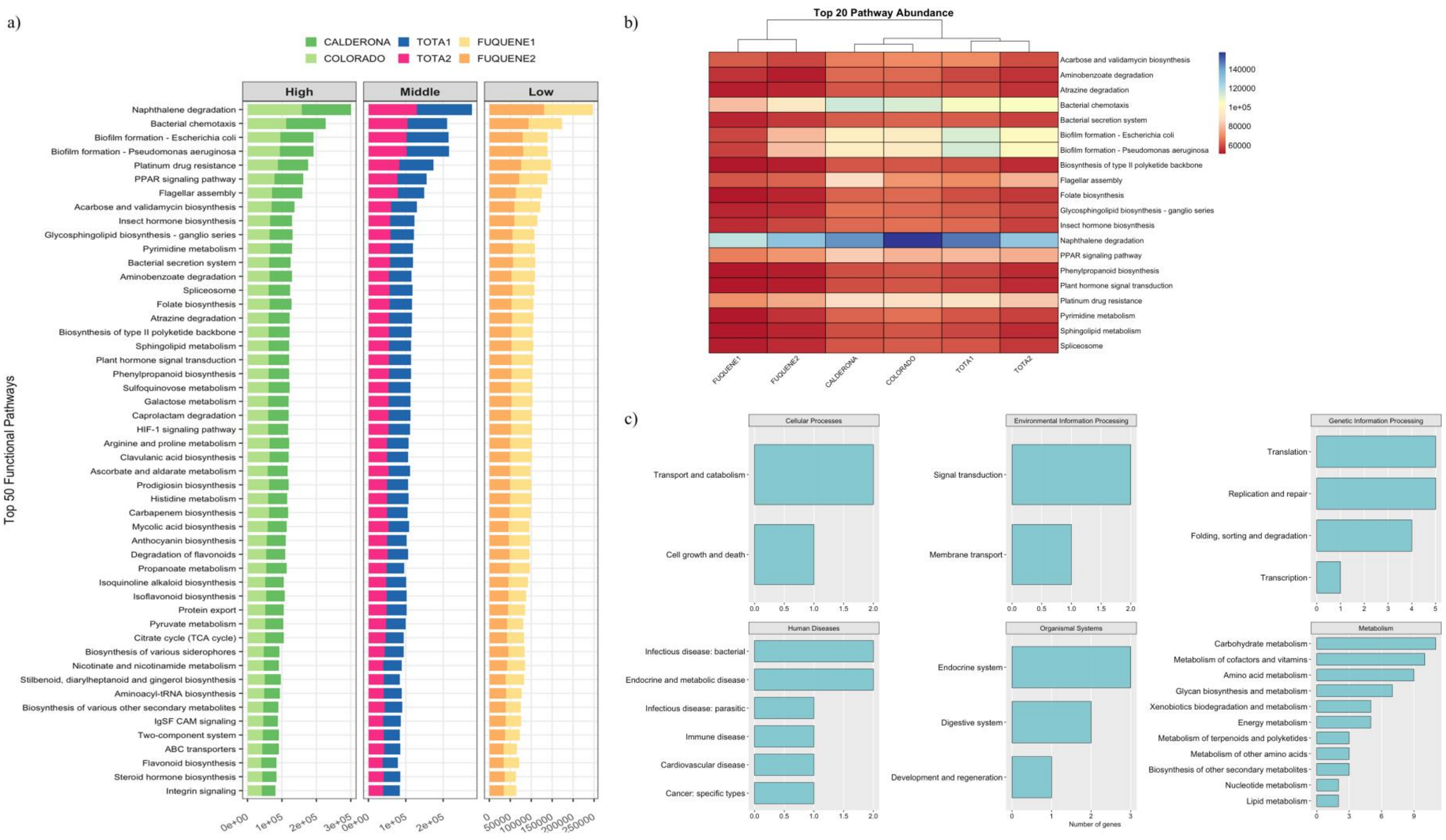
3.6. Functional prediction of bacterial communities presents in aquatic systems of Northern Andean Region of Colombia

The predicted functional potential of bacterial communities was inferred using PICRUSt2 across all samples. The Nearest Sequenced Taxon Index (NSTI), which quantifies the phylogenetic distance between observed taxa and reference genomes used for functional prediction, ranged from 0.159 to 0.181 (mean = 0.171). These relatively low values indicate that most ASVs are closely related to sequenced reference organisms, suggesting moderate to high confidence in the inferred functional profiles. A per-ASV NSTI cutoff of < 2 was applied to reduce uncertainty in functional predictions. Based on this threshold, 569 of 4 552 ASVs (12.5%) were excluded, and 3 983 ASVs (87.5%) were retained for downstream analyses. While more stringent thresholds (e.g., NSTI < 1) could further improve prediction accuracy, they would also remove a larger proportion of taxa and potentially underestimate community functional diversity. The NSTI values obtained in this study are consistent with those reported for other aquatic environments (Fang et al., 2020; Langille et al., 2013).

Analysis of inferred functional profiles revealed variation in predicted metabolic pathways across samples. Pathways associated with carbohydrate metabolism, xenobiotic degradation, and cellular processes were among the most represented functional categories (Fig. 4a). In addition, pathways related to bacterial chemotaxis and biofilm formation were identified among the predicted functions, although these represent inferred capabilities rather than direct evidence of activity. Clustering of samples based on predicted functional profiles showed differentiation among preservation statuses, separating lakes into groups corresponding to low (Fúquene), intermediate (Tota), and high (Calderona and Colorado) preservation levels (Fig. 4b).

Differential abundance analysis using ALDEx2 identified 98 KEGG Ortholog (KO) terms grouped into six functional categories that varied significantly across preservation statuses (Fig. 4c; Table S2). Most of these differences were associated with pathways related to carbohydrate metabolism, xenobiotic degradation, cofactors and vitamins metabolism, and general cellular processes. Additionally, some pathways annotated in databases as related to infectious diseases were detected; however, these annotations reflect gene homology within reference databases and should not be interpreted as evidence of pathogenic activity in these systems.

https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026



https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026

Figure 4. Predicted functional profiles of bacterial communities across lakes with contrasting preservation status. (a) Relative abundance of the top 50 predicted functional pathways, (b) clustering of samples based on predicted functional profiles, and (c) differentially abundant KEGG Ortholog (KO) groups across preservation statuses.

4. Discussion

In this research, the analysis of bacterial and archaeal diversity across different aquatic ecosystems in the Northern Andean Region of Colombia revealed significant insights into microbial composition and functional potential concerning preservation status, lake size and spatial heterogeneity. Although a few reads assigned to archaeal taxa were retrieved, our study employed universal 16S V3–V4 primers designed for Bacteria. Consequently, archaeal diversity is likely under-represented, and any Archaeal ASVs detected should be interpreted with caution. The most abundant bacterial phyla identified were Actinobacteria, Proteobacteria, Bacteroidetes, Cyanobacteria, and Verrucomicrobia, with Actinobacteria and Proteobacteria consistently dominating across all systems. Notably, the presence of methanogenic archaea, such as Methanobacterium and Methanosarcina, although in low abundance, underscores the potential for localized anaerobic processes influenced by trophic dynamics, particularly in specific systems like Colorado and Fúquene. The detection of these methanogens in lake sediments had suggested that localized anaerobic processes can occur, especially under certain physicochemical conditions conditions (Ward and Frea, 1980). In eutrophic lakes, the decomposition of organic matter can create anoxic conditions favorable for methanogenic activity (Kersti Kangro et al., 2007; Smith and Schindler, 2009). This microbial process plays a crucial role in the carbon cycle by converting organic carbon into methane, a potent greenhouse gas (Raymond et al., 2013; Tranvik et al., 2009; Ward and Frea, 1980).

