Biogeosciences
Discussions

Open Access

EGU

Interactive comment on "Microbial communities associated with sediments and polymetallic nodules of the Peru Basin" *by* Massimiliano Molari et al.

Beth Orcutt (Referee)

borcutt@bigelow.org

Review by Beth N. Orcutt and Tim D'Angelo, Bigelow Laboratory for Ocean Sciences

This study documents the composition and relative diversity of bacterial and archaeal microbial communities inhabiting polymetallic nodules and surrounding sediment of the Peru Basin collected in 2015. The motivations for this study are to determine if polymetallic nodules have unique microbial communities, as such seabed mineral deposits may be targeted for deep-sea mining. While there have been similar prior studies of microbial community composition of polymetallic nodules, those studies focused on areas in the northern and central Pacific Ocean where organic carbon deposition rates are

lower. Thus, the new study from closer to an equatorial region with higher organic carbon export rates allows an analysis of how broader oceanographic properties impact microbial community diversity.

The first major claim of the current study is that microbial diversity is higher in the surrounding sediment than in the polymetallic nodules. This finding is different from a recent survey of available data from polymetallic nodules and sediments of the comparable Clairon Clipperton Zone, which indicated that nodules and sediments had comparable levels of diversity: https://ran-s3.s3.amazonaws.com/isa.org.jm/s3fs-public/files/documents/deep_ccz_biodiversity_synthesis_workshop_report_-_final.pdf. We encourage the authors to consider the implications of these differences between studies, and if data processing steps could be part of this difference.

Related to this part of this study, we caution that the workflow described in the methods may lead to inflated diversity metrics. The workflow described in L143-144 may allow lower quality sequence reads to pass the QC step, as most published workflows don't allow for sliding window PHRED scores of less than 28-30. For example, Dorado Outcrop basalt samples have around 1500 OTUs after filtering out low abundance/prevalence OTUs (described in Lee et al., 2016). We would expect a similar diversity on nodule samples exposed to bottom seawater but the samples described in this study have 5 - 14K OTUs per sample. Low quality reads can result in artificially large number of OTUs when using clustering-based methods. This has been documented by the developers of MOTHUR as a problem with low quality reads associated with old problematic Illumina chemistry kits. Even if there are true biological differences between Dorado Outcrop basalts and the samples in the current study that translate to different alpha diversity patterns, the presence of 525,169 singletons (as seen in Table 2) is a sign that there are likely issues with the QC steps of this workflow. We recommend that the authors revisit the sequence processing steps and consider using higher quality thresholds, and also consider using an algorithm that produces unique sequence variants (i.e. ASVs) instead of OTU clustering. Moreover, we wonder if there

is a more streamlined way to present the information included in Figure 1, or if some of this information could be moved to supplemental materials? It seems like a bit of overkill to have 10 plots essentially showing the same information.

A second major effort of this work is to identify taxa that are differentially abundant between nodules and sediments. While the text in Lines 244-261 describes these differences, and Table 4 includes the result of Aldex2 analysis, we don't find that Figure 4 visually conveys these differences in an easily digestible way and suggest using differential log abundance plots to more clearly show which taxa vary between the sample types.

Another major focus of this work is the comparison of the microbial community structures between the Peru Basin nodules and those of the CCZ. I think that the paper could be improved by providing some kind of summary graphic or schematic that visually explains the differences, and their causes, as described in the text. For example, a cartoon illustrating that the lower OC flux in the CCZ leads to nodules that look like X with communities that look like Y and perform Z functions, versus how those conditions are different at the Peru Basin. Such a summary graphic could really help simplify the presentation of the major recommendations from this work in a way that is easy to grasp, which will be especially helpful for policy makers thinking about deep-sea mining.

A question: in the methods, there is mention of collecting samples for cell abundance determination, but such data are not presented in this paper. Is it possible to include such data? This would help to evaluate if the "hot spot" idea discussed in the paper correlates to cell biomass - i.e. is lower diversity correlated to higher biomass?

A suggestion: there is some mismatch between the 3 hypotheses posed in the introduction and the three objectives posed in the discussion section. The discussion text follows the outline of the objectives, but there is not explicit "testing" of the hypotheses proposed at the beginning of the paper, and also the discussion does exactly follow the

objectives as proposed. For example, discussion section 4.3 discusses metabolisms inferred from the amplicon data, not what environmental factors structure the community, as would be assumed by how objective three is worded. We recommend bringing better alignment between the hypotheses/objectives and what the data actually address.

minor suggestions:

L1: Title could be more descriptive of what the study discovered

L15 - consider removing "need to"

L22 - Acidomicrobia, only one "i". To update throughout the manuscript.

L78-79 - need consistency in the presentation of thousands of kilometers. In one instance, there is no punctuation; in the second instance, there is punctuation.

L80 - missing a decimal point in 0.2-0.6%?

L123 - were any negative DNA extraction controls included in this study, since low biomass might have been expected? If yes, please describe.

L141 - Is there a reference that shows why these trimmomatic SLIDINGWINDOW parameters were used? They seem relaxed and would allow for sub-par quality reads to pass the QC step. Most workflows don't allow for sliding window PHRED scores of less than 28-30.

L141 - recommendation to deposit your data processing pipeline to github or similar repository.

L144 - There is a comparative "while" statement describing the differences between how bacterial and archaeal sequences were merged, but the way it is worded, it appears to describe the same order of operations.

L174 - Transforming count matrices using the center-log ratio requires a strategy for re-

placing zeros with a pseudo count because the presence of zeros produces NA values. There is no zero-replacement strategy described in this workflow. The Bray-Curtis distance cannot be computed on data matrices that contain negative numbers. A Center-log-Ratio transformed count matrix contains negative numbers. CLR transformed data is usually ordinated using the Aitchison distance metric or the Euclidean distance. I am unclear on how these analyses were performed in the way that they are described. Was the data $\log10(x+1)$ transformed? That transformation is compatible with the bray-curtis distance. The resulting ordinations looks correct, but I think the description in the methods section is inaccurate. Could the authors provide a document with the code used to perform these steps?

L237 - these percentages are for all nodules in aggregate as an average, but does not show the variation between samples. I recommend including standard deviation plus/minus for each percentage.

L338 - could the differences in relative percentages of archaea between this study and prior studies be due to difference in DNA extraction, primers used, or sequencing approach?

L359 - suggestion to add a clause to the end of the sentence regarding nodules and sediments have distinct communities, stating that this observation is consistent with what has been found in earlier studies, and cite a few examples.

L413 - "reductive"