Anonymous Referee #3

The article presents a study analyzing the microbiome of four different lakes located in the Eastern Cordillera of Colombia. The underlying premise is that two of the lakes are relatively well preserved, whereas the other two are more heavily affected by human activities and environmental degradation, exhibiting higher levels of eutrophication and contamination.

The objective of the study is interesting because it focuses on the analysis of the 16S rRNA gene to characterize the bacterial communities inhabiting the lakes and compare their composition. The study also includes a functional analysis based on these amplicons. Although this analysis is presented as a prediction of potential functional activities, it might provide valuable insights into the possible ecological roles and metabolic functions of the bacterial communities identified.

> ***Response***:
> *We thank the reviewer for the positive assessment of the study objective and for recognizing the value of combining 16S rRNA gene-based bacterial community characterization with predicted functional profiling. In the revised manuscript, we have clarified that the study is exploratory and descriptive, based on six water samples from four Colombian Andean lakes. We also clarified that the preservation-status categories were assigned qualitatively from available environmental information and were not treated as statistically replicated explanatory factors.*
>
> *In addition, we revised the Abstract, Introduction, and Methods to emphasize that PICRUSt2 outputs represent predicted genomic potential inferred from 16S rRNA gene data, rather than direct evidence of gene presence, gene expression, metabolic activity, or ecosystem process rates. Thus, the functional analysis is now presented as a hypothesis-generating component of the study.*

Mayor comments:
The authors acknowledge the limited number of samples available for comparing the lakes. However, it is unclear why only one sample was collected from two of the lakes, whereas two samples were collected from the other two lakes.

> ***Response***:
> *We thank the reviewer for pointing out that the unequal number of samples among lakes required clarification. The original sampling effort was constrained by access, logistics, public-order conditions, and available resources. Subsequent plans to expand the dataset through additional balanced sampling were delayed by pandemic-related restrictions and continued access limitations.*
>
> *We also clarified that, for this reason, the dataset was treated as an exploratory collection of individual water samples rather than as a replicated lake-level comparison. Accordingly, the revised manuscript does not use the dataset to test lake-level effects or preservation-status effects, and all results are interpreted descriptively at the sample level.*

The manuscript includes an entire section devoted to a phylogenetic analysis of sequences assigned to the genus Mycobacterium. A clearer justification is needed for focusing specifcally on this genus, and, more importantly, for emphasizing this analysis when the authors themselves acknowledge that the results should be interpreted with caution due to the limited taxonomic resolution at the species level.

***Response**:*
We thank the reviewer for highlighting the need for a clearer rationale and a more proportionate treatment of this analysis. We revised the Introduction to explain that Mycobacterium was selected because nontuberculous mycobacteria are common inhabitants of natural and engineered aquatic environments and include lineages of potential relevance to human, veterinary, aquaculture, and wildlife health. However, we agree that the V3–V4 region of the 16S rRNA gene does not provide sufficient resolution for reliable species-level identification within this genus.

Accordingly, we substantially reduced the emphasis placed on this analysis. All definitive species-level assignments and species-specific interpretations were removed, and the analysis was reframed as an exploratory phylogenetic placement of Mycobacterium ASVs relative to named reference sequences. Species names are now used only to identify the reference sequences included in the tree and are not assigned to the environmental ASVs. The Results explicitly state that the observed placements indicate phylogenetic affinity but do not establish species identity, pathogenicity, viability, virulence, host infection, or health risk. In addition, the maximum-likelihood tree was moved from the main figures to Supplementary Figure S2, and the corresponding discussion was shortened. Definitive taxonomic identification would require future analysis using cultured isolates, whole-genome sequencing, or validated higher-resolution marker genes.

- Discussion contains several very general statements that do not provide specific insights. For example, in lines 423–428, the authors mention that "contributing factors such as nutrient fluctuations, temperature, light..." influence the microbial communities. Are these factors relevant only to Lake Tota or to all the lakes studied? If they apply to all lakes, do the authors have measurements for any of these environmental variables collected during sampling?

***Response**:*
We thank the reviewer for this observation and agree that the former wording could be interpreted as attributing the observed microbial patterns to environmental variables that were not measured. No physicochemical variables were collected concurrently with microbiological sampling, including nutrient concentrations, dissolved oxygen, chlorophyll-a, pH, temperature, conductivity, organic-matter indicators, or hydrological variables. Therefore, environmental–microbiological correlations could not be performed.

The Discussion has been revised extensively to remove unsupported explanations based on nutrient fluctuations, temperature, oxygen conditions, light availability, rainfall, evaporation, phytoplankton, zooplankton, or seasonal dynamics. These factors are now mentioned only as mechanisms documented in previous freshwater studies and are explicitly distinguished from variables evaluated in the present study. We also removed the former statement that seasonal changes influenced bacterial abundance and diversity in Tota. The two Tota samples are now described as individual observations from two sampling dates, one of which was a historical December 2018 sample, and not as replicated temporal or seasonal sampling.

The revised text states explicitly that differences among samples cannot be attributed to eutrophication, nutrient availability, anthropogenic pressure, preservation status, lake size, spatial heterogeneity, sampling period, water-column depth, or other unmeasured environmental drivers.

- Similarly, in line 499, the authors refer to the presence of infectious bacteria associated with diseases and their potential importance for public health. However, earlier in the manuscript

(lines 368–370), they state that these functional assignments should not be interpreted literally because they are inferred based on gene homology. This represents a contradiction that should be clarified.

***Response**:*
We agree and have removed the interpretation that KEGG pathways labelled as infectious diseases indicate the presence of infectious bacteria or public-health risk. The revised manuscript now explains that organism-, disease-, drug-resistance-, and eukaryote-associated KEGG labels reflect the structure and nomenclature of the reference database and the occurrence of homologous genes within annotated pathways.

These labels are no longer interpreted as evidence of the named organisms, pathogenicity, virulence, antimicrobial resistance, infection, or corresponding biological processes in the sampled lakes. The PICRUSt2 results are now described only as a descriptive approximation of predicted genomic potential based on ASV phylogenetic placement and available reference genomes, rather than as direct evidence of gene presence, expression, metabolic activity, or ecological function.

We also revised the interpretation of Leptospira, Legionella, and Mycobacterium. Their detection is now reported strictly at the taxonomic resolution supported by the short 16S rRNA gene region. The manuscript explicitly states that these assignments do not establish pathogenic species, viable cells, transmission, infection, or public-health risk.

- The Discussion ends with an entire paragraph describing the study's limitations. The points raised there essentially demonstrate that additional sampling is needed and that both the phylogenetic and functional results should be interpreted cautiously. However, throughout the manuscript, these analyses are given considerable emphasis. Rather than focusing so extensively on the study's limitations, the authors should instead emphasize the conclusions that can be robustly supported by the type of analysis performed, namely 16S rRNA gene sequencing.

***Response**:*
We agree and reorganized both the Discussion and Conclusions. The former extended limitations paragraph was removed, and the principal limitations are now integrated directly into the relevant interpretative sections. This allows each result to be discussed together with the specific constraint affecting its interpretation without ending the Discussion with an extensive catalogue of methodological weaknesses.

The revised Discussion places primary emphasis on findings directly supported by the 16S rRNA gene data: the predominance of common freshwater bacterial phyla, descriptive differences in taxonomic composition and coverage-standardized diversity among samples, genus-level detection of selected low-abundance taxa, and exploratory phylogenetic placement of Mycobacterium-related ASVs. Species-level claims based on the V3–V4 region were removed, the co-occurrence analysis was deleted entirely, and the functional analysis was reduced to descriptive PICRUSt2 predictions without differential-abundance testing or environmental attribution.

The Conclusions were also rewritten to focus on the descriptive baseline provided by the six samples. A concise final statement identifies the need for replicated sampling, concurrent physicochemical measurements, metagenomic sequencing, and targeted taxonomic validation, while avoiding unsupported conclusions about preservation status, temporal dynamics, ecological interactions, pathogenicity, or environmental causation..

Minor comments

- The Introduction would benefit from a more detailed description of the potential differences in the geological origins of the lakes, as this is stated to be an important factor in line 111. However, no information is provided to support this claim beyond Table 1. Could these geological differences explain the distinct microbial community compositions observed among the lakes?

> ***Response**:*
> *We thank the reviewer for noting that the geological context was insufficiently described. We expanded the Introduction to include all four water bodies. Fúquene and Tota are described using published geomorphological and paleolimnological evidence, Calderona is identified as a high-elevation glacial lake, and Colorado is described more cautiously as a lower-elevation, rainfall-influenced wetland because we found no sufficiently robust geological source supporting a formal "pluvial-origin" classification. We also clarify that these contrasting settings could influence microbial habitats indirectly through basin morphology, sediments, hydrology, residence time, and water chemistry. However, because these variables were not systematically measured and the study lacked adequate replication, geological or hydrological origin cannot be identified as the cause of the observed microbial differences.*

- In the sampling methodology, the materials used for sample collection are not described in sufficient detail. For example, the manufacturer of the "Bottle 1 Sequi disk" (line 123) is not specified, nor is the type of filter used for sample filtration (line 124).

> ***Response**:*
> *We thank the reviewer for identifying the insufficient description of the sampling materials. We revised the Methods to clarify that water was collected at a depth of 5 m using a custom-made, non-commercial Schindler–Patalas-type sampler constructed from a thoroughly cleaned plastic bottle; therefore, no commercial manufacturer or model applies. We also added the diameter and color of the Secchi disk, the diameter and opening size of the prefilter, and the membrane material, filtration area, manufacturer, and catalog number of the Sterivex units. The preservation reagent and transport conditions were also specified.*

- In the DNA extraction section, could the authors briefly explain the purpose of CTAB in the extraction protocol? (lines 140–141).

> ***Response**:*
> *We thank the reviewer for this suggestion. We revised the DNA extraction section to clarify that CTAB acts as a cationic detergent that contributes to cell lysis and, in the presence of high salt, facilitates the removal of polysaccharides and other co-extracted contaminants that could inhibit downstream molecular analyses.*

- Lines 145–146: Were the primers used adopted from previous studies, or were they specifically designed by the authors for this study? No reference is provided to support their use.

> ***Response**:*
> *We thank the reviewer for noting the missing reference. We have added the original DADA2 citation after the description of chimera removal using removeBimeraDenovo.*

- Line 160: please include a reference after "removeBimeraDenovo".

***Response**:*
We thank the reviewer for identifying the missing reference. The primers were not designed specifically for this study but were adopted from Klindworth et al. (2013). We have added the corresponding citation and clarified the primer names, sequences, targeted V3–V4 region, and basis for their selection.

- Line 162: Why was a threshold of 100 reads selected instead of a lower or higher value? This appears to be a rather arbitrary cutoff, and it may be relatively high, potentially affecting the detection of taxa that are present at low abundance and influencing the identification of taxa specific to each lake.

***Response**:*
We thank the reviewer for raising this concern. Upon re-examining the original DADA2 analysis script, we found that the manuscript incorrectly reported the application of a 100-read abundance threshold. No such ASV-level threshold was applied in the original pipeline. We have therefore removed this statement and revised the Methods to describe the actual filtering procedure. ASVs lacking a phylum-level assignment and non-target chloroplast or eukaryotic sequences were excluded, whereas ASVs unresolved at lower taxonomic ranks were retained as unclassified categories. We also revised the description of the DADA2 workflow to clarify the read-trimming parameters, error-model learning, concatenation of non-overlapping paired reads, and consensus-based chimera assessment.

- Line 173: Please rewrite this sentence, as it is difficult to understand what the word "respectively" refers to.

***Response**:*
We thank the reviewer for identifying the unclear wording. We rewrote the sentence to define each Hill-number order explicitly: q=0 as ASV richness, q=1 as exponential Shannon diversity, and q=2 as inverse Simpson diversity. We also separated the description of alpha-diversity estimation from the descriptive Bray–Curtis beta-diversity analysis.

- Lines 235-242: The authors compare ASV abundances among the lakes. Could they clarify whether the comparison is based on the abundances of the same ASVs across lakes, or whether it is simply an overall comparison to determine which lake contains ASVs with higher relative abundances?

***Response**:*
We thank the reviewer for requesting clarification. In the former version, the analysis compared the overall distributions of ASV read abundances among individual samples; it was not a matched comparison of the abundances of the same ASVs across lakes. We recognize that this approach treated ASVs within each sample as replicate observations and did not provide an appropriate test of differences among lakes. We therefore removed this analysis, including the associated pairwise comparisons, p-values, significance statements, and figure elements. The revised Results now describe ASV richness, taxonomic relative abundance, alpha diversity, and beta diversity only at the individual-sample level, without inferential comparisons among lakes or sampling dates.

- Line 238: Although the text states that no comparison was performed between Colorado and Fúquene, Figure 1 includes a comparison between Colorado and Fúquene 2. Is this a typographical error?

***Response**:*
We thank the reviewer for identifying this inconsistency in the former version. After reconsidering the statistical approach, we removed the ASV-abundance pairwise comparisons in their entirety because the limited sampling design did not support reliable inferential testing among lakes or sampling dates. Consequently, the revised Figure 1 contains no comparison brackets, p-values, or significance annotations, and the corresponding statements were removed from the Results. Community differences are now presented descriptively at the individual-sample level.

- Lines 235-242: The manuscript also states that statistically significant differences were detected at $p < 0.001$, whereas Figure 1 reports $p< 0.0001$. In addition, the number of asterisks varies from two to four across the different comparisons, but the corresponding p-value thresholds are not explained. The figure should either be revised or the statistical significance levels should be described more clearly.

***Response**:*
We thank the reviewer for identifying the inconsistency between the p-values reported in the text and those shown in the former Figure 1, as well as the absence of definitions for the significance symbols. After reconsidering the statistical design, we removed all inferential ASV-abundance comparisons from the manuscript because the limited and uneven sampling did not support reliable hypothesis testing. Accordingly, the revised Figure 1 contains no p-values, asterisks, significance brackets, or pairwise comparisons, and the corresponding statements were deleted from the Results. Differences among samples are now presented descriptively only.

- Regarding the diversity analysis, line 276 states that the inverse Simpson index was calculated, whereas the figure legend (line 312) refers to the Simpson index. Please clarify which diversity index was actually used.

***Response**:*
We thank the reviewer for identifying this inconsistency. The analysis used Hill numbers, for which $q=2$ corresponds to inverse Simpson diversity, not the Simpson index. We have revised the figure legend and the corresponding Methods and Results text to use the term "inverse Simpson diversity" consistently throughout the manuscript.

- Línea 280: The authors state that Calderona shows a gradual decline. However, the decline is still substantial, although not as pronounced as in the other lakes. Moreover, no statistical analysis is presented to support the claim that the pattern differs from those observed in the other lakes. Please clarify how this conclusion was reached.

***Response**:*
We thank the reviewer for identifying this overinterpretation. The statement that Calderona showed a more gradual decline was based only on visual inspection of the Hill-number profile and was not supported by a formal statistical comparison. We therefore removed this claim. The revised Results now describe the alpha-diversity patterns among individual samples descriptively and do not assert that Calderona differed significantly from the other lakes.

- Line 329-332: The authors mention the detection of the bacterial genera Leptospira and Legionella, but they do not indicate in which samples these genera were detected. This information can only be found in the supplementary table. Since these genera are highlighted in both the Results and the Discussion because of their potential significance, the manuscript

should clearly state in which samples they were detected and their relative abundances. Furthermore, line 462 states that Legionella proliferates under specific environmental conditions, but these conditions are not described. Do the environmental conditions of the lakes included in this study meet those requirements?

***Response**:*
We thank the reviewer for this observation. Although the sample-specific occurrence and relative abundances of Leptospira and Legionella were already provided in Table S1, we agree that they should also be reported explicitly in the Results because these genera were discussed in the main text. We therefore added the samples and corresponding relative abundances for both genera. Leptospira was detected in the two Fúquene and two Tota samples, whereas Legionella was detected in all six samples and reached its highest relative abundance in Calderona. We also removed the statement suggesting that environmental conditions in the sampled lakes favored Legionella proliferation, because the relevant environmental variables were not measured contemporaneously. The revised manuscript treats these findings only as genus-level taxonomic observations and does not infer species identity, viability, pathogenicity, proliferation, or health risk.

- Typos: In lines 67 and 69, the publication year is missing from the citation to Sessitsch et al. Additionally, in line 238, the word "Furthermore" appears to contain a typographical error.

***Response**:*
We thank the reviewer for identifying these typographical errors. They were corrected in the revised version.