Review of 'High-throughput screening of sediment bacterial communities from Oxygen Minimum Zones of the northern Indian Ocean' submitted to Biogeosciences, by Jovitha Lincy V.J.* and Cathrine S. Manohar

General:

The manuscript by Lincy and Manohar presents a study comparing one amplicon sequence dataset from the Arabian sea sediments to another amplicon dataset from the Bay of Bengal sediments. The authors performed a phylogenetic analysis and a diversity study using some of the classical indices used for community diversity studies.

The manuscript is very hard to read for two reasons, the first of which is the quality of the language (see specific comments), the second of which, and this is the more severe point, it lacks a clear storyline. The authors base their analysis on two samples and come up with various claims that could possibly also just be a sequencing artifact given the difference in sequencing quality and depth of the two (!) sampling points. First, they claim a significant difference between the samples which is statistically impossible to claim, second, they claim a variability within the Indian ocean, again based on two samples, third, the diversity indices are not reliable because of the difference in sequencing quality and finally, based on this, the functional diversity model doesn't give any reliable result. Another point is that the authors constantly compare their sediment data to water column metagenomes, which just doesn't make any sense and is very confusing.

I would suggest a major rewriting of the manuscript with a clearer focus: It could be very interesting to see those sediment amplicon datasets compared to other OMZ sediment data and then to obtain information on a core benthic OMZ community versus a 'flexible' OMZ community. This, rather than a PICRUST model, could then be used to explain the differences in benthic biogeochemistry and benthic-pelagic fluxes of different OMZ regions. I would further suggest to re-run the amplicon BLAST analysis on SILVA instead of Greengenes, SILVA is the only database constantly updated which will certainly give a way more informative result.

Specific comments

L 31: I would rephrase by *high throughput sequencing', because it's not one technique

L. 34 replace 'from' with 'in'

l. 35 ‘indicated’, ‘the bacterial diversity’

L. 37 I disagree with this statement

L. 38 not the site but the community is diverse, please rephrase

L. 39 it’s diverse compared to off Goa, or it’s unexplored as compared to off Goa?

l. 41 ‘at both sites’, also the presence of genes is not necessarily related to their activity, so I feel this statement is too far-fetched.

l. 58 ‘consists’ or ‘is comprised’

l. 64 This needs some re-thinking. There is a body of work now showing the plasticity of processes with overlaps between sulfur and nitrogen turnover

l. 65 Are these explanations relevant for the manuscript?

L. 67 what does this tell us in the context of the AS versus the BoB?

l. 69 this is somewhat in contrast to water column studies from OMZs showing high abundances of aerobic ammonia oxidizers. Bristow et al showed for the BoB aerobic organisms existing in the OMZ

l. 73 ‘hence it is interesting’

l. 74 This is the first time you are talking about benthic ecosystems. I think this should come straight away to better guide into the manuscript’s topic

l. 74-77 The sentences sound awkward, please rephrase.

l. 78 replace ‘a database’ with ‘data’

l. 79 ‘helped’

l. 82 explain

l. 83 ‘ the Northern Indian Ocean’

l. 84 ‘techniques’, replace ‘restricted’ by ‘been limited’, replace ‘pelagic OMZs’ by ‘the pelagic realm’ or ‘the water column’

L 85/86/87 Please rephrase, sounds awkward

L 89 here and elsewhere, please remove ‘technique’. It is not a technique unless you refer to pyrosequencing, but even then you wouldn’t say it like this.

l. 90 replace ‘long-read 454 pyro-sequencing technology’ with ‘pyrosequencing’. Also ‘the bacterial community composition’

l. 90 ‘the 16S ribosomal RNA’

l. 91 ‘in a perennial OMZ location’

l. 92 replace ‘utilizing’ by ’using the’. This is not what PICRUS does. You can predict functional diversity, you can not say anything about genetic distributions

l. 94 comparable to what?

l. 95 This sentence doesn’t have any informative value.

l. 101/ 102 replace ‘region of’ by ‘in the’. What is GS1A, what is PS1B?

l. 103 replace ‘and corresponding water column depth’ by ‘in waters with a depth of’. ‘A box corer’. What does undisturbed sediment sample mean?

l. 104 ‘liners’

l. 106 ‘containers’, replace ‘aseptically’ with ‘sterile’

l. 109 please rephrase, sounds awkward

l.113 ‘a CN analyzer’, also explain C and N; ‘described in’, Which standards did you use, how many blanks? Precision and detection limit?

l. 114 What means estimated here?

l. 117 reference?

l. 119 This needs more explanation.

l. 123 did you do blank extractions? ‘A Nanodrop…’

l. 129 company’s location? Replace ‘are’ with ‘were’

l. 133 sequencing chemistry and sequencer type?

l. 137 so you used the pre-analysis provided by the company?

l. 141 ‘the EzBioCloud 16S…’

l. 146/147 How, which commands?

