Biogeosciences Discussions

Biogeosciences Discuss., author comment AC1
https://doi.org/10.5194/bg-2021-165-AC1, 2021

Reply on RC1

Sania Arif et al.

Author comment on "Composition and niche-specific characteristics of microbial consortia colonizing Marsberg copper mine in the Rhenish Massif" by Sania Arif et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-165-AC1, 2021

The authors would like to thank referee for her constructive suggestions and comments

1. 14 and elsewhere: please use caution with the term "metagenome" when referring to your 16S rRNA gene amplicon data. Although, metagenomics in its strictest definition includes amplicon sequencing the prevailing use of the term in microbiology is in reference to genomic sequencing of multiple organisms in a single community. Change this throughout the paper to something like community amplicon sequencing. This change will also highlight the fact that you performed both metagenomics and amplicon-sequence based community characterization for your study.

Separated the amplicon sequencing from metagenomics in the manuscript.

2. 51 and l. 65: the concept of "cold" is mentioned here but not expanded upon or highlighted as a focus of the study until the discussion and conclusions. Without more explanation I don't see the justification for this to be a unique aspect of the site and why it would be novel at all. Why is studying a cold environment with heavy metal enriched waters important? How is "cold" quantified and relative to what? Comparisons to other heavy metal-contaminated sites and their temperatures/microbial communities would be valuable here.

Since, Marsberg Kilian Coppermine has a constant temperature of 10 °C, it is referred as cold environment. It was not expanded because 10 °C is not a good borderline between psychrophiles and mesophiles as both could grow on it. Ktedonobacteria cultured members are either mesophiles or thermophiles but psychrophiles are reported through amplicon sequencing only. In the scope of this study, it was not possible to culture Ktedonobacteria to validate its temperature range and particular role in a cold heavy-metal comminated site.

78: please provide (throughout the paper) the rational for performing metagenomics on a single sample and especially why this particular sample was selected.

Explained in line 83 that the MB1 biofilm was selected because of the high abundance of Chloroflexi (Ktedonobacteria) and line 225-229 of new version: Extracted DNA from one of the leachate biofilm samples, MB1 abundant in Ktedonobacteria (see fig. 4) as a representative of leachate group was submitted to the Göttingen Genomics Laboratory for shotgun metagenomic sequencing. The rationale behind selecting MB1 biofilm was to investigate the survival mechanisms that contributed to the high abundance of Ktedonobacteria around the toxic copper-rich leachate stream.

188-189: this sentence, especially the phrase "as a representative of leachate group" is not very clear. Can this sentence be revised for clarity? It might be valuable to provide

some more detail on the sample selection here or to refer the reader to the results where community data is used to justify selection of MB1.
corrected
3. 323-324: "MB1 as a representative of the leachate group" – looking at Figure 4 MB1 has a number of differences compared to the other leachate group samples. What was "representative" based upon? It actually has a lower abundance of Actinobacteria compared to the other leachate samples, while Chloroflexi are in higher abundance than 4 of the leachate samples. Regardless of how/why this sample was chosen please add details on the reasoning and selection process so that it is clear to the reader how the MAGs fit into our understanding of the overall community in the system.
Since Ktedonobacteria were abundant in leachate biofilms and mostly members of Ktedonobacteria are uncultured and unknown, we were interested to investigate its heavy metal resistance and aromatic compounds metabolism. That's why MB1 was selected as it has more Chloroflexi(Ktedonobacteria) reads. Moreover, it was possible from this sample to extract enough DNA for metagenomic sequencing.
4. 91: I think you are missing a "respectively" here
Corrected
5. Figure 1: can you add a symbol for the water table depth? It is unclear if the water depth is shown in the cross section or if the water table is being pumped down for mining activities
The water table depth of the streams is just few centimetres deep and wide. These narrow water streams are naturally flowing downwards and, in the NE, or SW directions without being pumped artificially.
6. 85: it is unclear to me if the "spring water" samples are meant to represent background, uncontaminated samples. The geochemistry suggest that they are uncontaminated from the mining activities but it is not easy to discern from Figure 1. Explicitly stating whether those are meant to be background samples would be valuable. Okay, I just re-read Section 3.1 and there it states that these are sources of fresh groundwater. I still suggest adding that detail to the methods.
Corrected in line number 94. The section 2.1 is rewritten.
7. 102: I would change this to genomic DNA to encompass both sequencing approaches used. (see comment 1 above)
Corrected
8. Figure 2: please add scale bars to all panels.
Corrected
9. 115-117: this sentence needs clarification to state which sample types were stored in which bottle type.
Corrected and explained in the manuscript
10. 122 and elsewhere: be consistent with use of charges for all anions and cations. Throughout the paper you switch between including or not including charges for elements. I usually only use charges for ions and not elements; make sure superscripts and subscripts for sulfate and nitrate are correctly formatted (e.g., L. 215 and 217; Figure 3 legend).
Corrected. The charges for the ions have been skipped to avoid confusion because we are talking about analysed total concentrations, not different specifically charged ions. With ICP-OES or ICP-MS we measured total amount of elements, e.g. Sr or Cu, not specific ions. Therefore, in tableS3 no charges are given, in fact everything is total concentration of dissolved element or parameter.
11. 135 and 149: omit "photometrically" as it is not necessary nor is it the best word choice.
Deleted.
12. 202: the abbreviation was already defined
Corrected
13. Section 3.1: the methods state that PHREEQC was used to look "calculation of ion

