Answers to comments from reviewer 1

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

Note: Since we only had a pdf of the reviewer suggestions we address the 3 paragraphs from the general comments without typing all text from the reviewer then address the specific comments below.

General Comments.

Surprise about the picoeukaryote.

This is a valid suggestion however at the time of the study the sequence data for the respective diatom hosts was unknown. A subsequent and current study in Foster lab is work developing new molecular methods (e.g. in situ hybridization, qPCR) for identifying on the diatom hosts by single cell PCRs, however this information was not concurrent with the work presented here so it is not included. Unfortunately we also have no more DNA template left to process as we used it all up on the 9 targets which we report on.

Discrepancies between 'at sea' and lab-based.

The comments and suggestions (see below) are also well noted, however we cannot address the extraction efficiencies given the nature of the samples (field based mixed communities. Although we attempted to be as uniform as possible in the sampling and archiving, it is likely that some heterogeneity in sampling resulted from that samples were take from different levels of the niskin. Perhaps all water could have been drained, homogenized prior to the subsampling of the volumes for the nucleic acids, however we were pressed to get the samples extracted and run on the qPCR in a timely manner to direct other points of interest for the cruise at large. We address some of the sources for the discrepancies in more detail in the revised version (Pgs 22-23, lines 13-4).

Restructuring and refining of introduction, results, etc.

We agree, and have made some reorganizing of the presented work.

It should be noted that this paper was submitted for a special issue of Biogeosciences which is entirely focused on the research cruise (OUTPACE) and so it is part of a collective body of work by all members of the expedition.

Specific Comments

Abstract:

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

Agreed, and have provided ranges in the nifH copies enumerated for the various targets.

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

Originally this was referring to the % of samples below detection, we have rewritten the sentence and this no longer applies (Pg 2, Lines 10-12).

Line 9-12. This is inconsistent-the hosts mirrored their respective symbionts yet were below detection?
We recognize that this was not very clear as we were referring to when the hosts were present they were only in samples where their respective symbionts were enumerated. We have re-written the result (Pg 2, Lines 13-15).

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.
Agreed and have followed the suggestions (Pg 2, Lines 19-20).

Lines 15-16. Could you expand on this briefly?
Added in a short comment on the most important environmental parameter in the meta-analysis: temperature (Pg 2, Line 18).

Line 18. 'free-living cyanobacterial diazotrophs'
Agreed and rephrased (Line 2, Line 20)

Introduction:
Pg 4 lines 18-19. There is no clear link between these 2 paragraphs. Suggest moving.
Agreed and have restructured the introduction following these helpful suggestion.

Pg 5 lines 18-23. Perhaps expand on this slightly to indicate <u>why</u> understanding these differences are important for our understanding of marine N_2 fixation, especially within a predicted hotspot of N_2 fixation.
Added a sentence to emphasize the importance and benefits of a better understanding (Pg 5, Lines 3-5).

Pg 5 lines 24-25. Could you briefly outline why performing 'at sea' quantitation is/would be a preferential application for qPCR studies.
We have added a brief explanation as to why it was necessary and useful for the expedition. For example, the results from the qPCR was used to make informed decisions about regions high for target diazotrophs. Subsequently these stations were selected for long duration stations so intensive sampling could be implemented to better characterize the community, the microbial activities and hydrographic conditions (e.g. rate measures, drifter arrays, sediment traps, etc.) (Pg 6, Lines 3-8).

Materials and Methods:

Pg 6 21-24. Could you indicate here briefly which conditions you were aiming for with these LD stations.
A brief explanation is provided and also two articles in the special issue are referenced (Pgs 6-7, Lines 25, 1-4).

Pg 7 line 3. Which make and model of CTD was used.
The Seabird 911 was added (Pg 7 Line 15).

Pg 7, lines 9-11. Were these samples also stored with the glass bead mixture. Was the same amount of seawater filtered?
Details have been provided. The 4 samples taken for the 'at sea' processing were not stored, these were immediately extracted and the beads were necessary for the agitation step. The same volume was always filtered (2.5L), except for a few cases with a lot of material clogging the filters (Pg 7 Lines 19-20; Pg 8, Lines 1-2).

Pg 7 line 16. How long were they stored?
The nutrient analysis samples were stored for a few months as all samples were re-analyzed in the land-based laboratory (Pg 9, Lines 7-8).

Pg 7 Line 22. Why are these methods included under the 'Nutrient analyses' sub-heading?
This was a mistake, the text has been replaced in an appropriate subheading (Pg 7, Line 11).

Pg 8 line 14. The *T. pelagicum* in Suppl. Table 5?
This was also a mistake, both should be K.pelagicum *and have now been corrected.*

Pg 8. Line 16. Were the host diatoms quantified too? If not why not?
The host diatoms were not quantified as their 18S rRNA sequences were unknown at the time of the study and hence qPCR primers and probes could not be designed. Subsequent and current work in Foster's Lab is working on these types of assays. Unfortunately all DNA templates were used up on the 9 targets (Trichodesmium, UCYNA-1, UCYNA-2, UCYNA-1 host, UCYNA-2 host, UCYN-B, het-1, het-2, het-3).

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 methods are here.
We wholly agree with this shortcoming in our approach, but unfortunately we do not have replicate samples to test. It should also be noted that replicate field samples are challenging to collect as one cannot be certain of uniform distribution in a niskin bottle. We have added a short summary of the latter in the discussion (Pg 22-23 Lines 13-4).

Pg 9 Line 14. What is the percent identity between the UCYN-A1 and UCYN-A2 host 18S rRNA sequences?
We have added this detail (97.95%) by determining a distance matrix on the 635bp fragment of the 18S rRNA for the following sequences: accession number JX291893 (UCYN-A1 Host) and accession numbers KF771248-KF771254 (UCYN-A2 host) (Pg 10 Lines 16-17).

Pg 9 Lines 13-21. This information might be better summarized in a table.
Agreed and have added an additional suppl. table 1 (1c).

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?

Earlier cross-reactivity tests reported in Foster et al. 2007 found no cross-reaction between the het-3 and the other het groups, so we felt it was redundant to repeat. This detail is added into the text. (Pg 11, Lines 6-8).

Pg 11 lines 8-10. Is there a particular reason why assays weren't performed for these stations?
Overall qPCR assays 'at sea' were limited by time. The main reason for not performing qPCR at these stations was that they were not possible LD station candidates due to geographical and/or hydrological reasons. A short sentence has been added in the text (Pg 12 Line 5).

Pg 11 lines 22-24. What were the efficiencies of the other assays?
The efficiencies were only tested on the het groups as we were trying to be conservative with the limited template (Pg 12 Line 19)

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.
Agreed and have moved the text. Although it should be noted that the intention of this paper was not to be a 'methods' paper, and we are subsequently working on developing better qPCR assays which will be summarized in a separate body of work at a later date.

Pg 13 Line 7. Table 1 contains values for DIN-should they be bq?
We acknowledge the inconsistency and have amended the table where the bq is defined.

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.
We agree with the statement that a comparison of the extraction method is necessary if one was to truly compare the results. However here our intention is to report the differences, which we attained. The differences in abundance could be derived from the difference in extraction method, the variation of extraction efficiency per target, the patchiness of plankton, etc. It is not clear how the reviewer suggests we correct this section.

Pg 14 Lines 1-5. Explanation for the differences in the at sea and the lab based. Can you explain these results? This needs to be discussed further on pages 20-21.
There are a few possibilities here, and are explained in the text (pg 22-23, Lines 20-25; Lines 1-11).

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.
Agreed and have modified the text throughout section 3.4. Note that the depth of maximum abundances is the average from the two regions (MA and SG).

Pg 15. Lines 4-6. Please revise this sentence for clarity.
Agreed and have made it several sentences. (Pg 16, Lines 16-19).

Pg 15 Lines 7-11. The confusion with the LD and SD stations within the MA and SG.
Agreed and have followed the suggestion of the reviewer to indicate on Figures and also in the suppl. table 3.

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.
Noted and rephrased where appropriate. Although, we include the detail on samples when referring to the UCYN-A as it was striking that these were absent and/or patchy in distribution and we lose this context if we only refer to station.

Pg. 16 Section 3.4 indicate number of observations included when presenting the significant correlations.
Noted and fixed in the text.

Pg. 17 line 18. Perhaps indicate the significant clustering for group 1 and 2 on Figure 3 for clarity.
Agreed and amended the Figure with this detail.

Pg 18 lines 1-11. The RDA is explained very nicely, perhaps you should color code the dots in figure 4a to reflect the different response variables.
Figure 4a, the dots in the RDA represent samples and cannot be color coded according to diazotrophs, since they are represented by a vector (red labeled dot).

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.
Agreed, we have added bar graphs (one for each diazotroph) as an additional figure (Figure 5).

Pg 19 Lines 14-24. I suggest moving this section to 3.2 of the results. Do you have the data for the "viceversa" e.g. UCYN-A2 host assay with the UCYN-A1 host target, and the het-2 assay with the het-1 target? This is not obvious from Supplementary Figure 1.
We did run the similar assays in the reverse ('vice versa') and have added the additional graphs in supplementary figure 1. Originally we limited the presentation since the results were similar and we were trying to keep things simplified. (Pg 14, *section 3.2)*

Discussion:

Pg 21 lines 2-8. Please discuss these results more thoroughly.
Agreed and have added text (Pg 22-23, Lines 13-4).

Specifically, can you comment on potential differences in DNA extraction efficiency between the two methods?
Agreed that this would be a valuable piece of information however we did not determine the DNA extraction efficiency, and it would also be quite difficult since

these are mixed community field samples. One would need a known abundance of a particular target e.g. lab based culture work could address this for a particular target that has been isolated e.g. UCYN-B and Trichodesmium.

Are you comparing the same diazotroph community (e.g. from the same Niskin bottle/homogenized samples).
Samples were always taken from the same niskin bottle, but their entire volume was not homogenized.

Were there any inhibitors?
No inhibitors were added.

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 be at least discussed.
Agreed and further details have been added to the text (Pg 22 Lines 19-23).

Pg 21. Lines 23-25. Do you have a hypothesis as to why you observed these surprising results?
Assuming that the reviewer is referring to the restriction of the UCYN-A to one depth of LD C, perhaps one could hypothesize that it was a small-scale bloom or entrainment along an isocline. However we prefer to be conservative since it was only one observation and leave the text as is without drawing a larger conclusion.

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 comparisons?
We are not sure how to address this suggestion. Is it that we should provide the exact nifH copy numbers for all the 11 datasets?

Pg 24 lines 1-8. Indicating that DDAs are important for export production in this region, like the NPSG.
Agreed, added a sentence (Pg 26 Lines 13-14).

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.
The cited studies for UCYN-C have used qPCR (Turk-Kubo et al., 2015) and isolation, then subsequent 16S rRNA and nifH sequencing (Taniuchi et al., 2012). The qPCR study, which was conducted within a lagoon of the Melanesian archipelago, found UCYN-C to partly dominate diazotroph abundances, and they used the same qPCR assay as we did (Foster et al., 2007). We have added a short summary statement about the UCYN-C assay, which was evaluated in Turk-Kubo et al. 2015 that reported the UCYN-C assay to quantify the majority (up to 85% of the sequences in their study) of UCYN-C phylotypes (Pg 26, Lines 23-24).

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.

UCYN-A has been shown to have a colder temperature optimum (and range) than the other cyanobacterial diazotrophs in this study, which could also drive a subsurface maximum. Moreover the distributions observed could be linked to viral infections and grazing by zooplankton, none of which was actually measured.

Pg 28-29 lines 23-19. It would be nice to see further discussion around these results of the meta-analysis: the similarities and differences to other regions and the local/environmental factors driving these patterns could be discussed.
Agreed. Added text (Pg 30-32 Lines 23-3).

Pg 30 Line 19. Could you provide the same context for the UCYN-A assay.
The cross reactivity for the two UCYN-A assays had a near perfect match when run as A2 assay with A1 standard, meaning that no matter the abundance of A1 (high or low), there would always be a risk of significant cross-reactivity and overestimation of A2 in the presence of A1.

Figures:

Figure 1. Does the white dashed line indicate the separation of the MA and SG? Please clarify. Would it be possible to overlay SST on the station map, as this was an important explanatory variable?
Yes, the dotted line separates the MA an SG and has been clarified in the caption. It is possible to overlay the map with SST but it will not be very informative in our relatively narrow region where the SST was mostly uniform (everything would be the same color). The large impact of temperature was mediated vertically rather than horizontally.

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).
Agreed

Figure 3. Indicate group 1 and group 2 on the hierarchical clustering for clarity.
Agreed. The two hierarchical clusters of group 1 (surface) and group 2 (subsurface) has been marked.

Figure 4. Color coordinated dots might help to support the text. Please include a y-axis label in 4b.
Agreed, however see previous reply in the results section regarding coloring the dots. Y-axis label (Variance) added.

I would also like to see the meta-analysis presented as a figure if possible.
Agreed and have added a new figure 5 which summarizes the meta-analysis.

A T-S plot as a supplementary figure would also help to distinguish the different water masses of the MA and SG.
This is a valid suggestion included a supplementary Figure 2 which has the upper 500m T-S plots for stations LD A and SD 15 (MA and SG, respectively). More details

on the water masses will likely be addressed by the physical oceanographers whom have submitted work on the hydrography in the special issue.

1 # Distribution and drivers of symbiotic and free-living

2 # diazotrophic cyanobacteria in the Western Tropical South

3 # Pacific

4 Marcus Stenegren[1], Andrea Caputo[1], Carlo Berg[2], Sophie Bonnet[3,4], Rachel A Foster[1]

5 [1] Stockholm University, Department of Ecology, Environment and Plant Sciences. Stockholm, Sweden

6 [2] Science for Life Laboratory/Department of Biology and Environmental Science, Linnaeus University, Kalmar,

7 Sweden

8 [3] Aix Marseille Université, Centre National de la Recherche Scientifique. Marseille/Noumea, New Caledonia,

9 France

10 [4] Mediterranean Institute of Oceanography, Institut de Recherche pour le Dévelopement. Marseille, France

11 *Correspondence to*: Marcus Stenegren (marcus.stenegren@su.se)

Abstract

The abundance and distribution of cyanobacterial diazotrophs were quantified in two regions (Melanesian archipelago, MA and subtropical gyre, SG) of the Western Tropical South Pacific using *nifH* qPCR assays. UCYN-A1 and A2 host populations were quantified using 18S rRNA qPCR assays including one newly developed assay. All phylotypes were detected in the upper photic zone (0-50 m), with higher abundances in the MA region. *Trichodesmium* and UCYN-B dominated, when detected *nifH* copies ranged 2.18×10^{2}-9.41×10^{6} and $1.10 \times 10^{2} - 2.78 \times 10^{6}$ copies L^{-1}, respectively. Het-1 (symbiont of *Rhizosolenia* diatoms) was the next most abundant (1.40×10^{1}-1.74×10^{5} *nifH* copies L^{-1}), and co-occurred with het-2 and het-3. UCYN-A1 and A2 were the least abundant diazotrophs and were below detection (bd) in 84 and 97, respectively, of 120 samples. UCYN-A1 and A2 were detected while their respective hosts were bd, suggesting a lower partner fidelity or free-living state. Pairwise comparisons of the *nifH* abundances and various environmental parameters supported two groups: deep dwelling (45 m) comprised of UCYN-A1 and A2 and surface group (0-15m) comprised of *Trichodesmium*, het-1 and het-2. Temperature and PAR were positively correlated the surface group; while UCYN-A1 and A2 were positively correlated with depth, salinity and oxygen. Similar results were identified in a meta-analysis of 11 external datasets, where temperature seemed to have a major impact. Combined, our results indicate that conditions favoring the UCYN-A symbiosis differ from those of diatom diazotroph associations and free-living cyanobacterial diazotrophs.

1 **1 Introduction**

2 Biological di-nitrogen (N_2) fixation is considered a major source of new nitrogen (N)
3 to oceanic ecosystems (Karl et al., 1997). N_2 fixation is an energetically expensive process,
4 where N_2 gas is reduced to bioavailable ammonia (Howard and Rees, 1996) and is performed
5 by a small but diverse group of bacteria and archaea. The nitrogenase enzyme, which is
6 encoded by a suite of *nif*-genes, mediates N_2 fixation (Jacobson et al., 1989; Young, 2005).
7 Nitrogenase has a high iron (Fe) requirement (Howard and Rees, 1996), and often N_2 fixers,
8 or diazotrophs, are Fe- limited (Kustka et al., 2003; Raven, 1988). Nitrogenase is also
9 sensitive to oxygen (O_2), which has been shown to negatively influence N_2 fixation efficiency
10 (Meyerhof and Burk, 1928; Stewart, 1969). Thus, autotrophic diazotrophs (e.g. cyanobacteria)
11 have evolved strategies, such as temporal and spatial separation of the fixation process, to
12 protect their nitrogenase from O_2 evolution during photosynthesis (Berman-Frank et al., 2001;
13 Haselkorn, 1978; Mitsui et al., 1986). N_2 fixation is widespread and occurs in marine, limnic
14 and terrestrial habitats. In marine ecosystems it mainly occurs in the photic zone, closest to
15 the surface, however, more recently, evidence has shown activity in deeper depths below the
16 photic zone, including oxygen minimum zones (Benavides et al., 2016; Bonnet et al., 2013;
17 Fernandez et al., 2011; Halm et al., 2009; Löscher et al., 2015).

18 N_2 fixation in the photic zone is often attributed to a diverse group of cyanobacteria.
19 Traditionally, marine, photic dwelling diazotrophs are divided into two groups based on cell
20 diameter, e.g. > 10 μm and < 10 μm size fractions. Diatom diazotroph associations (DDAs),
21 symbioses between heterocystous cyanobacteria and a variety of diatom genera, and large
22 filamentous non-heterocystous *Trichodesmium* spp., compose the larger size fraction (>10
23 μm). *Trichodesmium* spp. occurs as free filaments or often in two morphologies of colonies:
24 tufts/rafts and puffs. There are three defined lineages of symbionts of DDAs based on their
25 *nifH* phylogeny: het-1 and het-2 refers to the two the *Richelia intracellularis* lineages which

associate with diatom genera, *Rhizosolenia* and *Hemiaulus,* respectively, while the third lineage, het-3, is a symbiosis between the heterocystous *Calothrix rhizosoleniae* and *Chaetoceros compressus* diatoms (Foster et al., 2010; Foster and Zehr, 2006).

The unicellular diazotrophic cyanobacterial groups are divided into: UCYN-A, UCYN-B, and UCYN-C groups and are representatives of the <10 µm size fraction. The UCYN-A (*Candidatus Atelocyanobacterium thalassa*) group can be further delineated into 6 sublineages (Thompson et al., 2014; Turk-Kubo et al., 2017), two (UCYN-A1, UCYN-A2) are identified as symbiotic with small prymnesiophyte microalgae (reviewed by Farnelid et al., 2016, see references within). The UCYN-B group has its closest cultured relative as *Crocosphaera watsonii* and lives freely, colonially, and also in symbiosis with the diatom *Climacodium frauenfeldianum* (Bench et al., 2013; Carpenter and Janson, 2000; Webb et al., 2009; Zehr et al., 2001). Often overlooked, is the observation that UCYN-B, when colonial or symbiotic could also be associated with the > 10µm size fraction. Less is known about the UCYN-C, and given that its *nifH* nucleotide sequence is 90% similar (Foster et al., 2007) to *Cyanothece* spp. ATCC51142, it is assumed to be analogous, and thus co-occur with the other < 10 µm size fraction. A diverse group of free-living heterotrophic bacteria (e.g. gamma proteobacteria) (Berthelot et al., 2015; Bombar et al., 2016; Halm et al., 2012; Langlois et al., 2005) and archaea (Zehr et al., 2005) are also within the < 10 µm size fraction.

The distribution and activity of diazotrophs in open ocean ecosystems are governed by different ambient environmental factors, including macronutrient availability (Moutin et al., 2008; Sañudo-Wilhelmy et al., 2001) and temperature (Messer et al., 2016; Moisander et al., 2010). There are also simultanous influences by several factors (i.e. co-limitation of nutrients, Mills et al., 2004). Moreover, most oceanic models of N_2 fixation assume that all diazotrophs are equally controlled by the same environmental parameters (Deutsch et al., 2007; Hood et al., 2004; Landolfi et al., 2015), despite well recognized differences in genetic repertoires for

assimilating dissolved nutrient pools (e.g. dissolved organic phosphate, Dyhrman et al., 2006; Dyhrman and Ruttenberg, 2006), life histories (free, symbiotic, colonial), and cell sizes (µm to mm). A better understanding of these differences would benefit oceanic models of N_2 fixation, our understanding of marine N_2 budgets and the impact of N_2 fixation in oceanic surface waters and their communities.

The Tropical South Pacific Ocean (TSP) is considered one of the most oligotrophic regions in the World's oceans (Claustre and Maritorena, 2003) with a widespread N deficiency (Deutsch et al., 2007; Raimbault et al., 2007) and in the central SP gyre, some of the lowest concentrations of dissolved Fe in the world have been reported (Blain et al., 2008). One exception is the Western Tropical South Pacific (WTSP), harboring many islands with Fe rich sediments adding to an island mass effect (Shiozaki et al., 2014) and being influenced by multiple ocean currents, both surface and subsurface, that drive the distribution of dissolved nutrients, micronutrients, and the biota (Fitzsimmons et al., 2014; Gourdeau et al., 2008; Marchesiello and Estrade, 2010; Wells et al., 1999). The structure of these currents also promotes shearing instabilities and strong eddies (Qiu et al., 2009). Moreover, Van Den Broeck et al. (2004) suggested that the WTSP is phosphate limited, while Law et al. (2011) hypothesized that primary production and N_2 fixation in the WTSP follows the seasonality of cyclones, which in their wake, enrich surface waters with phosphate, and fuel primary and new production. An earlier investigation along a transect in the western equatorial Pacific estimated that 74% of the total N_2 fixation could be attributed to the <10 µm size fraction as abundances of unicellular cyanobacteria were high (17 cells mL^{-1}) (Bonnet et al., 2009). However, diazotroph quantification is lacking further South in tropical waters, despite being recently recognized as a hot spot of N_2 fixation, with average rates of ~570 µmol N m^{-2} d^{-1} (Bonnet et al., this issue), i.e. in the upper range (100-1000 µmol N m^{-2} d^{-1}) of rates gathered in the global N_2 fixation MAREDAT database (Luo et al., 2012).

1 The primary aim of this study was to quantify diazotroph abundance and distribution in
2 the WTSP with an emphasis on symbiotic N_2-fixing populations; both by ´at sea´ and lab
3 based quantitative approaches. Since earlier expeditions to the SW Pacific identified it as a
4 potential 'hotspot' for diazotrophy (Bonnet et al., 2015; 2016; 2017; Moisander et al., 2010),
5 the 'at sea' qPCR was performed in order to make informed decisions about the presence,
6 absence, and relative abundance of diazotrophs so that stations could be selected for other
7 assays and characterization of the hydrographic conditions (e.g. incubation experiments,
8 sediment and drifter deployments (see Bonnet et al. this issue). For a more comprehensive
9 investigation of the symbiotic diazotrophs we developed a new primer and probe set for
10 quantification of the UCYN-A1 host. We also identified key environmental parameters, both
11 biotic and abiotic, which influenced the distribution of diazotrophs in the WTSP and tested
12 the congruency of these parameters in an additional 11 publicly available datasets. We
13 hypothesized that the distribution and the underlying factors of the diazotrophic symbioses
14 should differ due to the major differences in host taxonomy (e.g. diatom vs. prymnesiophyte),
15 size (1-2 µm to 100's µm), and life history (free vs. symbiotic; chain forming). For
16 comparison and for similarly divergent characteristics (symbiotic vs. free; colonial vs. single),
17 several free-living (UCYN-B, *Trichodesmium* spp.) cyanobacterial diazotrophs were also
18 included.

19 ## 2 Materials and Methods

20 ### 2.1 Sampling

21 Sampling was conducted on a transect in the WTSP during austral summer (19 Feb-5 Apr,
22 2015), on board the R/V *L'Atalante* (Fig. 1a). Nucleic acid samples were taken from 18
23 stations: three long duration (LD A, B and C) stations (approximately eight days duration) and
24 15 short duration (SD 1-15) stations (approximately eight hours duration). The LD stations
25 were selected according to the following: regions with low advection to better estimatee

1 biogeochemical budgets, contrasting diazotrophic community (e.g. *Trichodesmium* spp.
2 dominated vs. UCYN dominated), and along a trophic and N_2 fixation gradient from west to
3 east (Moutin et al., this issue). The cruise transect was divided into two geographic regions
4 (Fig. 1a). The first region (Melanesian archipelago, MA) included SD 1-12, LD A and LD B
5 stations (160º E-178º E and 170º-175º W). The second region (subtropical gyre, SG) included
6 SD 13-15 and LD C stations (160º W-169º W). LD stations were chosen based on
7 hydrographic conditions, satellite imagery, microscopic analyses of >10 µm cyanobacterial
8 diazotrophs and the results of 'at sea' qPCR analyses of four unicellular diazotrophic targets
9 (UCYN-A1, UCYN-A2, UCYN-B and UCYN-C) (see below and Moutin et al., this issue).

10 **2.1.1 Nucleic acids**

11 Seawater (2.5 L) was collected into clean (10% bleach rinsed) 2.75 L polycarbonate bottles
12 from 6-7 discrete depths based on surface incident light intensity (100, 75, 54, 36, 10, 1, and
13 0.1%) once per station at both SD and LD stations using Niskin bottles (12 L) arranged on a
14 Conductivity Temperature Depth (CTD; Seabird 911) rosette.

15 After collection from the CTD rosette, 2.5L seawater was immediately filtered onto a
16 0.2 µm pore size Supor filter (Pall Corporation, Pall Norden AB, Lund, Sweden) held within a
17 25 mm diameter swinnex filter holder (Merck Millipore, Solna, Sweden) using a peristaltic
18 pump (Cole-Parmer, Masterflex, Easy-load II, USA). In case the filters clogged with too
19 much material, the remaining volume was measured and noted for later calculations. The
20 filters were placed in pre-sterilized bead beater tubes (Biospec Bartlesville, OK, USA)
21 containing 30 µL of 0.1 mm and 0.5 mm glass bead mixture, flash frozen in liquid nitrogen
22 and archived at -80 ºC. Four additional DNA samples were collected from 4 discrete depths,
23 (75, 50, 36, 10 % light), at 11 of the 18 stations, for the 'at sea' qPCR (see below) and filtered
24 as described above. The 'at sea' samples were extracted and processed for qPCR immediately
25 after collection (see below).

2.1.2 Cell abundances and microscopy observations

At the LD stations, 5 L of seawater was collected at the same depths in parallel with the nucleic acid samples from the CTD-rosette. Two sets of samples, one set each day, were taken on day 1 and 3 at each LD station and immediately filtered onto a 47 mm diameter Poretics (millipore) membrane filter with a pore size of 5 µm using a peristaltic pump.

