

***Interactive comment on “Differential analysis of prokaryotic communities from pristine mangrove tidal zone sediments reveal distinct structures and functional profiles” by Carolina Oliveira de Santana et al.***

**Anonymous Referee #2**

Received and published: 20 October 2020

General Comments: Clearly a lot of work went into this study, however there is still much work that needs to be done with the paper. Currently the paper reads like a draft that still needs several more rounds of circulation between authors. The introduction and discussion sections need better organization/flow with more specific, relevant background/literature pertaining to sediment microbes and mangroves, and importance of tidal zone (how that might influence microbial communities and mangroves). There are numerous sentences within the results section that belong under either the methods or discussion sections. The discussion lacks focus and synthe-

C1

sis. Overall, this paper needs a significant overhaul. It seems like it would be useful for the authors to clarify specific objectives, if not for the paper, for themselves, to achieve better focus and clarity in conveying this study and its findings. Another major consideration is that the authors should be very careful about what and how they convey findings and conclusions on metabolic pathways and functionality of microbes when only using 16S rRNA amplicon data. Particularly, they should be weary and cautious about using a tool like PICRUST2 to make any major conclusions with respect to microbial functionality. Personally, I think that if you are going to use PICRUST2 as a tool here you should be backing up as much of those findings with literature as possible. For example, you could compare your findings with mangrove metagenomics studies such as those done by Andreote et. al. 2012. Also, it is important to very explicitly state the pitfalls of PICRUST2 analyses (see Sun et. al. 2020, <https://link.springer.com/content/pdf/10.1186/s40168-020-00815-y.pdf>) and that these are just inferential findings and would need to be confirmed via additional analyses such as transcriptomics or experimental setups.

Specific Comments:

Abstract

Line 21: The term metagenomic at this point is used solely to describe shotgun or whole metagenomic sequencing, not amplicon sequencing.

Line 27: You say “significantly different prokaryotic communities but in Line

Line 31: Change metagenomics to “amplicon” or “16S rRNA”. I don’t think you should include “function in the keywords, as functionality is solely inferred indirectly via amplicon sequencing and Picrust2 analysis. Choose either “Mangrove” or “Pristine Mangrove Forest”.

Introduction

Line 37: Be more specific when you say “large portion”, how much do they constitute?

C2

Line 42: Instead of talking about studies which look at microbial communities and plant development, include specific background on microbes and mangroves that is relevant to your study.

Line 45: You haven't really established "dependency" of mangrove forest on sediment microbiome at this point. You can expand on how they can be considered dependent or remove this type of wording.

Line 47: I'm not sure what you mean by "single type of sediment." Since you don't discuss types of sediment or measure sediment characteristics such as grain size or grain type (silt, sand, etc.) in this paper you should not mention sediment type. I think you might mean that due to the fact that mangroves are highly influenced by tidal flow, which results in variations of "environmental conditions across small spatiotemporal scales."

Line 60: What is the "original area"? Do you have specific information on this?

Line 71: Not sure what you mean by "terrestrial processes." Do you mean biogeochemical processes?

Line 74: I would say something more like "understand the differences between microhabitats within mangrove systems" instead.

Line 75: By "mangrove regions" do you mean different tidal zones? If not, you might want to briefly explain what the different regions of mangroves are.

Line 78: Use something like "We identified taxa which may be driving different nutrient cycles between zones." I should note that you may want to rethink this sentence altogether as you don't really show that there are different nutrient regimes/cycles between different zones as of now. You could include literature that suggests this or data of your own to support it.

Lines 81 -85: You could circle back to this in the discussion, specifically you could theorize what changes in community structure you might expect based on your findings

C3

and the literature in contaminated mangrove sediments.

#### Methods

Line 100: What is meant by tidal influence? It would be good to include demarcations for this, i.e. distances, vegetation, etc. Based on Fig.1 it seems like you may have used sediment water content, or time of water coverage.

Line 104: When you write "disruption of rhizospheres" do you mean "contamination" of rhizospheres associated with vegetation, because you aren't wanting to include those communities?

Line 107: Were vegetation densities measured, if so, what was the metric?

Line 111: How was organic matter content measured?

Line 127: I don't think you need the "ILLUMINACLIP" section, especially if you already have your code published on Github. Additionally, you explain your code in text immediately following.

Line 131: You mention just QIIME, and QIIME2 in the following steps. QIIME is no longer supported or kept up, so if you used the original QIIME I would recommend using QIIME2 for that step.

Line 132 – 141: DADA2 calls ASVs and not OTUs. I have only superficially used QIIME2, as I typically use mothur or DADA2 directly in R, but my understanding is that QIIME2 and DADA2 primarily call ASVs, but gives the option to then cluster those ASVs after they have been called. If this is what you did, you should explain that process briefly. In reading on (Line 138), it would seem that there is either a miscommunication or misunderstanding of the bioinformatic steps with respect to clustering and taxonomic assignment. "Open reference" refers to a clustering method which can be done using Vsearch in QIIME2. It looks like when assigning taxonomy after OTU clustering, Q2 gives you 3 different options, I am thinking you used classify-consensus-vsearch? I would also combine sections 2.4.2 and

C4

2.4.3 in to one section where all bioinformatics workflows in QIIME 2 are discussed together. For reference: <https://docs.qiime2.org/2020.8/tutorials/otu-clustering/> and <https://docs.qiime2.org/2020.8/tutorials/overview/>

Line 149 – 151: What dissimilarity metric did you use for metaMDS, i.e., jaccard, bray-curtis etc.? I am wondering why you didn't run something like Canonical Analysis of Principle Coordinates (capscale in Vegan) to investigate correlations between environmental variables with community structure.

Line 153: Did you take the limitations in the link provided below into consideration when running these analyses?

PiCRUST2 Limitations Link: [http://picrust.github.io/picrust/tutorials/quality\\_control.html](http://picrust.github.io/picrust/tutorials/quality_control.html)

Line 159: Explicitly explain how you conducted the taxa enrichment analysis. I could not find the reference paper for Spealman et. al. 2020.

## Results

Line 186: Do you mean uncultured "prokaryote" not "eukaryote"? I don't believe there is any eukaryotic designations in the 16S SILVA taxonomy reference. You should also qualify further why you felt comfortable assigning an archaeal taxon to the uncultured "eukaryote".

Line 196: Figure 1B colors are difficult to differentiate. Consider adding a pattern to colors which are too close to differentiate. For figure 1B, consider changing the y-axis range to 32 so that we can see more of the other bars.

Line 201, 204, 205: QIIME2 sentences belong in methods

Line 207: P-value for Bray-Curtis is not significant

Line 211: Why did you use both distance metrics, i.e., Jaccard and Bray-Curtis? You should choose the one most appropriate to your data and study. Did you try hierarchical clustering to see how data might cluster without apriori considerations like tidal zone?

C5

Line 222: What is your organic matter (OM) metric, is it total OM? I don't think this is the best/clearest way to analyze these data with environmental variables. See methods comment Line 145-151.

Line 227-229: These sentences belong in the methods section

Line 231: What is "elevation" in this context? Do you mean zonation?

Line 236: What is an "icon" in this context?

Line 237: What specifically about "carbon metabolism"? As all microbes need a carbon source, this should be explained in a bit more detail.

Line 247: In figure 5 it would be good to use different colors for differentiating bacteria and archaea as you are already using green and blue in the figure legend.

Line 252: This sentence belongs in Methods

Lines 255 & 256: don't need the "above both" phrase, it is confusing.

Line 261/262: "Taken together, KO enrichment reinforces the previously observed trend of reduced abundance in the Intertidal site, and greatest abundance at the Sublittoral zone." This sentence seems like it should be in the discussion, especially if the "previous trend" you are referring to is one from the literature. Also, be consistent on whether you are capitalizing the tidal zones or not. I think it is more correct not to capitalize.

Line 268-276: I may have missed the results in this paragraph, but it seems like all of this belongs in the methods section as it is describing how something was done versus reporting the findings of what was done.

Line 270: Where is here? Is it this study or a figure?

Lines 277 – 327: I would combine all of the "cycling" sections under one section called "biogeochemical cycling" or something like this. You have several sentences throughout this section that would be more appropriate in the discussion section. Essentially

C6

any of the sentences with citations should probably be in the discussion. Examples: Lines 287, 291-293, etc.

Line 328: I appreciate the effort that went into this figure. Personally, I would like to see this figure with less taxa, only ones you specifically mention within the text, so that it is less busy. I am not sure why the # of nodes legend needs to be a tapered triangle. It makes me think I should be considering both the color and thickness of lines when I'm looking at nodes. I also think you should use a more differentiating way to denote taxa with an associated metabolic role in the literature and taxa with KO greater than 10%. You could use black and white circles, and add a gray circle for those with both if they exist.

#### Discussion

I decided to make an overarching comment here, instead of going line by line, because the discussion needs a lot of work and restructuring. I think one of the best ways to go about fixing the discussion will be to come up with a thesis statement for each paragraph and figure out what points you are trying to convey. This will help you to clarify and re-organize your thoughts. I would like to see in the discussion more synthesis of your findings, such as why you think you might find certain taxa with potential metabolic capabilities enriched in certain tidal zones. It seems to me that you set out to study the sediment microbiome of pristine mangrove environment across 3 tidal zones to serve as a baseline and to characterize differences in taxa and potential biogeochemical cycling that is predominant in those zones. However, neither your introduction or discussion provide enough clear, relevant background or support for your overarching goal. I would have also liked to see some discussion on anaerobic taxa, and where you find more anaerobic taxa with respect to tidal zone. You use the term microhabitat in both the intro and discussion, but it is unclear what this means in the context of your study. You should clearly define what your usage of microhabitat means. Are you talking about microbial habitat, are you talking about spatial scales, millimeters – meters?

C7

#### Technical Corrections:

Line 57: Latter not “later”

Line 62: I am assuming that the A in APA is referring to Ambiente, but just want to point out you use the English “Environmental” just before APA, so I'm not sure if you should write EPA or use the word Ambiente.

Line 117: Use protocol instead of “program”

Line 154: All KEGGs should be capitalized.

Line 193: Genera instead of “genus”.

Line 101: I don't know if superficial is the correct word, I typically see the use of “surficial”.

---

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2020-181>, 2020.

C8