General Comments

Stenegren et al. utilise *nifH* specific quantitative PCR (qPCR) to determine the abundances of a suite of cyanobacterial diazotrophs in the western tropical South Pacific Ocean. The author's present results from both 'at sea' and laboratory based qPCR assays for unicellular (UCYN) groups, and identified interesting differences in absolute quantification between the two. In addition, Stenegren et al. present a new qPCR assay to quantify the picoeukaryote host of UCYN-A1. Correlation analyses were used to determine environmental constraints on the abundances of the different diazotroph groups, and include a meta-analysis of other publicly available datasets to expand their findings to other ocean regions. This work will be of interest to the readership of Biogeosciences as it reports on diazotroph abundances in a relatively understudied marine province, which has recently been identified as a potential hotspot for biological N_2 fixation.

I am surprised that only the picoeukaryote hosts of UCYN-A were quantified by qPCR, but the hosts of the heterocystous diazotrophs were not, this could have strengthened the authors ability to address their hypotheses and test the underlying environmental factors influencing the different symbioses.

I think the discrepancies between the 'at sea' and lab based qPCR UCYN targets should be discussed further, some of the differences in *nifH* copies are quite large and this could have implications for other studies that only perform qPCR on DNA samples post-sampling, or vice versa. Could the two different DNA extraction methods that were used cause these differences? Were the samples that are being compared taken from the same Niskin bottles, or could the differences shown in Supplementary Table 2 be due to natural heterogeneity in microbial communities sampled at slightly different times while on station?

In general, I think that the manuscript would benefit from some re-structuring of the Introduction and the Results prior to publication, and could be refined to streamline the main purpose of the study and to highlight the main findings.

Specific comments

Abstract:

In general, the Abstract could be improved by including the quantitative *nifH* copy numbers rather than the percentage of *nifH* copies detected.

Line 9. What does the >47 % refer to when you say the UCYN-A lineages were poorly detected?

Lines 9-12. This is inconsistent- the hosts mirrored their respective symbionts yet were below detection?

Lines 14-15. Perhaps re-phrase to state that they correlated with the surface group. Include the parameters that were significantly correlated with the deep UCYN-A group too.

Lines 15-16. Could you expand on this briefly?

Line 18. ‘free-living cyanobacterial diazotrophs’

Introduction:

Pg 4 lines 18-19. There is no clear link between these two paragraphs. Suggest moving the next paragraph, from page 5 lines 14-23, up to provide the link between the diversity of cyanobacterial diazotrophs and the introduction to the environmental characteristics of the WTSP.

Pg 5 lines 18-23. Perhaps expand on this slightly to indicate why understanding these differences are important for our understanding of marine N_2 fixation, especially within a predicted hotspot for N_2 fixation.

Pg 5 lines 24-25. Could you briefly outline why performing ‘at sea’ quantitation is/would be a preferential application for qPCR studies?

Materials and Methods:

Pg 6 21-24. Could you indicate here (briefly) which conditions you were aiming for with these LD stations?

Pg 7 line 3. Which make and model of CTD was used?

Pg 7 lines 9-11. Where these samples also stored with the glass bead mixture? Was the same amount of seawater filtered?

Pg 7 line 16. How long were they stored for?

Pg 7 line 22. Why are these methods included under the ‘Nutrient analyses’ sub-heading?

Pg 8 line 14. This is T. pelagicum in Supplementary Table 5?

Pg 8 line 16. Were the host diatoms quantified too? If not, why not?

Pg 9 lines 2-4. Have you tested the two DNA extraction methods on identical samples to determine any potential differences between the two methods? It would be good to get a clear sense of how different these extraction methods are here.

Pg 9 line 14. What is the percent identity between the UCYN-A1 and UCYN-A2 host 18S rRNA sequences?

Pg 9 lines 13-21. This information might be better summarised in a table.

Pg 10 lines 10-12. It’s great to see this information here but why wasn’t het-3 included in the het cross reactivity tests?

Pg 11 lines 8-10. Is there a particular reason why assays weren’t performed for these stations?

Pg 11 lines 22-24. What were the efficiencies of the other assays?

Results:

I think the results of the cross-reactivity tests should be moved to become section 3.2 as this is important for the interpretation of the qPCR assays.

Pg 13 line 7. Table 1 contains values for DIN- should they be bq?

Pg 13 lines 18-19. A comparison of the two DNA extraction methods is required to determine if they could have affected the qPCR results.