At the SD stations, the same collection was implemented, however a 25 mm diameter Poretics membrane filter was used. The high densities of cells on the latter made it impossible to properly enumerate the various cyanobacterial diazotrophs and as such these samples were used only for qualtitative observations (see below). Immediately after filtration, samples were fixed in 1 % paraformaldehyde (v/v) for 30 min prior to storing at -20 ºC. For enumeration, the filter was mounted on a glass slide and examined under an Olympus BX60 microscope equipped with a filter for blue (460-490 nm) and green (545-580 nm) excitation wavelengths. Three areas (area = 0.94 mm^2) per filter were counted separately and values were averaged. When abundances were low, the entire filter (area = 1734 mm^2) was observed and cells enumerated. Due to weak fluorescence, only *Trichodesmium* colonies and free-filaments could be accurately estimated by microscopy and in addition, the larger cell diameter *Trichodesmium*, herafter referred to as *Katagynemene pelagicum*, was enumerated separately as these were often present albeit at lower cell densities. Other cyanobacterial diazotrophs, e.g *C. watsonii*-like, *C. rhizosoleniae* (het-3), and *R. intracellularis* (het-1, het-2) were also present on the larger 47 mm diameter samples, however fluorescence was weak and therefore difficult to enumerate. Pico-eukaryote populations, identified as round 1-3 µm diameter cells, with red excitation under the blue filter set, were also observed. For the latter populations, qualitative observations of presence and some details on cell integrity (e.g. fluorescence, diatom frustule, free-living or symbiotic form) are included.

1 **2.2 Nutrient analyses**

2 Seawater for nutrient analyses was collected from each station using the CTD rosette at the
3 same depths as those collected for the nucleic acids. Seawater for inorganic nutrient analysis
4 were collected in 20 mL high-density polyethylene HCL-rinsed bottles and poisoned with
5 $HgCl_2$ to a final concentration of 20 μg L^{-1} and stored at 4°C until analysis. The samples were
6 returned to the laboratory frozen within 4 months for analysis. Dissolved nitrate and nitrite
7 (NO_3^-+NO_2^-, DIN), phosphate (PO_4^{3-}, DIP) and silicate ($Si(OH)_4$, DiSi) concentrations were
8 determined by standard colorimetric techniques using a segmented flow analyzer according to
9 Aminot and Kérouel (2007) on a SEAL Analytical AA3 HR system (SEAL Analytica,
10 Serblabo Technologies, Entraigues Sur La Sorgue, France). Quantification limits for nitrate,
11 phosphate and silicate were all 0.05 μmol L^{-1}.

12 **2.3 DNA extraction**

13 The DNA from the 120 archived samples was extracted as described in Moisander et al.
14 (2008), with a 30 second reduction in the agitation step in a Fast Prep cell disrupter (Thermo,
15 Model FP120; Qbiogene, Inc. Cedex, France) and an elution volume of 70 µL. The nucleic
16 acid samples collected for the 'at sea' qPCR were extracted immediately after filtration using
17 a modified version of the DNAeasy plant kit (Qiagen) total DNA extraction protocol. The
18 modifications were an initial 2-minute agitation step using a bead beater (Biospec
19 MiniBeadBeater-16, Model 607EUR; Biospec), 10 min proteinase K incubation and final
20 elution volume was 25 μl.

21 **2.4 Oligonucleotide design**

22 A new primer and probe set was designed to amplify the UCYN-A1 host and was based on
23 published 18S rRNA sequence (accession number JX291893) reported from N. Pacific gyre
24 (station ALOHA) (Thompson et al., 2012). The design utilized the same 96 bp target region of
25 the 18S rRNA used to amplify UCYN-A2 hosts described in Thompson et al. 2014 (Suppl.

1 Table 1). The primers and probe for the UCYN-A1 host 18S rRNA gene assay are as follows:
2 Forward, 5' AGGTTTGCCGGTCTGCCGAT-3'; Reverse, 5'
3 GAGCGGGTGTCGGAGACGGAT-3'; Probe, 5'-FAM-CTGGTAGAACTGTCCT-
4 TAMRA-3'. The forward, reverse and probe contain 2-4, 1, and 5 mismatches, respectively,
5 to UCYN-A2 host sequences (accession number KF771248-KF771254) and the following
6 closely related sequences (98-100%): uncultured eukaryote clones (station ALHOA:
7 EU50069; Cariaco Basin: GU824119) *Chrysochromulina parkeae*: AM490994),
8 *Braarudospaera bigelowii* TP056a: AB250784 *B. bigelowii* Furue-15: AB478413; *B.*
9 *bigelowii* Funahama T3: AB478413; *B. bigelowii* Yastushiro-1 AB478414. The UCYN-A1
10 oligonucleotides specificity was tested *de nova* against the following closely related sequences
11 derived from uncultured eukaryotic clonal sequences (accession numbers: EU500067-68;
12 FJ537341; EU500138-39; EF695227; EU500141; EU499958; EF695229; EF695220). Only
13 one mismatch was found in the forward probe for one sequence (EU500138) (Suppl. Table 1).
14 The nucleotide sequence identity for UCYN-A1 and UCYN-A2 host sequences is 97.95%.
15 Finally, a cross reactivity test between the newly designed UCYN-A1 host oligonculeotides
16 and a dilution series of the UCYN-A2 host template was run (see below).

17 **2.5 Quantitative PCR**

18 Abundances of selected diazotrophs *nifH* gene copies (UCYN-A1, UCYN-A2, UCYN-B,
19 UCYN-C, het-1, het-2, het-3 and *Trichodesmium* spp.) and the 18S rRNA of UCYN-A1 and
20 A2 hosts were performed using previously published oligonucleotides and TaqMAN assays
21 (Church et al., 2005; Foster et al., 2007; Moisander et al., 2010; Thompson et al., 2014) and
22 the newly designed UCYN-A1 host oligonculeotides (Suppl. Table 1). The qPCRs were
23 conducted in a StepOnePlus system (Applied Biosystems, Life Technologies, Stockholm
24 Sweden) in fast (>40 min) mode with the following parameters: 95 ºC for 20 s, followed by
25 45 cycles of 95 ºC for 1 s and 60 ºC for 20 s.

1 Cross reactivity tests were run on two of the heterocystous symbiont (het-1 and het-2)
2 oligonucleotides, the UCYN-A1 and UCYNA-2 oligonucleotides, and the newly designed
3 UCYN-A1 host oligonucleotides and UCYN-A2 host primer and probe set. Cross reactivity
4 tests for het-3 were omitted as previously published tests (Foster et al., 2007) showed no
5 cross-reaction with het-1 or het-2. The standard curve for a particular target was run in
6 reactions with the other primers and probe sets. For example, the UCYN-A1 TaqMAN host
7 primers and probes were run in reactions with UCYN-A2 template DNA. The cross reactivity
8 for the het-1 and het-2 primer and probe sets has been previously reported (Foster et al. 2007),
9 however only when the assay is run in standard mode. Standard mode runs the holding,
10 denaturation and annealing stages at the following longer intervals than in Fast mode: 11 min
11 and 40 s, 14 s, and 40 s, respectively. Hence, we tested the cross-reactivity for the het primers
12 and probes when run in fast mode, as the fast mode was used in our study. Similarly, the
13 cross-reactivity between UCYN-A1 and UCYN-A2 were tested in fast mode at two annealing
14 temperatures 60 ºC and 64 ºC; 64 ºC is the recommended annealing temperature for the
15 UCYN-A2 assay (Thompson et al. 2014).

16 Reaction volume was 20 µL in all qPCRs and consisted of 10 µL of 2X TaqMan fast buffer
17 (Applied Biosystems, 5.5 µL of nuclease free water (Sigma Aldrich Sweden AB, Stockholm
18 Sweden), 1 µL each of the forward and reverse primers (10 µM), 0.5 µL of fluorogenic probe
19 (10 µM) and 2 µL of DNA extract. For standard mode runs, the latter master mix was
20 identical with the exception of replacing the fast 2X buffer with the standard 2X buffer. For
21 reactions quantifying *Trichodesmium* spp. *nifH* copies, SD 9 was excluded and 1 µL of DNA
22 template was used for the remaining stations due to low template volume, and total reaction
23 volume was adjusted by addition of 1 µL of nuclease free water. Reactions were performed in
24 duplicates for the 'at sea' qPCR and in triplicates for the archived samples and lab based
25 qPCR. For the 'at sea' qPCR, only four targets (UCYN-A1, UCYN-A2, UCYN-B, and

UCYN-C) were quantified and only at the SD stations. No assays were processed at SD 5-6, 10-12, and 14 for the 'at sea' qPCR due to no potential LD station selection for these sites. Two µL of nuclease free water was used as template in no template controls (NTCs); no *nifH* copies were detected in the NTCs.

Gene copy abundance was calculated from the mean Ct value of the 3 replicates and the standard curve for the appropriate oligonucleotides in the lab based qPCRs. For the 'at sea' qPCR, a mean Ct value of 2 replicates was used to maximize the number of samples run on one amplification plate (96 well). In samples where 1 or 2 out of 3 replicates produced an amplification, signals were noted as detectable, but not quantifiable (dnq) and no amplification was noted as below detection (bd).

2.6 Standard curves and PCR efficiency

Standard curves were plotted and analyzed in Excel for each target based on the qPCR cycle threshold (Ct) values from known dilutions of synthesized target gene fragments (gBlocks®; Integrated DNA Technologies, Leuven Belgium) (359 bp *nifH* and 733 bp 18S rRNA for UCYN-A hosts). Tenfold dilutions were made starting with 10^8 to 10^1 gene copies L^{-1}. The PCR efficiency, for identification of possible interfering contaminants in our samples, was determined as previously described (Short et al., 2004) for 12 samples run on the het-1, het-2, and het-3 primers and probe tests. The qPCR efficiency ranged from 90-99 % with an average of 94 % efficiency for the diazotroph targets het-1, het-2 and het-3.

2.7 Statistics and data analysis

Skewness and normal distribution tests by descriptive statistics was performed in IBM SPSS (ver. 23) on the following parameters recorded during sample collection in the WTSP from the CTD package: depth (m), oxygen (ml L^{-1}), temperature (ºC), chlorophyll fluorescence (µg L^{-1}), photosynthetically active radiation (PAR; µmol photons m^{-2} s^{-1}), salinity (PSU), and gene

1 copy abundances determined by qPCR. Significant skew was noted when skewness, divided
2 by its standard deviation, exceeded 1.95. All but three targets (het-1, UCYN-B and
3 *Trichodesmium* spp.) and three environmental parameters (temperature, salinity and oxygen)
4 were significantly skewed (not normally distributed) even after LOG10 transformation.
5 Therefore a non-parametric Spearman's rank correlation was conducted to test possible
6 correlations between the targets and environmental parameters, where we assume that the het
7 groups and UCYN-A clade is symbiotic, while UCYN-B is free living. The resulting
8 correlation matrices were visualized in the form of a heat map of hierarchical clustering in R
9 (ver. 3.2.2) using packages 'hmisc' and 'gplots'. Multivariate statistics by redundancy
10 analysis (RDA) was conducted using the R package 'vegan'. T-tests, in IBM SPSS (ver. 23)
11 were performed to characterize the different regions along the cruise transect based on
12 environmental parameters, including nutrients, measured between stations and was reported as
13 mean concentrations. For meta-analysis on the external dataset from 11 publically available
14 datasets, sampled in the Atlantic, Pacific and South China Sea, data was acquired from the
15 PANGAEA database and previous publications (Benavides et al., 2016; Bombar et al., 2011;
16 Church et al., 2005, 2008, Foster et al., 2007, 2009; Goebel et al., 2010; Kong et al., 2011;
17 Langlois et al., 2008; Moisander et al., 2008, 2010). We included only datasets with a
18 minimum of 10 datapoints on the previously mentioned diazotrophic targets. Note that in all
19 datasets the two UCYN-A phylotypes (A1 and A2) were not distinguished, and het-3 was
20 excluded since it was rarely quantified. The meta-analysis was conducted using the software
21 OpenMEE (Wallace et al., 2016) (based on R package 'metafor'), where correlation
22 coefficients from Spearman's rank were z-transformed (Fisher's) and tested using weighted
23 random effect models. Graphical visualization of the mean abundances of the most numerous
24 diazotrophs across the cruise transect was also performed in IBM SPSS (ver. 23).

3 Results

3.1 Hydrographic conditions

Near surface (0-5m) DIN concentrations were below the quantification limit (bq) in both the MA and SG regions, while the mean surface DIP and DiSi concentrations were below the quantification limit or low across all stations in the MA (bq-0.08 µM and 0.54-0.56 µM, respectively) and significantly ($p<0.001$; t-test) higher (0.18 ± 0.07 µM and 0.79 ± 0.04 µM, respectively) at the stations in the SG (Table 1). The upper 25-30 m of depth throughout the cruise transect had stable temperatures of 29-30 ºC, but declined differently in deeper waters of the MA compared to the SG (Suppl. Fig. 2). The depth of the deep chlorophyll maximum (DCM) was between 70-165 m, except for LD B (DCM at 35 m), which was sampled during a degrading surface phytoplankton bloom, and a 30-day composite of the surface chlorophyll *a* (Chl *a*) confirmed the decreasing level of surface fluorescence measured by the CTD package at LD B (data not shown).

3.2 Cross reactivity tests

No amplification was detected for the newly designed UCYN-A1 host oligonucleotides run with the UCYN-A2 as template DNA and vice versa (Suppl. Fig. 1a-b).

Running the het assay in fast mode showed a lower cross-reactivity between the het-1 assay and the het-2 template than vice versa (the het-2 assay and het-1 template) (Suppl. Fig. 1e-f). In fact, no amplification was detected in the last two template additions and the Ct differences were > 9 when het-1 assay was run with het-2 templates. The UCYN-A2 assay detected the UCYN-A1 template in all but the last template addition and with Ct differences >3 (1 order of magnitude) while there was a 18-20 difference in Ct value (less gene copies) when UCYN-A1 assay was run in fast mode with UCYN-A2 templates at either annealing temperature (60° C or 64° C) and only the first three template additions (10^8-10^6 *nifH* copies µL^{-1}) were detected (Suppl. Fig. 1c-d).

1 **3.3 Comparison of 'at sea' and lab-based qPCR**

2 In order expedite the sample processing for the 'at sea' qPCR, a shortened and modified DNA
3 extraction protocol was performed, 4 depths were sampled, and 4 targets run (UCYN groups).
4 In total, 44 samples can be compared with results from the parallel archive samples and we
5 considered only when there was at least one order of magnitude difference in detection. A
6 summary of the comparison, including the difference in *nifH* copy abundance is provided in
7 Suppl. Table 2.

8 In general, the 'at sea' and lab based qPCR were similar in quantifying the targets.
9 Discrepancies were noted in 7, 8 and 11 samples, which had higher detection in the 'at sea'
10 analyses for UCYN-A1, UCYN-A2 and UCYN-B, respectively. There were fewer instances
11 (3, 4, and 5, respectively) of samples processed in the lab with the full extraction that had
12 higher abundances for the UCYN-A1, UCYN-A2 and UCYN-B, respectively.

13 **3.4 Horizontal and vertical distributions**

14 *Trichodesmium* and UCYN-B were the most abundant diazotrophs and abundances ranged
15 10^4-10^6 *nifH* copies L^{-1} at multiple depths (4-6 depths) in the upper water column (0-35 m)
16 (Fig. 1-2; Suppl. Table 3). *Trichodesmium* represented 80-99% of total *nifH* genes detected at
17 9 out of 17 stations with highest detection in the MA and low to bd in the SG. Microscopy
18 observations and abundances of *Trichodesmium* spp. confirmed a high abundance of free
19 filaments of *Trichodesmium* and *C. watsonii*-like cells at LD B, while colonies were in
20 general rarely observed (Suppl. Table 5).

21 At stations where *Trichodesmium* was not the most abundant diazotroph (e.g. SD 2, 6,
22 7, 14, 15, and LD C), UCYN-B had the highest depth integrated *nifH* copy abundance.
23 UCYN-B was also the most consistently detected diazotroph, and was quantifiable from all
24 stations sampled accounting for for 81-100% of the total detected *nifH* gene copies in the SG.

1 There was also a depth dependency for maximum abundance such that the average depth
2 maximas of *Trichodesmium* (1.88 x 10^6 *nifH* copies L^{-1}) and UCYN-B (5.03 x 10^5 *nifH* copies
3 L^{-1}) at the stations in the MA were 10 and 25 m, respectively. In the SG, the average depth
4 maximum for UCYN-B (1.50 x 10^5 *nifH* copies L^{-1}) was the same (25 m), while the average
5 depth of the *Trichodesmium* maximum (1.30 x 10^4 *nifH* copies L^{-1}) deepened to 31m.

6 Of the three heterocystous cyanobacterial symbiont lineages (het-1, het-2, het-3), het-1
7 was the most dominant (60% detection in total samples, 72 of 120 samples), and similar to
8 *Trichodesmium*, had higher detection in the stations of the MA region. For example, at
9 stations SD 2, 4 and 9, het-1 represented 10-15% of the total *nifH* genes quantified in the
10 depth profiles, but in the total *nifH* genes quantified across the entire transect, het-1 only
11 represented 1.5 %. Abundances for het-1 ranged between 10^3-10^5 *nifH* copies L^{-1} (15 of the 18
12 stations) at multiple depths (0-90 m) and the average depth maximum at MA stations (3.11 x
13 10^4 *nifH* copies L^{-1}) was closer to the surface (15 m) compared to the SG stations (1.63 x 10^3
14 *nifH* copies L^{-1}; 60 m) (Fig. 1; Suppl. Table 3). Het-2 and het-3 co-occurred with het-1,
15 however at lower abundances (10^2-10^4 *nifH* copies L^{-1}) and often bd. For example, het 2 and
16 het-3 were bd at all depths sampled in 1 and 3 stations, respectively, located in the SG. The
17 averge depth of maximum abundance (17 m) for het-2 (3.89 x 10^3 *nifH* copies L^{-1}) was similar
18 to het-1 (15 m), while het-3 (1.53 x 10^3 *nifH* copies L^{-1}) was deeper at 33 m (considering only
19 the MA stations). Microscopy observations confirmed the presence of *R. intracellularis* at 5
20 SD stations of the MA and LD B and absence at the SD stations and LD C of the SG.
21 Noticeable was the co-occurence of free filaments of *R. intracellularis* and degrading diatom
22 cells (mainly belonging to the genus *Rhizosolenia*), especially at the SD 5, 6 and 7.

23 The unicellular symbiotic groups, UCYN-A1 and A2 (and their respective hosts), were
24 the least detected targets. For example, UCYN-A1 was bd in 53% (63 of 120 samples) and
25 UCYN-A2 was bd in 66% (79 of 120 samples) of samples. UCYN-A1 and A2 represented <

0.4 % of total *nifH* genes detected and UCYN-A symbionts were bd in the SG, except at LD C. When detected, average *nifH* abundance for UCYN-A1 and A2 were 8.60 x 10^4 and 4.60 x 10^4 *nifH* copies L^{-1}, respectively, and usually accounted for <1.0-1.5 % of the total *nifH* copies enumerated per station. One exception was at LD C in the SG, when UCYN-A1 and A2 accounted for 4 and 11%, respectively, of the total *nifH* gene copies, and were the second most abundant diazotroph (3.19 x 10^4 and 8.53 x 10^4 *nifH* copies L^{-1}). The average depth of maximum *nifH* abundance for the UCYN-A1 (1.60 x 10^4 *nifH* copies L^{-1}) and A2 (5.76 x 10^3 *nifH* copies L^{-1}) symbionts was consistently recorded at deeper depths (55 and 58 m, respectively; 10 % light level).

The detection of the UCYN-A1 and A2 hosts mirrored the detection of their respective symbionts. However, in 22 and 15 samples, respectively, the UCYN-A1 and A2 symbionts were quantified while their hosts were bd. The UCYN-A hosts were never detected in samples where their respective symbionts were bd or dnq. When both UCYN-A host and symbiont were present, the abundances of the hosts were always one order of magnitude less than their respective symbionts, with the exception of two samples for UCYN-A1 symbionts where their respective host abundances were half, or nearly equal in abundance. UCYN-C was the least abundant unicellular diazotroph and was only quantified in the 'at-sea' qPCR where detection was poor and limited to the MA region (3 of 11 stations: 1-3 of 4 depths sampled) and abundances never exceeded 10^2 *nifH* copies L^{-1} (Suppl. Table 3).

3.5 Diazotroph and UCYN-A host covariation

Several significant correlations between the target diazotrophs and hosts were identified (Fig. 3; Suppl. Table 4a). The *nifH* gene copy abundances of *Trichodesmium* and UCYN-B were significantly positively correlated with each other (N=108, $p<0.01$). In addition, UCYN-B *nifH* gene copy abundance was significantly positively correlated with those of both UCYN-A symbionts (A1 and A2; N=120, $p< 0.01$) and UCYN-A2 host abundance (N=118, $p<0.04$).

Abundances of UCYN-A1 and A2 were significantly positively correlated with each other, and in addition, with their respective host abundances (N=118-120, $p<0.01$). Lastly, the *nifH* copy abundances for het-1, het-2 and het-3 were significantly positively correlated with one another, and with the *nifH* copy abundances of *Trichodesmium* and UCYN-B (N=108-120, $p<0.01$). The only correlations that were not significant were between the UCYN-A (including their hosts) and *Trichodesmium* and the het-groups (with the exception of het-3, which correlated with the UCYN-A2 host (N=118, $p<0.04$)).

3.6 Influence of environmental conditions on diazotroph and UCYN-A host abundances in the WTSP

The abundances of UCYN-A1 and A2 were significantly positively correlated with salinity (N=107, $p<0.02$ and N=107, $p<0.03$, respectively) and depth (N=118, $p<0.01$ and N=118, $p<0.002$, respectively) (Fig. 3; Suppl. Table 4b). However, except for het-3 (regarding depth), all other diazotrophs were significantly negatively correlated with salinity (N=97-108, $p<0.006$) and depth (N=108-120, $p<0.001$). Moreover, *Trichodesmium*, UCYN-B, and the het-group (except het-3) were significantly positively correlated with PAR (N=63-72, $p<0.003$) and temperature (N=97-108, $p<0.001$) while UCYN-A1 and A2 were significantly negatively correlated (N=72-108, $p<0.02$) with the latter parameters. All diazotrophic targets, except UCYN-A1, UCYN-A2, and het-3, were significantly negatively correlated with DIN concentration (N=90-94, $p<0.02$). Similarly, all diazotrophs, except UCYN-A2, were significantly negatively correlated with DIP concentration (N=96-100, $p<0.02$), and all diazotrophs except UCYN-A1, A2 and het-3 were significantly negatively correlated with DiSi concentration (N=97-102, $p<0.001$). The abundances of UCYN-A hosts, UCYN-A1 and UCYN-A2, and UCYN-B were significantly correlated with dissolved oxygen (UCYN-A and hosts, N=106-108, $p<0.003$ and UCYN-B, N=108, $p<0.03$). In general, the correlations

between abundances and several hydrographic parameters divided the diazotrophs into two groups: the UCYN-A symbionts (and respective hosts) and all other diazotrophs.

Hierarchical clustering based on the Spearman's rank analyses resulted in the two major groups: (1) a shallow and (2) deeper euphotic zone, inferred from the negative and positive correlations, respectively, with depth (Fig. 3). For example, *Trichodesmium* and the symbiotic het-1 and het-2 lineages characterize an upper water column group 1 with significant clustering and positive correlations with temperature ($N=97\text{-}108$, $p<0.001$) and PAR ($N=63\text{-}72$, $p<0.003$), while only UCYN-A1 and A2 symbionts and their respective hosts represent group 2. UCYN-B was unique in an overlapping distribution, and resulted in positive significant correlations with both the shallow (group 1) and deep (group 2) euphotic zone diazotrophs (e.g. *Trichodesmium*, $N=108$, $p<0.001$ and UCYN-A1, $N=120$, $p<0.004$, respectively). The deeper dwelling group 2 significantly clustered and correlated positively with oxygen, depth, salinity and fluorescence ($N=108\text{-}120$, $p<0.03$, except for UCYN-A2 and fluorescence, $N=108$, $p=0.053$). Despite clustering with group 1, het-3 was less robust in a negative correlation with salinity ($N=108$, $p=0.005$).

The results from the Spearman's rank correlations were further confirmed and visualized in the RDA biplot (Fig. 4a), which explains parameter importance (Fig. 4b). Correlations with nutrients and PAR were omitted due to the limited number of data points. Fluorescence, depth and salinity correlated positively with each other and negatively with temperature, while oxygen was not significantly correlated with any other environmental parameters. The response variables UCYN-A1 and A2 and their respective hosts clustered with the explanatory variables: fluorescence, salinity and depth, with a dependency towards oxygen. On the other hand, the shallower euphotic group 1 (response variables *Trichodesmium*, het-1 and het-2) clustered closer to explanatory varaible temperature. In addition, most of the observed variance is explained by the two axes RDA1 (72 %) and RDA2

(22 %), indicative of depth and temperature, respectively, as the most important environmental parameters for diazotroph abundance in our study. Together they form a depth-temperature gradient (RDA1) where *Trichodesmium* occupies the warmest and shallowest waters, and UCYN-A occupies the coldest and deepest waters, among the investigated cyanobacterial diazotrophs.

3.7 Global drivers of diazotrophic abundance

We found consistency between our results in the WTSP and the correlations identified in the 11 external datasets by the non-parametric correlation analyses and meta-analyses (Fig. 5a-d; Suppl. Table 6). For example, in three of the external datasets, abundances of *Trichodesmium* spp., UCYN-B, and het-1, were significantly positively correlated with temperature and negatively correlated with the same three parameters as in our study in the WTSP: salinity, DIP, and DIN. The latter correlations were identified in two regions of the WTSP (tropical and subtropical) and in the northern South China Sea (NSCS). In contrast to a significant positive correlation between UCYN-A abundance and depth reported here in the WTSP, UCYN-A abundance was negatively correlated with depth in 4 of the 11 external datasets (two regions of the WTSP, Tropical Atlantic (TA), and NSCS). Moreover, and consistent with several of the other diazotrophs (*Trichodesmium*, UCYN-B, het-1), UCYN-A abundance was negatively correlated with DIP and DIN concentrations (5 and 3 additional external datasets, respectively) (Suppl. Table 6).

Meta-analysis revealed similar groupings (e.g. shallow and deep) as observed in the WTSP, however, the significance was less robust (Fig. 5a-d; Suppl. Table 6). For example abundances of *Trichodesmium* and het-1 and het-2 were significantly positively correlated with temperature and negatively correlated with salinity ($p<0.05$). No significance was found for UCYN-A abundance for the latter parameters, and UCYN-B abundance was not significantly correlated with salinity and significantly positively correlated with temperature

1 ($p<0.05$). In addition, UCYN-A was the only diazotroph that was not significantly correlated
2 with het-2, while all other diazotrophs had a significant positive correlation with het-2
3 ($p<0.05$). Similar to our findings reported for the WTSP, all diazotrophs, except UCYN-A,
4 correlated significantly negatively with depth, DIP and DIN concentrations ($p<0.05$) (except
5 het-2 with DIP which was not significant). Finally, UCYN-B and het-1 abundances were
6 significantly negatively correlated with chl *a* ($p<0.05$), while *Trichodesmium*, UCYN-A and
7 het-2 were not.

