Biogeosciences Discuss., 12, C10085–C10094, 2016
www.biogeosciences-discuss.net/12/C10085/2016/

Biogeosciences Discussions

Interactive comment on "Succession within the prokaryotic communities during the VAHINE mesocosms experiment in the New Caledonia lagoon" *by* U. Pfreundt et al.

U. Pfreundt et al.

wolfgang.hess@biologie.uni-freiburg.de

Received and published: 3 March 2016

1) In this study, Pfreundt et al. are presenting the bacterial community composition analysis results from the VAHINE mesocosm experiment. The VAHINE experiment deals with constructing mesocosms in the oligotrophic South Pacific, and fertilizing them with phosphorus. The manuscript details the changes in bacterial community in the mesocosm and the ambient waters (the lagoon samples) during the duration of this experiment. The results are presented from the perspective of phosphorus starvation, and whether this has an effect on the bacterial community. The paper is overall well written, although the language could use a bit of polishing here and there. Shortening

some sentences for clarity may help readers understand the study better.

Answer: Thank you for the overall positive and very helpful comments. We have addressed all concerns during revision and have made the presentation more concise at several places.

2) My biggest complaint about this paper is its constant referencing to the sister paper from the same authors "Global analysis of gene expression dynamics within the marine microbial community during the VAHINE meso- cosm experiment in the South West Pacific". I understand that the authors may want to publish the results separately. However, since this paper heavily relies on the transcript data for the discussion part, I'm asking myself why not make one big paper, and save the reader from going back and forth between two papers.

Answer: Yes, the companion paper on the metatranscriptome (Pfreundt et al., 2016) is very informative and both are important for understanding the processes within the investigated microbial communities. Unfortunately, it cannot be joined with the current one due to the extent of information. The companion paper was submitted one full month before this manuscript, but unfortunately it was dealt with quite slowly. However, it is publicly available now accessible as a Discussion Paper in Biogeosciences under doi:10.5194/bg-2015-564.

My specific comments to different sections are listed below. ==Abstract and Introduction== 3) Concise and to the point

Answer: Thank you!

==Materials and Methods== 4) I understand that in total, three mesocosms were deployed, but only one (M1) one was studied for bacterial community composition. Of course I understand the issue of cost, but it seems like a chance missed to have some replicated.

Answer: This is true, and sequencing another 20 samples from a second mesocosm

would have indeed been advantageous. The fact that we have a continuous time-series, and 37 samples in total, however, alleviates the lack of replicates to a certain extent, because it allowed us to do correlation studies and multivariate statistics to infer differences between M1 and the lagoon, as well as revealing dependencies of microbial groups to certain environmental variables.

4) Were there any other negative controls, other than the lagoon sampling? A mesocosm without DIP addition?

Answer: A non-fertilized mesocosm was not planned for this experiment, but would have been advantageous also for the other studies conducted during this experiment. With the data we have, we thus tried to tease apart the effects of mesocosm confinement from those of DIP fertilization through multivariate statistical methods. This is possible because DIP fertilization did not take place until the evening of day 4.

5) Did the authors account for different 16S copies while calculating the pseudo-absolute cell numbers?

Answer: No, as stated on page 20191, line 26 in the MS "We assumed equal 16S gene copy numbers." (no absolute assumption was made). While we know that this is not true, and that, for example, copiotrophs mostly have more copies than, for examples Synechococcus, it is impossible to use "true" numbers here, because our abundant OTUs are often not represented by a finished genome. However, we propose to substantially refine this calculation in the revision by using true 16S copy numbers as given in the Integrated Microbial Genome (IMG) database for the putative close relatives of our abundant OTUs. This will roughly halve the absolute numbers of, for example, Rhodobacteraceae and Alteromonadaceae, compared to Synechococcus.

6) How were the pearson correlations calculated, and why was pearson correlation selected specifically? Is the data normally distributed? Has the significance of these correlations been tested?