Pg 14 lines 1-5. But some of these differences are quite large, for example from Table S2 UCYN-A1 at LDB (10% irradiance) nifH copies at sea were 1.08×10^3 compared to bd in the lab quantified samples, and UCYN-B at SD1 was bd at sea and $> 1 \times 10^5$ for the lab based assays. This is a potentially major issue, with no clear pattern for as to why. Can you explain these results? This needs to be discussed further on pages 20-21.

Pg 14 lines 17-20. Here, and in other places throughout the results where you report depths of maximum abundances, please include the *nifH* copy numbers in the text.

Pg 15 lines 4-6. Please revise this sentence for clarity.

Pg 15 lines 7-11. The different LD and SD stations within the MA and SG become a bit confusing throughout the results. Perhaps indicate the different regions in Figure 1 and supplementary tables where applicable.

Pg 15 line 14. Sometimes you refer to number of stations and other times the number of samples when talking about prevalence of the different groups, please be consistent.

Pg 16. Section 3.4 indicate number of observations included when presenting the significant correlations.

Pg 17 line 18. Perhaps indicate the significant clustering for group 1 and 2 on Figure 3 for clarity.

Pg 18 lines 1-11. The RDA is explained very nicely, perhaps you should colour code the dots in figure 4a to reflect the different response variables.

Pg 18 Section 3.5. I think the results of the meta-analysis would be more compelling if represented as a figure in the main text, perhaps as a heatmap/correlogram like Figure 3.

Pg 19. Lines 14-24. I suggest moving this section to 3.2 of the results. Do you have the data for the “vice versa” e.g. for the UCYN-A2 host assay with UCYN-A1 host target, and the het-2 assay with het-1 target? This is not obvious from Supplementary Figure 1.

Discussion:

Pg 21 lines 2-8. Please discuss these results more thoroughly. Specifically, can you comment on potential differences in DNA extraction efficiency between the two methods? Are you comparing the same diazotroph community (e.g. from the same Niskin bottle/ homogenised samples)? Were there any inhibitors? There doesn't appear to be a clear pattern in over/under estimated of the at sea versus lab assays based on Table S2, so perhaps you can't explain the differences, but possibilities should at least be discussed.

Pg 21 lines 23-25. Do you have a hypothesis as to why you observed these surprising results?

Pg 22. Could you also compare the actual abundances throughout these paragraphs to give more context- perhaps also the seasonal timing of the different studies for comparison.

Pg 24 lines 1-8. Indicating that DDAs are important for export production in this region, like the NPSG.

Pg 24 lines 9-18 Could this also be due to a limited understanding/representation of UCYN-C diversity; how specific is the qPCR assay?

Pg 28 lines 14-22. Why do you think this was the case? What other factors (perhaps beyond what you measured) could have influenced the depth distributions of these groups.

Pg 28-29 lines 23-19. It would be nice to see further discussion around the results of the meta-analysis; the similarities and differences to other regions and the local/environmental factors driving these patterns could be discussed.

Pg 30 line 19. Could you provide the same context for the UCYN-A assays?

Figures

Figure 1. Does the white dashed line indicate the separation between the MA and SG? Please clarify. Would it be possible to overlay SST on the station map, as this was an important explanatory variable.

Figure 2. You mention specific depths in the text- perhaps indicate average depth on 2a, or include % surface irradiance in the text. Can you make 2b slightly larger as the station numbers are difficult to distinguish (perhaps also indicate the MA to SG transition).

Figure 3. Indicate group 1 and group 2 on the hierarchical clustering for clarity.

Figure 4. Colour coordinated dots might help to support the text. Please include a y-axis label in 4b.

I would also like to see the meta-analysis presented as a figure if possible.

A T-S plot as a supplementary figure would also help to distinguish the different water masses of the MA and SG.

Technical corrections

Pg 3 line 21. Insert comma after ‘genera’

Pg 3 line 25. Remove additional ‘the’

Pg 5 line 20. Typo ‘recognized’

Pg 7 line 23. ‘in’ rather than ‘and’

Pg 7 line 24. Remove ‘two different days’ and parentheses.

Pg 8 line 8. Remove ‘in’

Pg 8 line 16. Unclear if there is a comma missing or parentheses missing, please check.

Pg 8 line 7. Semi-colon missing

Pg 13 line 1. OpenMEE reference missing

Pg 14 line 6. ‘Horizontal and vertical distributions’ should be a new section

Pg 14 line 16. Remove extra ‘for’