8 **4 Discussion**

9 **4.1 Environmental conditions in the WTSP**

10 The SP is one of the most oligotrophic regions of the world's oceans with chronically low
11 dissolved nutrient concentrations, especially DIN, and thus, is considered an area primed for
12 N_2 fixation. Likewise, we encountered surface hydrographic conditions in the WTSP that
13 were consistently low in dissolved nutrient concentrations and similar to earlier reports for the
14 equatorial Pacific (Bonnet et al., 2009; Dufour et al., 1999; Moutin et al., 2008; Van Den
15 Broeck et al., 2004). The conservative tracers of temperature and salinity remained constant in
16 the surface between the MA and SG regions, hence the elevated nutrient concentrations in the
17 SG is likely not related to an eddy intrusion. The deviation away from a 16:1 relationship
18 (Redfield ratio) (data not shown) in the upper 125 m in both regions (MA and SG) was
19 indicative of DIN limitation. The low DIP concentrations in MA waters suggest utilization of
20 DIP by diazotrophs in the absence of DIN, and likely other sources of nitrogen were available,
21 e.g. dissolved organic nitrogen or N_2 fixation (Karl et al., 2001).

22 **4.2 Detection of diazotrophs and application of 'at sea' qPCR**

23 *Trichodesmium,* UCYN-B, and the het-groups are easily identifiable by standard epi-
24 fluorescence microscopy, and so these populations can readily be observed 'at sea'. However,
25 the UCYN-A1 and UCYN-A2, and their respective hosts, require a lengthy fluorescent *in situ*

1 hybridization (FISH) protocol that is difficult to implement in the field. On the other hand,
2 nowadays oceanographers have a suite of other molecular genetic tools, some of which are
3 also 'sea-going' and autonomous (e.g. Robidart et al. 2014; Ottesen et al. 2013; Preston et al.
4 2011), thereby making quantification of microscopically unidentified microorganisms
5 tangible by quantifying their genes, simultaneous with collection of hydrographic data. Here,
6 we showed a rather efficient, steadfast (within 3 hrs of sample collection), and 'sea-going'
7 nucleic acid extraction and qPCR to quantify diazotrophs by their *nifH* gene, which was used
8 in real time during the OUTPACE cruise to help locate the LD stations for the purpose of the
9 project (see Moutin et al., this issue). The comparisons of the 'at sea' assays to the lab-based
10 full extraction protocol and qPCR on archived samples indicated that the assays were
11 consistent (Suppl. Fig. 3a-c), and surprisingly the shortened DNA extraction performed 'at sea'
12 had higher abundances for all three targets (UCYN-A1, UCYN-A2 and UCYN-B) in 16-25 %
13 of the samples processed, depending on the target diazotroph.

14 The disparity between the enumerations made 'at sea' and in the lab-based samples could
15 result from technical and/or sampling differences. For example, we do not know the DNA
16 extraction efficiency of either method used ('at sea vs. lab based), nor if extraction efficiency
17 varies for a particular target e.g. UCYN-A or UCYN-B. Variations between samples or
18 replicates could be attested to different DNA extractions varying in their efficiencies based on
19 target organism (Boström et al., 2004). Some strains of UCYN-B are known to secrete thick
20 extracellular matrix (ECM) (Sohm et al. 2011), while others form dense colonies, and some
21 are symbiotic with diatoms (Carpenter and Janson, 1999). Moreover, the UCYN-A symbioses
22 are considered fragile and easily dismanteled (Thompson et al., 2012). The latter observations
23 could potentially result in differences in extraction efficiency, which is supported by
24 significant disparities in all samples where the UCYN-A hosts were enumerated. However we
25 find no clear pattern in higher detection for one method ('at sea' vs. lab-based assays) or one

particular target. Despite taking the samples from the same niskin bottle for the archived and 'at sea' assays, we cannot discount the natural heterogeneity of plankton and differences in settling, and that perhaps some samples were more enriched than others. Seldom are replicates processed in qPCR studies as those presented here, and so the disparity reported here highlights the need to address variation derived from sampling and/or extraction efficiencies. Nonetheless, 'at sea' (and lab-based) qPCRs could also be appended with a multi-plexing approach to both increase and broaden the number of metabolic pathways (e.g. *narB, rbcL, nirS*) and/or phylotypes quantified simultaneously.

4.3 Abundance and vertical distribution of diazotrophs in the WTSP

Earlier work based on N isotope ratios (delta ^{15}N) of suspended particulate matter and dissolved organic N (DON) in the WTSP suggested that new production is likely fueled by N_2 fixation in this region (Hansell and Feely, 2000; Yoshikawa et al., 2005). The SP is also an area where high abundances of the unicellular diazotrophs, in particular UCYN-A and UCYN-B, have been previously reported (Biegala and Raimbault, 2008; Bonnet et al., 2009, 2015; Moisander et al., 2010) and account for a significant (74%) portion of the areal N_2 fixation (Bonnet et al., 2009). Hence, it was likely to encounter the presence of diazotrophic populations.

Recently UCYN-A and its various lineages have been highlighted as one of the most widespread and abundant diazotrophs (Farnelid et al., 2016 and references therein), which has led to the dramatic shift in the canonical paradigm of *Trichodesmium* as the only significant diazotroph. Surprisingly, here, we report abundances of the UCYN-A1 and UCYN-A2 lineages that are comparatively lower than earlier reports. In fact, UCYN-A1 and A2 were the least detected diazotrophs. Both UCYN-A phylotypes were largely restricted to the MA, with the exception of high densities (3.2×10^4 and 8.5×10^4 *nifH* copies L^{-1}, respectively) found at one depth (60 m) of LD C, which borders the MA region. Consistent with higher UCYN-A

biomass at depth at LD C were microscopy observations of high abundances of picoeukaryotes similar in size and shape previously reported for the UCYN-A hosts (Krupke et al. 2013). The vertical distribution of UCYN-A1 (and A2) was similar to Moisander's et al. (2010) and others, including earlier studies in the North Pacific Ocean (NP) and NA, where maximum abundances of UCYN-A are common to deeper depths in the euphotic zone (below 45 m) (e.g. Bonnet et al., 2015; Foster et al., 2007; Goebel et al., 2010; Needoba et al., 2007). Likewise, we also observed as others (Cabello et al., 2016) that the UCYN-A based symbioses co-occur and typically have decreased abundance towards the DCM, and maximum abundances slightly above the nitracline.

Unlike UCYN-A phylotypes, UCYN-B and *Trichodesmium* were the most abundant diazotrophs in the WTSP, and UCYN-B in particular was the most detected phylotype (99% detection; dnq or higher in 119 of 120 samples). High abundances of *Trichodesmium* in the upper 10 m, including presence of surface slicks and free filaments, was widespread in the MA region and consistent with earlier observations of high surface densities further north in the SP (Moisander et al., 2010; Shiozaki et al., 2014). Surface slicks have also been reported elsewhere, e.g. the North Atlantic (NA) (Goebel et al., 2010; Langlois et al., 2005). The depth of maximum abundance for *Trichodesmium* deepened from the MA (10 m) region to the open gyre (SG, 31 m), which was similar to earlier reports in the equatorial Pacific (Bonnet et al., 2009). A niche partitioning has been suggested for *Trichodesmium* and unicellular diazotrophs in the SP (Bonnet et al., 2015; Moisander et al., 2010) and elsewhere (Goebel et al., 2010; Langlois et al., 2005; Messer et al., 2015). However, here in the WTSP, *Trichodesmium* abundance was correlated with UCYN-B, which is consistent with previous studies in other ocean basins, e.g. Atlantic Ocean (Foster et al., 2007, 2009; Langlois et al., 2008), and the South China Sea (Moisander et al., 2008). UCYN-B co-occurred with *Trichodesmium* in the surface samples, although at lesser *nifH* copy abundances, and more often UCYN-B had

1 subsurface maxima (35-70 m) in both regions (MA and SG) of the transect. The latter is also
2 consistent with Moisander et al. (2010) who observed maximum abundances of UCYN-B
3 north of the Fijian islands at 37m.

4 All 3 heterocystous symbiont phylotypes co-occurred and were widespread in the MA,
5 with het-1 as the most abundant and most highly detected het group (70% detection or 84 of
6 120 samples). The early work of Moisander et al. (2010) detected het-1 in all but one of 26
7 stations sampled (56% detected, or 56 of 100 samples), and highest *nifH* copy densities were
8 reported north east of our cruise transect. Moreover, Bonnet et al. (2015) detected het-1 and
9 het-2 at the surface of one out of 10 stations west (approximately 10 degrees W) of our cruise
10 transect. Het-2 and het-3 were not quantified by Moisander et al. (2010) and het-3 was not
11 quantified by Bonnet et al. (2015). Therefore our study is among the first to report on the
12 abundances and distributions for all 3 heterocystous diazotrophs in a large expanse of the SP.
13 The 3 het phylotypes were however recently reported from a mesocosm (enclosed design)
14 experiment in the Noumea lagoon, a low nutrient low chlorophyll (LNLC) region located
15 along the New Caledonian coast (Turk-Kubo et al., 2015). In fact, het-1 and het-2 were among
16 the most abundant diazotrophs in the first half of the experiment (Turk-Kubo et al., 2015).
17 Two additional earlier studies have also reported microscopic observations of free-living
18 *Richelia* in the same lagoon (Biegala and Raimbault, 2008; Garcia et al., 2007).

19 Highest densities (10^4-10^6 *nifH* copies L^{-1}) of the *Richelia* phylotypes were restricted to
20 the western region of the MA, and in the upper 12 m, which is shallower than the subsurface
21 maximum (25 m) commonly reported for het-1 (and het-2) in the Western Tropical North
22 Atlantic (WTNA) and NP (Church et al., 2005; Foster et al., 2007; Goebel et al., 2010). Our
23 microscopy observations from SD 5-7 and LD A indicated that near surface *Rhizosolenia*
24 populations were in a moribund state since frustules were broken and free filaments of
25 *Richelia* were observed. Our observations also coincide with a region of high backscattering

1 measurements in the upper water column (5-30 m) (Dupouy et al., this issue). Het-1 *nifH*
2 copies were 4 orders of magnitude higher in abundance in the moored sediment traps of LD A
3 (325 m: 2.0 x 10^7 *nifH* copies L^{-1}) and LD B (325 and 500m: 5.8 x 10^6 and 1.10 x 10^7 *nifH*
4 copies L^{-1}, respectively) (Caffin et al., this issue) than the *nifH* copies detected in the
5 overlying waters (3.11 x 10^3 *nifH* copies L^{-1} and 4.1 x 10^2 *nifH* copies L^{-1}, respectively).
6 Combined, the latter observations suggest that a higher density of the het-1 population was
7 likely present prior to our sampling and perhaps derived from a 'seed' population originating
8 in the coastal regions of New Caledonia, and that they play an important role for export
9 production in this region, as has previously been shown in e.g. the NP (Karl et al., 2012).

10 The UCYN-C phylotype was poorly detected in the 'at sea' assays (61% samples were
11 bd and maximum abundance was 5.0x10^2 *nifH* copies L^{-1}), and as such was not enumerated in
12 the archived samples. The low detection of UCYN-C is consistent with Taniuchi et al. (2012),
13 who estimated that UCYN-C only represented a small portion of diazotrophs detected in the
14 western Pacific (Kuroshio Current). However, a recent study reported relatively high UCYN-
15 C abundances in the open waters of the Solomon Sea (north of the MA) (Berthelot et al.,
16 submitted). UCYN-C has also been observed in the New Caledonian lagoon (Turk-Kubo et al.,
17 2015), where it was the most dominant diazotroph in the first part of the aforementioned
18 mesocom experiment (Turk-Kubo et al., 2015). Moreover, Turk-Kubo et al. (2015) reviewed
19 the specificity of the UCYN-C assay used in our study (Foster et al., 2007) and concluded that
20 it does quantify a majority of UCYN-C phylotypes. Hence, like most plankton, abundances
21 can be patchy as was observed with UCYN-C in our study.

22 **4.4 UCYN-A and host (co)-occurrence**

23 Earlier and recent work has suggested a high host dependency (e.g. smaller and streamlined
24 genomes), and selectivity in the UCYN-A based symbioses (Cabello et al., 2016; Cornejo-
25 Castillo et al., 2016; Farnelid et al., 2016; Krupke et al., 2013, 2014; Thompson et al., 2012;

Tripp et al., 2010). Moreover, the UCYN-A partnerships are also considered mutualistic, where the host and symbiont both benefit by exchange of metabolites (e.g. reduced C and N, respectively) (Krupke et al., 2014; Thompson et al., 2012); hence one would expect parallel distributions for both partners. Some have argued that the partnership is also obligatory since few observations of free-living hosts have been reported and abundances of free symbionts assumed to be derived from disruption during sample preparation are always correlated with their hosts (Cabello et al., 2016; Krupke et al., 2014; Thompson et al., 2012). Thus, by use of our newly designed oligonucleotides for the UCYN-A1 host and previously designed oligonucleotides for the UCYN-A2 host (Thompson et al., 2014), we unexpectedly found that both UCYN-A1 and A2 were often (89% and 59%, respectively; not considering dnq) detected in the absence (or bd) of their respective hosts, while the hosts, when detected, always coincided with increased UCYN-A abundance. Our observations could result if the UCYN-A lineages can live freely, or in either a loose association, or perhaps with a wider range of hosts than previously thought and detected by the UCYN-A host assays. Presence of UCYN-A in the absence of their respective hosts could also indicate that the growth of symbiont and host is asynchronous, a pattern reported once in the het-1 or *Rhizosolenia-Richelia* symbioses (Villareal 1989).

The number of cells per partner lineage is considered specific as well, such that 1-2 UCYN-A1 cell is associated with a prymnesiophyte partner (UCYN-A1 host) and the larger *B. bigelowii* (UCYN-A2 host) host associates with multiple and variable numbers of UCYN-A2 cells to compensate for its higher N requirement (Cornejo-Castillo et al., 2016). On the contrary, we found evidence that there are multiple UCYN-A1 and A2 symbionts in both host types, which is somewhat surprising given that the host target gene (18S rRNA) is a multiple copy gene, meaning that we would expect higher gene copy numbers for each host. Nonetheless, we consistently observed higher abundances for the UCYN-A1 and A2

symbionts than their respective hosts. UCYN-A1 and A2 were 2-10 and 6-34 times, respectively, more abundant than their hosts. A symbiosome-like compartment has also been described attached to the UCYN-A2 host or residing free (Cornejo-Castillo et al., 2016). Thus, one plausible explanation for the higher abundances of the UCYN-A2, in particular, in the absence of their respective host, could result if our assays quantified UCYN-A2 residing in a dislodged free-floating symbiosome, or an overestimate of the UCYN-A2 due to cross-reactivity with UCYN-A3 lineage as expected by *in silico* tests (Farnelid et al. 2016). It is less likely that the UCYN-A2 was overquantified due to cross-reaction with UCYN-A1 templates since our cross-reactivity tests showed a weak cross reaction (see below).

4.5 Environmental influence on diazotroph abundances and distributions

The annual N inputs through biological N_2 fixation in the oceans is considered high, ranging 100-200 Tg N (Eugster and Gruber, 2012; Luo et al., 2012), yet large uncertainties remain in what factor(s) influence the abundance, distribution, and activity of marine diazotrophs. Initially, we hypothesized that conditions favoring a particular cyanobacterial diazotroph would differ given the contrasting life histories (free-living, colonial, and symbiotic). Moreover, we also suspected that the conditions promoting DDAs would differ from those influencing the UCYN-A based symbioses given the vast differences in the symbionts and hosts (e.g. genome content of symbiont, cell size of symbiont and hosts in the two systems; expected number of symbionts/host; host phylogeny: diatom vs. prymnesiophyte). Thus, determining the condition or sets of conditions that drive cyanobacterial diazotroph distribution, abundance, and activity is of great interest.

Hydrographic conditions and dissolved nutrient concentrations measured at the time of sampling were used to correlate diazotrophic abundance with various environmental parameters. Consistently, in two independent statistical tests, two groups emerged in the WTSP: 1) UCYN-A1 and A2 and their respective hosts 2) het-1, het-2 and het-3, UCYN-B

and *Trichodesmium*. Thus, agreeing with our initial hypothesis that conditions favoring the UCYN-A based symbioses does differ from the conditions for DDAs, and in addition for the free-living cyanobacterial diazotrophs.

Temperature is often cited as the most important driver of diazotroph abundance and distribution (Messer et al., 2016; Moisander et al., 2010). As shown earlier in the WTSP, both *Trichodesmium* spp. and UCYN-B were most abundant in warmer surface waters (> 27 ºC) in the north, while UCYN-A dominated in the cooler (24-26 ºC) southern waters of WTSP (Bonnet et al., 2015; Moisander et al., 2010). Likewise, we found similar abundances and temperature optima for the latter three diazotrophs and significant correlations between the various diazotrophs and temperature. In fact, all diazotrophs, except the UCYN-A lineages were significantly positively correlated with temperature in the WTSP. In addition to temperature, environmental parameters PAR, salinity and depth were also significantly influencing abundance and distribution. Moreover, the latter two variables drove the abundances of UCYN-A symbioses (A1 and A2) apart from the rest of the diazotrophs in the WTSP, including both free-living phylotypes and the symbiotic heterocystous lineages.

The maximum abundances at depth for UCYN-A1 and UCYN-A2 were slightly above or at the nitracline and coincided with higher measures of fluorescence from the CTD. The latter is consistent with observations of high UCYN-A abundances in coastal habitats (Bombar et al., 2014), estuaries (Messer et al., 2015), or in waters that are recently entrained with new nutrients (Moisander et al., 2010). Increased *nifH* copies and/or *nifH* gene expression for UCYN-A have also been reported from bioassay experiments amended with nutrients, including DIN, phosphate and iron (Krupke et al., 2015; Langlois et al., 2012; Moisander et al., 2012). The latter is in contrast with the data reported here in the WTSP (including the meta-analysis) and several of the external datasets (e.g. WTSP, TA, NA, NSCS), which finds a negative correlation between DIN and DIP concentrations and

abundance of most of the diazotrophs, including UCYN-A. In the WTNA, waters with high DiSi concentration and low N:P ratios, driven by a disproportionate utilization of N relative to P, results in consistent and widespread blooms of the *Hemiaulus-Richelia* symbioses (het-2) (Foster et al., 2007; Subramaniam et al., 2008). Across the cruise transect, DIP and DiSi concentrations were considered not limiting (Thierry Moutin, this issue), while DIN was below detection, hence conditions favoring symbiotic diatoms, and as reported here, the higher abundances of het-1 *nifH* gene copies and observations of *Rhizosolenia* hosts in the MA.

All the diazotrophs described here are either photoautotrophic or associated with photoautotrophic partners (UCYN-A, het-group). Therefore, light irradiance (e.g PAR) and availability will impact the abundance and distribution of the diazotrophic populations. Moreover, and related to light availability is the influence of day length or changes in the photoperiod which can influence diazotroph distribution, in particular the symbiotic diatoms (Karl et al., 2012). Results from CARD-FISH observations of the UYCN-A1 and A2 symbioses have reported a strong dependency on light intensity, which results in higher abundances nearer to the surface (Cabello et al., 2016). Presence in shallower waters is also thought to be strategic for avoiding competition (Cabello et al. 2016). However, in the WTSP, in 11 of the 14 stations where UCYN-A1 and A2 were detected at sub-surface depth maxima, the same lineages (and corresponding hosts) were undetected at the surface and a negative correlation was found with PAR. Microscopy observations also confirmed higher numbers of pico-eukaryotes at depth. Hence, it would appear that low light was a pre-requisite for high abundances of UCYN-A; while the other free-living diazotrophs and symbiotic het-1 and het-2 were positively correlated with PAR, and had maxima closer to the surface with higher PAR. Interestingly and unexpected was the lack of correlation between PAR and the UCYN-A host lineages, especially since it is the host partners that require light for photosynthesis.

In an attempt to identify the consistency in the correlation patterns identified in the WTSP with other regions of the world's ocean, the same statistical analyses were performed on 11 publically available datasets and subsequently run through a meta-analysis. Our statistical analyses provided coefficients and p-values for easy evaluation and comparisons between data sets for the influence of environmental parameter(s) and diazotrophs abundance. It confirmed that UCYN-A indeed stands out from the other diazotrophs in terms of environmental parameter influence, mainly by not being significantly correlated with temperature, which for all other diazotrophs was a significant positive correlation. In terms of temperature, UCYN-A has the lowest temperature optimum among the cyanobacterial diazotrophs, which will influence their distributions and subsequent correlation with temperature both regionally and vertically. The studies in the meta-analysis was conducted in a wide geographical range where a majority, in all ocean basins included, had no significant correlation for UCYN-A and temperature. For most other environmental parameters the pattern for UCYN-A in the WTSP does not hold true in the meta-analysis. However, for the other diazotrophs depth and salinity follow the same pattern as observed in the WTSP (except for UCYN-B which was not significantly correlated with salinity). Interestingly, UCYN-B seems to be mostly detected in the Pacific, where it also was found to be significantly negatively correlated with salinity, while in the Atlantic, perhaps due to lower detections or riverine impact, it was not significantly correlated resulting in no significant correlation in the meta-analysis overall. Furthermore, what did unify all diazotrophs in the meta-analysis were their consistent significant positive correlations with each other and significant negative correlations between abundance and concentrations of DIP and DIN, which was also observed in the WTSP, and again UCYN-A was the exception.

In summary, the correlations observed in the WTSP were not always consistent with the meta-analysis of the external datasets. We attribute the inconsistencies in part to seasonal

1 differences in sample collections, and the impact of an individual environmental parameter or
2 sets of parameters on a local and regional scale, particularly for coastal studies, that make it
3 difficult to unambiguously explain the abundance and distribution patterns. Unlike our initial
4 hypotheses, determining the condition or sets of conditions favoring one diazotroph or life
5 history strategy (free-living vs. symbiotic) is complex and likely not all diazotrophs are
6 influenced by the same condition in time and space.

7 **4.6 Estimation of diazotrophs by nifH qPCR**

8 When interpreting abundance estimates by qPCR there are a few assumptions to keep in mind.
9 A caveat of qPCR assays assumes that there is one gene copy per cell. However, recent
10 evidence in filamentous and heterocystous cyanobacteria reports evidence of polyploidy
11 dependent on cell cycle (Griese et al., 2011; Sargent et al., 2016; Sukenik et al., 2012).
12 Moreover, *Trichodesmium* may contain up to 100 genome copies per cell (Sargent et al.,
13 2016), thus a potential for overestimation. On the other hand, underestimation by qPCR is
14 also plausible if one considers that DNA extraction efficiency is not 100% and can vary
15 between species and DNA extraction kits (Mumy and Findlay, 2004), and if high probe
16 specificity favors exclusion of closely related phylotypes for a particular target or lineage.

17 A final consideration with qPCR as shown here, is the degree of cross-reactivity in
18 assays targeting closely related lineages (e.g. UCYN-A and het). Oligonucleotide specificity
19 as a source of underestimation of the UCYN-A lineages was recently reviewed by a *de nova*
20 analyses (Farnelid et al., 2016) showing the potential to underestimate UCYN-A sublineages
21 since the widely used oligonucleotides for UCYN-A1 contains several mismatches to the
22 other UCYN-A sublineages. The latter becomes important when the sublineages co-occur.
23 Here, however, we highlight the potential to overestimate. For example, UCYN-A2
24 oligonucleotides amplified the UCYN-A1 templates, indicating a tendency to overquantify
25 UCYN-A2 in the presence of A1. Moreover, when the annealing temperature was set to 64 ^{o}C,

1 to distinguish between UCYN-A1 and A2 as recommended by Thompson et al. (2014), the
2 assay still failed to separate the two sub-lineages when run in fast mode. Thus, the fast mode
3 feature has a shortcoming that could influence a wider range of targets than the ones presented
4 here. We observed the same cross-reactivity reported earlier (Foster et al., 2007) for het-1 and
5 het-2 when run in fast mode and highlights the potential to overestimate het-2 if het-1 co-
6 occurs at densities approximately 10^6 *nifH* copies L^{-1}. The latter observation has never been
7 reported.

8 **Conclusions**

9 Consistent with earlier observations in the WTSP, we found diazotrophic cyanobacteria
10 to be abundant. The most abundant cyanobacterial diazotrophs were UCYN-B,
11 *Trichodesmium* and the symbiotic *Richelia* lineage het-1. Although the cell integrity and
12 detection of het-1 in water column samples and those from depth (e.g. sediment traps)
13 indicated that the populations were in a senescent state, our work represents one of the first
14 documentation of the three DDA populations in a wide expanse of the WTSP. In contrast to
15 earlier work in the SP and other recent reports from global ocean surveys (Farnelid et al.,
16 2016; Martínez-Pérez et al., 2016), we observed low abundances and poor detection of both
17 UCYN-A (A1 and A2) lineages. According to our qPCR results, UCYN-A was also
18 enumerated when their respective hosts were below detection, which contrasts to the assumed
19 high fidelity and dependency in the partnerships; however, we cannot discount that the
20 disparity in host-symbiont detection was not a result from qPCR oligonucleotide assay bias
21 and/or overestimations indicated by our cross-reactivity tests.

22 Our initial hypothesis was that the condition or sets of conditions, which promote the
23 distribution of one diazotroph, would differ. Moreover, the parameters for symbiotic
24 diazotrophs should also differ from that of free-living phylotypes, and given the vast
25 difference in hosts (diatoms and prymnesiophyte, respectively) and genome content for the het

1 and UCYN-A symbionts, we further hypothesized divergent conditions favoring one
2 symbiosis over another. In the WTSP, the same conditions favored abundances of both the
3 free-living phylotypes and the diatom (het groups) symbioses. However, the same conditions
4 impacted the abundance of UCYN-A based symbiosis negatively, hence, somewhat
5 supporting our intial hypothesis that conditions for one symbiosis type would differ. In the
6 external datasets, however, we observed differences in environmental conditions favoring
7 abundances of the investigated diazotrophs compared to the WTSP, which underscores that
8 diazotrophs are not similarly influenced by the same condition in time and space.

9 Multivariate approaches on numerous parameters and with high spatial resolution are
10 required to understand the complex and often indirect effects that govern species distribution.
11 Finally, this study highlights reliable quantification of *nifH* genes for various N_2 fixing
12 cyanobacteria 'at sea' in the tropical open ocean and how environmental parameters influence
13 distribution and abundance of diazotrophs differently both regionally and across ocean basins.
14 However, it is of great interest to know, if the same parameters influence gene expressions
15 (e.g. *nifH*), and ultimately N_2 fixation rates, in the same manner, thus, understanding the
16 weight of environmental parameters influencing diazotrophic abundance and distribution.
17 Given the global significance of N_2 fixation as a major new source of N to the oceans, the
18 metanalysis presented here could be directly applicable to improving parameter constraints on
19 model-based approaches for predicting areas prone to diaztrophy.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The participation (MS and AC), sample processing, and work presented here is supported by Knut and Alice Wallenberg foundation (to RAF). This is a contribution to the OUTPACE (Oligotrophy from Ultra-oligoTrophy PACific Experiment) project funded by the French research national agency (ANR-14-CE01-0007-01), the LEFE-CyBER program (CNRS-INSU), the GOPS program (IRD) and the CNES (BC T23, ZBC 4500048836). The OUTPACE cruise (http://dx.doi.org/10.17600/15000900) was managed by T. Moutin and S. Bonnet from the MIO (Mediterranean Institute of Oceanography). T. Moutin is specifically thanked for invaluable input and feedback on this paper. The authors thank the crew of the R/V *L'Atalante* for outstanding shipboard operation. G. Rougier and M. Picheral are thanked for their efficient help in CTD rosette management and data processing, as is Catherine Schmechtig for the LEFE CYBER database management. All data and metadata are available at the following web address: http://www.obs-vlfr.fr/proof/php/outpace/outpace.php. Special acknowledgement to Andreas Krupke (formerly of WHOI), Kyle Frischkorn (WHOI), Mar Benavides (MIO/IRD), Hugo Berthelot (MIO), and Mathieu Caffin (MIO). Olivier Grosso (MIO) and Sandra Hélias (MIO) are acknowledged for nutrient analyses. Additionally we thank Dr. Lotta Berntzon for assisting in sample processing and Konrad Karlsson for assistance with the multivariate statistics. Lastly, we thank Dr. Bethany Jenkins and Prof. Anders Bignert for invaluable feedback and suggestions on improvement.

