

Answers to comments from reviewer 2

Here we respond to the reviewer comments/suggestions beneath and in italicize. We have updated the appropriate line and page numbers.

Specific comments

P2L11-L12: Your assertion that the detection of UCYN-A *nifH* genes but not the host 18S rRNA genes (via your specific qPCR primer sets) may imply a free-living state for UCYN-A is highly speculative and inappropriate for the abstract. Remove this statement (I suggest removing this entire sentence).

We have deleted the speculation, and state the percent of samples where UCYN-A1 and A2 were detected and their respective hosts below detection (Pg. 2, lines 11-12)

P2L18: “temperature seemed to have a major impact”: please clarify/rephrase.

We have clarified our statement. (Pg. 2, Line 17-20)

P5L21: Is 17 cells per mL really a high concentration? Perhaps replace “high” with “moderate.”

We have followed the reviewer suggestion and modified our text. (Pg. 5, line 21)

P6L13: Rephrase “underlying factors.” Environmental drivers?

Changed according to reviewer's suggestion. (Pg. 6, line 13).

P6L17: Didn't you also target UCYN-C?

We have added the UCYN-C back in the text here, but it should be noted that the UCYN-C was only quantified in the ‘at-sea qPCR’ hence the comparison is difficult since we have far fewer abundance estimates, and therefore it wasn't included in any statistical analyses. (Pg. 6, line 17)

P13L1: Did you use data from both the lab-based and ship-based qPCR assays for your correlations? I find this concerning since you saw such large differences between lab and field assays.

The data generated by the ‘at sea’ qPCR assays was not used for any subsequent statistical analyses since only a few targets were quantified (UCYN-A1, UCYN-A2, UCYN-B, and UCYN-C) and only 4 depths at a limited number of stations. We included the ‘at sea’ qPCR for a comparison with the lab-based (archived) sample processing.

P15L4: “we considered only when there was at least one order of magnitude difference in detection” —please clarify. I counted 38 rows in your Supp. Table 2. Does this mean that 38 out of the 44 samples for which you can make the lab-based/sea-based qPCR comparisons had over an order of magnitude difference in *nifH* copy numbers? I find this very concerning if you are combining the 2 datasets for your statistical tests.

*We noted in the Suppl. Table 2 only when a sample was at least 1 order higher/lower for a particular target (UCYN-A1, UCYN-A2 and UCYN-B), since these were the only targets processed in both the ‘at sea’ and lab-based qPCR). So for example, we detect 18 *nifH* copies L^{-1} for UCYN-A1 in the ‘at sea’ sample from SD3 35 m, and in the parallel separate sample filtered at sea and stored until processing (full extraction and qPCR) in the lab was dnq. In addition, there is a suppl. Figure illustrates the comparison; this is Suppl. Fig. 3 (see attached excel file) and one can more easily see that for UCYN-B and UCYN-A1, often values fall on a 1:1 line.*

As stated above the 2 datasets were not pooled, since the extraction protocol was not identical it was not appropriate to pool the values.

We have clarified this paragraph and included the correct reference to the Suppl. Figure. (Pg. 15, lines 4-9).

Discussion overall: The discussion could be greatly streamlined, particularly section 4.3.

We have followed the suggestion of the reviewer and have tried to shorten the discussion. Many of our results concur with previous findings, and so we try to only highlight the new contributions of this work.

P22L9-P23L8: I am concerned about the large differences you observed between the qPCR performed in the lab and at sea. The supplement to this manuscript only included Supp. Fig 1 and Supp. Tables S1-S6, so I cannot see the Supplementary Figure 3 referred to in the text, which apparently addresses the inconsistencies. You say that you cannot discount the “natural heterogeneity of plankton,” but it seems you could easily distinguish between natural heterogeneity and differences due to extraction/qPCR method by looking at the variability in nifH copy number among biological replicates processed in the same way. Did you include any biological replicates, taken from the same net, and process the samples using the same methods? If you saw the same variability among replicates as you do between the two different methods, then you could attribute the differences you see to natural variability. But if the difference in methods is the reason you see such large differences in samples processed in lab vs at sea, then perhaps you should only use one or the other dataset instead of combining them.

Unfortunately biological replicates were not taken during this cruise, as we were pressed to process the ‘at sea’ qPCR in a timely manner to inform the cruise at large. Moreover, replication in sampling (e.g. 2-3 replicate samples) for qPCR is not typical (see previous works from multiple groups) and is often related to time and/or water budgets. Although, here it seems that it should be a consideration in future samplings.

