Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2021-165-RC1, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2021-165

Denise Akob (Referee)

Referee 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-RC1, 2021

The manuscript by Arif et al. sought to characterize microbial communities in mine drainage of the Marsburg Copper Mine. Waters in the copper mine span a gradient of heavy metal concentrations with samples collected in the copper precipitation flume having mg/L quantities of heavy metals. Microbial communities were characterized in all samples collected and these samples grouped into leachate, spring water, and unconsolidated rocks. The various samples were analyzed to look at the gradient in chemistry and distribution of microbial taxa. The authors selected a single leachate sample for metagenomic sequencing which yielded 8 MAGs. The combination of amplicon-based and metagenomic sequencing was a nice strategy as the MAGs gave the ability to assess potential metabolic function better than 16S rRNA gene sequencing alone. The authors found that members of the Actinobacteria and Chloroflexi were numerically dominant and that the MAGs contained a number of heavy metal survival mechanisms. Overall, this was a really interesting study with a good approach. However, I have a number of suggestions listed below to improve clarity. I also encourage the authors to consider additional comparison of their results to other acid mine drainage systems.

- 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.
- 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.

- 78: please provide (throughout the paper) the rational for performing metagenomics on a single sample and especially why this particular sample was selected.
- 91: I think you are missing a "respectively" here
- 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
- 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.
- 102: I would change this to genomic DNA to encompass both sequencing approaches used. (see comment 1 above)
- Figure 2: please add scale bars to all panels.
- 115-117: this sentence needs clarification to state which sample types were stored in which bottle type.
- 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).
- 135 and 149: omit "photometrically" as it is not necessary nor is it the best word choice.
- 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.
- 202: the abbreviation was already defined
- 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.
- 216-218 and elsewhere: make sure the units are written correctly with superscripts and spacing; check figures too.
- Figure 3: are the samples presented in the direction of water flow? Consider plotting the copper flume leachate heavy metal concentrations in mg/L.
- 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.
- 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
- 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.
- 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.

- 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.
- 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.
- 313: correct the reference style.
- 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.
- 349: this is a great idea to publish these maps and make the details available to the reader.
- 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.
- 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:
  - 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
  - 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
- 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.
- 427-429: the previous microbial work at the site would be a good addition to the introduction.
- 439: omit metagenomic here and use a broader term/phrase
- 474: omit "etc"
- 476: add "had" before specialized
- 477-481: this sentence is much too long and really difficult to follow. Revise
- 487: change to Deltaproteobacteria
- 492: correct the units
- 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.
- 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.

- 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.
- 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.
- 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.
- 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?
- 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.