References

Aminot, A. and Kérouel, R.: Dosage automatique des nutriments dans les eaux marines: méthodes en flux continu, in Méthodes d'analyse en milieu marin, edited by Ifremer, p. 188. [online] Available from: http://cat.inist.fr/?aModele=afficheN&cpsidt=18706171 (Accessed 1 December 2016), 2007.

Benavides, M., Bonnet, S., Hernández, N., Martínez-Pérez, A. M., Nieto-Cid, M., Álvarez-Salgado, X. A., Baños, I., Montero, M. F., Mazuecos, I. P., Gasol, J. M., Osterholz, H., Dittmar, T., Berman-Frank, I. and Arístegui, J.: Basin-wide N2 fixation in the deep waters of the Mediterranean Sea, Glob. Biogeochem. Cycles, 30(6), 2015GB005326, doi:10.1002/2015GB005326, 2016.

Bench, S. R., Heller, P., Frank, I., Arciniega, M., Shilova, I. N. and Zehr, J. P.: Whole genome comparison of six Crocosphaera watsonii strains with differing phenotypes, J. Phycol., 49(4), 786–801, doi:10.1111/jpy.12090, 2013.

Berman-Frank, I., Lundgren, P., Chen, Y.-B., Küpper, H., Kolber, Z., Bergman, B. and Falkowski, P.: Segregation of Nitrogen Fixation and Oxygenic Photosynthesis in the Marine Cyanobacterium Trichodesmium, Science, 294(5546), 1534–1537, doi:10.1126/science.1064082, 2001.

Berthelot, H., Moutin, T., L'Helguen, S., Leblanc, K., Hélias, S., Grosso, O., Leblond, N., Charrière, B. and Bonnet, S.: Dinitrogen fixation and dissolved organic nitrogen fueled primary production and particulate export during the VAHINE mesocosm experiment (New Caledonia lagoon), Biogeosciences, 12(13), 4099–4112, doi:10.5194/bg-12-4099-2015, 2015.

Berthelot, H., Benavides, M., Moisander, P., Grosso, O., Bonnet, S.: Extremely high Nitrogen fixation rates in the particulate and dissolved pools in the Southwestern Pacific. Submitted to Geophysical Research Letters.

Biegala, I. C. and Raimbault, P.: High abundance of diazotrophic picocyanobacteria (<3 µm) in a Southwest Pacific coral lagoon, Aquat. Microb. Ecol., 51(1), 45–53, doi:10.3354/ame01185, 2008.

Blain, S., Sarthou, G. and Laan, P.: Distribution of dissolved iron during the natural iron-fertilization experiment KEOPS (Kerguelen Plateau, Southern Ocean), Deep Sea Res. Part II Top. Stud. Oceanogr., 55(5–7), 594–605, doi:10.1016/j.dsr2.2007.12.028, 2008.

Bombar, D., Moisander, P. H., Dippner, J. W., Foster, R. A., Voss, M., Karfeld, B. and Zehr, J. P.: Distribution of diazotrophic microorganisms and nifH gene expression in the Mekong River plume during intermonsoon, Mar. Ecol. Prog. Ser., 424, 39–52, doi:10.3354/meps08976, 2011.

Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J. and Zehr, J. P.: Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria, ISME J., 8(12), 2530–2542, doi:10.1038/ismej.2014.167, 2014.

Bombar, D., Paerl, R. W. and Riemann, L.: Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection, Trends Microbiol., 24(11), 916–927, doi:10.1016/j.tim.2016.07.002, 2016.

1 Bonnet, S., Biegala, I. C., Dutrieux, P., Slemons, L. O. and Capone, D. G.: Nitrogen fixation
2 in the western equatorial Pacific: Rates, diazotrophic cyanobacterial size class distribution,
3 and biogeochemical significance, Glob. Biogeochem. Cycles, 23(3), GB3012,
4 doi:10.1029/2008GB003439, 2009.

5 Bonnet, S., Dekaezemacker, J., Turk-Kubo, K. A., Moutin, T., Hamersley, R. M., Grosso, O.,
6 Zehr, J. P. and Capone, D. G.: Aphotic N2 Fixation in the Eastern Tropical South Pacific
7 Ocean, PLOS ONE, 8(12), e81265, doi:10.1371/journal.pone.0081265, 2013.

8 Bonnet, S., Rodier, M., Turk-Kubo, K. A., Germineaud, C., Menkes, C., Ganachaud, A.,
9 Cravatte, S., Raimbault, P., Campbell, E., Quéroué, F., Sarthou, G., Desnues, A., Maes, C.
10 and Eldin, G.: Contrasted geographical distribution of N2 fixation rates and nifH phylotypes
11 in the Coral and Solomon Seas (southwestern Pacific) during austral winter conditions, Glob.
12 Biogeochem. Cycles, 29(11), 2015GB005117, doi:10.1002/2015GB005117, 2015.

13 Bonnet, S., Berthelot, H., Turk-Kubo, K., Cornet-Barthaux, V., Fawcett, S., Berman-Frank, I.,
14 Barani, A., Grégori, G., Dekaezemacker, J., Benavides, M. and Capone, D.G.: Diazotroph
15 derived nitrogen supports diatom growth in the South West Pacific: a quantitative study using
16 nanoSIMS. Limnol. Oceanogr., 61(5), pp.1549-1562, 2016.

17 Bonnet, S., Caffin, M., Berthelot, H. and Moutin, T.: Hot spot of N_2 fixation in the western
18 tropical South Pacific pleads for a spatial decoupling between N_2 fixation and denitrification.
19 PNAS, 114(14), E2800-E2801, 2017.

20 Bonnet, S., Caffin, M., Berthelot, H., Grosso, O., Guieu, C. and Moutin, T.: Contribution of
21 dissolved and particulate fractions to the Hot Spot of N_2 fixation in the Western Tropical
22 South Pacific Ocean (OUTPACE cruise). This issue.

23 Boström, K. H., Simu, K., Hagström, Å. and Riemann, L.: Optimization of DNA extraction
24 for quantitative marine bacterioplankton community analysis. Limnol. Oceanogr-Meth., 2(11),
25 365-373, 2004.

26 Cabello, A. M., Cornejo-Castillo, F. M., Raho, N., Blasco, D., Vidal, M., Audic, S., de Vargas,
27 C., Latasa, M., Acinas, S. G. and Massana, R.: Global distribution and vertical patterns of a
28 prymnesiophyte–cyanobacteria obligate symbiosis, ISME J., 10(3), 693–706,
29 doi:10.1038/ismej.2015.147, 2016.

30 Caffin, M., Foster, R., Berthelot, H., Stenegren, M., Caputo, A. and Bonnet, S.: Fate of
31 diazotroph-derived nitrogen across trophic gradients in the Western Tropical South Pacific.
32 This issue

33 Carpenter, E. J. and Janson, S.: Intracellular Cyanobacterial Symbionts in the Marine Diatom
34 Climacodium Frauenfeldianum (bacillariophyceae), J. Phycol., 36(3), 540–544,
35 doi:10.1046/j.1529-8817.2000.99163.x, 2000.

36 Church, M. J., Jenkins, B. D., Karl, D. M. and Zehr, J. P.: Vertical distributions of nitrogen-
37 fixing phylotypes at Stn Aloha in the oligotrophic North Pacific Ocean, Aquat. Microb. Ecol.,
38 38(1), 3–14, doi:10.3354/ame038003, 2005.

39 Church, M. J., Björkman, K. M., Karl, D. M., Saito, M. A. and Zehr, J. P.: Regional
40 distributions of nitrogen-fixing bacteria in the Pacific Ocean, Limnol. Oceanogr., 53(1), 63–
41 77, doi:10.4319/lo.2008.53.1.0063, 2008.

1 Claustre, H. and Maritorena, S.: The Many Shades of Ocean Blue, Science, 302(5650), 1514–
2 1515, doi:10.1126/science.1092704, 2003.

3 Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sánchez-Baracaldo, P., Lima-Mendez,
4 G., Hingamp, P., Alberti, A., Sunagawa, S., Bork, P., Vargas, C. de, Raes, J., Bowler, C.,
5 Wincker, P., Zehr, J. P., Gasol, J. M., Massana, R. and Acinas, S. G.: Cyanobacterial
6 symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories
7 in single-celled phytoplankton, Nat. Commun., 7, 11071, doi:10.1038/ncomms11071, 2016.

8 Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N. and Dunne, J. P.: Spatial coupling of
9 nitrogen inputs and losses in the ocean, Nature, 445(7124), 163–167,
10 doi:10.1038/nature05392, 2007.

11 Dufour, P., Charpy, L., Bonnet, S. and Garcia, N.: Phytoplankton nutrient control in the
12 oligotrophic South Pacific subtropical gyre (Tuamotu Archipelago), Mar. Ecol. Prog. Ser.,
13 179, 285–290, doi:10.3354/meps179285, 1999.

14 Dupouy, C., Frouin, R., Maillard, M., Tedetti, M., Charriere, B., Roettgers, R., Martias, C.,
15 Rodier, M., Pujo-Pay, M., Duhamel, S., Guidi, L. and Sempere, R.: Influence of cyanobacteria
16 on Longitudinal and short-term variations in UV-Vis optical properties in the Western
17 Tropical South Pacific Ocean. This issue.

18 Dyhrman, S. T. and Ruttenberg, K. C.: Presence and regulation of alkaline phosphatase
19 activity in eukaryotic phytoplankton from the coastal ocean: Implications for dissolved
20 organic phosphorus remineralization, Limnol. Oceanogr., 51(3), 1381–1390,
21 doi:10.4319/lo.2006.51.3.1381, 2006.

22 Dyhrman, S. T., Chappell, P. D., Haley, S. T., Moffett, J. W., Orchard, E. D., Waterbury, J. B.
23 and Webb, E. A.: Phosphonate utilization by the globally important marine diazotroph
24 Trichodesmium, Nature, 439(7072), 68–71, doi:10.1038/nature04203, 2006.

25 Eugster, O. and Gruber, N.: A probabilistic estimate of global marine N-fixation and
26 denitrification, Glob. Biogeochem. Cycles, 26(4), GB4013, doi:10.1029/2012GB004300,
27 2012.

28 Farnelid, H., Turk-Kubo, K., Muñoz-Marín, M. del C. and Zehr, J. P.: New insights into the
29 ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A, Aquat.
30 Microb. Ecol., 77(3), 125–138, doi:10.3354/ame01794, 2016.

31 Fernandez, C., Farías, L. and Ulloa, O.: Nitrogen Fixation in Denitrified Marine Waters,
32 PLOS ONE, 6(6), e20539, doi:10.1371/journal.pone.0020539, 2011.

33 Fitzsimmons, J. N., Boyle, E. A. and Jenkins, W. J.: Distal transport of dissolved
34 hydrothermal iron in the deep South Pacific Ocean, Proc. Natl. Acad. Sci., 111(47), 16654–
35 16661, doi:10.1073/pnas.1418778111, 2014.

36 Foster, R. A. and Zehr, J. P.: Characterization of diatom–cyanobacteria symbioses on the
37 basis of nifH, hetR and 16S rRNA sequences, Environ. Microbiol., 8(11), 1913–1925,
38 doi:10.1111/j.1462-2920.2006.01068.x, 2006.

39 Foster, R. A., Subramaniam, A., Mahaffey, C., Carpenter, E. J., Capone, D. G. and Zehr, J. P.:
40 Influence of the Amazon River plume on distributions of free-living and symbiotic

1 cyanobacteria in the western tropical north Atlantic Ocean, Limnol. Oceanogr., 52(2), 517–
2 532, doi:10.4319/lo.2007.52.2.0517, 2007.

3 Foster, R. A., Subramaniam, A. and Zehr, J. P.: Distribution and activity of diazotrophs in the
4 Eastern Equatorial Atlantic, Environ. Microbiol., 11(4), 741–750, doi:10.1111/j.1462-
5 2920.2008.01796.x, 2009.

6 Foster, R. A., Goebel, N. L. and Zehr, J. P.: Isolation of Calothrix Rhizosoleniae
7 (cyanobacteria) Strain Sc01 from Chaetoceros (bacillariophyta) Spp. Diatoms of the
8 Subtropical North Pacific Ocean1, Journal of Phycology, 46(5), 1028–1037,
9 doi:10.1111/j.1529-8817.2010.00885.x, 2010.

10 Garcia, N., Raimbault, P. and Sandroni, V.: Seasonal nitrogen fixation and primary
11 production in the Southwest Pacific: nanoplankton diazotrophy and transfer of nitrogen to
12 picoplankton organisms, Mar. Ecol. Prog. Ser., 343, 25–33, doi:10.3354/meps06882, 2007.

13 Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., Montoya,
14 J. P., Edwards, C. A. and Zehr, J. P.: Abundance and distribution of major groups of
15 diazotrophic cyanobacteria and their potential contribution to N2 fixation in the tropical
16 Atlantic Ocean, Environ. Microbiol., 12(12), 3272–3289, doi:10.1111/j.1462-
17 2920.2010.02303.x, 2010.

18 Gourdeau, L., Kessler, W. S., Davis, R. E., Sherman, J., Maes, C. and Kestenare, E.: Zonal
19 Jets Entering the Coral Sea, J. Phys. Oceanogr., 38(3), 715–725, doi:10.1175/2007JPO3780.1,
20 2008.

21 Griese, M., Lange, C. and Soppa, J.: Ploidy in cyanobacteria, FEMS Microbiol. Lett., 323(2),
22 124–131, doi:10.1111/j.1574-6968.2011.02368.x, 2011.

23 Halm, H., Musat, N., Lam, P., Langlois, R., Musat, F., Peduzzi, S., Lavik, G., Schubert, C. J.,
24 Singha, B., LaRoche, J. and Kuypers, M. M. M.: Co-occurrence of denitrification and
25 nitrogen fixation in a meromictic lake, Lake Cadagno (Switzerland), Environ. Microbiol.,
26 11(8), 1945–1958, doi:10.1111/j.1462-2920.2009.01917.x, 2009.

27 Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., D'Hondt, S. and
28 Kuypers, M. M.: Heterotrophic organisms dominate nitrogen fixation in the South Pacific
29 Gyre, ISME J., 6(6), 1238–1249, doi:10.1038/ismej.2011.182, 2012.

30 Hansell, D. A. and Feely, R. A.: Atmospheric Intertropical Convergence impacts surface
31 ocean carbon and nitrogen biogeochemistry in the western tropical Pacific, Geophys. Res.
32 Lett., 27(7), 1013–1016, doi:10.1029/1999GL002376, 2000.

33 Haselkorn, R.: Heterocysts, Annu. Rev. Plant Physiol., 29(1), 319–344,
34 doi:10.1146/annurev.pp.29.060178.001535, 1978.

35 Hood, R. R., Coles, V. J. and Capone, D. G.: Modeling the distribution of Trichodesmium and
36 nitrogen fixation in the Atlantic Ocean, J. Geophys. Res. Oceans, 109(C6), C06006,
37 doi:10.1029/2002JC001753, 2004.

38 Howard, J. B. and Rees, D. C.: Structural Basis of Biological Nitrogen Fixation, Chem. Rev.,
39 96(7), 2965–2982, doi:10.1021/cr9500545, 1996.

Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E. and Dean, D. R.: Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii., J. Bacteriol., 171(2), 1017–1027, 1989.

Karl, D., Letelier, R., Tupas, L., Dore, J., Christian, J. and Hebel, D.: The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean, Nature, 388(6642), 533–538, doi:10.1038/41474, 1997.

Karl, D. M., Björkman, K. M., Dore, J. E., Fujieki, L., Hebel, D. V., Houlihan, T., Letelier, R. M. and Tupas, L. M.: Ecological nitrogen-to-phosphorus stoichiometry at station ALOHA, Deep Sea Res. Part II Top. Stud. Oceanogr., 48(8–9), 1529–1566, doi:10.1016/S0967-0645(00)00152-1, 2001.

Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M. and Mahaffey, C.: Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation, Proc. Natl. Acad. Sci., 109(6), 1842–1849, doi:10.1073/pnas.1120312109, 2012.

Kong, L., Jing, H., Kataoka, T., Sun, J. and Liu, H.: Phylogenetic diversity and spatio-temporal -distribution of nitrogenase genes (nifH) in the northern South China Sea, Aquat. Microb. Ecol., 65(1), 15–27, doi:10.3354/ame01531, 2011.

Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B. M., Amann, R. I., Kuypers, M. M. M. and Foster, R. A.: In situ identification and N2 and C fixation rates of uncultivated cyanobacteria populations, Syst. Appl. Microbiol., 36(4), 259–271, doi:10.1016/j.syapm.2013.02.002, 2013.

Krupke, A., Lavik, G., Halm, H., Fuchs, B. M., Amann, R. I. and Kuypers, M. M. M.: Distribution of a consortium between unicellular algae and the N2 fixing cyanobacterium UCYN-A in the North Atlantic Ocean, Environ. Microbiol., 16(10), 3153–3167, doi:10.1111/1462-2920.12431, 2014.

Krupke, A., Mohr, W., LaRoche, J., Fuchs, B. M., Amann, R. I. and Kuypers, M. M.: The effect of nutrients on carbon and nitrogen fixation by the UCYN-A–haptophyte symbiosis, ISME J., 9(7), 1635–1647, doi:10.1038/ismej.2014.253, 2015.

Kustka, A. B., Sañudo-Wilhelmy, S. A., Carpenter, E. J., Capone, D., Burns, J. and Sunda, W. G.: Iron requirements for dinitrogen- and ammonium-supported growth in cultures of Trichodesmium (IMS 101): Comparison with nitrogen fixation rates and iron: carbon ratios of field populations, Limnol. Oceanogr., 48(5), 1869–1884, doi:10.4319/lo.2003.48.5.1869, 2003.

Landolfi, A., Koeve, W., Dietze, H., Kähler, P. and Oschlies, A.: A new perspective on environmental controls of marine nitrogen fixation, Geophys. Res. Lett., 42(11), 2015GL063756, doi:10.1002/2015GL063756, 2015.

Langlois, R., Mills, M. M., Ridame, C., Croot, P. and LaRoche, J.: Diazotrophic bacteria respond to Saharan dust additions, Mar. Ecol. Prog. Ser., 470(doi:10.3354/meps10109), 1–14, doi:Langlois, Rebecca, Mills, M. M., Ridame, Celine, Croot, Peter and LaRoche, Julie (2012) Diazotrophic bacteria respond to Saharan dust additions Marine Ecology Progress Series, 470 . pp. 1-14. DOI 10.3354/meps10109 <http://dx.doi.org/10.3354/meps10109>., 2012.

Langlois, R. J., LaRoche, J. and Raab, P. A.: Diazotrophic Diversity and Distribution in the Tropical and Subtropical Atlantic Ocean, Appl. Environ. Microbiol., 71(12), 7910–7919, doi:10.1128/AEM.71.12.7910-7919.2005, 2005.

Langlois, R. J., Hümmer, D. and LaRoche, J.: Abundances and Distributions of the Dominant nifH Phylotypes in the Northern Atlantic Ocean, Appl. Environ. Microbiol., 74(6), 1922–1931, doi:10.1128/AEM.01720-07, 2008.

Law, C. S., Woodward, E. M. S., Ellwood, M. J., Marriner, A., Bury, S. J. and Safi, K. A.: Response of surface nutrient inventories and nitrogen fixation to a tropical cyclone in the southwest Pacific, Limnol. Oceanogr., 56(4), 1372–1385, doi:10.4319/lo.2011.56.4.1372, 2011.

Löscher, C. R., LaRoche, J. and Schmitz, R. A.: Underestimation of Marine Dinitrogen Fixation: A Novel Method and Novel Diazotrophic Habitats, in Biological Nitrogen Fixation, edited by F. J. de Bruijn, pp. 891–896, John Wiley & Sons, Inc. [online] Available from: http://onlinelibrary.wiley.com/doi/10.1002/9781119053095.ch87/summary (Accessed 29 November 2016), 2015.

Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K. H., Böttjer, D., Capone, D. G., Carpenter, E. J., Chen, Y. L., Church, M. J., Dore, J. E., Falcón, L. I., Fernández, A., Foster, R. A., Furuya, K., Gómez, F., Gundersen, K., Hynes, A. M., Karl, D. M., Kitajima, S., Langlois, R. J., LaRoche, J., Letelier, R. M., Marañón, E., McGillicuddy, D. J., Moisander, P. H., Moore, C. M., Mouriño-Carballido, B., Mulholland, M. R., Needoba, J. A., Orcutt, K. M., Poulton, A. J., Rahav, E., Raimbault, P., Rees, A. P., Riemann, L., Shiozaki, T., Subramaniam, A., Tyrrell, T., Turk-Kubo, K. A., Varela, M., Villareal, T. A., Webb, E. A., White, A. E., Wu, J. and Zehr, J. P.: Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates, Earth Syst. Sci. Data, 4(1), 47–73, doi:Luo, Y.-W.; Doney, S.C.; Anderson, L.A.; Benavides, M.; Berman-Frank, I.; Bode, A.; Bonnet, S.; Boström, K.H.; Böttjer, D.; Capone, D.G.; Carpenter, E.J.; Chen, Y.L.; Church, M.J.; Dore, J.E.; Falcón, L.I.; Fernández, A.; Foster, R.A.; Furuya, K.; Gómez, F.; Gundersen, K.; Hynes, A.M.; Karl, D.M.; Kitajima, S.; Langlois, R.J.; LaRoche, J.; Letelier, R.M.; Marañón, E.; McGillicuddy, D.J.; Moisander, P.H.; Moore, C.M.; Mouriño-Carballido, B.; Mulholland, M.R.; Needoba, J.A.; Orcutt, K.M.; Poulton, A.J.; Rahav, E.; Raimbault, P.; Rees, A.P.; Riemann, L.; Shiozaki, T.; Subramaniam, A.; Tyrrell, T.; Turk-Kubo, K.A.; Varela, M.; Villareal, T.A.; Webb, E.A.; White, A.E.; Wu, J.; Zehr, J.P.. 2012 Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. Earth System Science Data, 4 (1). 47-73. 10.5194/essd-4-47-2012 <http://dx.doi.org/10.5194/essd-4-47-2012>, 2012.

Marchesiello, P. and Estrade, P.: Upwelling limitation by onshore geostrophic flow, J. Mar. Res., 68(1), 37–62, doi:10.1357/002224010793079004, 2010.

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B. M., Lavik, G., Schmitz, R. A., LaRoche, J. and Kuypers, M. M. M.: The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle, Nat. Microbiol., 1, 16163, doi:10.1038/nmicrobiol.2016.163, 2016.

Messer, L. F., Doubell, M., Jeffries, T. C., Brown, M. V. and Seymour, J. R.: Prokaryotic and diazotrophic population dynamics within a large oligotrophic inverse estuary, Aquat. Microb. Ecol., 74(1), 1–15, doi:10.3354/ame01726, 2015.

Messer, L. F., Mahaffey, C., Robinson, C. M., Jeffries, T. C., Baker, K. G., Bibiloni Isaksson, J., Ostrowski, M., Doblin, M. A., Brown, M. V. and Seymour, J. R.: High levels of heterogeneity in diazotroph diversity and activity within a putative hotspot for marine nitrogen fixation, ISME J., 10(6), 1499–1513, doi:10.1038/ismej.2015.205, 2016.

Meyerhof, O. and Burk, D.: Über die Fixation des Luftstickstoffs durch Azotobacter, Z. Phys. Chem. Abt A, (139), 117–142, 1928.

Mills, M. M., Ridame, C., Davey, M., La Roche, J. and Geider, R. J.: Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic, Nature, 429(6989), 292–294, doi:10.1038/nature02550, 2004.

Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S. and Arai, T.: Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically, Nature, 323, 720–722, doi:10.1038/323720a0, 1986.

Moisander, P. H., Beinart, R. A., Voss, M. and Zehr, J. P.: Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon, ISME J., 2(9), 954–967, doi:10.1038/ismej.2008.51, 2008.

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P. and Zehr, J. P.: Unicellular Cyanobacterial Distributions Broaden the Oceanic N2 Fixation Domain, Science, 327(5972), 1512–1514, doi:10.1126/science.1185468, 2010.

Moisander, P. H., Zhang, R., Boyle, E. A., Hewson, I., Montoya, J. P. and Zehr, J. P.: Analogous nutrient limitations in unicellular diazotrophs and Prochlorococcus in the South Pacific Ocean, ISME J., 6(4), 733–744, doi:10.1038/ismej.2011.152, 2012.

Moutin, T., Karl, D. M., Duhamel, S., Rimmelin, P., Raimbault, P., Mooy, V., S, B. A. and Claustre, H.: Phosphate availability and the ultimate control of new nitrogen input by nitrogen fixation in the tropical Pacific Ocean, , doi:10.5194/bg-5-95-2008, 2008.

Moutin, T., A. Doglioli, A. deVerneil, S. Bonnet.: The Oligotrophy to the UlTra-oligotrophy PACific Experiment (OUTPACE cruise, Feb. 18 to Apr. 3, 2015). This issue.

Mumy, K. L. and Findlay, R. H.: Convenient determination of DNA extraction efficiency using an external DNA recovery standard and quantitative-competitive PCR, J. Microbiol. Methods, 57(2), 259–268, doi:10.1016/j.mimet.2004.01.013, 2004.

Needoba, J. A., Foster, R. A., Sakamoto, C., Zehr, J. P. and Johnson, K. S.: Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean, Limnol. Oceanogr., 52(4), 1317–1327, doi:10.4319/lo.2007.52.4.1317, 2007.

Ottesen, E. A., Young, C. R., Eppley, J. M., Ryan, J. P., Chavez, F. P., Scholin, C. A. and DeLong, E. F.: Pattern and synchrony of gene expression among sympatric marine microbial populations, PNAS, 110(6), E488–E497, doi:10.1073/pnas.1222099110, 2013.

Preston, C. M., Harris, A., Ryan, J. P., Roman, B., Iii, R. M., Jensen, S., Everlove, C., Birch, J., Dzenitis, J. M., Pargett, D., Adachi, M., Turk, K., Zehr, J. P. and Scholin, C. A.: Underwater Application of Quantitative PCR on an Ocean Mooring, PLOS ONE, 6(8), e22522, doi:10.1371/journal.pone.0022522, 2011.

1 Qiu, B., Chen, S. and Kessler, W. S.: Source of the 70-Day Mesoscale Eddy Variability in the
2 Coral Sea and the North Fiji Basin, J. Phys. Oceanogr., 39(2), 404–420,
3 doi:10.1175/2008JPO3988.1, 2009.

4 Raimbault, P., Garcia, N. and Cerutti, F.: Distribution of inorganic and organic nutrients in the
5 South Pacific Ocean - evidence for long-term accumulation of organic matter in nitrogen-
6 depleted waters, Biogeosciences Discuss., 4, 3041–3087, 2007.

7 Raven, J. A.: The iron and molybdenum use efficiencies of plant growth with different energy,
8 carbon and nitrogen sources, New Phytol., 109(3), 279–287, doi:10.1111/j.1469-
9 8137.1988.tb04196.x, 1988.

10 Robidart, J. C., Church, M. J., Ryan, J. P., Ascani, F., Wilson, S. T., Bombar, D., Marin, R.,
11 Richards, K. J., Karl, D. M., Scholin, C. A. and Zehr, J. P.: Ecogenomic sensor reveals
12 controls on N2-fixing microorganisms in the North Pacific Ocean, ISME J, 8(6), 1175–1185,
13 doi:10.1038/ismej.2013.244, 2014.

14 Sañudo-Wilhelmy, S. A., Kustka, A. B., Gobler, C. J., Hutchins, D. A., Yang, M., Lwiza, K.,
15 Burns, J., Capone, D. G., Raven, J. A. and Carpenter, E. J.: Phosphorus limitation of nitrogen
16 fixation by Trichodesmium in the central Atlantic Ocean, Nature, 411(6833), 66–69,
17 doi:10.1038/35075041, 2001.

18 Sargent, E. C., Hitchcock, A., Johansson, S. A., Langlois, R., Moore, C. M., LaRoche, J.,
19 Poulton, A. J. and Bibby, T. S.: Evidence for polyploidy in the globally important diazotroph
20 Trichodesmium, FEMS Microbiol. Lett., 363(21), fnw244, doi:10.1093/femsle/fnw244, 2016.

21 Shiozaki, T., Kodama, T. and Furuya, K.: Large-scale impact of the island mass effect through
22 nitrogen fixation in the western South Pacific Ocean, Geophys. Res. Lett., 41(8),
23 2014GL059835, doi:10.1002/2014GL059835, 2014.

24 Short, S. M., Jenkins, B. D. and Zehr, J. P.: Spatial and Temporal Distribution of Two
25 Diazotrophic Bacteria in the Chesapeake Bay, Appl. Environ. Microbiol., 70(4), 2186–2192,
26 doi:10.1128/AEM.70.4.2186-2192, 2004.

27 Sohm, J. A., Webb, E. A. and Capone, D. G.: Emerging patterns of marine nitrogen fixation.
28 Nat. Rev. Microbiol., 9(7), 499-508, 2011.

29 Stewart, W. D. P.: Biological and Ecological Aspects of Nitrogen Fixation by Free-Living
30 Micro-Organisms, Proc. R. Soc. Lond. B Biol. Sci., 172(1029), 367–388, 1969.

31 Subramaniam, A., Yager, P. L., Carpenter, E. J., Mahaffey, C., Björkman, K., Cooley, S.,
32 Kustka, A. B., Montoya, J. P., Sañudo-Wilhelmy, S. A., Shipe, R. and Capone, D. G.:
33 Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic
34 Ocean, Proc. Natl. Acad. Sci., 105(30), 10460–10465, doi:10.1073/pnas.0710279105, 2008.

35 Sukenik, A., Kaplan-Levy, R. N., Welch, J. M. and Post, A. F.: Massive multiplication of
36 genome and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum
37 (Cyanobacteria), ISME J., 6(3), 670–679, doi:10.1038/ismej.2011.128, 2012.

38 Taniuchi, Y., Chen, Y. L., Chen, H.-Y., Tsai, M.-L. and Ohki, K.: Isolation and
39 characterization of the unicellular diazotrophic cyanobacterium Group C TW3 from the

tropical western Pacific Ocean, Environ. Microbiol., 14(3), 641–654, doi:10.1111/j.1462-2920.2011.02606.x, 2012.

Thompson, A., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F. and Zehr, J. P.: Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host, Environ. Microbiol., 16(10), 3238–3249, doi:10.1111/1462-2920.12490, 2014.

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., Kuypers, M. M. M. and Zehr, J. P.: Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga, Science, 337(6101), 1546–1550, doi:10.1126/science.1222700, 2012.

Tripp, H. J., Bench, S. R., Turk, K. A., Foster, R. A., Desany, B. A., Niazi, F., Affourtit, J. P. and Zehr, J. P.: Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium, Nature, 464(7285), 90–94, doi:10.1038/nature08786, 2010.

Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B. and Zehr, J. P.: Distinct ecological niches of marine symbiotic N2-fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa* sublineages. J. Phycol., 53(2), 451-461, 2017.

Turk-Kubo, K. A., Frank, I. E., Hogan, M. E., Desnues, A., Bonnet, S. and Zehr, J. P.: Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia lagoon), Biogeosciences, 12(24), 7435–7452, doi:10.5194/bg-12-7435-2015, 2015.

Van Den Broeck, N., Moutin, T., Rodier, M. and Bouteiller, A. L.: Seasonal variations of phosphate availability in the SW Pacific Ocean near New Caledonia, Mar. Ecol. Prog. Ser., 268, 1–12, doi:10.3354/meps268001, 2004.

Wallace, B. C., Lajeunesse, M. J., Dietz, G., Dahabreh, I.J., Trikalinos, T.A., Schmid, C.H. and Gurevitch, J.: OpenMEE: Intuitive, open-source software for meta analysis in ecology and evolutionary biology. Methods Ecol. Evol., 2016.

Webb, E. A., Ehrenreich, I. M., Brown, S. L., Valois, F. W. and Waterbury, J. B.: Phenotypic and genotypic characterization of multiple strains of the diazotrophic cyanobacterium, Crocosphaera watsonii, isolated from the open ocean, Environ. Microbiol., 11(2), 338–348, doi:10.1111/j.1462-2920.2008.01771.x, 2009.

Wells, M. L., Vallis, G. K. and Silver, E. A.: Tectonic processes in Papua New Guinea and past productivity in the eastern equatorial Pacific Ocean, Nature, 398(6728), 601–604, doi:10.1038/19281, 1999.

Yoshikawa, C., Nakatsuka, T. and Kawahata, H.: Transition of low-salinity water in the Western Pacific Warm Pool recorded in the nitrogen isotopic ratios of settling particles, Geophys. Res. Lett., 32(14), L14615, doi:10.1029/2005GL023103, 2005.

Young, J. P. W.: The Phylogeny and Evolution of Nitrogenases, in Genomes and Genomics of Nitrogen-fixing Organisms, edited by R. Palacios and W. E. Newton, pp. 221–241, Springer Netherlands. [online] Available from: http://link.springer.com/chapter/10.1007/1-4020-3054-1_14 (Accessed 29 November 2016), 2005.

Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., Hansen, A. and Karl, D. M.: Unicellular cyanobacteria fix N2 in the subtropical North Pacific Ocean, Nature, 412(6847), 635–638, doi:10.1038/35088063, 2001.

Zehr, J. P., Methe, B. and Foster, R.: New Nitrogen-Fixing Microorganisms from the Oceans: Biological Aspects and Global Implications, in Biological Nitrogen Fixation, Sustainable Agriculture and the Environment, pp. 361–365, Springer Netherlands. [online] Available from: http://link.springer.com/chapter/10.1007/1-4020-3570-5_87 (Accessed 29 November 2016), 2005.

Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., Tripp, H. J. and Affourtit, J. P.: Globally Distributed Uncultivated Oceanic N2-Fixing Cyanobacteria Lack Oxygenic Photosystem II, Science, 322(5904), 1110–1112, doi:10.1126/science.1165340, 2008.

1 Table 01. Summary of environmental conditions in the surface along the cruise transect.

Region	Stations	surface* DIN‡ µM	surface* DIP µM	surface* DiSi µM	surface* salinity (PSU)	surface* temp. ˚C
Melanesian archipelago (MA) 160˚ E-178˚ E 170 ˚W - 175 ˚W	SD1-12 LDA LDB	0.02 ± 0.01	0.03 ± 0.02	0.55 ± 0.10	35.13 ± 0.27	29.33 ± 0.45
Subtropical gyre (SG) 160 ˚W- 169˚W	SD13-15 LDC	0.01 ± 0.01	0.18 ± 0.07	0.79 ± 0.04	35.12 ± 0.10	29.34 ± 0.18

2 *5m depth, ‡NO_2+NO

1 **Figure 1.** Sampling locations and the horizontal and vertical distributions of diazotrophs and
2 the UCYN-A1 and UCYN-A-2 hosts in the study area. Sampling depths are indicated as black
3 dots (white for LD stations) and the abundances are the log *nifH* gene copy L^{-1} for the
4 diazotrophs and 18S rRNA gene copies L^{-1} for the UCYN-A host lineages. The boundary
5 between the Melanesian archipelago and subtropical gyre in the transect map is marked with a
6 dotted line.

7 **Figure 2 a-b.** LOG10 transformed mean abundances for the 4 most abundant diazotrophs
8 across the transect: het-1 (blue), *Trichodesmium* (green), UCYN-A (yellow) and UCYN-B
9 (red). The mean *nifH* abundance values (log n*ifH* copies L^{-1}) shown as a function of (a)
10 percent (%) surface irradiance, depth (m) on the secondary y axis and (b) at each station. The
11 solid black line in (b) indicates the transition between the Melanesian archipelago (MA) and
12 the subtropical gyre (SG).

13 **Figure 3.** Hierarchical clustering heat map of Spearman's Rho results. The histogram shows
14 negative (blue) and positive (green) values of correlation strength between parameters. Stars
15 within cells mark significant correlations ($p<0.05$). In brackets are the two distinct groups in
16 the WTSP.

17 **Figure 4 a-b.** Multivariate RDA biplot (a), which also depicts variance of included
18 parameters (b). As can be seen, a majority of the variance in the dataset is explained by the
19 RDA1 and RDA2 axes meaning that most of the variance observed is explained by the
20 included environmental parameters. The arrows are the constrained explanatory vectors with
21 the dots representing the superimposed unconstrained response variables. PAR and nutrients
22 (DIP and DIN) were omitted due to limited data.

23 **Figure 5 a-d.** Meta-analysis bar graphs, a) *Trichodesmium*, b) UCYN-B, c) UCYN-A, d) het-
24 1, with the significant ($p<0.05$) parameters for each diazotroph arranged as the strongest effect

1 to the left and weakest to the right (either positive or negative). Each parameter is color coded,

2 where the cyanobacterial diazotrophs have been assigned a spectrum of orange from

3 *Trichodesmium* (darkest) to het-2 (lightest). Red=temperature, blue=salinity, black=depth,

4 green=chlorophyll *a*, yellow=DIN and purple=DIP.

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 UCYNA1 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 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.
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 asses 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: UCYNA-1 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 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 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)

1 # Distribution and drivers of symbiotic and free-living
2 # diazotrophic cyanobacteria in the Western Tropical South
3 # Pacific

4 Marcus Stenegren[1], Andrea Caputo[1], Carlo Berg[2], Sophie Bonnet[3,4], Rachel A Foster[1]

5 [1] Stockholm University, Department of Ecology, Environment and Plant Sciences. Stockholm, Sweden

6 [2] Science for Life Laboratory/Department of Biology and Environmental Science, Linnaeus University, Kalmar,
7 Sweden

8 [3] Aix Marseille Université, Centre National de la Recherche Scientifique. Marseille/Noumea, New Caledonia,
9 France

10 [4] Mediterranean Institute of Oceanography, Institut de Recherche pour le Dévelopement. Marseille, France

11 *Correspondence to*: Marcus Stenegren (marcus.stenegren@su.se)

Abstract

The abundance and distribution of cyanobacterial diazotrophs were quantified in two regions (Melanesian archipelago, MA and subtropical gyre, SG) of the Western Tropical South Pacific using *nifH* qPCR assays. UCYN-A1 and A2 host populations were quantified using 18S rRNA qPCR assays including one newly developed assay. All phylotypes were detected in the upper photic zone (0-50 m), with higher abundances in the MA region. *Trichodesmium* and UCYN-B dominated, and ranged $2.18 \times 10^{2} – 9.41 \times 10^{6}$ and $1.10 \times 10^{2} – 2.78 \times 10^{6}$ *nifH* copies L^{-1}, respectively. Het-1 (symbiont of *Rhizosolenia* diatoms) was the next most abundant (1.40×10^{1}-1.74×10^{5} *nifH* copies L^{-1}), and co-occurred with het-2 and het-3. UCYN-A1 and A2 were the least abundant diazotrophs and were below detection (bd) in 63 and 79, respectively, of 120 samples. In addition, in up to 39% of samples where UCYN-A1 and A2 were detected, their respective hosts were bd. Pairwise comparisons of the *nifH* abundances and various environmental parameters supported two groups: a deep-dwelling group (45 m) comprised of UCYN-A1 and A2 and a surface group (0-15m) comprised of *Trichodesmium*, het-1 and het-2. Temperature and PAR were positively correlated with the surface group; while UCYN-A1 and A2 were positively correlated with depth, salinity and oxygen. Similarly, in a meta-analysis of 11 external datasets, all diazotrophs, except UCYN-A were correlated with temperature. Combined, our results indicate that conditions favoring the UCYN-A symbiosis differ from those of diatom diazotroph associations and free-living cyanobacterial diazotrophs.

Deleted: detected *nifH* copies of the two

Deleted: Similar results were identified in aA

Deleted: determined temperature and depth were most commonly correlated with diazotrophs abundance

Deleted: . , where temperature and depth had a major impact on the diazotrophs in the ocean basins investigated.

1 Introduction

Biological di-nitrogen (N_2) fixation is considered a major source of new nitrogen (N) to oceanic ecosystems (Karl et al., 1997). N_2 fixation is an energetically expensive process, where N_2 gas is reduced to bioavailable ammonia (Howard and Rees, 1996) and is performed by a small but diverse group of bacteria and archaea. The nitrogenase enzyme, which is encoded by a suite of *nif*-genes, mediates N_2 fixation (Jacobson et al., 1989; Young, 2005). Nitrogenase has a high iron (Fe) requirement (Howard and Rees, 1996), and often N_2 fixers, or diazotrophs, are Fe- limited (Kustka et al., 2003; Raven, 1988). Nitrogenase is also sensitive to oxygen (O_2), which has been shown to negatively influence N_2 fixation efficiency (Meyerhof and Burk, 1928; Stewart, 1969). Thus, autotrophic diazotrophs (e.g. cyanobacteria) have evolved strategies, such as temporal and spatial separation of the fixation process, to protect their nitrogenase from O_2 evolution during photosynthesis (Berman-Frank et al., 2001; Haselkorn, 1978; Mitsui et al., 1986). N_2 fixation is widespread and occurs in marine, limnic and terrestrial habitats. In marine ecosystems it mainly occurs in the photic zone, closest to the surface; however, more recently, evidence has shown activity in deeper depths below the photic zone, including oxygen minimum zones (Benavides et al., 2016; Bonnet et al., 2013; Fernandez et al., 2011; Halm et al., 2009; Löscher et al., 2015).

N_2 fixation in the photic zone is often attributed to a diverse group of cyanobacteria. Traditionally, marine, photic-zone dwelling diazotrophs are divided into two groups based on cell diameter, e.g. > 10 µm and < 10 µm size fractions. Diatom diazotroph associations (DDAs), symbioses between heterocystous cyanobacteria and a variety of diatom genera, and large filamentous non-heterocystous *Trichodesmium* spp., compose the larger size fraction (>10 µm). *Trichodesmium* spp. occurs as free filaments or often in two morphologies of colonies: tufts/rafts and puffs. There are three defined lineages of symbionts of DDAs based on their *nifH* phylogeny: het-1 and het-2 refers to the two the *Richelia intracellularis* lineages

which associate with diatom genera, *Rhizosolenia* and *Hemiaulus,* respectively, while the third lineage, het-3, associates with the heterocystous *Calothrix rhizosoleniae* and *Chaetoceros compressus* diatoms (Foster et al., 2010; Foster and Zehr, 2006).

The unicellular diazotrophic cyanobacterial groups are divided into: UCYN-A, UCYN-B, and UCYN-C groups and are representatives of the <10 µm size fraction. The UCYN-A (*Candidatus Atelocyanobacterium thalassa*) group can be further delineated into 6 sublineages (Thompson et al., 2014; Turk-Kubo et al., 2017), two (UCYN-A1, UCYN-A2) are identified as symbiotic with small prymnesiophyte microalgae (reviewed by Farnelid et al., 2016, see references within). The UCYN-B group has its closest cultured relative as *Crocosphaera watsonii* and lives freely, colonially, and also in symbiosis with the diatom *Climacodium frauenfeldianum* (Bench et al., 2013; Carpenter and Janson, 2000; Webb et al., 2009; Zehr et al., 2001). Often overlooked, is the observation that UCYN-B, when colonial or symbiotic could also be associated with the > 10µm size fraction. Less is known about the UCYN-C group, and given that its *nifH* nucleotide sequence is 90% similar (Foster et al., 2007) to *Cyanothece* spp. ATCC51142, it is assumed to be analogous, and thus co-occur with the other < 10 µm size fraction. A diverse group of free-living heterotrophic bacteria (e.g. gamma proteobacteria) (Berthelot et al., 2015; Bombar et al., 2016; Halm et al., 2012; Langlois et al., 2005) and archaea (Zehr et al., 2005) are also within the < 10 µm size fraction.

The distribution and activity of diazotrophs in open ocean ecosystems are governed by different ambient environmental factors, including macronutrient availability (Moutin et al., 2008; Sañudo-Wilhelmy et al., 2001) and temperature (Messer et al., 2016; Moisander et al., 2010). There are also simultaneous influences by several factors (i.e. co-limitation of nutrients, Mills et al., 2004). Moreover, most oceanic models of N_2 fixation assume that all diazotrophs are equally controlled by the same environmental parameters (Deutsch et al., 2007; Hood et al., 2004; Landolfi et al., 2015), despite well recognized differences in genetic

repertoires for assimilating dissolved nutrient pools (e.g. dissolved organic phosphate, Dyhrman et al., 2006; Dyhrman and Ruttenberg, 2006), life histories (free, symbiotic, colonial), and cell sizes (µm to mm). A better understanding of these differences would benefit oceanic models of N_2 fixation, our understanding of marine N_2 budgets and the impact of N_2 fixation in oceanic surface waters and their communities.

The Tropical South Pacific Ocean (TSP) is considered one of the most oligotrophic regions in the World's oceans (Claustre and Maritorena, 2003) with a widespread N deficiency (Deutsch et al., 2007; Raimbault et al., 2007) and the central SP gyre has some of the lowest reported concentrations of dissolved Fe (Blain et al., 2008). One exception is the Western Tropical South Pacific (WTSP), which harbors many islands with Fe rich sediments adding to an island mass effect (Shiozaki et al., 2014) and is influenced by multiple ocean currents, both surface and subsurface, that drive the distribution of dissolved nutrients, micronutrients, and the biota (Fitzsimmons et al., 2014; Gourdeau et al., 2008; Marchesiello and Estrade, 2010; Wells et al., 1999). The structure of these currents also promotes shearing instabilities and strong eddies (Qiu et al., 2009). Moreover, Van Den Broeck et al. (2004) suggested that the WTSP is phosphate limited, while Law et al. (2011) hypothesized that primary production and N_2 fixation in the WTSP follows the seasonality of cyclones, which in their wake, enrich surface waters with phosphate, and fuel primary and new production. An earlier investigation along a transect in the western equatorial Pacific estimated that 74% of the total N_2 fixation could be attributed to the <10 µm size fraction although abundances of unicellular cyanobacteria were moderate (17 cells mL^{-1}) (Bonnet et al., 2009). However, diazotroph quantification is lacking further South in tropical waters, despite being recently recognized as a hot spot of N_2 fixation, with average rates of ~570 µmol N m^{-2} d^{-1} (Bonnet et al., this issue), i.e. in the upper range (100-1000 µmol N m^{-2} d^{-1}) of rates gathered in the global N_2 fixation MAREDAT database (Luo et al., 2012).

The primary aim of this study was to quantify diazotroph abundance and distribution in the WTSP, with an emphasis on symbiotic N_2-fixing populations, both by ´at sea´ and lab based quantitative approaches. Since earlier expeditions to the SW Pacific identified it as a potential 'hotspot' for diazotrophy (Bonnet et al., 2015; 2016; 2017; Moisander et al., 2010), the 'at sea' qPCR was performed in order to make informed decisions about the presence, absence, and relative abundance of diazotrophs so that stations could be selected for other assays and characterization of the hydrographic conditions (e.g. incubation experiments, sediment and drifter deployments (see Bonnet et al. this issue). For a more comprehensive investigation of the symbiotic diazotrophs we developed a new primer and probe set for quantification of the UCYN-A1 host. We also identified key environmental parameters, both biotic and abiotic, which influenced the distribution of diazotrophs in the WTSP and tested the congruency of these parameters in an additional 11 publicly available datasets. We hypothesized that the distribution and the environmental drivers of the diazotrophic symbioses should differ due to the major differences in host taxonomy (e.g. diatom vs. prymnesiophyte), size (1-2 μm to 100's μm), and life history (free vs. symbiotic; chain forming). For comparison and for similarly divergent characteristics (symbiotic vs. free; colonial vs. single), several free-living (UCYN-B, *Trichodesmium* spp. and UCYN-C) cyanobacterial diazotrophs were also included.

2 Materials and Methods

2.1 Sampling

Sampling was conducted on a transect in the WTSP during austral summer (19 Feb-5 Apr, 2015), on board the R/V *L'Atalante* (Fig. 1a). The cruise transect and stations were selected according to the following: regions with low advection to better estimate biogeochemical budgets, contrasting diazotrophic community (e.g. *Trichodesmium* spp. dominated vs. UCYN dominated), and along a trophic and N_2 fixation gradient from west to east (Moutin et al., this

Deleted: Nucleic acid samples were taken from 18 stations: three long duration (LD A, B and C) stations (approximately eight days duration) and 15 short duration (SD 1-15) stations (approximately eight hours duration).

issue). The cruise is divided into two geographic regions: Melanesian archipelago (MA; SD 1-12, LD A and LD B stations, 160º E-178º E and 170º-175º W) and the subtropical gyre (SG; SD 13-15 and LD C stations, 160º W-169º W). Long duration stations (LD) of 8 days were chosen based on hydrographic conditions, satellite imagery, microscopic analyses of >10 µm cyanobacterial diazotrophs and the results of our 'at sea' qPCR analyses of four unicellular diazotrophic targets (UCYN-A1, UCYN-A2, UCYN-B and UCYN-C) (see below and Moutin et al., this issue).

Deleted: transect was divided
Deleted: (Fig. 1a). The first region (
Deleted: ,
Deleted:) included
Deleted: (
Deleted: . The second region (
Deleted: ,
Deleted:) included
Deleted: (
Deleted: D

2.1.1 Nucleic acids

Seawater (2.5 L) was collected from 18 stations (LD A-C; SD 1-15) into clean (10% bleach rinsed) 2.75 L polycarbonate bottles from 6-7 discrete depths based on surface incident light intensity (100, 75, 54, 36, 10, 1, and 0.1%) once per station at both short duration (SD) and LD stations using Niskin bottles (12 L) arranged on a Conductivity Temperature Depth (CTD; Seabird 911) rosette.

After collection from the CTD rosette, 2.5L seawater was immediately filtered onto a 0.2 µm pore size Supor filter (Pall Corporation, Pall Norden AB, Lund, Sweden) held within a 25 mm diameter swinnex filter holder (Merck Millipore, Solna, Sweden) using a peristaltic pump (Cole-Parmer, Masterflex, Easy-load II, USA). In case the filters clogged with too much material, the remaining volume was measured and noted for later calculations. The filters were placed in pre-sterilized bead beater tubes (Biospec Bartlesville, OK, USA) containing 30 µL of 0.1 mm and 0.5 mm glass bead mixture, flash frozen in liquid nitrogen and archived at -80 ºC. Four additional DNA samples were collected from 4 discrete depths, (75, 50, 36, 10 % light), at 11 of the 18 stations, for the 'at sea' qPCR (see below) and filtered as described above. The 'at sea' samples were extracted and processed for qPCR immediately after collection (see below).

2.1.2 Cell abundances and microscopy observations

At the LD stations, 5 L of seawater was collected at the same depths in parallel with the nucleic acid samples from the CTD-rosette. Two sets of samples, one set each day, were taken on day 1 and 3 at each LD station and immediately filtered onto a 47 mm diameter Poretics (millipore) membrane filter with a pore size of 5 µm using a peristaltic pump.

At the SD stations, the same collection was implemented, however a 25 mm diameter Poretics membrane filter was used. The high densities of cells on the latter made it impossible to properly enumerate the various cyanobacterial diazotrophs and as such these samples were used only for qualitative observations (see below). Immediately after filtration, samples were fixed in 1 % paraformaldehyde (v/v) for 30 min prior to storing at -20 ºC. For enumeration, the filter was mounted on a glass slide and examined under an Olympus BX60 microscope equipped with a filter for blue (460-490 nm) and green (545-580 nm) excitation wavelengths. Three areas (area = 0.94 mm^2) per filter were counted separately and values were averaged. When abundances were low, the entire filter (area = 1734 mm^2) was observed and cells enumerated. Due to weak fluorescence, only *Trichodesmium* colonies and free-filaments could be accurately estimated by microscopy and in addition, the larger cell diameter *Trichodesmium*, hereafter referred to as *Katagynemene pelagicum*, was enumerated separately as these were often present albeit at lower cell densities. Other cyanobacterial diazotrophs, e.g. *C. watsonii*-like, *C. rhizosoleniae* (het-3), and *R. intracellularis* (het-1, het-2) were also present on the larger 47 mm diameter samples, however fluorescence was weak and therefore difficult to enumerate. Pico-eukaryote populations, identified as round 1-3 µm diameter cells, with red excitation under the blue filter set, were also observed. For the latter populations, qualitative observations of presence and some details on cell integrity (e.g. fluorescence, diatom frustule, free-living or symbiotic form) are included.

1 **2.2 Nutrient analyses**

2 Seawater for nutrient analyses was collected from each station using the CTD rosette at the
3 same depths as those collected for the nucleic acids. Seawater for inorganic nutrient analysis
4 were collected in 20 mL high-density polyethylene HCL-rinsed bottles and poisoned with
5 $HgCl_2$ to a final concentration of 20 µg L^{-1} and stored at 4°C until analysis. The samples were
6 frozen and returned to the laboratory for analysis within 4 months. Dissolved nitrate and
7 nitrite (NO_3^-+NO_2^-, DIN), phosphate (PO_4^{3-}, DIP) and silicate (Si $(OH)_4$, DiSi) concentrations
8 were determined by standard colorimetric techniques using a segmented flow analyzer
9 according to Aminot and Kérouel (2007) on a SEAL Analytical AA3 HR system (SEAL
10 Analytica, Serblabo Technologies, Entraigues Sur La Sorgue, France). Quantification limits
11 for nitrate, phosphate and silicate were all 0.05 µmol L^{-1}.

12 **2.3 DNA extraction**

13 The DNA from the 120 archived samples was extracted as described in Moisander et al.
14 (2008), with a 30 second reduction in the agitation step in a Fast Prep cell disrupter (Thermo,
15 Model FP120; Qbiogene, Inc. Cedex, France) and an elution volume of 70 µL. The nucleic
16 acid samples collected for the 'at sea' qPCR were extracted immediately after filtration using
17 a modified version of the DNAeasy plant kit (Qiagen) total DNA extraction protocol. The
18 modifications were an initial 2-minute agitation step using a bead beater (Biospec
19 MiniBeadBeater-16, Model 607EUR; Biospec), 10 min proteinase K incubation and final
20 elution volume was 25 µl.

21 **2.4 Oligonucleotide design**

22 A new primer and probe set was designed to amplify the UCYN-A1 host and was based on a
23 published 18S rRNA sequence (accession number JX291893) reported from N. Pacific gyre
24 (station ALOHA) (Thompson et al., 2012). The design utilized the same 96 bp target region of
25 the 18S rRNA used to amplify UCYN-A2 hosts described in Thompson et al. 2014 (Suppl.

1 Table 1). The primers and probe for the UCYN-A1 host 18S rRNA gene assay are as follows:
2 Forward, 5' AGGTTTGCCGGTCTGCCGAT-3'; Reverse, 5'
3 GAGCGGGTGTCGGAGACGGAT-3'; Probe, 5'-FAM-CTGGTAGAACTGTCCT-
4 TAMRA-3'. The forward, reverse and probe contain 2-4, 1, and 5 mismatches, respectively,
5 to UCYN-A2 host sequences (accession number KF771248-KF771254) and the following
6 closely related sequences (98-100%): uncultured eukaryote clones (station ALHOA:
7 EU50069; Cariaco Basin: GU824119) *Chrysochromulina parkeae*: AM490994),
8 *Braarudospaera bigelowii* TP056a: AB250784 *B. bigelowii* Furue-15: AB478413; *B.*
9 *bigelowii* Funahama T3: AB478413; *B. bigelowii* Yastushiro-1 AB478414. The UCYN-A1
10 oligonucleotides specificity was tested *de nova* against the following closely related sequences
11 derived from uncultured eukaryotic clonal sequences (accession numbers: EU500067-68;
12 FJ537341; EU500138-39; EF695227; EU500141; EU499958; EF695229; EF695220). Only
13 one mismatch was found in the forward probe for one sequence (EU500138) (Suppl. Table 1).
14 The nucleotide sequence identity for UCYN-A1 and UCYN-A2 host sequences is 97.95%.
15 Finally, a cross reactivity test between the newly designed UCYN-A1 host oligonucleotides
16 and a dilution series of the UCYN-A2 host template was run (see below).

17 **2.5 Quantitative PCR**

18 Abundances of *nifH* gene copies from selected diazotrophic groups (UCYN-A1, UCYN-A2,
19 UCYN-B, UCYN-C, het-1, het-2, het-3 and *Trichodesmium* spp.) and the 18S rRNA of
20 UCYN-A1 and A2 hosts were quantified using previously published oligonucleotides and
21 TaqMAN assays (Church et al., 2005; Foster et al., 2007; Moisander et al., 2010; Thompson
22 et al., 2014) and the newly designed UCYN-A1 host oligonucleotides (Suppl. Table 1). The
23 qPCRs were conducted in a StepOnePlus system (Applied Biosystems, Life Technologies,
24 Stockholm Sweden) in fast (>40 min) mode with the following parameters: 95 ºC for 20 s,
25 followed by 45 cycles of 95 ºC for 1 s and 60 ºC for 20 s.

1 Cross reactivity tests were run on two of the heterocystous symbiont (het-1 and het-2)
2 oligonucleotides, the UCYN-A1 and UCYNA-2 oligonucleotides, and the newly designed
3 UCYN-A1 host oligonucleotides and UCYN-A2 host primer and probe set. Cross reactivity
4 tests for het-3 were omitted as previously published tests (Foster et al., 2007) showed no
5 cross-reaction with het-1 or het-2. The standard curve for a particular target was run in
6 reactions with the other primers and probe sets. For example, the UCYN-A1 TaqMAN host
7 primers and probes were run in reactions with UCYN-A2 template DNA. The cross reactivity
8 for the het-1 and het-2 primer and probe sets has been previously reported (Foster et al. 2007),
9 however only when the assay is run in standard mode. Standard mode runs the holding,
10 denaturation and annealing stages at the following longer intervals than in Fast mode: 11 min
11 and 40 s, 14 s, and 40 s, respectively. Hence, we tested the cross-reactivity for the het primers
12 and probes when run in fast mode, as the fast mode was used in our study. Similarly, the
13 cross-reactivity between UCYN-A1 and UCYN-A2 were tested in fast mode at two annealing
14 temperatures 60 ºC and 64 ºC; 64 ºC is the recommended annealing temperature for the
15 UCYN-A2 assay (Thompson et al. 2014).

16 Reaction volume was 20 µL in all qPCRs and consisted of 10 µL of 2X TaqMan fast buffer
17 (Applied Biosystems), 5.5 µL of nuclease free water (Sigma Aldrich Sweden AB, Stockholm
18 Sweden), 1 µL each of the forward and reverse primers (10 µM), 0.5 µL of fluorogenic probe
19 (10 µM) and 2 µL of DNA extract. For standard mode runs, the latter master mix was
20 identical with the exception of replacing the fast 2X buffer with the standard 2X buffer. For
21 reactions quantifying *Trichodesmium* spp. *nifH* copies, SD 9 was excluded and 1 µL of DNA
22 template was used for the remaining stations due to low template volume, and total reaction
23 volume was adjusted by addition of 1 µL of nuclease free water. Reactions were performed in
24 duplicates for the 'at sea' qPCR and in triplicates for the archived samples and lab based
25 qPCR. For the 'at sea' qPCR, only four targets (UCYN-A1, UCYN-A2, UCYN-B, and

1 UCYN-C) were quantified and only at the SD stations. No assays were processed at SD 5-6,
2 10-12, and 14 for the 'at sea' qPCR due to no potential LD station selection for these sites.
3 Two µL of nuclease free water was used as template in no template controls (NTCs); no *nifH*
4 copies were detected in the NTCs.

5 Gene copy abundance was calculated from the mean Ct value of the 3 replicates and the
6 standard curve for the appropriate oligonucleotides in the lab based qPCRs. For the 'at sea'
7 qPCR, a mean Ct value of 2 replicates was used to maximize the number of samples run on
8 one amplification plate (96 well). In samples where 1 or 2 out of 3 replicates produced an
9 amplification, signals were noted as detectable, but not quantifiable (dnq) and no
10 amplification was noted as below detection (bd).

11 **2.6 Standard curves and PCR efficiency**

12 Standard curves were plotted and analyzed in Excel for each target based on the qPCR cycle
13 threshold (Ct) values from known dilutions of synthesized target gene fragments (gBlocks®;
14 Integrated DNA Technologies, Leuven Belgium) (359 bp *nifH* and 733 bp 18S rRNA for
15 UCYN-A hosts). Tenfold dilutions were made starting with 10^8 to 10^1 gene copies L^{-1}. The
16 PCR efficiency, for identification of possible interfering contaminants in our samples, was
17 determined as previously described (Short et al., 2004) for 12 samples run on the het-1, het-2,
18 and het-3 primers and probe tests. The qPCR efficiency ranged from 90-99 % with an average
19 of 94 % efficiency for the diazotroph targets het-1, het-2 and het-3.

20 **2.7 Statistics and data analysis**

21 Skewness and normal distribution tests by descriptive statistics was performed in IBM SPSS
22 (ver. 23) on the following parameters recorded during sample collection in the WTSP from
23 the CTD package: depth (m), oxygen (ml L^{-1}), temperature (ºC), chlorophyll fluorescence (µg
24 L^{-1}), photosynthetically active radiation (PAR; µmol photons m^{-2} s^{-1}), salinity (PSU), and gene

copy abundances determined by qPCR. Since fewer samples and targets were processed in the 'at sea' qPCR, only the lab based qPCR results from the archived samples was used for statistical analysis. Significant skew was noted when skewness, divided by its standard deviation, exceeded 1.95. All but three targets (het-1, UCYN-B and *Trichodesmium* spp.) and three environmental parameters (temperature, salinity and oxygen) were significantly skewed (not normally distributed) even after LOG10 transformation. Therefore a non-parametric Spearman's rank correlation was conducted to test possible correlations between the targets and environmental parameters, where we assume that the het groups and UCYN-A clade is symbiotic, while UCYN-B is free living. The resulting correlation matrices were visualized in the form of a heat map of hierarchical clustering in R (ver. 3.2.2) using packages 'hmisc' and 'gplots'. Multivariate statistics by redundancy analysis (RDA) was conducted using the R package 'vegan'. T-tests, in IBM SPSS (ver. 23), were performed to characterize the MA and SG regions and used the mean concentrations of nutrients at each station as input. For meta-analysis on the external data from 11 publically available datasets, sampled in the Atlantic, Pacific and South China Sea, data was acquired from the PANGAEA database and previous publications (Benavides et al., 2016; Bombar et al., 2011; Church et al., 2005; 2008, Foster et al., 2007, 2009; Goebel et al., 2010; Kong et al., 2011; Langlois et al., 2008; Moisander et al., 2008, 2010). We included only datasets with a minimum of 10 data points on the previously mentioned diazotrophic targets. Note that in all datasets the two UCYN-A phylotypes (A1 and A2) were not distinguished, and het-3 was excluded since it was rarely quantified. The meta-analysis was conducted using the software OpenMEE (Wallace et al., 2016) (based on R package 'metafor'), where correlation coefficients from Spearman's rank were z-transformed (Fisher's) and tested using weighted random effect models. Graphical visualization of the mean abundances of the most numerous diazotrophs across the cruise transect was also performed in IBM SPSS (ver. 23).

Deleted: two different regions

Deleted: were

Deleted: along the cruise transect based on

Deleted: environmental parameters, including

Deleted: ,

1 **3 Results**

2 **3.1 Hydrographic conditions**

3 Near surface (0-5m) DIN concentrations were below the quantification limit (bq) in both the
4 MA and SG regions, while the mean surface DIP and DiSi concentrations were below the
5 quantification limit or low across all stations in the MA (bq-0.08 μM and 0.54-0.56 μM,
6 respectively) and significantly ($p<0.001$; t-test) higher (0.18 ± 0.07 μM and 0.79 ± 0.04 μM,
7 respectively) at the stations in the SG (Table 1). The upper 25-30 m of the euphotic zone had
8 stable temperatures of 29-30 ºC, and the thermocline was deeper in the SG compared to the
9 MA. The depth of the deep chlorophyll maximum (DCM) was between 70-165 m throughout
10 the cruise, except for LD B (DCM at 35 m), which was sampled during a degrading surface
11 phytoplankton bloom. A 30-day composite of the surface chlorophyll *a* (Chl *a*) confirmed the
12 decreasing level of surface fluorescence measured by the CTD package at LD B (data not
13 shown).

Deleted: depth

Deleted: throughout the cruise transect

Deleted: but declined differently in deeper waters for individual stations of the MA compared to the SG as a whole (Suppl. Fig. 2)

14 **3.2 Cross reactivity tests**

15 No amplification was detected for the newly designed UCYN-A1 host oligonucleotides run
16 with the UCYN-A2 as template DNA and vice versa (Suppl. Fig. 1a-b).

17 Running the het assay in fast mode showed a lower cross-reactivity between the het-1 assay
18 and the het-2 template than vice versa (the het-2 assay and het-1 template) (Suppl. Fig. 1e-f).
19 In fact, no amplification was detected in the last two template additions and the Ct differences
20 were > 9 when het-1 assay was run with het-2 templates. The UCYN-A2 assay detected the
21 UCYN-A1 template in all but the last template addition and with Ct differences >3 (1 order of
22 magnitude) while there was a 18-20 difference in Ct value (less gene copies) when UCYN-A1
23 assay was run in fast mode with UCYN-A2 templates at either annealing temperature (60° C
24 or 64° C) and only the first three template additions (10^8-10^6 *nifH* copies μL^{-1}) were detected
25 (Suppl. Fig. 1c-d).

1 **3.3 Comparison of 'at sea' and lab-based qPCR**

2 In order expedite the sample processing for the 'at sea' qPCR, a shortened and modified DNA
3 extraction protocol was performed, 4 depths were sampled, and 4 targets run (UCYN groups).
4 In total, 44 samples can be compared with results from the parallel archived samples
5 processed in the lab. Moreover, we considered the differences to be noteworthy only when
6 there was at least one order of magnitude difference in detection. A summary of the
7 comparison, including the difference in *nifH* copy abundance is provided in Suppl. Table 2.
8 In addition, a graphical comparison shows a 1:1 relationship between the *nifH* gene copies
9 detected in the 'at sea' and archived samples (Suppl. Fig. 3).

10 In general, the 'at sea' and lab based qPCR were similar in quantifying the targets.
11 Discrepancies were noted in 7, 8 and 11 samples, which had higher detection in the 'at sea'
12 analyses for UCYN-A1, UCYN-A2 and UCYN-B, respectively. There were fewer instances
13 (3, 4, and 5, respectively) of samples processed in the lab with the full extraction that had
14 higher *nifH* gene abundances for the UCYN-A1, UCYN-A2 and UCYN-B, respectively.

15 **3.4 Horizontal and vertical distributions**

16 *Trichodesmium* and UCYN-B were the most abundant diazotrophs and *nifH* gene abundances
17 ranged 10^4-10^6 *nifH* copies L^{-1} at multiple depths (4-6 depths) in the upper water column (0-
18 35 m) (Fig. 1-2; Suppl. Table 3). *Trichodesmium* represented 80-99% of total *nifH* genes
19 detected at 9 out of 17 stations with highest detection in the MA and low to bd in the SG.
20 Microscopy observations and abundances of *Trichodesmium* spp. confirmed a high abundance
21 of free filaments of *Trichodesmium* and *C. watsonii*-like cells at LD B, while colonies were in
22 general rarely observed (Suppl. Table 5).

23 At stations where *Trichodesmium* was not the most abundant diazotroph (e.g. SD 2, 6,
24 7, 14, 15, and LD C), UCYN-B had the highest depth integrated *nifH* copy abundance.

UCYN-B was also the most consistently detected diazotroph, and was quantifiable from all stations sampled accounting for 81-100% of the total detected *nifH* gene copies in the SG. There was also a depth dependency for maximum *nifH* gene abundance such that the average depth maxima's of *Trichodesmium* (1.88 x 10^6 *nifH* copies L^{-1}) and UCYN-B (5.03 x 10^5 *nifH* copies L^{-1}) at the stations in the MA were 10 and 25 m, respectively. In the SG, the average depth maximum for UCYN-B (1.50 x 10^5 *nifH* copies L^{-1}) was the same (25 m), while the average depth of the *Trichodesmium* maximum (1.30 x 10^4 *nifH* copies L^{-1}) deepened to 31m.

Of the three heterocystous cyanobacterial symbiont lineages (het-1, het-2 and het-3), het-1 was the most dominant (60% detection in total samples, 72 of 120 samples), and similar to *Trichodesmium*, had higher detection in the stations of the MA region. For example, at stations SD 2, 4 and 9, het-1 represented 10-15% of the total *nifH* genes quantified in the depth profiles, but in the total *nifH* genes quantified across the entire transect, het-1 only represented 1.5 %. Abundances for het-1 ranged between 10^3-10^5 *nifH* copies L^{-1} (15 of the 18 stations) at multiple depths (0-90 m) and the average depth maximum at MA stations (3.11 x 10^4 *nifH* copies L^{-1}) was closer to the surface (15 m) compared to the SG stations (1.63 x 10^3 *nifH* copies L^{-1}; 60 m) (Fig. 1; Suppl. Table 3). Het-2 and het-3 co-occurred with het-1, however at lower abundances (10^2-10^4 *nifH* copies L^{-1}) and often bd. For example, het 2 and het-3 were bd at all depths sampled in 1 and 3 stations, respectively, located in the SG. The average depth of maximum abundance (17 m) for het-2 (3.89 x 10^3 *nifH* copies L^{-1}) was similar to het-1 (15 m), while het-3 (1.53 x 10^3 *nifH* copies L^{-1}) was deeper at 33 m (considering only the MA stations). Microscopy observations confirmed the presence of *R. intracellularis* at 5 SD stations of the MA and LD B and absence at the SD stations and LD C of the SG. Noticeable was the co-occurrence of free filaments of *R. intracellularis* and degrading diatom cells (mainly belonging to the genus *Rhizosolenia*), especially at the SD 5, 6 and 7.

1 The unicellular symbiotic groups, UCYN-A1 and A2 (and their respective hosts), were
2 the least detected targets. For example, UCYN-A1 was bd in 53% (63 of 120 samples) and
3 UCYN-A2 was bd in 66% (79 of 120 samples) of samples. UCYN-A1 and A2 represented <
4 0.4 % of total *nifH* genes detected and UCYN-A symbionts were bd in the SG, except at LD
5 C. When detected, average *nifH* gene abundance for UCYN-A1 and A2 were 8.60 x 10^4 and
6 4.60 x 10^4 *nifH* copies L^{-1}, respectively, and usually accounted for <1.0-1.5 % of the total *nifH*
7 gene copies enumerated per station. One exception was at LD C in the SG, when UCYN-A1
8 and A2 accounted for 4 and 11%, respectively, of the total *nifH* gene copies, and were the
9 second most abundant diazotroph (3.19 x 10^4 and 8.53 x 10^4 *nifH* copies L^{-1}). The average
10 depth of maximum *nifH* gene abundance for the UCYN-A1 (1.60 x 10^4 *nifH* copies L^{-1}) and
11 A2 (5.76 x 10^3 *nifH* copies L^{-1}) symbionts was consistently recorded at deeper depths (55 and
12 58 m, respectively; 10 % light level).

13 The detection of the UCYN-A1 and A2 hosts mirrored the detection of their respective
14 symbionts. However, in 22 and 15 samples, respectively, the UCYN-A1 and A2 symbionts
15 were quantified while their hosts were bd. The UCYN-A hosts were never detected in samples
16 where their respective symbionts were bd or dnq. When both UCYN-A host and symbiont
17 were present, the abundances of the host's 18S rRNA genes were always one order of
18 magnitude less than their respective symbiont's *nifH* genes, with the exception of two samples
19 for UCYN-A1 symbionts where their respective host 18S rRNA gene abundances were half,
20 or nearly equal in abundance. UCYN-C was the least abundant unicellular diazotroph and was
21 only quantified in the 'at-sea' qPCR where detection was poor and limited to the MA region
22 (3 of 11 stations: 1-3 of 4 depths sampled) and abundances never exceeded 10^2 *nifH* copies L^{-1}
23 (Suppl. Table 3).

3.5 Diazotroph and UCYN-A host covariation

Several significant correlations between the target diazotrophs and hosts were identified (Fig. 3; Suppl. Table 4a). The *nifH* gene abundances of *Trichodesmium* and UCYN-B were significantly positively correlated with each other ($N=108$, $p<0.01$). In addition, UCYN-B *nifH* gene copy abundance was significantly positively correlated with those of both UCYN-A symbionts (A1 and A2; $N=120$, $p< 0.01$) and UCYN-A2 host abundance ($N=118$, $p<0.04$). Abundances of UCYN-A1 and A2 *nifH* genes were significantly positively correlated with each other, and in addition, with their respective host 18S rRNA gene abundances ($N=118$-120, $p<0.01$). Lastly, the *nifH* gene abundances of het-1, het-2 and het-3 were significantly positively correlated with one another, and with the *nifH* gene abundances of *Trichodesmium* and UCYN-B ($N=108$-120, $p<0.01$). The only correlations that were not significant were between the UCYN-A (including their hosts) and *Trichodesmium* and the het-groups (with the exception of het-3, which correlated with the UCYN-A2 host ($N=118$, $p<0.04$)).

3.6 Influence of environmental conditions on diazotroph and UCYN-A host abundances in the WTSP

The *nifH* gene abundances of UCYN-A1 and A2 were significantly positively correlated with salinity ($N=107$, $p<0.02$ and $N=107$, $p<0.03$, respectively) and depth ($N=118$, $p<0.01$ and $N=118$, $p<0.002$, respectively) (Fig. 3; Suppl. Table 4b). However, except for het-3 (regarding depth), all other diazotrophs were significantly negatively correlated with salinity ($N=97$-108, $p<0.006$) and depth ($N=108$-120, $p<0.001$). Moreover, *Trichodesmium*, UCYN-B, and the het-group (except het-3) were significantly positively correlated with PAR ($N=63$-72, $p<0.003$) and temperature ($N=97$-108, $p<0.001$) while UCYN-A1 and A2 were significantly negatively correlated ($N=72$-108, $p<0.02$) with the latter parameters. All diazotrophic targets, except UCYN-A1, UCYN-A2, and het-3, were significantly negatively correlated with DIN concentration ($N=90$-94, $p<0.02$). All diazotrophs, except UCYN-A2, were significantly

Deleted: Similarly, all

negatively correlated with DIP concentration (N=96-100, p<0.02), and all diazotrophs except UCYN-A1, A2 and het-3 were significantly negatively correlated with DiSi concentration (N=97-102, p<0.001). The gene copy abundances for the UCYN-A hosts' and UCYN-A1 and UCYN-A2, and UCYN-B were significantly correlated with dissolved oxygen (UCYN-A and hosts, N=106-108, p<0.003 and UCYN-B, N=108, p<0.03). In general, the correlations between *nifH* gene abundances and several hydrographic parameters divided the diazotrophs into two groups: the UCYN-A symbionts (and respective hosts) and all other target diazotrophs.

Hierarchical clustering based on the Spearman's rank analyses resulted in two major groups: (1) a shallow group and (2) a deep group. The clustering in the euphotic zone was inferred from the negative and positive correlations for shallow and deep groups, respectively, with depth (Fig. 3). For example, *Trichodesmium* and the symbiotic het-1 and het-2 lineages characterize an upper water column (shallow) group with significant clustering and positive correlations with temperature (N=97-108, p<0.001) and PAR (N=63-72, p<0.003), while only UCYN-A1 and A2 symbionts and their respective hosts represent the deep group. UCYN-B was unique in an overlapping distribution, and resulted in positive significant correlations with both the shallow and deep groups euphotic zone diazotrophs (e.g. *Trichodesmium*, N=108, p<0.001 and UCYN-A1, N=120, p<0.004, respectively). Diazotroph targets in the deep dwelling group significantly clustered and correlated positively with oxygen, depth, salinity and fluorescence (N=108-120, p<0.03, except for UCYN-A2 and fluorescence, N=108, p=0.053). Despite clustering with the shallow group, het-3 was less robust in a negative correlation with salinity (N=108, p=0.005).

The results from the Spearman's rank correlations were further confirmed and visualized in the RDA biplot (Fig. 4a), which explains parameter importance (Fig. 4b). Correlations with nutrients and PAR were omitted due to the limited number of data points.

Deleted: of
Deleted: 18S rRNA genes,
Deleted: *nifH* genes
Deleted: ;
Deleted: er
Deleted: ,
Deleted: ,
Deleted: ir
Deleted: 1
Deleted: 2
Deleted: (group 1)
Deleted: (group 2)
Deleted: er
Deleted: 2
Deleted: 1

Fluorescence, depth and salinity correlated positively with each other and negatively with temperature, while oxygen was not significantly correlated with any other environmental parameters. The response variables UCYN-A1 and A2 and their respective hosts clustered with the explanatory variables: fluorescence, salinity and depth, with a dependency towards oxygen. On the other hand, the shallow euphotic group (response variables *Trichodesmium*, het-1 and het-2) clustered closer to explanatory variable temperature. In addition, most of the observed variance is explained by the two axes RDA1 (72 %) and RDA2 (22 %), indicative of depth and temperature, respectively, as the most important environmental parameters for diazotroph abundance in our study. Together they form a depth-temperature gradient (RDA1) where *Trichodesmium* occupies the warmest and shallowest waters, and UCYN-A1 and A2 occupies the coldest and deepest waters, among the investigated cyanobacterial diazotrophs.

Deleted: er

Deleted: 1

3.7 Global drivers of diazotrophic abundance

We found consistency between our results in the WTSP and the correlations identified in the 11 external datasets by the non-parametric correlation analyses and meta-analyses (Fig. 5a-d; Suppl. Table 6). For example, in three of the external datasets, *nifH* gene abundances of *Trichodesmium* spp., UCYN-B, and het-1, were significantly positively correlated with temperature and negatively correlated with the same three parameters as in our study in the WTSP: salinity, DIP, and DIN. The latter correlations were identified in two regions of the WTSP (tropical and subtropical) and in the northern South China Sea (NSCS). In contrast to a significant positive correlation between UCYN-A *nifH* gene abundance and depth reported here in the WTSP, UCYN-A *nifH* gene abundance was negatively correlated with depth in 4 of the 11 external datasets (two regions of the WTSP, Tropical Atlantic (TA), and NSCS). Moreover, and consistent with several of the other diazotrophs (*Trichodesmium*, UCYN-B, het-1), UCYN-A *nifH* gene abundance was negatively correlated with DIP and DIN concentrations (5 and 3 additional external datasets, respectively) (Suppl. Table 6).

Meta-analysis revealed similar groupings (e.g. shallow and deep) as observed in the WTSP, however, the significance was less robust (Fig. 5a-d; Suppl. Table 6). For example *nifH* gene abundances of *Trichodesmium* and het-1 and het-2 were significantly positively correlated with temperature and negatively correlated with salinity ($p<0.05$). No significance was found for UCYN-A *nifH* gene abundance for the latter parameters, and UCYN-B *nifH* gene abundance was not significantly correlated with salinity, but was significantly positively correlated with temperature ($p<0.05$). In addition, UCYN-A was the only diazotroph that was not significantly correlated with het-2, while all other diazotrophs had a significant positive correlation with het-2 ($p<0.05$). Similar to our findings reported for the WTSP, all diazotrophs, except UCYN-A, correlated significantly negatively with depth, DIP and DIN concentrations ($p<0.05$) (except het-2 with DIP which was not significant). Finally, UCYN-B and het-1 *nifH* gene abundances were significantly negatively correlated with Chl *a* ($p<0.05$), while *Trichodesmium*, UCYN-A and het-2 were not.

4 Discussion

4.1 Environmental conditions in the WTSP

The SP is one of the most oligotrophic regions of the world's oceans with chronically low dissolved nutrient concentrations, especially DIN, and thus, is considered an area primed for N_2 fixation. Likewise, we encountered surface hydrographic conditions in the WTSP that were consistently low in dissolved nutrient concentrations and similar to earlier reports for the equatorial Pacific (Bonnet et al., 2009; Dufour et al., 1999; Moutin et al., 2008; Van Den Broeck et al., 2004). The conservative tracers of temperature and salinity remained constant in the surface between the MA and SG regions, hence the elevated nutrient concentrations in the SG is likely not related to an eddy intrusion. The deviation away from a 16:1 relationship (Redfield ratio) (data not shown) in the upper 125 m in both regions (MA and SG) was indicative of DIN limitation. The low DIP concentrations in MA waters suggest utilization of

DIP by diazotrophs in the absence of DIN, however, other sources of nitrogen were likely available to the phytoplankton community, e.g. dissolved organic nitrogen and N_2 fixation (Karl et al., 2001).

4.2 Detection of diazotrophs and application of 'at sea' qPCR

Trichodesmium, UCYN-B, and the het-group are easily identifiable by standard epi-fluorescence microscopy, and so these populations can readily be observed 'at sea'. However, the UCYN-A1 and UCYN-A2, and their respective hosts, require a lengthy fluorescent *in situ* hybridization (FISH) protocol that is difficult to implement in the field. On the other hand, modern oceanographers have a suite of other molecular genetic tools, some of which are also 'sea-going' and autonomous (e.g. Robidart et al. 2014; Ottesen et al. 2013; Preston et al. 2011), thereby making quantification of microscopically unidentified microorganisms tangible by quantifying their genes, simultaneous with collection of hydrographic data. Here, we describe a rather effective, steadfast (within 3 hrs of sample collection), and 'sea-going' nucleic acid extraction and qPCR technique to quantify diazotrophs by their *nifH* gene, which was used in real time during the OUTPACE cruise to help locate the LD stations for the purpose of the project (see Moutin et al., this issue). The comparisons of the 'at sea' assays to the lab-based full extraction protocol and qPCR on archived samples indicated that the assays were consistent (Suppl. Fig. 3a-c), and surprisingly the shortened DNA extraction performed 'at sea' had higher abundances for all three targets (UCYN-A1, UCYN-A2 and UCYN-B) in 16-25 % of the samples processed, depending on the target diazotroph.

The disparity between the enumerations made 'at sea' and in the lab-based samples could result from technical and/or sampling differences. For example, we do not know the DNA extraction efficiency of either method used ('at sea' vs. lab based), nor if extraction efficiency varies for a particular target e.g. UCYN-A or UCYN-B. Variations between samples or replicates could be attested to different DNA extractions varying in their efficiencies based on

1 target organism (Boström et al., 2004). Some strains of UCYN-B are known to secrete thick
2 extracellular matrix (ECM) (Sohm et al. 2011), while others form dense colonies, and some
3 are symbiotic with diatoms (Carpenter and Janson, 2000). Moreover, the UCYN-A symbioses
4 are considered fragile and easily dismantled (Thompson et al., 2012). The latter observations
5 could potentially result in differences in extraction efficiency, which is supported by
6 significant disparities in all samples where the UCYN-A hosts were enumerated. However we
7 find no clear pattern in higher detection for one method ('at sea' vs. lab-based assays) or one
8 particular target. Despite taking the samples from the same Niskin bottle for the archived and
9 'at sea' assays, we cannot discount the natural heterogeneity of plankton and differences in
10 settling, and that perhaps some samples were more enriched than others. Seldom are replicates
11 processed in qPCR studies as those presented here, and so the disparity reported here
12 highlights the need to address variation derived from sampling and/or extraction efficiencies.
13 Nonetheless, 'at sea' (and lab-based) qPCRs could also be appended with a multi-plexing
14 approach to both increase and broaden the number of metabolic pathways (e.g. *narB, rbcL,*
15 *nirS*) and/or phylotypes quantified simultaneously.

16 **4.3 Abundance and vertical distribution of diazotrophs in the WTSP**

17 Earlier work based on N isotope ratios (delta ^{15}N) of suspended particulate matter and
18 dissolved organic N (DON) in the WTSP suggested that new production is likely fueled by N_2
19 fixation in this region (Hansell and Feely, 2000; Yoshikawa et al., 2005). The SP is also an
20 area where high *nifH* gene abundances of the unicellular diazotrophs, in particular UCYN-A
21 and UCYN-B, have been previously reported (Biegala and Raimbault, 2008; Bonnet et al.,
22 2009, 2015; Moisander et al., 2010) and account for a significant (74%) portion of the areal
23 N_2 fixation (Bonnet et al., 2009). Hence, it was likely to encounter the presence of
24 diazotrophic populations.

1 Recently UCYN-A and its various lineages have been highlighted as one of the most
2 widespread and abundant diazotrophs (Farnelid et al., 2016 and references therein), which has
3 led to the dramatic shift in the canonical paradigm of *Trichodesmium* as the only significant
4 diazotroph. Surprisingly, here, we report *nifH* gene abundances of the UCYN-A1 and UCYN-
5 A2 lineages that are comparatively lower than earlier reports. In fact, UCYN-A1 and A2 were
6 the least detected diazotrophs of our targets. Consistent with higher UCYN-A1 and A2
7 biomass (3.2 x 10^4 and 8.5 x 10^4 *nifH* copies L^{-1}, respectively) at depth (60 m) at LD C were
8 microscopy observations of high abundances of picoeukaryotes similar in size and shape
9 previously reported for the UCYN-A hosts (Krupke et al. 2013). The vertical distribution of
10 UCYN-A1 (and A2) was similar to that reported by Moisander et al. (2010) and others,
11 including earlier studies in the North Pacific Ocean (NP) and North Atlantic (NA), where
12 maximum abundances of UCYN-A are common to deeper depths in the euphotic zone (below
13 45 m) (e.g. Bonnet et al., 2015; Foster et al., 2007; Goebel et al., 2010; Needoba et al., 2007).
14 Likewise, we also observed as others (Cabello et al., 2016) that the UCYN-A1 and A2
15 symbioses co-occur and typically have decreased abundance towards the DCM, and
16 maximum abundances slightly above the nitracline.

17 Unlike UCYN-A phylotypes, high *nifH* gene abundances of *Trichodesmium* was
18 common in the upper 10 m of the MA region, and included both surface slicks and free
19 filaments, which was consistent with earlier observations of high surface densities further
20 north in the SP (Moisander et al., 2010; Shiozaki et al., 2014), as well as regional satellite
21 observations of blooms (Dupouy et al., 2011). Surface slicks have also been reported
22 elsewhere, e.g. the NA (Goebel et al., 2010; Langlois et al., 2005). The depth of maximum
23 *nifH* gene abundance for *Trichodesmium* deepened from the MA (10 m) region to the open
24 gyre (SG, 31 m), which was similar to earlier reports in the equatorial Pacific (Bonnet et al.,
25 2009). A niche partitioning has been suggested for *Trichodesmium* and unicellular diazotrophs

Deleted: Both UCYN-A phylotypes were largely restricted to the MA, with the exception of high densities (3.2 x 10^4 and 8.5 x 10^4 *nifH* copies L^{-1}, respectively) found at one depth (60 m) of LD C, which borders the MA region.

Deleted: UCYN-B and *Trichodesmium* were the most abundant diazotrophs in the WTSP, and UCYN-B in particular was the most detected phylotype (99% detection; dnq or higher in 119 of 120 samples). H

Deleted: ing

Deleted: presence of

Deleted: was widespread in the MA region and

Deleted: i

in the SP (Bonnet et al., 2015; Moisander et al., 2010) and elsewhere (Goebel et al., 2010; Langlois et al., 2005; Messer et al., 2015). However, here in the WTSP, *Trichodesmium* abundance was positively correlated with UCYN-B, which instead is consistent with previous studies in other ocean basins, e.g. the Atlantic Ocean (Foster et al., 2007, 2009; Langlois et al., 2008), and the South China Sea (Moisander et al., 2008). UCYN-B co-occurred with *Trichodesmium* in the surface samples, although at lower *nifH* gene abundances, and more often UCYN-B had subsurface maxima (35-70 m) in both regions (MA and SG) of the transect. The latter is also consistent with Moisander et al. (2010) who observed maximum *nifH* gene abundances of UCYN-B north of the Fijian islands at 37m.

All 3 heterocystous symbiont phylotypes co-occurred and were widespread in the MA, with het-1 as the most abundant and most highly detected het group (70% detection or 84 of 120 samples). The early work of Moisander et al. (2010) detected het-1 in all but one of 26 stations sampled (56% detected, or 56 of 100 samples), and highest *nifH* gene densities were reported north east of our cruise transect. Moreover, Bonnet et al. (2015) detected het-1 and het-2 at the surface of one out of 10 stations west (approximately 10 degrees W) of our cruise transect. Het-2 and het-3 were not quantified by Moisander et al. (2010) and het-3 was not quantified by Bonnet et al. (2015). Therefore our study is among the first to report on the abundances and distributions for all 3 heterocystous diazotrophs in a large expanse of the SP. The three het phylotypes were however recently reported from a mesocosm (enclosed design) experiment in the Noumea lagoon, a low nutrient low chlorophyll (LNLC) region located along the New Caledonian coast (Turk-Kubo et al., 2015). In fact, het-1 and het-2 were among the most abundant diazotrophs in the first half of the experiment (Turk-Kubo et al., 2015). Two additional earlier studies have also reported microscopic observations of free-living *Richelia* in the same lagoon (Biegala and Raimbault, 2008; Garcia et al., 2007).

1 The highest densities (10^4-10^6 *nifH* copies L^{-1}) of the *Richelia* phylotypes (het-1 and
2 het-2) were restricted to the western region of the MA, and in the upper 12 m, which is
3 shallower than the subsurface maximum (e.g. ≥ 25 m) previously reported for het-1 (and het-
4 2) in the Western Tropical North Atlantic (WTNA) and NP (Church et al., 2005; Foster et al.,
5 2007; Goebel et al., 2010). Our microscopy observations from SD 5-7 and LD A indicated
6 that near surface *Rhizosolenia* populations were in a moribund state since frustules were
7 broken and free filaments of *Richelia* were observed. Our observations also coincide with a
8 region of high backscattering measurements in the upper water column (5-30 m) (Dupouy et
9 al., this issue). Het-1 *nifH* genes were four orders of magnitude higher in abundance in the
10 moored sediment traps of LD A (325 m: 2.0 x 10^7 *nifH* copies L^{-1}) and LD B (325 and 500m:
11 5.8 x 10^6 and 1.10 x 10^7 *nifH* copies L^{-1}, respectively) (Caffin et al., this issue) than the *nifH*
12 genes detected in the overlying waters (3.11 x 10^3 *nifH* copies L^{-1} and 4.1 x 10^2 *nifH* copies L^{-1}
13 , respectively). Combined, the latter observations suggest that a higher density of the het-1
14 population was likely present prior to our sampling and perhaps derived from a 'seed'
15 population originating in the coastal regions of New Caledonia, and that they play an
16 important role for export production in this region, as has previously been shown in e.g. the
17 NP (Karl et al., 2012).

Deleted: commonly

18 The UCYN-C phylotype was poorly detected in the 'at sea' assays (61% samples were
19 bd and maximum abundance was $5.0x10^2$ *nifH* copies L^{-1}), and as such was not enumerated in
20 the archived samples. The low detection of UCYN-C is consistent with Taniuchi et al. (2012),
21 who estimated that UCYN-C only represented a small portion of diazotrophs detected in the
22 western Pacific (Kuroshio Current). However, a recent study reported relatively high UCYN-
23 C abundances in the open waters of the Solomon Sea (north of the MA) (Berthelot et al.,
24 submitted). UCYN-C has also been observed in the New Caledonian lagoon (Turk-Kubo et
25 al., 2015), where it was the most dominant diazotroph in the first part of the aforementioned

1 mesocosm experiment (Turk-Kubo et al., 2015). Moreover, Turk-Kubo et al. (2015) reviewed
2 the specificity of the UCYN-C assay used in our study (Foster et al., 2007) and concluded that
3 it does quantify a majority of UCYN-C phylotypes. Hence, like most plankton, abundances
4 can be patchy as was observed with UCYN-C in our study.

5 **4.4 UCYN-A and host (co)-occurrence**

6 Earlier and recent work has suggested a high host dependency (e.g. smaller and streamlined
7 genomes), and selectivity in the UCYN-A based symbioses (Cabello et al., 2016; Cornejo-
8 Castillo et al., 2016; Farnelid et al., 2016; Krupke et al., 2013, 2014; Thompson et al., 2012;
9 Tripp et al., 2010). Moreover, the UCYN-A partnerships are also considered mutualistic,
10 where the host and symbiont both benefit by exchange of metabolites (e.g. reduced C and N,
11 respectively) (Krupke et al., 2014; Thompson et al., 2012) and based on streamlined genome
12 content of UCYN-A, it has a photo-fermentative metabolism dependent on other organisms
13 (Tripp et al., 2010; Zehr et al., 2008). Hence one would expect parallel distributions for both
14 partners. Some have argued that the partnership is also obligatory since few observations of
15 free-living hosts have been reported and abundances of free symbionts assumed to be derived
16 from disruption during sample preparation are always correlated with their hosts (Cabello et
17 al., 2016; Krupke et al., 2014; Thompson et al., 2012). Thus, by use of our newly designed
18 oligonucleotides for the UCYN-A1 host and previously designed oligonucleotides for the
19 UCYN-A2 host (Thompson et al., 2014), we unexpectedly found that both UCYN-A1 and A2
20 were often (89% and 59%, respectively; not considering dnq) detected in the absence (or bd)
21 of their respective hosts, while the hosts, when detected, always coincided with increased
22 UCYN-A abundance. Our observations could result if the UCYN-A lineages live in either a
23 loose association, or perhaps with a wider diversity of hosts than detected by the UCYN-A
24 host assays.

The number of cells per partner lineage is considered specific as well, such that 1-2 UCYN-A1 cell(s) is associated with a prymnesiophyte partner (UCYN-A1 host) and the larger *B. bigelowii* (UCYN-A2 host) host associates with multiple and variable numbers of UCYN-A2 cells to compensate for its higher N requirement (Cornejo-Castillo et al., 2016). We consistently observed higher abundances of the *nifH* genes for the UCYN-A1 and A2 symbionts compared to their respective hosts' 18S rRNA genes (2-10 and 6-34 times more abundant, for A1 and A2, respectively). Similar results were found by Thompson et al. (2014) who observed symbiont/host ratios of 0.2-11. However, a higher detection of the UCYN-A 1 (and A2) in absence of the host detection would result if there is a broader diversity of hosts associated with the UCYN-A symbionts. Ultimately a CARD-FISH approach using the same 18S rRNA oligonucleotide as in the qPCR would help distinguish the discrepancy between the partner gene copy abundances.

A symbiosome-like compartment has also been described attached to the UCYN-A2 host or residing free (Cornejo-Castillo et al., 2016). Thus, another plausible explanation for the higher abundances of the UCYN-A2, in particular, in the absence of their respective host, could result if our assays quantified UCYN-A2 residing in a dislodged free-floating symbiosome, or an overestimate of the UCYN-A2 due to cross-reactivity with UCYN-A3 lineage as expected by *in silico* tests (Farnelid et al. 2016). It is less likely that the UCYN-A2 was overestimated due to cross-reaction with UCYN-A1 templates since our cross-reactivity tests showed a weak cross-reaction (see below).

4.5 Environmental influence on diazotroph abundances and distributions

The annual N inputs through biological N_2 fixation in the oceans is considered high, ranging from 100-200 Tg N (Eugster and Gruber, 2012; Luo et al., 2012), yet large uncertainties remain in what factor(s) influence the abundance, distribution, and activity of marine diazotrophs. Initially, we hypothesized that the conditions favoring a particular cyanobacterial

1 diazotroph would be unique given the contrasting life histories of different diazotrophic
2 groups (free-living, colonial, and symbiotic). Moreover, we also suspected that the conditions
3 promoting DDAs would differ from those influencing the UCYN-A based symbioses given
4 the vast differences in the symbionts and hosts (e.g. genome content of symbiont, cell size of
5 symbiont and hosts in the two systems; expected number of symbionts/host; host phylogeny:
6 diatom vs. prymnesiophyte). Thus, determining the condition or sets of conditions that drive
7 cyanobacterial diazotroph distribution, abundance, and activity is of great interest.

Deleted: differ

8 Hydrographic conditions and dissolved nutrient concentrations measured at the time of
9 sampling were used to correlate diazotrophic abundances with various environmental
10 parameters. Consistently, in two independent statistical tests, two groups emerged in the
11 WTSP; 1) UCYN-A1 and A2 and their respective hosts 2) het-1, het-2 het-3, UCYN-B and
12 *Trichodesmium*. Thus, conditions favoring the UCYN-A based symbioses differ from the
13 conditions for DDAs, and free-living cyanobacterial diazotrophs.

Deleted: agreeing with our initial hypothesis that

14 Temperature is often cited as the most important driver of diazotroph abundance and
15 distribution (Messer et al., 2016; Moisander et al., 2010), and recently marine nitrogen
16 fixation was shown to be controlled by surface irradiance and dissolved oxygen concentration
17 (Luo et al., 2014). As shown earlier in the WTSP, both *Trichodesmium* spp. and UCYN-B
18 were most abundant in warmer surface waters (> 27 ºC) in the north, while UCYN-A
19 dominated in the cooler (24-26 ºC) southern waters of WTSP (Bonnet et al., 2015; Moisander
20 et al., 2010). Likewise, we found similar abundances and temperature optima for the latter
21 three diazotrophs and significant correlations between the various diazotrophs and
22 temperature. In fact, all diazotrophs, except the UCYN-A lineages were significantly
23 positively correlated with temperature in the WTSP. In addition to temperature, the
24 environmental parameters of PAR, salinity and depth were also significantly influencing *nifH*
25 gene abundance and distribution. Moreover, the latter two variables separated the abundances

of the UCYN-A symbioses (A1 and A2) from the rest of the diazotrophs in the WTSP, including both free-living phylotypes and the symbiotic heterocystous lineages.

The maximum abundances at depth for UCYN-A1 and A2 were slightly above or at the nitracline and coincided with higher measures of fluorescence from the CTD. The latter is consistent with observations of high UCYN-A abundances in coastal habitats (Bombar et al., 2014), estuaries (Messer et al., 2015), or in waters that have been recently entrained with new nutrients (Moisander et al., 2010). Increased *nifH* copies and/or *nifH* gene expression for UCYN-A have also been reported from bioassay experiments amended with nutrients, including DIN, phosphate and iron (Krupke et al., 2015; Langlois et al., 2012; Moisander et al., 2012). The latter is in contrast with the data reported here in the WTSP (including the meta-analysis) and several of the external datasets (e.g. WTSP, TA, NA, NSCS), which finds a negative correlation between DIN and DIP concentrations and *nifH* gene abundance of most of the diazotrophs, including UCYN-A. Across the cruise transect conditions were instead favoring diatoms in symbiosis with diazotrophs, since surface DIP and DiSi concentrations were considered not limiting (Thierry Moutin, this issue), while DIN was below detection, and as reported here, resulted in higher abundances of het-1 *nifH* gene copies and observations of *Rhizosolenia* hosts in the MA. In the WTNA, waters with high DiSi concentration and low N:P ratios, driven by a disproportionate utilization of N relative to P, results in consistent and widespread blooms of the *Hemiaulus-Richelia* symbioses (het-2) (Foster et al., 2007; Subramaniam et al., 2008).

All the diazotrophs described here are either photoautotrophic or associated with photoautotrophic partners (UCYN-A, het-group). Therefore, light irradiance (e.g. PAR) and availability will impact the abundance and distribution of the diazotrophic populations. Results from CARD-FISH observations of the UYCN-A1 and A2 symbioses have reported a strong dependency on light intensity, which results in higher abundances nearer to the surface

Deleted: Moreover, and related to light availability is the influence of day length or changes in the photoperiod which can influence diazotroph distribution, in particular the symbiotic diatoms (Karl et al., 2012).

(Cabello et al., 2016). Presence in shallower waters is also thought to be strategic for avoiding competition (Cabello et al. 2016). However, in the WTSP, in 11 of the 14 stations where UCYN-A1 and A2 were detected at sub-surface depth maxima, the same lineages (and corresponding hosts) were undetected at the surface and displayed a negative correlation with PAR. Microscopy observations also confirmed higher numbers of pico-eukaryotes at depth. Hence, it would appear that low light correlates with high abundances of UCYN-A; while the other free-living diazotrophs and symbiotic het-1 and het-2 were positively correlated with PAR, and had maxima closer to the surface with higher PAR. Interestingly, we found no significant correlation between PAR and the UCYN-A host lineages, which is unexpected, since the host partners require light for photosynthesis.

In an attempt to identify the consistency in the correlation patterns identified in the WTSP with other regions of the world's ocean, the same statistical analyses were performed on 11 publically available datasets and subsequently run through a meta-analysis. Our statistical analyses provided coefficients and p-values for easy evaluation and comparisons between data sets for the influence of environmental parameter(s) and diazotroph abundance. It confirmed that UCYN-A indeed stands out from the other diazotrophs in terms of environmental parameter influence, mainly by not being significantly correlated with temperature, which for all other diazotrophs was a significant positive correlation. In terms of temperature, UCYN-A has the lowest temperature optimum among the target cyanobacterial diazotrophs, which will influence their distributions and subsequent correlation with temperature both regionally and vertically. The studies in the meta-analysis include a wide geographical range and independent of ocean basin a majority of the datasets found no significant correlation between UCYN-A and temperature. For most other environmental parameters the pattern for UCYN-A in the WTSP does not hold true in the meta-analysis. For example a significant negative correlation with depth was found for UCYN-A and perhaps

Deleted: especially
Deleted: it is
Deleted: that
Deleted: was conducted in
Deleted: where a majority of the datasets
Deleted: , in all ocean basins included, showed

1 related to high surface temperatures, given that Moisander et al. (2010) observed UCYN-A

2 closer to the surface in the southern SP where surface temperatures were typically cooler.

3 However, for the other target diazotrophs, depth and salinity follow the same pattern as

4 observed in the WTSP (except for UCYN-B which was not significantly correlated with

5 salinity). Interestingly, UCYN-B seems to be mostly detected in the Pacific, where it also was

6 found to be significantly negatively correlated with salinity, while in the Atlantic, perhaps due

7 to lower detections or riverine impact, it was not significantly correlated resulting in no

8 significant correlation in the meta-analysis overall. Furthermore, what did unify all

9 diazotrophs in the meta-analysis were their consistent significant positive correlations with

10 each other and significant negative correlations between abundance and concentrations of DIP

11 and DIN, which was also observed in the WTSP, and again UCYN-A was the exception.

12 In summary, the correlations observed in the WTSP were not always consistent with the

13 meta-analysis of the external datasets. We attribute the inconsistencies in part to seasonal

14 differences in sample collections, and the impact of an individual environmental parameter or

15 sets of parameters on a local and regional scale, particularly for coastal studies, that make it

16 difficult to unambiguously explain the abundance and distribution patterns. Determining the

17 condition or sets of conditions favoring one diazotroph or life history strategy (free-living vs.

18 symbiotic) is complex and likely not all diazotrophs are influenced by the same condition in

19 time and space.

20 **4.6 Estimation of diazotrophs by nifH qPCR**

21 When interpreting abundance estimates by qPCR there are a few assumptions to keep in mind.

22 A limitation of qPCR assays is that gene copies do not equal cell abundances. Recent

23 evidence in filamentous and heterocystous cyanobacteria reports evidence of polyploidy

24 dependent on cell cycle (Griese et al., 2011; Sargent et al., 2016; Sukenik et al., 2012).

25 Moreover, *Trichodesmium* may contain up to 100 genome copies per cell (Sargent et al.,

Deleted: caveat

Deleted: assumes that there is one gene copy per

Deleted: However, r

2016), thus a potential for overestimation. On the other hand, underestimation of gene copies by qPCR is also plausible if one considers that DNA extraction efficiency is not 100% and can vary between species and DNA extraction kits (Mumy and Findlay, 2004), and if high probe specificity favors exclusion of closely related phylotypes for a particular target or lineage.

A final consideration with qPCR as shown here, is the degree of cross-reactivity in assays targeting closely related lineages (e.g. UCYN-A and het). Oligonucleotide specificity as a source of underestimation of the UCYN-A lineages was recently reviewed by a *de nova* analyses (Farnelid et al., 2016) showing the potential to underestimate UCYN-A sublineages since the widely used oligonucleotides for UCYN-A1 contains several mismatches to the other UCYN-A sublineages. The latter becomes important when the sublineages co-occur. Here, however, we highlight the potential to overestimate. For example, UCYN-A2 oligonucleotides amplified the UCYN-A1 templates, indicating a tendency to overquantify UCYN-A2 in the presence of A1. Moreover, when the annealing temperature was set to 64 °C, to distinguish between UCYN-A1 and A2 as recommended by Thompson et al. (2014), the assay still failed to separate the two sub-lineages when run in fast mode. Thus, the fast mode feature has a shortcoming that could influence a wider range of targets than the ones presented here. We observed the same cross-reactivity reported earlier (Foster et al., 2007) for het-1 and het-2 when run in fast mode and highlights the potential to overestimate het-2 if het-1 co-occurs at densities approximately 10^6 *nifH* copies L^{-1}. The latter observation has never been reported.

Conclusions

Consistent with earlier observations in the WTSP, we found diazotrophic cyanobacteria to be abundant, with the most abundant cyanobacterial diazotrophs being UCYN-B, *Trichodesmium* and the symbiotic *Richelia* lineage het-1. Although the cell integrity and

detection of het-1 in water column samples and those from depth (e.g. sediment traps) indicated that the populations were in a senescent state, our work represents one of the first documentation of the three DDA populations in a wide expanse of the WTSP. In contrast to earlier work in the SP and other recent reports from global ocean surveys (Farnelid et al., 2016; Martínez-Pérez et al., 2016), we observed low abundances and poor detection of both UCYN-A (A1 and A2) lineages. According to our qPCR results, UCYN-A was also enumerated when their respective hosts were below detection, which contrasts to the assumed high fidelity and dependency in the partnerships; however, we cannot discount that the disparity in host-symbiont detection was not a result from qPCR oligonucleotide assay bias and/or overestimations indicated by our cross-reactivity tests.

Our initial hypothesis was that the condition or sets of conditions, which promote the distribution of one diazotroph, would differ. Moreover, the parameters for symbiotic diazotrophs should also differ from that of free-living phylotypes, and given the vast difference in hosts (diatoms and prymnesiophyte, respectively) and genome content for the het and UCYN-A symbionts, we further hypothesized divergent conditions favoring one symbiosis over another. In the WTSP, the same conditions favored abundances of both the free-living phylotypes and the diatom (het groups) symbioses. However, the same conditions impacted the abundance of UCYN-A based symbiosis negatively, hence, somewhat supporting our initial hypothesis that conditions for one symbiosis type would differ. In the external datasets, however, we observed differences in environmental conditions favoring abundances of the investigated diazotrophs compared to the WTSP, which underscores that diazotrophs are not similarly influenced by the same condition in time and space.

Multivariate approaches on numerous parameters and with high spatial resolution are required to understand the complex and often indirect effects that govern species distribution. Finally, this study highlights the application and considerations of 'at sea' qPCR for steadfast

quantification of microbial populations. Moreover, a meta-analysis was used to determine the influence of environmental parameters on the distribution and abundance of diazotrophs in the SWP region and in other ocean basins. However, it is of great interest to know, if the same parameters influence gene expressions (e.g. *nifH*), and ultimately N_2 fixation rates, in the same manner, thus, understanding the weight of environmental parameters influencing diazotrophic abundance and distribution. Given the global significance of N_2 fixation as a major new source of N to the oceans, the meta-analysis presented here could be directly applicable to improving parameter constraints on model-based approaches for predicting areas prone to diazotrophy.

1 **Competing interests**

2 The authors declare that they have no conflict of interest.

3 **Acknowledgements**

4 The participation (MS and AC), sample processing, and work presented here is supported by
5 Knut and Alice Wallenberg foundation (to RAF). This is a contribution to the OUTPACE
6 (Oligotrophy from Ultra-oligoTrophy PACific Experiment) project funded by the French
7 research national agency (ANR-14-CE01-0007-01), the LEFE-CyBER program (CNRS-
8 INSU), the GOPS program (IRD) and the CNES (BC T23, ZBC 4500048836). The
9 OUTPACE cruise (http://dx.doi.org/10.17600/15000900) was managed by T. Moutin and S.
10 Bonnet from the MIO (Mediterranean Institute of Oceanography). T. Moutin is specifically
11 thanked for invaluable input and feedback on this paper. The authors thank the crew of the
12 R/V *L'Atalante* for outstanding shipboard operation. G. Rougier and M. Picheral are thanked
13 for their efficient help in CTD rosette management and data processing, as is Catherine
14 Schmechtig for the LEFE CYBER database management. All data and metadata are available
15 at the following web address: http://www.obs-vlfr.fr/proof/php/outpace/outpace.php. Special
16 acknowledgement to Andreas Krupke (formerly of WHOI), Kyle Frischkorn (WHOI), Mar
17 Benavides (MIO/IRD), Hugo Berthelot (MIO), and Mathieu Caffin (MIO). Olivier Grosso
18 (MIO) and Sandra Hélias (MIO) are acknowledged for nutrient analyses. Additionally we
19 thank Dr. Lotta Berntzon for assisting in sample processing and Konrad Karlsson for
20 assistance with the multivariate statistics. Lastly, we thank Dr. Bethany Jenkins and Prof.
21 Anders Bignert for invaluable feedback and suggestions on our manuscript.

Field Code Changed

1 **References**

2 Aminot, A. and Kérouel, R.: Dosage automatique des nutriments dans les eaux marines:
3 méthodes en flux continu, in Méthodes d'analyse en milieu marin, edited by Ifremer, p. 188.
4 [online] Available from: http://cat.inist.fr/?aModele=afficheN&cpsidt=18706171 (Accessed 1
5 December 2016), 2007.

6 Benavides, M., Bonnet, S., Hernández, N., Martínez-Pérez, A. M., Nieto-Cid, M., Álvarez-
7 Salgado, X. A., Baños, I., Montero, M. F., Mazuecos, I. P., Gasol, J. M., Osterholz, H.,
8 Dittmar, T., Berman-Frank, I. and Arístegui, J.: Basin-wide N2 fixation in the deep waters of
9 the Mediterranean Sea, Global Biogeochem. Cycles, 30(6), 2015GB005326,
10 doi:10.1002/2015GB005326, 2016.

11 Bench, S. R., Heller, P., Frank, I., Arciniega, M., Shilova, I. N. and Zehr, J. P.: Whole genome
12 comparison of six Crocosphaera watsonii strains with differing phenotypes, J. Phycol., 49(4),
13 786–801, doi:10.1111/jpy.12090, 2013.

14 Berman-Frank, I., Lundgren, P., Chen, Y.-B., Küpper, H., Kolber, Z., Bergman, B. and
15 Falkowski, P.: Segregation of Nitrogen Fixation and Oxygenic Photosynthesis in the Marine
16 Cyanobacterium Trichodesmium, Science, 294(5546), 1534–1537,
17 doi:10.1126/science.1064082, 2001.

18 Berthelot, H., Moutin, T., L'Helguen, S., Leblanc, K., Hélias, S., Grosso, O., Leblond, N.,
19 Charrière, B. and Bonnet, S.: Dinitrogen fixation and dissolved organic nitrogen fueled
20 primary production and particulate export during the VAHINE mesocosm experiment (New
21 Caledonia lagoon), Biogeosciences, 12(13), 4099–4112, doi:10.5194/bg-12-4099-2015, 2015.

22 Biegala, I. C. and Raimbault, P.: High abundance of diazotrophic picocyanobacteria (<3 µm)
23 in a Southwest Pacific coral lagoon, Aquat Microb Ecol, 51(1), 45–53,
24 doi:10.3354/ame01185, 2008.

25 Blain, S., Sarthou, G. and Laan, P.: Distribution of dissolved iron during the natural iron-
26 fertilization experiment KEOPS (Kerguelen Plateau, Southern Ocean), Deep Sea Research
27 Part II: Topical Studies in Oceanography, 55(5–7), 594–605, doi:10.1016/j.dsr2.2007.12.028,
28 2008.

29 Bombar, D., Moisander, P. H., Dippner, J. W., Foster, R. A., Voss, M., Karfeld, B. and Zehr,
30 J. P.: Distribution of diazotrophic microorganisms and nifH gene expression in the Mekong
31 River plume during intermonsoon, Mar Ecol Prog Ser, 424, 39–52, doi:10.3354/meps08976,
32 2011.

33 Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J. and Zehr, J. P.: Comparative
34 genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-
35 A cyanobacteria, ISME J, 8(12), 2530–2542, doi:10.1038/ismej.2014.167, 2014.

36 Bombar, D., Paerl, R. W. and Riemann, L.: Marine Non-Cyanobacterial Diazotrophs: Moving
37 beyond Molecular Detection, Trends in Microbiology, 24(11), 916–927,
38 doi:10.1016/j.tim.2016.07.002, 2016.

39 Bonnet, S., Biegala, I. C., Dutrieux, P., Slemons, L. O. and Capone, D. G.: Nitrogen fixation
40 in the western equatorial Pacific: Rates, diazotrophic cyanobacterial size class distribution,
41 and biogeochemical significance, Global Biogeochem. Cycles, 23(3), GB3012,
42 doi:10.1029/2008GB003439, 2009.

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K. A., Moutin, T., Hamersley, R. M., Grosso, O., Zehr, J. P. and Capone, D. G.: Aphotic N2 Fixation in the Eastern Tropical South Pacific Ocean, PLOS ONE, 8(12), e81265, doi:10.1371/journal.pone.0081265, 2013.

Bonnet, S., Rodier, M., Turk-Kubo, K. A., Germineaud, C., Menkes, C., Ganachaud, A., Cravatte, S., Raimbault, P., Campbell, E., Quéroué, F., Sarthou, G., Desnues, A., Maes, C. and Eldin, G.: Contrasted geographical distribution of N2 fixation rates and nifH phylotypes in the Coral and Solomon Seas (southwestern Pacific) during austral winter conditions, Global Biogeochem. Cycles, 29(11), 2015GB005117, doi:10.1002/2015GB005117, 2015.

Bonnet, S., Berthelot, H., Turk-Kubo, K., Cornet-Barthaux, V., Fawcett, S., Berman-Frank, I., Barani, A., Grégori, G., Dekaezemacker, J., Benavides, M. and Capone, D. G.: Diazotroph derived nitrogen supports diatom growth in the South West Pacific: A quantitative study using nanoSIMS, Limnol. Oceanogr., 61(5), 1549–1562, doi:10.1002/lno.10300, 2016.

Bonnet, S., Caffin, M., Berthelot, H. and Moutin, T.: Hot spot of N2 fixation in the western tropical South Pacific pleads for a spatial decoupling between N2 fixation and denitrification, PNAS, 114(14), E2800–E2801, doi:10.1073/pnas.1619514114, 2017.

Cabello, A. M., Cornejo-Castillo, F. M., Raho, N., Blasco, D., Vidal, M., Audic, S., de Vargas, C., Latasa, M., Acinas, S. G. and Massana, R.: Global distribution and vertical patterns of a prymnesiophyte–cyanobacteria obligate symbiosis, ISME J, 10(3), 693–706, doi:10.1038/ismej.2015.147, 2016.

Carpenter, E. J. and Janson, S.: Intracellular Cyanobacterial Symbionts in the Marine Diatom Climacodium Frauenfeldianum (bacillariophyceae), Journal of Phycology, 36(3), 540–544, doi:10.1046/j.1529-8817.2000.99163.x, 2000.

Church, M. J., Jenkins, B. D., Karl, D. M. and Zehr, J. P.: Vertical distributions of nitrogen-fixing phylotypes at Stn Aloha in the oligotrophic North Pacific Ocean, Aquat Microb Ecol, 38(1), 3–14, doi:10.3354/ame038003, 2005.

Church, M. J., Björkman, K. M., Karl, D. M., Saito, M. A. and Zehr, J. P.: Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean, Limnol. Oceanogr., 53(1), 63–77, doi:10.4319/lo.2008.53.1.0063, 2008.

Claustre, H. and Maritorena, S.: The Many Shades of Ocean Blue, Science, 302(5650), 1514–1515, doi:10.1126/science.1092704, 2003.

Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sánchez-Baracaldo, P., Lima-Mendez, G., Hingamp, P., Alberti, A., Sunagawa, S., Bork, P., Vargas, C. de, Raes, J., Bowler, C., Wincker, P., Zehr, J. P., Gasol, J. M., Massana, R. and Acinas, S. G.: Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton, Nature Communications, 7, 11071, doi:10.1038/ncomms11071, 2016.

Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N. and Dunne, J. P.: Spatial coupling of nitrogen inputs and losses in the ocean, Nature, 445(7124), 163–167, doi:10.1038/nature05392, 2007.

Dufour, P., Charpy, L., Bonnet, S. and Garcia, N.: Phytoplankton nutrient control in the oligotrophic South Pacific subtropical gyre (Tuamotu Archipelago), Mar Ecol Prog Ser, 179, 285–290, doi:10.3354/meps179285, 1999.

1 Dupouy, C., Benielli-Gary, D., Neveux, J., Dandonneau, Y. and Westberry, T. K.: An
2 algorithm for detecting Trichodesmium surface blooms in the South Western Tropical Pacific,
3 Biogeosciences, 8, 3631–3647, doi:10.5194/bg-8-3631-2011, 2011.

4 Dyhrman, S. T. and Ruttenberg, K. C.: Presence and regulation of alkaline phosphatase
5 activity in eukaryotic phytoplankton from the coastal ocean: Implications for dissolved
6 organic phosphorus remineralization, Limnol. Oceanogr., 51(3), 1381–1390,
7 doi:10.4319/lo.2006.51.3.1381, 2006.

8 Dyhrman, S. T., Chappell, P. D., Haley, S. T., Moffett, J. W., Orchard, E. D., Waterbury, J. B.
9 and Webb, E. A.: Phosphonate utilization by the globally important marine diazotroph
10 Trichodesmium, Nature, 439(7072), 68–71, doi:10.1038/nature04203, 2006.

11 Eugster, O. and Gruber, N.: A probabilistic estimate of global marine N-fixation and
12 denitrification, Global Biogeochem. Cycles, 26(4), GB4013, doi:10.1029/2012GB004300,
13 2012.

14 Farnelid, H., Turk-Kubo, K., Muñoz-Marín, M. del C. and Zehr, J. P.: New insights into the
15 ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A, Aquat
16 Microb Ecol, 77(3), 125–138, doi:10.3354/ame01794, 2016.

17 Fernandez, C., Farías, L. and Ulloa, O.: Nitrogen Fixation in Denitrified Marine Waters,
18 PLOS ONE, 6(6), e20539, doi:10.1371/journal.pone.0020539, 2011.

19 Fitzsimmons, J. N., Boyle, E. A. and Jenkins, W. J.: Distal transport of dissolved
20 hydrothermal iron in the deep South Pacific Ocean, PNAS, 111(47), 16654–16661,
21 doi:10.1073/pnas.1418778111, 2014.

22 Foster, R. A. and Zehr, J. P.: Characterization of diatom–cyanobacteria symbioses on the
23 basis of nifH, hetR and 16S rRNA sequences, Environmental Microbiology, 8(11), 1913–
24 1925, doi:10.1111/j.1462-2920.2006.01068.x, 2006.

25 Foster, R. A., Subramaniam, A., Mahaffey, C., Carpenter, E. J., Capone, D. G. and Zehr, J. P.:
26 Influence of the Amazon River plume on distributions of free-living and symbiotic
27 cyanobacteria in the western tropical north Atlantic Ocean, Limnol. Oceanogr., 52(2), 517–
28 532, doi:10.4319/lo.2007.52.2.0517, 2007.

29 Foster, R. A., Subramaniam, A. and Zehr, J. P.: Distribution and activity of diazotrophs in the
30 Eastern Equatorial Atlantic, Environmental Microbiology, 11(4), 741–750,
31 doi:10.1111/j.1462-2920.2008.01796.x, 2009.

32 Foster, R. A., Goebel, N. L. and Zehr, J. P.: Isolation of Calothrix Rhizosoleniae
33 (cyanobacteria) Strain Sc01 from Chaetoceros (bacillariophyta) Spp. Diatoms of the
34 Subtropical North Pacific Ocean1, Journal of Phycology, 46(5), 1028–1037,
35 doi:10.1111/j.1529-8817.2010.00885.x, 2010.

36 Garcia, N., Raimbault, P. and Sandroni, V.: Seasonal nitrogen fixation and primary
37 production in the Southwest Pacific: nanoplankton diazotrophy and transfer of nitrogen to
38 picoplankton organisms, Mar Ecol Prog Ser, 343, 25–33, doi:10.3354/meps06882, 2007.

39 Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., Montoya,
40 J. P., Edwards, C. A. and Zehr, J. P.: Abundance and distribution of major groups of
41 diazotrophic cyanobacteria and their potential contribution to N2 fixation in the tropical
42 Atlantic Ocean, Environmental Microbiology, 12(12), 3272–3289, doi:10.1111/j.1462-
43 2920.2010.02303.x, 2010.

1 Gourdeau, L., Kessler, W. S., Davis, R. E., Sherman, J., Maes, C. and Kestenare, E.: Zonal
2 Jets Entering the Coral Sea, J. Phys. Oceanogr., 38(3), 715–725, doi:10.1175/2007JPO3780.1,
3 2008.

4 Griese, M., Lange, C. and Soppa, J.: Ploidy in cyanobacteria, FEMS Microbiology Letters,
5 323(2), 124–131, doi:10.1111/j.1574-6968.2011.02368.x, 2011.

6 Halm, H., Musat, N., Lam, P., Langlois, R., Musat, F., Peduzzi, S., Lavik, G., Schubert, C. J.,
7 Singha, B., LaRoche, J. and Kuypers, M. M. M.: Co-occurrence of denitrification and
8 nitrogen fixation in a meromictic lake, Lake Cadagno (Switzerland), Environmental
9 Microbiology, 11(8), 1945–1958, doi:10.1111/j.1462-2920.2009.01917.x, 2009.

10 Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., D'Hondt, S. and
11 Kuypers, M. M.: Heterotrophic organisms dominate nitrogen fixation in the South Pacific
12 Gyre, ISME J, 6(6), 1238–1249, doi:10.1038/ismej.2011.182, 2012.

13 Hansell, D. A. and Feely, R. A.: Atmospheric Intertropical Convergence impacts surface
14 ocean carbon and nitrogen biogeochemistry in the western tropical Pacific, Geophys. Res.
15 Lett., 27(7), 1013–1016, doi:10.1029/1999GL002376, 2000.

16 Haselkorn, R.: Heterocysts, Annual Review of Plant Physiology, 29(1), 319–344,
17 doi:10.1146/annurev.pp.29.060178.001535, 1978.

18 Hood, R. R., Coles, V. J. and Capone, D. G.: Modeling the distribution of Trichodesmium and
19 nitrogen fixation in the Atlantic Ocean, J. Geophys. Res., 109(C6), C06006,
20 doi:10.1029/2002JC001753, 2004.

21 Howard, J. B. and Rees, D. C.: Structural Basis of Biological Nitrogen Fixation, Chem. Rev.,
22 96(7), 2965–2982, doi:10.1021/cr9500545, 1996.

23 Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L.,
24 Beynon, J., Newton, W. E. and Dean, D. R.: Physical and genetic map of the major nif gene
25 cluster from Azotobacter vinelandii., J. Bacteriol., 171(2), 1017–1027, 1989.

26 Karl, D., Letelier, R., Tupas, L., Dore, J., Christian, J. and Hebel, D.: The role of nitrogen
27 fixation in biogeochemical cycling in the subtropical North Pacific Ocean, Nature, 388(6642),
28 533–538, doi:10.1038/41474, 1997.

29 Karl, D. M., Björkman, K. M., Dore, J. E., Fujieki, L., Hebel, D. V., Houlihan, T., Letelier, R.
30 M. and Tupas, L. M.: Ecological nitrogen-to-phosphorus stoichiometry at station ALOHA,
31 Deep Sea Research Part II: Topical Studies in Oceanography, 48(8–9), 1529–1566,
32 doi:10.1016/S0967-0645(00)00152-1, 2001.

33 Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M. and Mahaffey, C.: Predictable and
34 efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen
35 fixation, PNAS, 109(6), 1842–1849, doi:10.1073/pnas.1120312109, 2012.

36 Kong, L., Jing, H., Kataoka, T., Sun, J. and Liu, H.: Phylogenetic diversity and spatio-
37 temporal distribution of nitrogenase genes (nifH) in the northern South China Sea, Aquat
38 Microb Ecol, 65(1), 15–27, doi:10.3354/ame01531, 2011.

39 Krupke, A., Musat, N., LaRoche, J., Mohr, W., Fuchs, B. M., Amann, R. I., Kuypers, M. M.
40 M. and Foster, R. A.: In situ identification and N2 and C fixation rates of uncultivated
41 cyanobacteria populations, Systematic and Applied Microbiology, 36(4), 259–271,
42 doi:10.1016/j.syapm.2013.02.002, 2013.

Krupke, A., Lavik, G., Halm, H., Fuchs, B. M., Amann, R. I. and Kuypers, M. M. M.: Distribution of a consortium between unicellular algae and the N2 fixing cyanobacterium UCYN-A in the North Atlantic Ocean, Environ Microbiol, 16(10), 3153–3167, doi:10.1111/1462-2920.12431, 2014.

Krupke, A., Mohr, W., LaRoche, J., Fuchs, B. M., Amann, R. I. and Kuypers, M. M.: The effect of nutrients on carbon and nitrogen fixation by the UCYN-A–haptophyte symbiosis, ISME J, 9(7), 1635–1647, doi:10.1038/ismej.2014.253, 2015.

Kustka, A. B., Sañudo-Wilhelmy, S. A., Carpenter, E. J., Capone, D., Burns, J. and Sunda, W. G.: Iron requirements for dinitrogen- and ammonium-supported growth in cultures of Trichodesmium (IMS 101): Comparison with nitrogen fixation rates and iron: carbon ratios of field populations, Limnol. Oceanogr., 48(5), 1869–1884, doi:10.4319/lo.2003.48.5.1869, 2003.

Landolfi, A., Koeve, W., Dietze, H., Kähler, P. and Oschlies, A.: A new perspective on environmental controls of marine nitrogen fixation, Geophys. Res. Lett., 42(11), 2015GL063756, doi:10.1002/2015GL063756, 2015.

Langlois, R., Mills, M. M., Ridame, C., Croot, P. and LaRoche, J.: Diazotrophic bacteria respond to Saharan dust additions, Marine Ecology Progress Series, 470(doi:10.3354/meps10109), 1–14, doi:Langlois, Rebecca, Mills, M. M., Ridame, Celine, Croot, Peter and LaRoche, Julie (2012) Diazotrophic bacteria respond to Saharan dust additions Marine Ecology Progress Series, 470 . pp. 1-14. DOI 10.3354/meps10109 <http://dx.doi.org/10.3354/meps10109>., 2012.

Langlois, R. J., LaRoche, J. and Raab, P. A.: Diazotrophic Diversity and Distribution in the Tropical and Subtropical Atlantic Ocean, Appl. Environ. Microbiol., 71(12), 7910–7919, doi:10.1128/AEM.71.12.7910-7919.2005, 2005.

Langlois, R. J., Hümmer, D. and LaRoche, J.: Abundances and Distributions of the Dominant nifH Phylotypes in the Northern Atlantic Ocean, Appl. Environ. Microbiol., 74(6), 1922–1931, doi:10.1128/AEM.01720-07, 2008.

Law, C. S., Woodward, E. M. S., Ellwood, M. J., Marriner, A., Bury, S. J. and Safi, K. A.: Response of surface nutrient inventories and nitrogen fixation to a tropical cyclone in the southwest Pacific, Limnol. Oceanogr., 56(4), 1372–1385, doi:10.4319/lo.2011.56.4.1372, 2011.

Löscher, C. R., LaRoche, J. and Schmitz, R. A.: Underestimation of Marine Dinitrogen Fixation: A Novel Method and Novel Diazotrophic Habitats, in Biological Nitrogen Fixation, edited by F. J. de Bruijn, pp. 891–896, John Wiley & Sons, Inc. [online] Available from: http://onlinelibrary.wiley.com/doi/10.1002/9781119053095.ch87/summary (Accessed 29 November 2016), 2015.

Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K. H., Böttjer, D., Capone, D. G., Carpenter, E. J., Chen, Y. L., Church, M. J., Dore, J. E., Falcón, L. I., Fernández, A., Foster, R. A., Furuya, K., Gómez, F., Gundersen, K., Hynes, A. M., Karl, D. M., Kitajima, S., Langlois, R. J., LaRoche, J., Letelier, R. M., Marañón, E., McGillicuddy, D. J., Moisander, P. H., Moore, C. M., Mouriño-Carballido, B., Mulholland, M. R., Needoba, J. A., Orcutt, K. M., Poulton, A. J., Rahav, E., Raimbault, P., Rees, A. P., Riemann, L., Shiozaki, T., Subramaniam, A., Tyrrell, T., Turk-Kubo, K. A., Varela, M., Villareal, T. A., Webb, E. A., White, A. E., Wu, J. and Zehr, J. P.: Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates,

1 Earth System Science Data, 4(1), 47–73, doi:Luo, Y.-W.; Doney, S.C.; Anderson, L.A.;
2 Benavides, M.; Berman-Frank, I.; Bode, A.; Bonnet, S.; Boström, K.H.; Böttjer, D.; Capone,
3 D.G.; Carpenter, E.J.; Chen, Y.L.; Church, M.J.; Dore, J.E.; Falcón, L.I.; Fernández, A.;
4 Foster, R.A.; Furuya, K.; Gómez, F.; Gundersen, K.; Hynes, A.M.; Karl, D.M.; Kitajima, S.;
5 Langlois, R.J.; LaRoche, J.; Letelier, R.M.; Marañón, E.; McGillicuddy, D.J.; Moisander,
6 P.H.; Moore, C.M.; Mouriño-Carballido, B.; Mulholland, M.R.; Needoba, J.A.; Orcutt, K.M.;
7 Poulton, A.J.; Rahav, E.; Raimbault, P.; Rees, A.P.; Riemann, L.; Shiozaki, T.; Subramaniam,
8 A.; Tyrrell, T.; Turk-Kubo, K.A.; Varela, M.; Villareal, T.A.; Webb, E.A.; White, A.E.; Wu,
9 J.; Zehr, J.P.. 2012 Database of diazotrophs in global ocean: abundance, biomass and
10 nitrogen fixation rates. Earth System Science Data, 4 (1). 47-73. 10.5194/essd-4-47-2012
11 <http://dx.doi.org/10.5194/essd-4-47-2012>, 2012.

12 Luo, Y.-W., Lima, I. D., Karl, D. M., Deutsch, C. A. and Doney, S. C.: Data-based
13 assessment of environmental controls on global marine nitrogen fixation, , doi:10.5194/bg-11-
14 691-2014, 2014.

15 Marchesiello, P. and Estrade, P.: Upwelling limitation by onshore geostrophic flow, Journal
16 of Marine Research, 68(1), 37–62, doi:10.1357/002224010793079004, 2010.

17 Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P.,
18 Lehnen, N., Fuchs, B. M., Lavik, G., Schmitz, R. A., LaRoche, J. and Kuypers, M. M. M.:
19 The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen
20 cycle, Nature Microbiology, 1, 16163, doi:10.1038/nmicrobiol.2016.163, 2016.

21 Messer, L. F., Doubell, M., Jeffries, T. C., Brown, M. V. and Seymour, J. R.: Prokaryotic and
22 diazotrophic population dynamics within a large oligotrophic inverse estuary, Aquat Microb
23 Ecol, 74(1), 1–15, doi:10.3354/ame01726, 2015.

24 Messer, L. F., Mahaffey, C., Robinson, C. M., Jeffries, T. C., Baker, K. G., Bibiloni Isaksson,
25 J., Ostrowski, M., Doblin, M. A., Brown, M. V. and Seymour, J. R.: High levels of
26 heterogeneity in diazotroph diversity and activity within a putative hotspot for marine
27 nitrogen fixation, ISME J, 10(6), 1499–1513, doi:10.1038/ismej.2015.205, 2016.

28 Meyerhof, O. and Burk, D.: Über die Fixation des Luftstickstoffs durch Azotobacter,
29 Zeitschrift physikalische Chemie (Abt. A), (139), 117–142, 1928.

30 Mills, M. M., Ridame, C., Davey, M., La Roche, J. and Geider, R. J.: Iron and phosphorus co-
31 limit nitrogen fixation in the eastern tropical North Atlantic, Nature, 429(6989), 292–294,
32 doi:10.1038/nature02550, 2004.

33 Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S. and Arai, T.: Strategy by
34 which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically, Nature, 323, 720–
35 722, doi:10.1038/323720a0, 1986.

36 Moisander, P. H., Beinart, R. A., Voss, M. and Zehr, J. P.: Diversity and abundance of
37 diazotrophic microorganisms in the South China Sea during intermonsoon, ISME J, 2(9),
38 954–967, doi:10.1038/ismej.2008.51, 2008.

39 Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A.,
40 Montoya, J. P. and Zehr, J. P.: Unicellular Cyanobacterial Distributions Broaden the Oceanic
41 N2 Fixation Domain, Science, 327(5972), 1512–1514, doi:10.1126/science.1185468, 2010.

42 Moisander, P. H., Zhang, R., Boyle, E. A., Hewson, I., Montoya, J. P. and Zehr, J. P.:
43 Analogous nutrient limitations in unicellular diazotrophs and Prochlorococcus in the South
44 Pacific Ocean, ISME J, 6(4), 733–744, doi:10.1038/ismej.2011.152, 2012.

1 Moutin, T., Karl, D. M., Duhamel, S., Rimmelin, P., Raimbault, P., Mooy, V., S, B. A. and
2 Claustre, H.: Phosphate availability and the ultimate control of new nitrogen input by nitrogen
3 fixation in the tropical Pacific Ocean, , doi:10.5194/bg-5-95-2008, 2008.

4 Mumy, K. L. and Findlay, R. H.: Convenient determination of DNA extraction efficiency
5 using an external DNA recovery standard and quantitative-competitive PCR, Journal of
6 Microbiological Methods, 57(2), 259–268, doi:10.1016/j.mimet.2004.01.013, 2004.

7 Needoba, J. A., Foster, R. A., Sakamoto, C., Zehr, J. P. and Johnson, K. S.: Nitrogen fixation
8 by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean,
9 Limnol. Oceanogr., 52(4), 1317–1327, doi:10.4319/lo.2007.52.4.1317, 2007.

10 Ottesen, E. A., Young, C. R., Eppley, J. M., Ryan, J. P., Chavez, F. P., Scholin, C. A. and
11 DeLong, E. F.: Pattern and synchrony of gene expression among sympatric marine microbial
12 populations, PNAS, 110(6), E488–E497, doi:10.1073/pnas.1222099110, 2013.

13 Preston, C. M., Harris, A., Ryan, J. P., Roman, B., Iii, R. M., Jensen, S., Everlove, C., Birch,
14 J., Dzenitis, J. M., Pargett, D., Adachi, M., Turk, K., Zehr, J. P. and Scholin, C. A.:
15 Underwater Application of Quantitative PCR on an Ocean Mooring, PLOS ONE, 6(8),
16 e22522, doi:10.1371/journal.pone.0022522, 2011.

17 Qiu, B., Chen, S. and Kessler, W. S.: Source of the 70-Day Mesoscale Eddy Variability in the
18 Coral Sea and the North Fiji Basin, J. Phys. Oceanogr., 39(2), 404–420,
19 doi:10.1175/2008JPO3988.1, 2009.

20 Raimbault, P., Garcia, N. and Cerutti, F.: Distribution of inorganic and organic nutrients in the
21 South Pacific Ocean - evidence for long-term accumulation of organic matter in nitrogen-
22 depleted waters, Biogeosciences Discussions, 4, 3041–3087, 2007.

23 Raven, J. A.: The iron and molybdenum use efficiencies of plant growth with different
24 energy, carbon and nitrogen sources, New Phytologist, 109(3), 279–287, doi:10.1111/j.1469-
25 8137.1988.tb04196.x, 1988.

26 Robidart, J. C., Church, M. J., