

Biogeosciences Discuss., referee comment RC1
<https://doi.org/10.5194/bg-2021-25-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2021-25

Anonymous Referee #1

Referee comment on "Sedimentation rate and organic matter dynamics shape microbiomes across a continental margin" by Sabyasachi Bhattacharya et al.,
Biogeosciences Discuss., <https://doi.org/10.5194/bg-2021-25-RC1>, 2021

General comments

Bhattacharya and colleagues obtained a very impressive set of metagenomic and metatranscriptomic data in combination with physicochemical and isotope geochemical measurements. The authors investigate differences in the sediment communities of permanent and seasonal oxygen minimum zones and present shifts in community structure and gene transcripts with depth as well as between the sites. The sampling, the generation, processing and analysis of the data, as well as the sequence archiving was performed carefully. Yet the presentation of the results needs to be improved. In my opinion, the manuscript needs to be shortened (substantially), better structured, and jargon needs to be avoided. The interesting and relevant messages of the study are buried in details, percentages and values. The study is difficult to access in part because the manuscript lacks visualizations (e.g. of the gene transcript abundance). It is recommended to structure all manuscript sections based on compelling objectives, find a red thread that guides the reader through these amazing data, and present only those results that are important to the major findings and conclusions and remove everything else. Please find detailed suggestions below.

Specific comments:

A picture speaks a thousand words! That is also true for figures. Instead of listing all the taxa, concentrations, abundances etc, please try to visualize. It is so much easier to grasp the critical information that way. For example, the transcript data: The authors have such an incredible dataset. I am wondering why the impressive gene transcript abundance data is not shown in a dedicated figure. The trends become much clearer when all these abundances are shown along the depth profile, or across samples. For inspiration on

representation of transcriptomic data check, e.g., these papers:
<https://sfamjournals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.14806>

<https://sfamjournals.onlinelibrary.wiley.com/doi/full/10.1111/1462-2920.13895>

<https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0664-z>

L557ff: Can those methane concentrations be shown in a figure? L583ff: For me, and I think most other readers, it is very difficult to grasp all these percentages and make sense of them, compare them, interpret them. Can this be visualized as well?

Please try to avoid all terms that are very specific for a small area of science (jargon). L479-L489: This section is full of jargon: discontinuous reduction, Gaussian distribution, exponential-decay zone, probability density function, etc. Please avoid using these terms, because most of the readers will not know what they mean. Text sections become very difficult if not impossible to understand when they are full of jargon, and then the entire manuscript becomes hard to digest. In addition, when a reader gets lost in these details the overall message gets lost too. The rule of thumb here is: If for example the Gaussian distribution is really important for the study then it needs to be elaborated and explained with words, concepts and examples everybody understands. If it is not critical for the study, remove it, or move it to the SI for the experts that may be interested in the details. L490-517: This section too is very hard to digest. Is all this detail really necessary? To be honest, I got lost in the equations right away, and was unable to say what the message of the paragraph is even after reading it repeatedly (I couldn't see the proverbial forest from all the trees). Please consider a broad audience. L520-545: Same in this paragraph

In addition to an amazing dataset the authors have some very interesting findings. E.g. (L635: "coexistence and covariance of sulfate-reducers with methanogens, ANME and acetogens"). However, those too get lost in all the details. It is highly recommended to shorten and declutter the manuscript. Find out what are the most interesting findings, what is the red thread that keeps the reader's (and writers) attention. Then start by introducing compelling objectives, and then follow these objectives in the results and discussion. Everything that is not needed for the core story should be removed or moved to the SI. There is a lot of potential to shorten the manuscript, I would say it should be at least a third shorter, if not half.

The authors often include results in the discussion section. These sections should either be separated, i.e. the discussion only refers to the results, but contains very little values and numbers anymore. Or they should be combined in a Results and Discussion section. In any case, it would greatly help if the discussion was structured based on objectives that were introduced earlier. Presently the discussion is very broad, and I could not see where the readers attention/focus is drawn towards, i.e. what are the major take home messages. L615-630: This reads like a summary of the results section. In general, this is welcomed, as now the trends become clear, but it is still a results part, a discussion goes beyond that. L700ff: This contains a lot of results again. L715-742: Now this section even contains isotope geochemical results that have not yet been mentioned before. There is no mention of a context, so this entire section needs to be moved to the results section. Also, there is way too much information, too many numbers, percentages, etc. L757ff: Results

Technical Corrections:

L38: the influence of bottom water oxygen on benthic microbial communities is not entirely unknown (e.g. Jessen et al. Sci Adv 2017), so it is recommended to write "are less understood" or "are still unclear" etc.

L81: This section should include mention of the study by Jessen et al 2017 Sci Adv, as it is presenting findings on microbial communities and organic matter in sediments along an oxygen gradient in the Black Sea

L105: what is meant by hectic? Do you mean substantial?

L119: ANME only refers to the ANaerobic MEthane oxidizing archaea.

L167: Thank you for depositing the sequence data prior to submission, and also thanks for including all raw data and not just MAGs! That is best practice!

L227ff: These lists are a good place to start shortening the MS. It could be included in a supplementary table. The idea of summing up the taxa of a target guild may be clear to the reader without the need to present an exhaustive list.

L251: What about ANME-2d (Methanoperedens). It becomes increasingly apparent that this clade may be important at marine systems as well. E.g. in this very recent publication <https://www.nature.com/articles/s41396-021-00918-w>

L270: The equations for the reported Simpson and Shannon indices seem to be the standard equation used for these respective indices. Is there a reason why the authors chose to present the math in detail? Here too it may suffice to cite the respective papers and maybe elaborate on these metrics in the SI. For the reader it would be more helpful to explain the meaning of these metrics, their end members (min/max values) and how to interpret them. In my opinion it is more necessary to explain the terms, .e.g. evenness/equitability than show the equation. As an example, for me it helps to interpret a given index, if I am aware of the max/min values. E.g. that the minimum value Inverse Simpson can be equals 1 (a pure culture) and the maximum value equals the number of observed species (which is the case when all species occur at exactly the same relative proportion).

L312: From which studies did the two ANME genomes come from? Please reference. There should be quite a number of available ANME MAGs by now. Maybe these can be used as a database for your next studies.

L328: Was hydrogen added as an electron donor for hydrogenotrophic methanogens?

L356ff: A lot of the results in this section are redundant to the Sampling section in the Materials and Methods. Please shorten, or remove, at either section.

L364: What does BP stand for, please write out each abbreviation at first use.

L379: Which database and version was used for the taxonomic classification? I might have missed this in the methods, but if it is not mentioned, please include.

L387: Please refer to "relative sequence abundance" to avoid confusion with other measures of relative abundance.

L392f: Generally, it is recommended to phrase it the other way around: "...xx % of

metagenomic reads affiliated with Alphaproteobacteria...”

L393ff: Personally, I think it is not necessary to list all these taxa. It makes the manuscript very long and voluminous and disrupts the reading flow. Unless a specific taxon needs to be highlighted, it will suffice to point the reader to a figure showing the clades and their relative sequence abundances.

L480: I think microbial guilds would be a more widespread and better term than “metabolic-types”

L579: The number of reads is important, but not here. 1. Large numbers break the reading flow and 2.) including the read numbers does not add info that is necessary for the argument at this point. This section is about mcrA transcripts and so it should be focused on this.

L580ff: redundant to materials section. Please remove.

L652: Could the absence of methane not be explained also by aerobic methane oxidation? Did you check if you have aerobic methane oxidizers in the dataset (eg Methylococcales).

L659: Is there a reason why a database like GTDB (<https://www.nature.com/articles/nbt.4229?testing>) was not used? It should contain plenty of ANME genomes/MAGs.