activities, pCO2 (partial pressure of CO2) of samples and mineral saturation states" but no data are presented. Please include those results or amend the methods section.
Modified. Supplementary information Excel Table S3 is uploaded at the Göttingen Research Online Database and unnecessary parts were deleted, which includes the whole calculation as mmol/l, calculated charge balance, saturation states and some minor important parameters (e.g. Cs, PO4, silica etc.). Table 3 includes basics (pH etc.) and the analytic values as mg/L or µg/L .
14. 216-218 and elsewhere: make sure the units are written correctly with superscripts and spacing; check figures too.
Corrected
15. Figure 3: are the samples presented in the direction of water flow? Consider plotting the copper flume leachate heavy metal concentrations in mg/L.
Information added about the direction of water flow in the figure 3 captions. Concentration plotted in mg/L.
16. 243-244 and 254-257: It is not clear what data are being tested and compared here. Were the "actual abundances" (Fig. S3) or the "relative abundance" (Fig. 4) used in the statistical tests? What cutoff was used to designate what an "abundant taxa was"? I'm also wondering if an ANOVA is the best statistical analysis here. If you are only comparing relative abundance values, it doesn't take into account overall changes in total biomass. An analysis like DESeq2 or LEfSe would be more appropriate as it would take into account the differences in biomass and identify specific taxa that are changing in differential abundance.
The actual abundance values after normalization were used for ANOVA. As recommended by reviewer, DESeq2 has been applied for the differential abundant analysis of taxa at the phylum and class level and described in main text.
17. Figure 4: consider using the same colors in panels A and B for taxa that have the same name at both taxonomic levels. Check spelling in panel B for Oxyphotobacteria
Corrected
18. 271, 278-284 (and elsewhere): It is really important to be cautious with assigning causality with observational data. Your data didn't allow you to monitor enrichment or replacement of taxa so I would tone down the language and state these things more as hypotheses.
Corrected
19. 291-298: in this section it is not clear how the nematode and alga were identified to species level. Either here or in the methods a description of those methods would be valuable. Further, the identification of Actinobacteria hyphae to genus level is also not clear, although there is 16S rRNA gene data indicating that these organisms are highly abundant.
They were identified as a result of 18S amplicon sequencing. The highest frequency sequences matched 100% to several Chloropyta sequences according to the Blast search and when a species is given, most of them were identified as Coccomyxa subellipsoidea. Abundant actinobacterial genera and a phylogenetic tree are provided in the supplementary files.
20. Results: I found the order of presentation to be confusing, especially for section 3.4. Sections 3.2 and 3.3 present the 16S rRNA gene sequencing for community analysis then 3.4 jumps to microscopy then back to 16S rRNA gene-based phylogenetic analysis. It would be much easier to follow the story if the diversity analyses (section 3.3) were presented before distribution of taxa (section 3.2) with the microscopy and phylogenetic analysis (section 3.4) last. The jump from 16S rRNA gene data to microscopy would not be as severe as it seems that these data sets support one another.
Corrected and rearranged
21. 302-313: the phylogenetic analysis was based on 350 bp amplicon sequences which doesn't provide a lot of information for robust taxonomic affiliation or phylogenetic inference. Making definitive statements about an OTU being rare or a novel class/species is a big reach based on limited sequence read length.