Microbial community structure displayed notable variations between lakes, especially in terms of preservation status. Calderona and Colorado, characterized by higher basin preservation, exhibited lower microbial diversity but more stable communities, as reflected in the consistent dominance of *Polynucleobacter* and *Flavobacterium*. The observed patterns suggest that higher preservation status in lakes leads to a reduction in microbial diversity but an increase in community stability as reported in diverse freshwater ecosystems (Diao et al., 2017; Hoetzinger et al., 2019; Nuy et al., 2020). This phenomenon may be influenced by factors beyond anthropogenic disturbances, as natural processes also shape community composition (Sessitsch et al., 2023). In small lakes, strong interactions between sediments, the littoral zone, and limnetic waters can significantly influence microbial dynamics (Søndergaard et al., 2003). These interactions often lead to shifts in nutrient availability and microbial community composition, particularly in response to sediment resuspension and littoral-pelagic coupling (Jones and Lennon, 2010). Additionally, the presence of Andean forest surrounding these lakes may serve as an indicator of conservation status and influence the input of organic matter with varying quality (Aranguren-Riaño et al., 2018). In the case of Tota, seasonal hydroclimatic variations likely play a role in regulating both natural and anthropogenic stressors (Aranguren-Riaño et al., 2018; Torres-Barrera et al., 2017). Such dynamics are critical for understanding the ecological balance within freshwater systems and their responses to environmental changes. In contrast, Fúquene, with a lower preservation status, exhibited the highest microbial richness and diversity, likely due to increased nutrient availability and disturbance events that create new ecological niches. These disturbances, whether natural (such as flooding or seasonal turnover) or anthropogenic (like runoff or pollution), disrupt established communities and enable the colonization of diverse microbial species (Hoang et al., 2023). Sampling variation was also evident in Tota, where seasonal changes influenced shifts in bacterial abundance and diversity, highlighting a dynamic microbial ecosystem. Contributing factors include seasonal nutrient fluctuations, as changes in temperature, light, and organic matter alter nutrient dynamics; hydrological changes, such as rainfall or evaporation affecting water levels and salinity; and biological interactions, like phytoplankton blooms or zooplankton grazing that influence resource availability for bacteria (Andrade-Sossa et al., 2023; Giongo et al., 2023; Hoang et al., 2023; Másmela-Mendoza et al., 2019).

Correlations in our co-occurrence network likely reflect shared niches or indirect links rather than direct interactions. These correlations underscore that microbial communities are dynamic entities shaped by environmental conditions, emphasizing their critical role in ecosystem resilience and functioning (Paver et al., 2020; Sessitsch et al., 2023; Shade et al., 2012). Although these patterns suggest preservation status and nutrient gradients shape community structure in impacted lakes, such inferences further require experimental validation. Moreover, the specialized relationship observed between *Mycobacterium* and *Polynucleobacter* underscores the adaptability and resilience of these communities in response to environmental changes. These insights support the conclusion's assertion that understanding microbial dynamics is crucial for effective freshwater ecosystem management, as recognizing co-occurrence patterns allows for better predictions of how these communities will respond to environmental changes and anthropogenic impacts, thereby informing conservation strategies (Sarkar et al., 2022).

Phylogenetic analysis of the *Mycobacterium* sequences (ASV29) revealed a diverse composition of *Mycobacterium* species in the Tota and Fúquene aquatic systems. Ten of the analyzed sequences were grouped with known species, including *M. celatum, M. noviomagense, M. szulgai, M. pseudokansasii, and M. cookii*, indicating close relationships among these taxa. The remaining sequences were categorized into polytopic clades, highlighting connections with multiple species groups, such as *M. attenuatum, M. innocens, M. arosiense, M. haemophilum, M. negraskense,* and *M. cookie*. *Mycobacteria* are known to inhabit various aquatic environments, including freshwater and marine systems, where they can affect the health of aquatic organisms and contribute to biogeochemical processes (Gcebe et al., 2018). For instance, studies have reported the isolation of non-tuberculous mycobacteria (NTM) from a range of aquatic animals, indicating their role in mycobacteriosis, which can impact fish health and aquaculture productivity (Gcebe et al., 2018; Hashish et al., 2018). The interactions between these mycobacterial species and other microbial communities can shape the ecological balance within aquatic ecosystems, highlighting the importance of monitoring their presence and understanding their ecological roles (Barletta et al., 2010; Nuy et al., 2020). In addition to *Mycobacterium*, sequences belonging to other genera of human health concern, such as *Leptospira* and *Legionella*, were also identified. *Leptospira* species, known for causing leptospirosis, can persist in freshwater environments and pose zoonotic risks through contact with contaminated water (Bradley and Lockaby, 2023). *Legionella*, the causative agent of Legionnaires' disease, is commonly associated with water systems and can proliferate under specific environmental conditions (Gattuso et al., 2022). The presence of these bacteria underscores the need for further investigation into potential public health risks and their ecological interactions. The interplay between these pathogenic and environmental microorganisms may influence microbial community dynamics, affecting both ecosystems functioning and water quality.

