Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-529-RC1, 2019

Biogeosciences Discussions

Interactive comment on "Diversity, distribution and nitrogen use strategies of bacteria in the South China Sea basin" *by* Yuan-Yuan Li et al.

Anonymous Referee #1

Received and published: 3 April 2019

The manuscript presents data from amplicon sequencing of a subunit of 16S rDNA and the nifH gene, the functional molecular marker gene for dinitrogen (N2) fixation. Further, a set of quantitative molecular data is presented on genes involved in nitrogen turnover and (surprisingly firstly mentioned in detail in the discussion) of iron acquisition. A statistical analysis was applied to explore the parameters to which the microbial community responds. The paper, as presented, provides a mere description of a lot of more or less expected facts, in line with previous studies from the South China Sea (SCS). Besides the fact that references are missing for various claims in the introduction and discussion parts, a clear storyline is missing. The manuscript will need major rewriting and focusing on one topic in order to make its significance visible. I would recommend to focus on either the community dynamics in the SCS, or on nitrogen re-

generation in those waters, to apply a more thorough description of the methods and a true discussion instead of a repetition of other studies' results. In this context, I have very strong doubts on the validity of the nifH dataset because the primers used in this study have been described to have an extreme bias in Ocean environments. I am not convinced that this dataset is representative. Similarly, the methods for the 16S rDNA amplicon analysis make no sense the way they are described. This needs substantial clarification. The conclusion does not provide additional insights, so also here re-writing is necessary. In addition to this, some aspects of the writing come across sloppy, including gene names, which are commonly written in italics, the use of the term N fixation, while N2 fixation is meant, or the awkward differentiation between bacteria and diazotrophs. Diazotrophs are mostly bacteria, besides bacteria only methanogenic archaea possess the genes for N2 fixation. A language editing is needed.

Below I have some more specific suggestions for the different parts of the manuscript:

Abstract

The expression 'layers' may appear intuitive, however, it took me a while to figure out if we are talking about the water column or sediment.

There is no line of reasoning or storyline visible, just a collection of facts. The last sentence is out of place and it is unclear on what this claim is based.

Introduction

l. 59 sentence says nothing

l. 61: not only for bacteria, but for life in general. Also, the sentence sounds awkward. References are missing

l. 64 I don't understand the sentence

l. 66 N2 fixation is not an adaptation but an evolutionary development

l. 73 How does this show up?

l. 73 What do you mean by metagenomics, here?

l. 74 revealed; generally please check the use of the correct tenses.

l. 87 coastal

l. 88 failed; key or new- what do you mean with this? Using metagenomics without single cell rates there is no way to do this. Also references are missing here.

l. 101 PICRUSt- please explain

Methods l. 111 Why do you use was? Has it been moved?

l. 112, 113 seawater from. . .

l. 114 was instead of were

l. 116 Do you want to say you collected the cells on the 0.2um filter?

l. 117 You stored on board until analysis?

l. 119 Which sensor did you use?

l. 121 an instead of the

l. 122 Germany was almost certainly not the company you bought the analyzer from, please give the correct name, what does this reference indicate here, this doesn't seem to be correct.

l. 124 this is a GF/F filter, and not a membrane. Please also provide the extraction method and the calibration details, including the type of standard you used.

l. 127 So you are saying you put the filters directly in the tubes in the Fast DNA spin kit? This can't work. Please describe how you extracted from the filters.

l. 128 replicates, not repeats

l. 129 reagent instead of regent; please give details on the Trizol-chloroform extraction.

This is insufficient to reproduce it. How was the leftover DNA removed, and how was the purity checked?

l. 138 f Here, we are running into a problem. Gaby et al. 2012 showed that this primer set is covering only 5% of nifH diversity in the sea. So, this makes this part of the dataset actually non-credible. You are coming back to this in l 382, where you show inconsistencies with other datasets, and you are discussing the potential of contamination- this could have been outruled by non-template controls. If you did them please provide the results. If not the dataset has almost no credibility.

Gaby JC, Buckley DH (2012) A Comprehensive Evaluation of PCR Primers to Amplify the nifH Gene of Nitrogenase. PLoS ONE 7(7): e42149. https://doi.org/10.1371/journal.pone.0042149

l. 144 replace the by a, provide details on the kit and protocol for the library construction, plus on the length of the reads

l. 146 With which program, how?

l. 157 ff: Gene names start with small letters

l. 159 Provide a reference, details on the quality checks. What do you mean with control genes? Is this to quantify against? If so I am not convinced of the quantification either. A quantification of a Prochlorococcs gene against a general cyanobacterial 16S amplicon does not make sense to me.

l. 170 remove 'levels of'

l. 172 What do you mean by holes?

l. 181 this needs a reference

l. 191 ff How exactly do you do a functional prediction based on 16S data, and somehow on the metagenome? Then again you use the 16S-only Greengenes database. Then you transfer to EggNOG, which is a functional annotation pipeline- what did you

transfer. This is completely unclear to me. Also, it is necessary to put references here. Why did you not just Kegg-map directly from the metagenomes? This, to me, would be more credible.

l. 197 Which bioinformatics analysis?

l. 200 Hydrochemical?

l. 201 What is an upper mixed layer, could that be substantiated ?

l. 212 What was the aim of this characterization?

l. 213 I would like to see a table with reads per sample.

l. 215 How do you know about the coverage?

l. 227 and throughout the results If you use bacteria in plural then use were instead of was

From l. 227 to l. 259: this is a listing of numbers, that are in parts contradictory. It is a bit unclear why you are reporting on all the groups without saying much. Please streamline this.

l. 262 What are qualified reads?

l. 263 this means a 99% identity?

l. 265 How was the richness determined?

l. 272 to instead of with

l. 288 they didn't emerge, they were detected. Also it's not a subset but a clade.

l. 289 this is not remarkable but rather expected, too.

l. 304 I frankly don't believe that nifH is the second most abundant gene anywhere.

l. 322 ff what you are saying here is that all parameters you included were significantly

correlated to the community structure. I am not quite sure how this informs us.

l. 388 How to infer a proportion here?

l. 395 This is most likely because only one group of euryarchaeota is able to fix N2- however, they have never been shown to actively do so in the water column. Thus the statement in line 401 is way too speculative.

l. 424 How do you know this? The statistics show you a correlation to the selection of parameters you used in the test. This needs more of a discussion. Also it is somewhat contradicting your conclusion

---

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-529, 2019.