L. 150 ‘The PICRUS…’

l. 154 Awkward sentence. Please replace. How did you decide on what genes are significant?

l. 161 On what do you base this prediction?, The OM was highest- what does that mean, could there be a number?

l. 162 The TOC/TN ratio was slightly high for PS1B sample- what does that mean?

How is the term ‘significantly’ explainable?

l. 163 remove comma after ‘water’

l. 168 This technical information could go into the methods part, where you should also describe the output of a rarefaction analysis.

l. 170 Explain ACE

l. 171 You have to convince with numbers and comparable numbers from other studies if you want to make such a statement.

l. 173 A rarefaction curve doesn’t give you information on the total diversity but on the sequencing saturation. As such you can only determine if your samples are quality-wise useable and if they can be compared. In your case you did not sequence in saturation and the saturation status between your two samples differs quite a bit. Thus it is questionable to compare the diversity.

I recommend a re-writing of l. 168-182. There is a lot of repetition and information that is not particularly useful the way it is presented.

l. 185 What does ‘slightly different’ refer to? To the diversity indices? This wouldn’t be a surprise as those are two thing which you can’t compare.

L. 187 I don’t think Greengenes is reliable anymore. There are huge discrepancies to the SILVA database. I would recommend a re-mapping to SILVA.

l. 188 What means ‘specific to both sites’? Present at both sites?

l. 192 I distrust both databases, as mentioned I suggest re-analyzing using SILVA

l. 193 Which literature?

l. 197 ‘contribute’, ‘the total bacterial…’

l. 202 It may result from the time of the last update of the databases

l. 203 ‘refer to’

l. 204-208 I don’t understand what you are trying to say with this sentence.

l. 210 ‘at the two sites’

l. 211 How does this look compared to other OMZs?

l. 219 ‘orders’

l. 220 Bacteria, which…

l. 221 ‘Planococcaceae_uc’ and similar expressions. Those are database output names, please directly replace with the proper taxonomy. What does identical mean here?

l. 225 This doesn’t emphasize spatial variability. It is one spot per ocean basin which has been sampled, here. To see variability two samples from very different regions are not enough.

l. 220 -232 This is a pure listing, which would benefit from some context to guide the reader to what should be learned, here.

l. 236 Which genome?

l. 237 Above you claimed a high percentage of known phyla.

’Firmicutes were’

l. 238 number are commonly spelled out up to twelve

The statement about carbon fixation is neither clear nor well-founded, please rephrase or remove.

l. 240 at the GS1A site

l. 241 remove ‘the’

l. 242 replace ‘too’ with ’to’

l. 243 xenobiotic compound degradation pathways: What is that?

l. 244 This sentence doesn’t add any information and is not understandable.

l. 250 What do the molarities refer to?

l. 256 replace ‘its’ by ‘it is’. ‘Also, the BB…the AS…’ (also in L. 266)

l. 258 What does ‘as well’ refer to?

l. 261 What does early nitrogen burial mean? Rapid burial?

l. 270 Remove this sentence. It doesn’t add any information.

l. 272 Pyrosequencing is not more useful, it has a lower coverage and depending on the pipeline way longer OTUs can be assembled from Illumina sequencing. In addition, V1-V3 is not a long fragment so this is no good point here and does anyway not contribute any information to the data discussion.

l. 273, 274 This statement is based on one publication which I don’t agree with and which ignores a significant body of work saying exactly the opposite. In any case, you have been working on bacterial data which do only represent a subset of the total diversity so more cautiousness would be recommended, here.

l. 280 what is now the H Index? Where does this come from?

This section contains only information which has been presented, before, and some evaluation of sequencing methods, which to me appear unnecessary and are partly incorrect.

l. 297 ecosystems

l. 298, replace ‘utilizing’ with ‘using’, again this is not a technique

l. 303 replace ‘alter’ with ‘determine’

l. 305 First of all, what are those phyla? Second, how does their presence in the Pacific mean anything about their metabolic pathways?

l. 308 13 phyla are not a vast diversity, but rather a very small diversity

l. 310 Remove, this doesn’t add anything.

l. 312-317 Why is this reported if the clades are meaningless in the present dataset?

l. 317 it’s not the 454 sequencing but the selected regions that are possibly not well representative, please check the primer bias if you want to make such a statement

l. 320 Which ones?

l. 325 What would ultra-low be?

l. 330 This would now give you a good chance to compare your data to data from the Eastern tropical Pacific and eastern tropical Atlantic OMZs, the Baltic and other environments (Gier et al, Bertics et al, work by Fullweiler and colleagues, by Jørgensen, Treude etc)

l 348 ‘anammox bacteria’, also introduce what this is

l. 350 ‘are known’

l. 354 explain what those genes do

l. 372 ‘the Black Sea’

l. 380 do you think they are active or are they detritus?

l. 382 ‘Chroococcales are’, ‘a low-light…’

l. 403 this is again a far-fetched statement

l. 406 I disagree. This is pure speculation.

Also the Conclusion part needs major re-writing