Answer: The referee probably refers to Suppl. Fig S2. We calculated Pearson correlations using the cor() function in R on the different taxonomic groups in M1 and the lagoon. This was added to the Methods section. We applied the Shapiro-Wilks normality test to the distributions of all taxonomic groups separately and found that 22 out of 60 distributions significantly deviated from normality. The trends tend to be the same whether Pearson or Spearman rank correlation is used. We attached the same figure as our original Fig. S2 to this answer (attached Fig. 1), but with Spearman instead of Pearson correlation coefficients. We only use these correlations to describe trends in the temporal dynamics of the displayed taxon groups, i.e. to investigate which groups behaved similarly in M1 and the lagoon, which is why we used Pearson, but we are also happy to display Spearman coefficients instead. We did not compute significance values for the correlations initially, because it is visible from the heatmap plots in Fig. S2 that the strong correlations (showing strong relationships between taxonomic groups in M1 and the lagoon) are rare and thus likely significant. Nevertheless, we now calculated all p-values for these correlations and added a Figure to this answer (Fig. 2) showing the same heatmap plots as in Fig. S2, but with p-values of the correlation instead of the correlation coefficient. As the referee will see by comparing Fig. S2 with this new Figure, all but 2 strong correlations (>0.6 or >-0.6) have a p-value <0.05, and are thus significant.

7) The authors make plenty use of the correlation values in the results and discussion. Another interesting analysis that the authors might consider adding here would be tests of significant associations between the taxa and groups of sites (lagoon vs. mesocosm, or periods in the mesocosm). These tests are implemented in R package indicspecies.

Answer: Thank you for this suggestion, we did a similar analysis, but not with this R package. The multivariate statistical analyses including significance tests (PERMANOVA) of the effects of the site, time, and depth on OTU distribution is presented in section 3.2.4 / Fig. 5 of the Discussion paper. Additionally, we did a constrained ordination to infer associations with environmental variables (section 3.2.5 / Fig.6), and

believe that together, these analyses are sufficient for this paper.

==Results and Discussion== 8) Figures are really hard to read, I think this is also partly because of the pdf format from BGD, but it would greatly help if the authors could increase symbol sizes in figures.

Answer: Sure, we will increase symbol sizes.

9) If I understand correctly, PP, APA, Chla and BP were measured for all mesocosms. If so, it would be good to see the data from M2 and M3 as well.

Answer: This data is provided in several companion (and cited) manuscripts (Van Wambeke et al., 2015, Berthelot et al, 2015). In this MS, we only included M1 and lagoon data to make our data (which is also M1 and lagoon) directly comparable.

10) The comparison and usage of both pseudo-abundance and relative abundance is confusing, especially since the figures display rather different trends.

Answer: We will constrain results and discussion to the absolute data in the revised MS, but still supply the relative figures in the Supplement, because the relative data is the original data and the absolute numbers are derived.

11) I suspect that the Deferribacteres is just mostly SAR406 clade, please indicate this in the text and figure captions

Answer: Indeed, we checked this and it is almost entirely SAR406. As suggested, we added this information to the MS.

11) In figures 4 and 6, it would be good to mark DIP addition as well.

Answer: We do not understand, why the referee suggests this for Fig. 4 (relative abundances of taxa in the lagoon) and Fig. 6 (canonical correspondence analysis). We will, however, indicate the DIP fertilization for Fig. 3 (which will become Fig. S3), and in the figures showing the absolute taxon abundances in M1.

12) UCYN-C and A should also marked in the figure captions, the manuscript text and figure captions do not match

Answer: Done.

13) I fail to see this suggested correlation between SAR11 and SAR86 in the data presented by the authors. I also fail to understand why they should be correlated – they occupy a similar niche, but they have different nutrient preferences, but I don't really see how that would lead to a correlation between the two groups.

Answer: Please refer to our comprehensive answer to point 14 by referee Danny Ionescu. Briefly, we do indeed see very strong correlation (r=0.8-0.9) between the 2 groups when comparing total transcript abundance (Pfreundt et al, 2016) over the full three weeks. It is not known why, but this has also been observed over a diel cycle (Aylward et al, 2015). We did not specifically test 16S abundance correlation between the two groups.