As stated above, the datasets were not pooled, and the comparison of the qPCR from the ‘at sea’ and archived samples was shown to be as transparent as possible. More efforts are underway to address many of the technical details (e.g. at sea vs. archive processing, the cross-reactivity) unveiled in this body of work, however for the purposes of this manuscript and limitations in that we do not have biological replicates, we care to focus on the results in the context of the special issue-which is to provide the abundance for the diazotrophs and the influence on the measured parameters. We do acknowledge the issues here and that is why we wanted to highlight some of the discrepancies; however we would like to stress that the two datasets were not combined for statistical analysis. A sentence has been added to clarify this. (Pg 13, line 1-3)

P23L23: Here and elsewhere, clarify that these were the least detected diazotrophs of those targeted (since you did not assess total diazotroph diversity).

We agree that this is the case and have clarified according to reviewer suggestion. (Pg. 24, line 6).

P25L20: “12m, which is shallower than the subsurface maximum”— Did these studies really all compare 12m to 25m? If not, this statement should be removed.

Here we are trying to highlight that the depth of maximum abundance for the Richelia symbiont groups (het-1 and het-2) in the MA region of the transect was shallower than in the other works. In the earlier works the upper photic zone was also sampled (e.g. WTNA the 100%-0.1% light level) and the depth of maximum abundance is reported to illustrate a niche partitioning (e.g. Moisander et al. 2010 for the UCYN groups and Trichodesmium). We have replaced 'commonly' with 'previously'. (Pg. 26, line 3).

P27L4-7: And because the UCYN-A genome suggests that it does not have the genetic capacity for independent carbon metabolism.

We have modified the text to address the dependency of the UCYN-A on other organisms based on its genome content and appropriate references. (Pg. 27, lines 11-13).

P27L7-17: You have already described reasons why we do not think that “the UCYN-A lineages can live freely.” As you explain, the most likely reason that you found higher abundances of UCYN-A1 nifH genes than the host 18S rRNA genes is that the qPCR primers used to not cover the full diversity of the hosts. Also, you don’t know that the hosts were “absent” from your samples, they were just below detection. I think it is inappropriate to speculate that UCYN-A may be free-living when you are only presenting qPCR data. You should present microscopic evidence (CARD-FISH) if you are going to make a claim that UCYN-A can exist in a free-living state.

We agree and have removed the speculation on the free-living nature of UCYN-A. (Pg. 28, lines 5-12).

P27L22: “we found evidence that there are multiple UCYN-A1 and A2 symbionts in both host types”— again, you need microscopic evidence to make these types of statements. The fact that you found higher abundances of UCYN-A nifH genes than the host 18S rRNA genes likely reflects that the qPCR primer/probe set does not hit the full diversity of hosts. The discussion on numbers of UCYN-A per host is entirely speculative when you only have qPCR data, so this entire paragraph should be removed or greatly shortened.

We agree that microscopic evidence is required in addition to the qPCR results and have deleted our interpretation. We highlight that the broader diversity is a possibility and state that CARD-FISH is necessary. (Pg. 28, lines 5-12)

P28L1: Here and elsewhere: UCYN-A1 and A2 nifH genes were 2-10... inefficient DNA extractions, polyploidy, etc mean that nifH gene copies do not correspond to cell concentrations (as you discuss later).

We agree, and have modified the text throughout. (Pg. 28, line 5)

P29L4-5: Also see Luo et al. (2014), Biogeosciences. I find it curious that you do not discuss this paper.

We have added this reference and also a summarizing sentence to the discussion. (Pg. 29, lines 15-17).

P30L21: “it would appear that low light was a pre-requisite”— this is an over-statement. You just found a correlation.

We agree, and have changed the wording to 'correlates with'. (Pg. 31, line 6).

P31L13-14: Comment on the negative correlation of UCYN-A with depth in the meta-analysis?

We have amended our text (Pg. 31-32, lines 24-2).

P32L9: You don't have to assume one gene copy per cell when you discuss qPCR data, as long as you refer to gene copies instead of cell abundances (e.g. UCYN-A1 nifH gene abundances instead of UCYN-A1 abundances). But throughout the manuscript, you talk about the concentrations of diazotrophic groups, not their gene copies. I think you should either make changes throughout the manuscript to refer to gene copies instead of cells, or else here (page 32) be explicit that YOU are assuming one gene copy per cell in this manuscript, though you realize that this assumption is likely not valid because of problems including polyploidy and inefficient extraction efficiency.

As suggested, we have made changes throughout the text, when appropriate, to reflect that we're not assuming one gene copy per cell. The limitations, caveats of qPCR has been modified as well. (Pg. 32, lines 21-22)

P34L11: "reliable quantification"— really?

We have modified this paragraph about the use and considerations of 'at sea' qPCR. (Pg. 34-35, lines 25-1).