The predicted functional potential of bacterial communities across all samples provides insight into possible ecological roles within these aquatic ecosystems. However, these inferences are derived from 16S rRNA gene data and should be interpreted with caution, as PICRUSt2 predictions rely on phylogenetic relationships with reference genomes and may not capture strain-level functional variability. Consequently, the inferred metabolic profiles represent potential functional capacity rather than direct evidence of gene presence or activity.

Within this framework, variation in predicted functional pathways appears to be associated with differences in lake preservation status, likely reflecting underlying gradients in nutrient availability, microbial community composition, and environmental conditions (Battin et al., 2003, 2016; Paver et al., 2020; Shade et al., 2012). In systems with higher nutrient loads, such as those undergoing eutrophication, microbial communities are often characterized by shifts in diversity and metabolic potential that can influence nutrient cycling processes (Carpenter et al., 1998; Cotner and Biddanda, 2002; Paerl and Huisman, 2009; Smith and Schindler, 2009). For instance, the predicted enrichment of pathways related to naphthalene degradation and biofilm formation suggests potential metabolic capabilities within these microbial communities, reflecting their adaptability to fluctuating nutrient conditions and their possible role in organic matter transformation (Mohapatra and Phale, 2021). However, these inferences are based on taxonomic predictions and cannot be directly attributed to specific taxa, such as *Escherichia coli* or *Pseudomonas aeruginosa*, nor interpreted as direct evidence of their roles in organic matter decomposition or adaptation to fluctuating nutrient conditions. Furthermore, distinct clustering based on metabolic potential underscores the influence of preservation status, categorizing Fúquene as low, Tota as middle, and both Calderona and Colorado as high. Understanding this functional potential is

https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026

crucial for developing effective management strategies to preserve aquatic ecosystems amid environmental changes and anthropogenic pressures. A differential abundance analysis revealed distinct variations in predicted genes across the three preservation statuses, highlighting significant differences in genes related to carbohydrate metabolism, xenobiotic degradation, cofactors and vitamins metabolism, and various cellular processes. These findings are ecologically important, particularly in the context of eutrophication, where changes in nutrient dynamics can significantly impact the functional roles of microbial communities (Forero-Pineda et al., 2021; Shade et al., 2012; Smith and Schindler, 2009). Furthermore, the presence of pathways associated with bacterial infectious diseases adds another layer of complexity to these ecosystems. Understanding these variations is crucial for assessing the health and resilience of aquatic systems, especially in the face of environmental changes and human impacts, as they can greatly influence the balance and sustainability of these habitats.