14) There is a big increase in relative abundance of SAR11 clade in the lagoon sample on day16, do the authors have a suggestion as to why that might be?

Answer: No, there is no indication in the transcriptional data that might suggest this strong increase in cell number. A similar increase is seen in M1 on day 8.

15) Figure 6 is an incredibly busy figure, and it is really hard to find anything in it. The authors should consider splitting the CCA biplot into two complementary figures showing sites and species separately.

Answer: This is a good suggestion. Thank you. We will create panels A and B, that show sites and OTUs separately, and increase the text size.

16) Defluviicoccus issue is interesting, although it's not surprising a bacterium associated with biological phosphorus removal systems would respond to DIP input. The authors are right that the "genus" is rather broadly defined. Looking at the sequences of this genus, I noticed that there is "marine" and "wastewater" of Defluviicoccus group.

It would be interesting if the authors could show that their 16S amplicons associated with Defluviicoccus are indeed falling into this marine branch.

Answer: This is a good suggestion. We will revise our statement accordingly and shorten this section, also in in agreement with the comment from reviewer 1. We used the SINA aligner and built a phylogenetic tree with Mr. Bayes for placement of our Defluviicoccus OTUs. Indeed, as seen in Fig. 3 attached to this answer, our OTUs fall into a tight cluster with 16S sequences from marine samples.

17) I searched the Bioproject database with the given accession number, which returned no results. I persisted and searched for the manuscript title, authors and other things, but still got no results. Please make sure your sequences are available to public.

Answer: This issue has been resolved with NCBI, and the data is now accessible under the given Bioproject ID PRJNA304389.

Interactive comment on Biogeosciences Discuss., 12, 20179, 2015.

Response Letter Figure 1. For all investigated phylogenetic groups, the relative abundances in M1 and the Noumea lagoon were correlated, this was presented in Fig S2 of our MS. Complementary to this, we here present Spearman correlations instead of Pearson correlations to show that the two methods exhibit the same trends and thus do not strongly influence our interpretations.

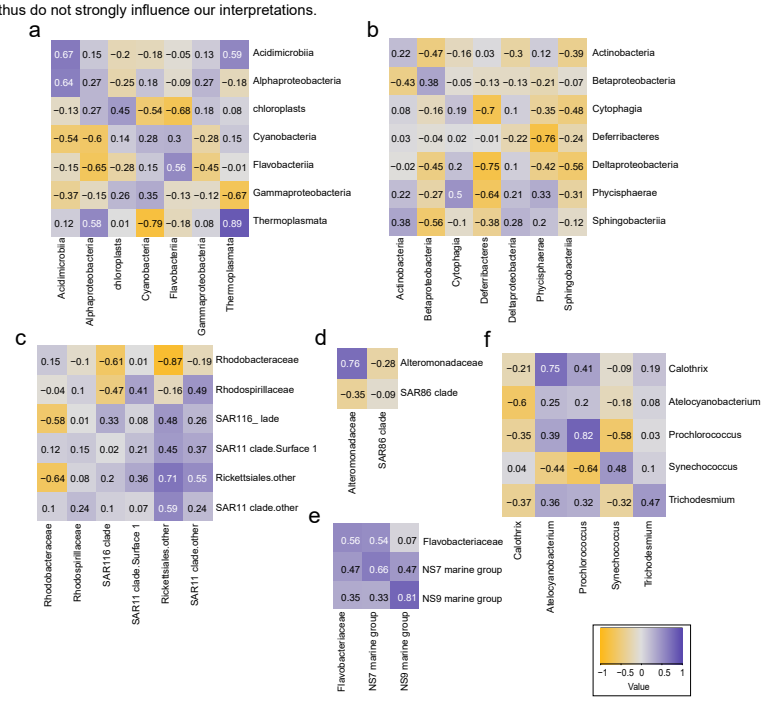

Fig. 1.

Response Letter Figure 2. For all investigated phylogenetic groups, the relative abundances in M1 and the Noumea lagoon were correlated, this was presented in Fig S2 of our MS. Complementary to this, we here present the p-values for the Pearson correlations presented in Fig S2 to show that these are indeed <0.05 for all strong correlations (> 0.6, < -0.6).