Corrected
22. 313: correct the reference style.
Corrected
23. 349: this is a great idea to publish these maps and make the details available to the reader.
Thank you
24. Figure 7: I find this figure hard to interpret. Is this really the best way to present these results? Could the figure be revised so that the text doesn't wrap or overlap the Venn diagram or the sample names? Caption: GO should be in all caps.
GO terms capitalized and simplified for readers.
25. 391: the topic of Mo and W uptake is of interest to me so I looked up the Markovich, 2001 reference. The ModABC molybdate system is not mentioned in this paper and the reference only assesses sulfate transporters in mammals. While Markovich does address membrane uptake of molybdate and tungstate along with sulfate I don't see how this is an appropriate reference for the ModAB system. This manuscript: https://www.frontiersin.org/articles/10.3389/fmicb.2018.03030/full, references a number of papers that would be more appropriate sources for that uptake system:
321. Self, W. T., Grunden, A. M., Hasona, A., and Shanmugam, K. T. (2001). Molybdate transport. Res. Microbiol. 152, 311–321. doi: 10.1016/S0923-2508(01)01202-5
322. Maupin-Furlow, J. A., Rosentel, J. K., Lee, J. H., Deppenmeier, U., Gunsalus, R. P., and Shanmugam, K. T. (1995). Genetic analysis of the modABCD (molybdate transport) operon of Escherichia coli. J. Bacteriol. 177, 4851–4856. doi: 10.1128/jb.177.17.4851-4856.1995
Corrected and added references.
26. Discussion: the discussion is highly focused on the MAG results and little about the overall system and community dynamics. Combining the results and discussion sections might be a stronger approach for the paper.
The biofilms details and their abundant taxa have already been briefly discussed in https://journals.asm.org/doi/full/10.1128/MRA.01315-20. Combining both sections at this stage is again a lot of work, therefore the overall community dynamics is being discussed in discussion section again
27. 427-429: the previous microbial work at the site would be a good addition to the introduction.
Added in line 75
28. 439: omit metagenomic here and use a broader term/phrase
Corrected.
29. 474: omit "etc"
Corrected.
30. 476: add "had" before specialized
Corrected.
31. 477-481: this sentence is much too long and really difficult to follow. Revise

Corrected
32. 487: change to Deltaproteobacteria
Changed to Gammaproteobacteria as it is gamma symbol γ.
33. 492: correct the units
Deleted.
34. 493-500: these sentences are very confusing and it is unclear how these affinities fit with your metagenomic data sets. The information takes away from, instead of strengthening, your study.
Deleted.
35. All Figures: Please verify that all figures meet formatting standards to be color blind friendly. In many cases the only way to tell different data sets apart is via color (e.g. Figure 5). Using different symbols and colors would make your figures accessible to a larger audience.
Some figures were changed to black and white color.

36. Bar charts are challenging to format to be colorblind friendly. So, if you cannot find an appropriate color scheme consider including the data presented as a table in the supplemental information or as another dataset at the Göttingen Research Online Database.
The excel files are uploaded in the Göttingen Research Online Database https://data.goettingen-research-online.de/dataset.xhtml?persistentId=doi:10.25625/DFFZ9R.
37. References: check all references for correct formatting, capitalization, italics, and use of consistent journal names vs. abbreviations. Also, the final reference is out of order.
Corrected.
38. Figure S3: the y-axis and caption should be changed to number of reads instead of "actual abundance" since that describes the data presented more accurately.
Corrected Figure S3.
39. Figure S3, Inset pie charts: Do these show the overall abundance for all samples presented in the figure? Are these number of reads or relative abundance?
Yes, it is overall abundance for all samples and based on number of reads. This information is added in line 35 of supplementary file.
40. Figure S5: I suggest making panels A and B larger such that each tree fits on a whole page. Since this is supplemental information its fine to take up a lot of space.
Corrected in line 41 and 42.