Interactive comment on "C$_5$ glycolipids of heterocystous cyanobacteria track symbiont abundance in the diatom *Hemiaulus hauckii* across the tropical north Atlantic" *by* Nicole J. Bale et al.

Anonymous Referee #2

Received and published: 24 September 2017

Bale and colleagues provide a very valuable data set aimed at interrogating the biomarker potential of heterocyst glycolipids with a pentose headgroup (C5 HGs) for diatom-diazotroph associations (DDAs). To establish such a diagnostic relationship is relevant for the reconstruction of the occurrence of DDAs in the fossil record. Thus the topic of the manuscript is definitely of great interest, and both the samples selected and analysis performed are ideally suited.

My main concern with the manuscript is that description and discussion of results are sometimes not as clear as they could be:

1. In the result section I would recommend to group the different stations, as already done to a certain degree in paragraph 3.4. ("high-salinity open ocean sites", "coastal-shelf stations", etc.). The naming of individual stations with only their number, paired with a very detailed description of concentrations at each of them, makes it hard to filter out the most relevant trends.

2. Short chain (C26) C5 HGs were initially described by Wörmer et al. (2012) in freshwater systems and a culture, before the description of longer chain C5 HGs by Schouten et al. (2013) in symbionts. As Wörmer et al (2012) only described C26 HGS, I would generally recommend to clearly differentiate between long- and short-chain C5 HGs throughout the text, e.g. in the conclusions "long-chain C5 HGs provide a robust, reliable method for detecting DDAs". Such a differentiation would make the authors' statements much more robust, as it eliminates potential interference from the short chain C5 HGs.

3. More importantly, I think that the discussion of the correlation between long-chain C5 HGs and different DDAs and free-living cyanobacteria needs to be improved to solidify the claim of a diagnostic relationship. For example, cross-plots and regression curves should be shown, instead of only stating r and p values. Based on these values alone, actually a strong correlation of C5 HGs is also observed with Trichodesmium colonies, and this harms the proposed biomarker potential. Even though I may share the authors' opinion that this regression might be coincidental and due to the co-ocurrence of Trichodesmium and DDAs at certain stations, a better effort to demonstrate the specific correlation between DDAs and C5 HGs is mandatory. In this sense I would for example recommend to pay special attention to the values which deviate from the regression. Assuming that several source organisms for C5 HGs exist, the fact that one potential producer (e.g. Rhizoselenia symbionts) is not abundant when C5 HGs are highly concentrated does not imply that it is not a potential source organism, as other producers may be present (e.g. Hemiaulus symbionts). This concept is hinted at in l. 281-290, but should be expanded. On the other hand, abundance of an organism

without corresponding HG abundance (e.g. maxima of Trichodesmium colonies) is a much more robust factor to rule out a potential source organism. In this sense it might also be interesting to plot a combined regression line for all DDAs vs C5 HGs.

4. Finally, the authors claim that the analyzed compounds are ribose-containing, but I couldn't find any description of how the sugar moiety has been characterized. If they haven't been described I would rather use the term pentose (as hexose is used for the C6 compounds).

Another issue is that the manuscript preparation sometimes seems a little careless, and I would appreciate a thorough revision. Some minor comments and edits include:

Text is sometimes indented, sometimes not.

l.12-13: "have a thickened cell walls" please correct use of singular/plural

l.14: use singular form "cyanobacterium" or plural verb form "make"

l.43: please specify that you are referring to heterocystous cyanobacteria "all hetero-cystous, non-symbiotic cyanobacteria"

l.45: It might be better to already mention here that short chain C5 HGs have been described in a non-symbiotic cyanobacterial culture, not only C6 HGs.

l.52: It is confusing to state that the "first study of the C5 HGs in the natural environment" was Bale et al. 2015 while providing the fact that "HGs with a C5 sugar moiety" were identified in freshwater environments three years earlier.

l.111: Station number is missing, maybe 10?

l.114: add "each", "For each sediment"

l.120: "freeze dried filtered seawater". I guess the lipids are extracted from the filters, not from the filtered seawater, right?

l.145: Just out of interest, have the authors tried to increase flow to shorten analysis

time? 0.2 ml/min seems slow for a UHPLC system

l.149: at which m/z is resolution measured?

l.178: is 36.3 a value for salinity?

l.180: "(Fig. 3c,d))" close parentheses

l.180: "NO3+NO2" (no subscript for "+")

l.212: may be rephrased: "Free Trichodesmium trichomes were broadly distributed (Fig. 4d) and often occurred. . ."

l.266: delete space: "Hemiaulus hauckii-Richelia"

l.290-293: The separation of DDAs depending on salinity with the current data is unclear, as the authors state. Therefore I would delete this topic and also the corresponding figure.

l.304-308: I think the sampling-volume explanation is a little confusing. Couldn't the authors just state that sensitivity of che chemical biomarker method is much more sensitive than the microscopic approach?

l.347: please rephrase to avoid the term "vegetal"

l.352: add "regarding" or similar: "difference regarding the limit of detection"

l.600: "Trichodesmium" should be in italics

l.603: use "dashed" instead of "broken"?

l.611-612, Table 1: "*" is not defined. Please define BMWL and DCM, even though they are already defined in the text, table should be informative on its own. Same for figure S1, where actually BML is used.

l.629: are Trichodesmium colonies expressed as colonies or trichomes/ml?

Figures: Please use larger fonts.

Figure 2-4: I think it would be better to place the axis legend (e.g. Salinity in figure 2) to the right, instead of on top of the figure. Especially in fig 3 and 4 this makes it easier to identify what is shown.

Figure 4: (d) is used twice, for panel (d) and what should be panel (e). Why is C32 C5 HG shown as %? Wouldn't it be more informative to show concentration?

C5