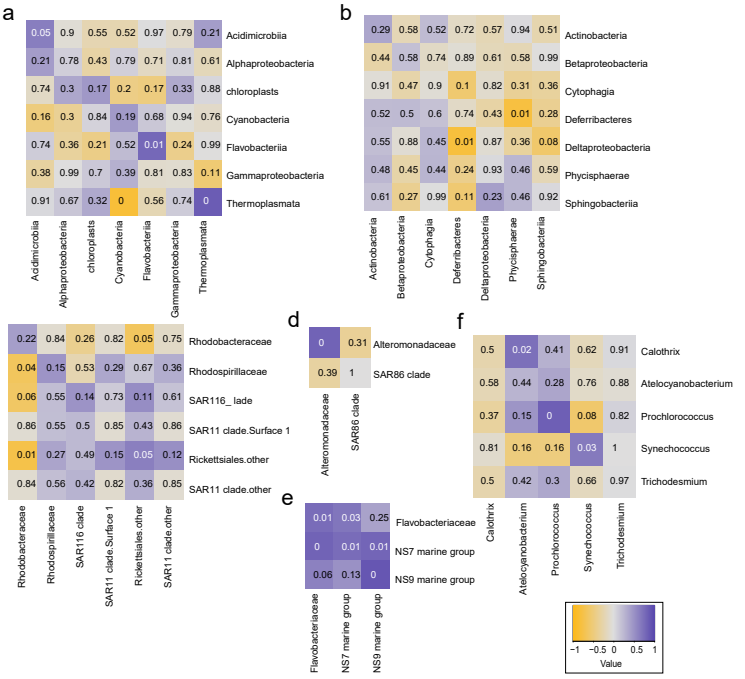

Fig. 2.

Response Letter Figure 3. Defluviicoccus sequences were downloaded with the SILVA browser from SSU version r123. Only sequences >1000 nt were selected, and from these, all sequences including `marine´,`deep-sea´, `soil´, or `symbiont´ were extracted plus 40 sequences that were not further classified (uncultured Defluviicoccus). These were aligned together with our Defluviicoccus OTUs using the SINA aligner (Pruesse et al, 2012), common gaps removed, and the alignment shortened to the length of the OTUs. The tree was built with Mr Bayes v3.2.5 (Ronquist et al, 2012), visualized in FigTree v1.4.2 (Rambaut & Drummond, 2009). The branches are weighted by posterior probability and the probablility of the basic nodes given in %. The summed up read count (normalized to sample total) is displayed together with the respective OTU ID. Our OTUs fall into a very distinct subcluster containing only marine Defluviicoccus 16S with one exception, which is annotated as coming from a soil sample.

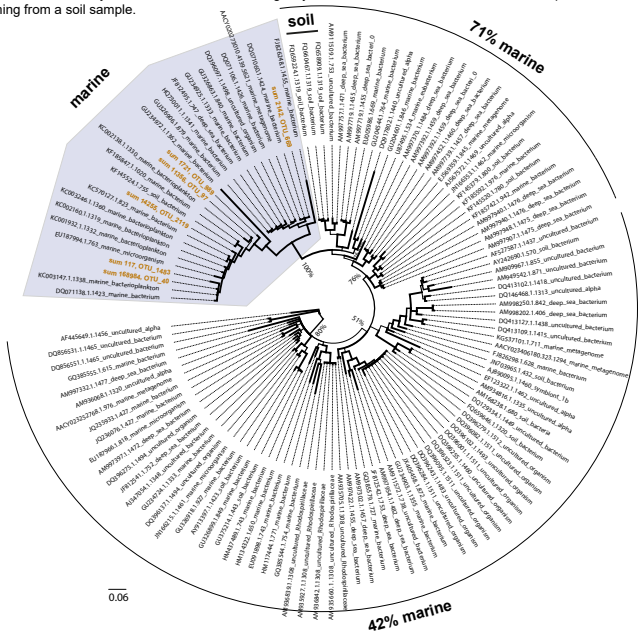

Fig. 3.