**Biogeosciences
Discussions**

Interactive
Comment

Interactive comment on "Microbial community diversity of the eastern Atlantic Ocean reveals geographic differences" *by* C. J. Friedline et al.

Anonymous Referee #2

Received and published: 28 February 2012

This paper is another in a series of studies from the ICoMM project which supported pyrosequencing of the V6 region of the 16S rRNA gene extracted from bacterial communities in various marine habitats around the world. The authors here present data from a transect in the Atlantic Ocean, from about 50 N to 31 S, covering about 7700 km. The main message of the study is that bacterial communities in the surface layer, deep chlorophyll maximum (DCM), and deep waters are all different. The study provides important new information about bacterial biogeography in the oceans.

The paper could be improved in several respects. First, the authors overstate the novelty of their study in the Introduction, although the Results/Discussion has a bit more complete review. The authors should work harder at identifying the truly novel aspects of their study. Second, the presentation of the study's data in the figures and

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

tables could be improved. Specific problems are discussed below. Finally, I question some of the approaches used by the authors to analyze their results, and they need to explore other techniques commonly used to analyze relationships between community structure and environmental properties. Again, specifics are given below.

Another general issue is that the authors don't compare their results with other papers from the ICoMM project. A complete review is beyond the scope of this paper, but it seems the authors could touch on a few issues. Do they see the same level of diversity as in other waters? What about the relative abundance of rare vs. abundant taxa? Although the "rare biosphere" may be overhyped, there are some interesting aspects to the question that the authors could address. The rare biosphere question is hardly touched on in this paper. Are the same taxa found in these waters as elsewhere? The paper mentions SAR11 but that's about it.

Specific comments

Abstract: The Abstract is a bit on the long side, with too much on data which don't need to be highlighted in an Abstract (e.g. the total number of sequences) and on results that are not particularly novel. The text starting at line 11 through 19 can be reduced substantially, with some things deleted completely. The potentially novel results start on Line 28.

P110, line 4 and elsewhere: "Eubacteria" is archaic and should not be used. "Bacteria" is sufficient.

Introduction: There is a lot of general information in this Introduction that any reader of this journal and article will know already. This includes much of the first paragraph. Many readers will quibble with the percentages (they seem too high) and the references, detracting from the novel aspects of the paper. The authors overstate the lack of information about the biogeography of bacteria in the oceans; what about the GOS studies and others? The Results/Discussion section has a bit more comprehensive review of what is known about these issues. The authors should minimize the general

stuff (readers will know bacteria are important) and get to the specific questions of this study.

Page 112, line 3: The sentence here hints that the authors don't use the term "metagenomic" in the way most others do. Contrary to what is implied by the authors, the approach does not rely on PCR. The authors' tag sequencing approach is not a metagenomic one.

Methods and Materials: Overall, this section is too long as the authors have included lots of extraneous details. Some specific candidates for deletion are given below.

Page 116, line 26: It seems that the authors did not "normalize" their sequences per sample, i.e. by calculating the relative frequency of the OTUs (actual read numbers/total sequences in the sample). Rather, it seems they randomly resampled their data so that each sample had same number of sequences as the sample with the lowest number of sequences. How was this done? How many times did they resample?

Page 117, line 1-5: This can be deleted. It's not appropriate for this section, and readers will already know all this.

Page 117, line 16: The authors don't need to give reason for doing a PCA (PCA is probably a more common abbreviation than PCoA). Again, not appropriate and not necessary.

Rather than a PCA, which depends on a parametric measure of distance, the authors should do (at least try) a nonmetric multidimensional scaling analysis which depends on fewer assumptions.

Page 117, line 23: The authors should also do a Canonical Correspondence Analysis (CCA) before getting into Mantel and partial Mantel tests.

Page 118, "Bayesian Inference". The authors drew up a similarity tree using an approach based on numerical taxonomy. First, other cluster type analyses are commonly done with these data. This web site has a succinct description of them: http://cran.r-

project.org/web/views/Environmetrics.html. The authors need a few words to argue why this "new" approach is necessary, but I don't see why it is necessary. It seems inappropriate to use an approach designed for taxonomy (and evolution) to examine similarities among samples, as the authors do here. Other approaches have been devised specifically for what the authors want to do. Second, it's not clear if anything really new is learned from any cluster or tree type of analysis. The PCA says it all.

Page 120, line 14: The authors begin their Results/Discussion with a lengthy paragraph that basically outlines the studyâĂŤagain. This should have been done already in the Introduction. This paragraph should be cut by 50%.

Page 122, line 2: "to down bias" is not a verb and it's not clear. Please rewrite.

Page 122, line 5: The authors mention that "Analysis of the rarefaction curves suggests that some of the deep-water communities" are more diverse (e.g. 2, 8, and 12) are more diverse than the rest of the communities". This is an important observation, but it's unclear whether overall deep-water communities are more diverse or whether samples 2, 8, and 12 are exceptions. It's impossible to see the sample depth in Figure 2. The comparison between shallow and deep-water community diversity should be tested statistically.

Page 123-124: The authors here talk about the overall composition of the three parts of the water column, but no figure or table summarizes these data. To do this, the authors should either modify Table 2 or devise a new table. That is, it would look like Table 2, but with data for each part of the water column. See further comments below about Table 2.

Page 124 line 10-17: This definition and overview of the deep chlorophyll maximum can be reduced by 50%.

Table 2: What is "n" versus "Mean"? These should be explained.

A few entries in this table occur more than once for unclear reasons. For example,

"Pelagibacter" occurs five times. What are these different "Pelagibacters"? The authors should give more information. Even less clear are the two entries called "Bacteria." If they can't be classified to a finer taxonomic level, why aren't they lumped together? Similarly for "Proteobacteria". I suppose they are different OTUs which cannot be classified to a finer level. If so, this should be said. I wonder if they should be lumped together (e.g. the two Bacteria entries) with an explanation in the table caption.

Although I can see an argument for this table, an argument against it is that it lumps together very different samples. To make it more useful, this table could be modified to summarize the composition of the most abundant taxa found in the three parts of the water column. That is, the table could give the top five taxa in each part of the water column and give the percentages for the relative abundance of the taxa in the other parts. In the extreme case of each water column section having a different top five, there would be 15 taxa listed in the table. To fit these data, some of the columns now given in Table 2 can be deleted without any loss. These include the range and the mean (or N or both, depending on what these are). The percentages (+/- SD) are enough.

Figure 2: Why use the "normalized" data and cut off samples with more sequences than the minimum? An advantage of this analysis is that all of the data can be plotted for looking at the shape of individual curves. To compare samples, the reader then can go to the sample with the fewest sequences and read up and down the figure.

Figure 3: This figure should be deleted or replace with something that can be seen without magnification. This is pretty much impossible to digest as a figure, even if readers expand it several fold on their computer screens. A table would have a slightly better chance of showing the data more effectively. But the authors could concentrate on the most abundant taxa or lump taxa together and present totals of higher phylogenetic levels.

Figure 4: More taxonomic information should be given for "Family 2".

Figure 5: The solid and dashed lines should be explained.

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper