Biogeosciences Discuss., 10, C134–C137, 2013
www.biogeosciences-discuss.net/10/C134/2013/

Biogeosciences
Discussions

Interactive comment on "DNA from lake sediments reveals the long-term dynamics and diversity of *Synechococcus* assemblages" *by* I. Domaizon et al.

Anonymous Referee #1

Received and published: 25 February 2013

General comments:

The authors focus on the long-term variability of Picocyanobacteria (PC) (especially focusing on Synechococcus=Syn.) utilizing molecular genetics to reveal changes in the PC diversity that is putatively linked to environmental changes in Lake Bourget over the last 100 years.

This study fits well into the scope of BG as it provides an interdisciplinary study combining molecular tools, geochemistry and paleolimnological analysis.

The author describe well the advantages and pitfalls of DNA analysis on environmental and degraded DNA and interpret their results critically.

Here are some remarks, which I think could improve the manuscript:

Please provide information on the length of the different PCR products amplified (for qPCR and PCR) (maybe you could use a little illustration, showing where the primers bind and the approx. length of the different fragments).

Indeed your amplified fragment is very long but surprisingly the number of putatively chimeric sequences seems to be relatively low. For the reader it would be even more convincing (I mean the authenticity of the sequences) if you could give a more detailed description of the previous tests that were performed to assess the level of DNA preservation (p. 2524, l. 17-19) - do you mean the qPCRs or any additional methods?

To estimate the levels of genetic variation within and between the defined 16SrRNA OTUs obtained from the different sediment layers - I would suggest calculating pairwise FST values to look for genetic differentiation between the different sediment layers you analysed (eg using Arlequin). Although your are not looking at real populations of Syn., Fst values and the net nucleotide divergence (comparing all pairs of sediment layers analysed) would give a good estimate of the genetic differentiation between the different sediment layers and these results could be compared to the UniFrac analysis.

Although you have identified the highest diversity in the ITS Fragment, I miss some phylogenetically/statistical analysis for ITS, like the ones which you applied to the 16SrRNA data set. This genetic marker might offer the possibility to show higher genetic differentiation between your genetic lineages. The results of pairwise Fst values for this data set is possibly also very interesting. Could you give an explanation why you did not perform any tests with ITS sequences?

In your study you describe summer air temperature as the driving factor for the abundance of Cyanobacteria and especially Syn., but the analysis of Syn. strains (Unifrac analysis) indicates also that possibly phylogenetically related groups of Syn. occurred at different trophic periods during the lakes's history, which would mean that although the trophic level of a lake has a minor influence on the presence of Syn. , that the

genetic diversity suggests a correlation between similar genotypes at a certain trophic state of the lakes. Have you thought of something like this? I think this result is very interesting as it shows that beyond the pure proof of Cyanobacteria /Syn.DNA through time (and its changes correlated to temperature) also genetic differentiation within specific groups of Syn. might be linked to other environmental factors which have not taken into consideration before. If you agree, maybe you could state on this in the manuscript.

Figure1: Maybe you could indicate the major environmental events that influenced Lake Bourget over the last 100 years.

Figure 3: This figure has very much information, but this information is partly not explained (e.g. the groups which sometimes have numbers or letters and the description of clades which is not really visible anymore). I would suggest to try giving all the clusters /clades informative names like, e.g. subalpine (if that is possible) and to give a short explanation in the caption (although you mention some of the formerly defined groups in the text, it would be helpful to see them in the figure caption once more).

Figure 3: Some OTUs (e.g. OTUb 16,17,18) only occur as single sequence types in just one sediment layer and are not unambiguously associated to another reference, how do you explain their authenticity?

Figure 3: I would also remove the diversity estimates and would present them in another table, which should also indicate the ranges of these estimations.

Figure S1: the rarefaction curves (except number BF24) do not really show that the diversity was fully covered, as all the curves do not reach saturation. You should indicate this in the text and consider that the real diversity is probably higher but the genetic approach (different DNA preservation, specificity of the primers, cloning etc.) influences the obtained species diversity in your data set.

Minor comments:

p. 2522, l. 24, please indicate the approx. length of the PCR product

p. 2523, l. 7, please indicate the approx. length of the PCR product

p. 2528, l. 26, please give another sentence to explain the UniFrac method (does this method only use the branch length and not the differences in the nucleotide sequences?)

p. 2527, l. 4, give the full name for TCC

p. 2529, l. 7, a hyphen is missing (Giguet-Covex)

Table 2: please indicate at the bottom of the table the total number of clones identified per time interval (=sediment layer)

Figure 1: please title the first column of the figure with e.g. years AD

Figure 3: Could it be that you indicated the wrong clade with "Cyanobium gracile", because this clade does not include any Cyanobium sequences, whereas group O and N includes Cyanobium sequences? If I am wrong, please indicate why that cluster is named like this?

Figure 4: please indicate the BF abbreviations for the different time intervals in the caption and/or in the figure.

Table S1. Here you could include the Fst values (plus pvalues in the upper part of the marix), if you would add the pvalues below the calculated distances of the unifrac results (both numbers in the lower part of the matrix).

Interactive comment on Biogeosciences Discuss., 10, 2515, 2013.