

## **General comments**

This manuscript discusses the distribution and environmental drivers of cyanobacterial diazotrophs on a transect in the Western Tropical South Pacific. The *nifH* genes of major cyanobacterial diazotroph groups (and the UCYN-A prymnesiophyte hosts) were quantified via qPCR and correlated to environmental parameters. Additionally, a meta-analysis was performed to test whether the environmental drivers identified in this study were found in other ocean regions.

The authors provide useful data on the distributions of cyanobacterial diazotrophs in a historically understudied ocean region, along with the environmental context of the stations sampled. I found the environmental correlations and cluster analysis particularly compelling, and was interested in the finding that the two UCYN-A sublineages occupied a deeper zone in the water column than the other diazotroph groups investigated. However, I have a few serious concerns about this manuscript. First, the authors suggest that their finding of lower abundances of UCYN-A host 18S rRNA genes than cyanobacterial UCYN-A *nifH* genes may imply that the UCYN-A in their samples were in a free-living state. This is completely speculative without microscopic evidence, and these statements should be removed from the abstract and conclusion. Second, I am concerned about the large discrepancies in qPCR abundance data between separate lab-based and ship-based methods. I elaborate on all of these concerns in the specific comments below. The manuscript also contained numerous grammar errors; I correct some but not all of these errors in the technical comments below. The language in this manuscript should be improved before publication.

The page and line numbers below refer to the revised manuscript that the authors submitted after incorporating responses from the first reviewer.

## **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).

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

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

P6L13: Rephrase “underlying factors.” Environmental drivers?

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

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.

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.

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

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 niskin, 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.

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

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.

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

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.

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.

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

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

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

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

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 discuss 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 ploidy and inefficient extraction efficiency.

P34L11: “reliable quantification”— really?

Fig. 2:

- Did the light really attenuate the same at all of the stations?
- Clarify in the legend whether 2b depicts surface concentrations. If so, can you add error bars from biological replicates
- 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

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.
- Rephrase “variance of included parameters” in the figure legend

Fig. 5

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

**Technical comments**

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

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

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

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

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

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

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

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

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

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

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

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

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

P10L20: Replace “performed” with “quantified”

P11L17: Include the end parentheses after “Biosystems”

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

P13L13: Replace "dataset" with "data"

P14L8 "but declined...compared to the SG" Be more specific.

P14L9-13: Rephrase this sentence.

P19L3-5: Rephrase this sentence.

P19L12: Replace "The deeper dwelling" with "Diazotrophic targets in the deeper dwelling"

P20L25: Replace "and significantly" with "but was significantly"

P21L20-21: "and likely... N2 fixation" please rephrase.

P22L2: Rephrase "nowadays"

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

P22L7: Replace "qPCR" with "qPCR technique"

P24L3: Replace "to Moisander's" with "to that reported by Moisander"

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

P24L25: Replace "lesser" with "lower"

P25L19: Replace "Highest" with "The highest"

P28L11: Replace "ranging" to "ranging from"

P28L14: Replace "conditions" with "the conditions"

P28L15: Replace "life histories" with "life histories of different diazotrophic groups"

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

P29L9: Here and elsewhere: replace "diazotrophs" with "diazotrophic groups"

P29L12: Replace "environmental parameters PAR" with "the environmental parameters of PAR"

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

P29L19: Replace “are” with “have been”

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

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

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

P30L24: “Interesting and unexpected was”— rephrase.

P31L5: Replace “diazotrophs” with “diazotroph”

P31L11-13: “The studies...temperature” correct the grammar errors in this sentence.

P32L3-6: “Unlike...space” perhaps delete this sentence.

P33L9: “Consistent...abundant”— rephrase, this statement is meaningless out of context.

P33L17-21: “According...tests”— remove this sentence.