Fig. 2:

- Did the light really attenuate the same at all of the stations?

No, there were slight differences except for LD B and three stations in the SG. The caption has been modified. (Pg. 47, lines 7-13).

- Clarify in the legend whether 2b depicts surface concentrations. If so, can you add error bars from biological replicates?

The figure caption states that both a) and b) are LOG10 transformed mean concentrations across the entire cruise at a) depth and b) station. Due to heterogeneity of the diazotrophs abundances, especially with depth (where there can be 10^5 gene copies/L at the surface and 0 at 80 m), error bars would be large and not very informative in this context and so these were omitted. (Pg. 47, lines 7-13)

- Capitalize depth, station etc.

- Rotate the text in 2b

- 1b is missing its panel label

- I think this figure would be easier to digest if you switched the axes in 2b and lined up the two panels vertically.

We agree and have modified the figure according to reviewer suggestions. (see author comment pdf)

Fig. 4

- It is not apparent to me what the individual points on this plot represent. Perhaps you could elaborate on the meaning of "unconstrained response variables." Or else just realize that not everyone will follow.

We tried to amend the figure caption for clarity. (Pg. 47, lines 18-24)

- Rephrase "variance of included parameters" in the figure legend.

We have followed the suggestion and amended the figure caption for clarity. (Pg. 47, lines 18-24).

Fig. 5

Please clarify whether this analysis used all of the data included in Supp. Table 6.

We have modified the figure caption and include a statement that the figure is based on Suppl. Table 6. (Pg. 48, lines 1-6).

Technical comments

P2L6-8: “*Trichodesmium*...respectively”: Rephrase this sentence to improve grammar.

The sentence has been slightly changed to improve clarity and grammar (Pg. 2, lines 6-8).

P2L14: Replace “deep dwelling” with “a deep-dwelling group”; replace “surface group” with “a surface group”

Replaced according to reviewer suggestion. (Pg. 2, lines 13-14).

P3L15: Replace the comma after “surface” with a semicolon.

Replaced according to reviewer suggestion. (Pg. 3, line 15).

P3L19: Replace “photic” with “photic-zone”

Replaced according to reviewer suggestion. (Pg. 3, line 19).

P4L2: Replace “is a symbiosis between” with “associates with”

Replaced according to reviewer suggestion. (Pg. 4, line 2).

P4L12: Replace “the UCYN-C” with “the UCYN-C group”

Replaced according to reviewer suggestion. (Pg. 4, line 14).

P5L9: Replace “lowest concentrations” with “lowest reported concentrations” and delete “in the world have been reported”

Replaced and deleted according to reviewer suggestions. (Pg. 5, lines 8-9).

P5L10-11: replace “harboring” with “which harbors” and replace “being” with “is”

Replaced according to reviewer suggestions. (Pg. 5, line 10).

P6L2: Place a comma after “WTSP” and replace the semicolon with a comma.

Replaced and added according to reviewer suggestion. (Pg. 6, line 2).

P7L6-9: This seems to repeat the sentence P6L24-P7L3.

We have modified the sampling section to limit redundancy. (Pg. 6-7, lines 21-7).

P9L6: Returned to the laboratory AND frozen? Please clarify.

This sentence was unclear and we have rephrased it for clarity. (Pg. 9, line 6).

P9L22: Replace “on published 18S rRNA sequence” with “on a published” or “on published...sequences”

Replaced according to reviewer’s first suggestion. (Pg. 9, line 22-23).

P10L18: Replace “selected diazotrophs *nifH* gene copies” with “*nifH* gene copies from selected diazotrophic groups”

Replaced according to reviewer suggestion. (Pg. 10, line 18).

P10L20: Replace “performed” with “quantified”

Replaced according to reviewer suggestion. (Pg. 10, line 20).

P11L17: Include the end parentheses after “Biosystems”

Added according to reviewer suggestion. (Pg. 11, line 17).

P13L10-13: “T-tests...concentrations” I find this sentence confusing.

The sentence has been amended for clarity (Pg. 13, lines 12-13).

P13L13: Replace “dataset” with “data”

Replaced according to reviewer suggestion. (Pg. 13, line 14).

P14L8 “but declined...compared to the SG” Be more specific.

We have modified the sentence to describe the deepening of the thermocline in the SG compared to the MA (Pg. 14, lines 7-9)

P14L9-13: Rephrase this sentence.

The sentence has been split and amended for clarity. (Pg. 14, lines 9-13).

P19L3-5: Rephrase this sentence.

The sentence has been amended for clarity. (Pg. 19, lines 10-12).

