

Review about manuscript: “Bacterial community structure and predicted functional potential across Andean lakes with contrasting preservation status based on 16S rRNA gene profiles”.

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.

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.
-
- 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 specifically 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.
- 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?
- 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.
- 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.

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?

- 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).
- In the DNA extraction section, could the authors briefly explain the purpose of CTAB in the extraction protocol? (lines 140–141).
- 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.
- Line 160: please include a reference after "*removeBimeraDenovo*".
- 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.
- Line 173: Please rewrite this sentence, as it is difficult to understand what the word "*respectively*" refers to.
- 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?
- 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?
- 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.
- 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.
- 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.
- 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?
- 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.