5. Conclusions

In conclusion, our study reveals complex links between microbial taxonomic diversity, predicted functional potential, and preservation status across Andean lakes in Colombia. Co-occurrence networks suggest potential ecological associations among dominant genera, and phylogenetic analyses of *Mycobacterium* expand our view of lineage diversity (but both are based on predictive or correlational methods). The observed variation in community composition and inferred metabolic pathways across lakes provides useful hypotheses and candidate indicators for future monitoring, yet these patterns require targeted validation before they can underpin system-specific conservation or management actions.

https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026

Study limitations statement. Several methodological and interpretative constraints temper our findings. First, all functional profiles derive from PICRUSt2 predictions rather than metagenomic or transcriptomic measurements and thus represent hypothetical potentials, not actual gene expression or activity. Second, species-level assignments—especially for *Mycobacterium* and other genera based on the V3–V4 16S region—are inherently tentative due to insufficient phylogenetic resolution. Third, with only one or two samples per lake and no true longitudinal sampling, any claims of temporal variation or community stability remain speculative. Fourth, our small sample size limits the robustness of "core microbiome" definitions and broader generalizations about diversity patterns. Fifth, the absence of concomitant environmental metadata (e.g., nutrient loads, dissolved oxygen) restricts our ability to link microbial shifts directly to physicochemical drivers. Finally, co-occurrence networks, built on correlation thresholds and null-model assumptions, may capture shared environmental preferences or indirect associations rather than genuine ecological interactions. Together, these limitations underscore the need for expanded sampling, direct functional assays, and integrated environmental measurements in future work.

https://doi.org/10.5194/egusphere-2026-2137
Preprint. Discussion started: 21 May 2026

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.

Reproduction rights

The authors confirm that all figures and tables included in this manuscript were created by the authors or derived from data generated in the present study. No third-party copyrighted material requiring permission has been reproduced. Where applicable, all materials are distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplementary information

Supplementary table 1. Complete metataxonomic description and relative abundance in the freshwater lentic systems analyzed.

Supplementary table 2. Multiple differential comparisons of predicted gene abundances in KEGG Orthology (KO) functions and annotated pathways.

Figure S1: Summary for rarefaction analysis of (a) observed *vs.* rarified number of ASVs, and (b) rarefaction curves for the aquatic samples studied.

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.

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.

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.

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.

Fang, J., Yang, R., Cao, Q., Dong, J., Li, C., Quan, Q., Huang, M., and Liu, J.: Differences of the microbial community structures and predicted metabolic potentials in the lake, river, and wetland sediments in Dongping Lake Basin, Environ. Sci. Pollut. Res., 27, 19661–19677, https://doi.org/10.1007/s11356-020-08446-4, 2020.

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.

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.

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.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F. O.: Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic Acids Res., 41, e1–e1, https://doi.org/10.1093/nar/gks808, 2013.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Vega Thurber, R. L., Knight, R., Beiko, R. G., and Huttenhower, C.: Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences, Nat. Biotechnol., 31, 814–821, https://doi.org/10.1038/nbt.2676, 2013.

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 pp., 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, 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.

Paulson, J. N., Stine, O. C., Bravo, H. C., and Pop, M.: Differential abundance analysis for microbial marker-gene surveys, Nat. Methods, 10, 1200–1202, https://doi.org/10.1038/nmeth.2658, 2013.

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-Paría, 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.

Ren, Z., Wang, F., Qu, X., Elser, J. J., Liu, Y., and Chu, L.: Taxonomic and Functional Differences between Microbial Communities in Qinghai Lake and Its Input Streams, Front. Microbiol., 8, https://doi.org/10.3389/fmicb.2017.02319, 2017.

Sarkar, S., Ward, K., Kamke, A., Ran, Q., Feehan, B., Richie, T., Reese, N., and Lee, S. T. M.: Perspective: Simple State Communities to Study Microbial Interactions: Examples and Future Directions, Front. Microbiol., 13, https://doi.org/10.3389/fmicb.2022.801864, 2022.

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.

Veech, J. A.: A probabilistic model for analysing species co-occurrence, Glob. Ecol. Biogeogr., 22, 252–260, https://doi.org/10.1111/j.1466-8238.2012.00789.x, 2013.

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, C., Mai, J., Cao, X., Burberry, A., Cominelli, F., and Zhang, L.: ggpicrust2: an R package for PICRUSt2 predicted functional profile analysis and visualization, Bioinformatics, 39, btad470, https://doi.org/10.1093/bioinformatics/btad470, 2023.

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