P19L12: Replace “The deeper dwelling” with “Diazotrophic targets in the deeper dwelling”

We have rephrased the sentence and refer to the 2 groups as shallow and deep to avoid confusion. So we refrain from using deeper and shallower, etc. (Pg. 19, lines 9-22)

P20L25: Replace “and significantly” with “but was significantly”

Replaced according to reviewer suggestion. (Pg. 21, line 6).

P21L20-21: “and likely... N2 fixation” please rephrase.

Rephrased according to reviewer suggestion. (Pg. 21-22, lines 25-3).

P22L2: Rephrase “nowadays”

Wording changed to ‘modern’. (Pg. 22, line 9).

P22L6: Replace “showed” with “describe.” Also, I think the term efficient is inappropriate, as you did not measure DNA extraction efficiency.

We agree and have replaced the words according to reviewer suggestion. (Pg. 22, line 13).

P22L7: Replace “qPCR” with “qPCR technique”

Replaced according to reviewer suggestion. (Pg. 22, line 14).

P24L3: Replace “to Moisander’s” with “to that reported by Moisander”

Replaced according to reviewer suggestion. (Pg. 24, line 10).

P24L8: “symbioses”—you mean both A1 and A2? Please clarify.

Yes, we mean both A1 and A2 and have amended this to the sentence. (Pg. 24, line 14).

P24L25: Replace “lesser” with “lower”

Replaced according to reviewer suggestion. (Pg. 25, line 6).

P25L19: Replace “Highest” with “The highest”

Replaced according to reviewer suggestion. (Pg. 26, line 1).

P28L11: Replace “ranging” to “ranging from”

Replaced according to reviewer suggestion. (Pg. 28, lines 22-23).

P28L14: Replace “conditions” with “the conditions”

Replaced according to reviewer suggestion. (Pg. 28, line 25).

P28L15: Replace “life histories” with “life histories of different diazotrophic groups”

Replaced according to reviewer suggestion. (Pg. 29, line 1).

P29L1-3: Rephrase this sentence to fix grammar errors.

This sentence has been modified and streamlined. (Pg. 29, Lines 12-13).

P29L9: Here and elsewhere: replace “diazotrophs” with “diazotrophic groups”

Since in our statistical analyses we find two groups (deep and shallow) we refrain from using diazotrophic group to not confuse.

P29L12: Replace “environmental parameters PAR” with “the environmental parameters of PAR”

We agree and have modified as suggested. (Pg. 29, lines 23-24).

P29L13: “influencing”, “drove”—here and elsewhere, rephrase so you are not inferring causation.

We agree and have modified the text when appropriate (Pg. 29, line 25)

P29L19: Replace “are” with “have been”

We agree and have replaced the wording (Pg. 30, line 6).

P30L1-8: These sentences don’t fit with the rest of the paragraph.

This paragraph has been modified to highlight that the nutrient conditions in this region favored DDAs over the UCYN-A. (Pg. 30, lines 3-20)

P30L12-14: “Moreover...Karl et al. 2012” - This sentence doesn’t fit with the rest of the paragraph.

Our intention of including Karl et al. 2012 was to highlight that day length which is in the context of light, could influence the symbiotic diatom populations. For the sake of streamlining, the sentence has been removed.

P30L19: Replace “and a negative” with “and displayed a negative”

We agree and have modified accordingly (Pg. 31, line 4).

P30L24: “Interesting and unexpected was”— rephrase.

The sentence has been modified. (Pg. 31, lines 8-10).

P31L5: Replace “diazotrophs” with “diazotroph”

Replaced according to reviewer suggestion. (Pg. 31, line 15).

P31L11-13: “The studies...temperature” correct the grammar errors in this sentence.
We agree that this sentence was broken, and have amended and changed for clarity (Pg. 31, lines 21-23)

P32L3-6: “Unlike...space” perhaps delete this sentence.
We prefer to keep this sentence since it highlights inherent difficulties in determining environmental parameter impact on diazotrophs. (Pg. 32, lines 16-19).

P33L9: “Consistent...abundant”— rephrase, this statement is meaningless out of context.
We agree and have rephrased and merged it with the following sentence. (Pg. 33, lines 23-25).

P33L17-21: “According...tests”— remove this sentence.
We did not remove the sentence since we feel that our results are important to highlight that there was a disconnect in the detection. A similar result was reported in a qPCR study by Thompson et al. 2014 (one of the first to describe the UCYN-A2 symbiosis in detail) that found symbiont/host ratios of 0.2-11 during 3 days of sampling. The same possible limitations applied, which they also state, and hence we felt it was still valid to include this in our summary of conclusions. A sentence on this study has been added to the discussion. (Pg. 28 lines 7-8)