

## Author response

We would like to thank the editor and reviewer for reviewing our manuscript and for the comments. The manuscript has been revised and all changes are highlighted in the text. The Reviewer's comments are shown at the end of the text.

- 1) We agree. We have tried to emphasize the fact throughout the text that these metagenomes are only predicted from 16S rRNA gene profiles and not real metagenomes.
- 2) Yes, and this is probably the case. A reference to microbial dark matter has been added (P9,L2; P13,L3; P22,L24.25).
- 3) a&b. we are comparing two different methods, which both give the same conclusion that there is no specific physicochemical parameter that could explain the clustering of the archaeal or bacterial communities. We used the UPGMA cladogram together with a heatmap in Fig 5 and the PCoA plotting in Fig 6. In the archaeal PCoA we can see a clustering of samples with intermediate salinity, >10 but <20 g TDS/L, but this could not be seen in the bacterial analysis. Thus we draw the conclusion that the clustering of the samples is more dependent on the community composition than on physicochemical parameters, especially for the bacteria. However, this is so speculative that the sentence has been omitted for clarity. (P9, L28-31)
- 4) Section 3.5 and the discussion has been edited according to the reviewer's comments.
- 5) It means that if a sequence read is very short it is more likely in the taxonomic assignment analysis to be assigned to several different taxa. The text has been clarified. (12, L29-32)
- 6) The text has been clarified. (P13, L19 ->)
- 7) The number of samples was 12 and this has been added to the abstract text and to the legend of Figure 4. (P1,L17; P28,L2)
- 8) The heatmap colours have been changed. (P28, Fig5)
- 9) The figure was edited and figures c-f were moved to the supplements (Figure S2a-d). The sizes of the PCoA plots in Fig 6ab were increased for clarity. (P29, Fig6)
- 10) The figure was changed to a table (Table 3) for clarity. (P26)

## Typos:

- 1) Corrected allows to allow
- 2) At removed
- 3) Corrected to previously
- 4) Every zink has been corrected to zinc
- 5) Estimated changed to predicted
- 6) Units placed in () and the abbreviations are explained.

## Reviewer comments

The submitted manuscript "Microbial communities and their predicted metabolic characteristics in deep fracture groundwaters of the crystalline bedrock at Olkiluoto, Finland" provides a complete description of the microbial communities (and predicted functions) of 12 samples ranging from 296 to 798 m depth. The methodology and analyses used are sound and I only have minor comments (see below):

- 1) In general, when talking about the predicted metabolisms, word choice should be adjusted. For example
- 2) pg 8 lines 28-29, could the inflation of "Bacteria, Other" be also due to novel sequences (i.e.

Microbial Dark Matter)?

3a) pg 9 lines 25-26, I'm not sure how this conclusion was made. Additional clarification would be helpful. The conclusion about physicochemical parameters is then repeated in lines 27-28.

3b) I don't understand how the conclusion that the number of "OTUs was the strongest clustering force" or what exactly this means. This part should be re-written for additional clarity

4) Section 3.5 should be rewritten in a language that emphasizes the metabolic functions are predicted. The first paragraph does a good job at conveying the metabolisms are predicted by the second paragraph and onward says things like "Nitrate is reduced", "Ammonia is then used" --- which could be a bit misleading despite the section title. I suggest clarifying this paragraph with either an additional sentence that the predicted community function could be described as XX or changing things like "Nitrate is reduced" to "Nitrate was predicted to reduce"

--- Clarification that the functions/metabolisms are predicted should be further clarified here:

4b) pg 15 lines 30 to 31

4c) pg 16 line 3

4d) pg 16 section "Amino acid metabolism"

5) pg 12, lines 26-27. I'm not sure what this sentence means. Please clarify

6) pg 13 lines 12-25. This paragraph is difficult to follow and should be rewritten for clarity

7) Abstract, Figure 4. When introducing or describing shared OTUs as a \*percentage\* (axis label is fraction) of samples, it would be useful to be reminded how many samples there were

8) Figure 5. The heatmaps could be made more interpretable via a better color scheme a single label on the y-axis saying "OTUs" rather than the individual OTU names.

9a) Figure 6. C-F are not necessary in the main text, they can be moved to the supplement. If moved to the supplement, the % variance described by PCoA axis 1 and 2 could be added to the axis labels of A and B.

9b) It is hard to distinguish the shapes in A and B

10) Figure 7 is pixelated. A different image type would look better in publication.

Typos:

1) pg 3 line 10 allows should be allows

2) pg 4 line 14, remove "at"

3) pg 12 line 20, the use of "so far hidden" should be changed. Perhaps the authors meant previously?

4) pg 17, "Membrane transport" section: zink should be zinc

5) pg 17 line 30, estimated to predicted

6) Table 1: put units in (). Define NPOC, DIC, TDS, TNC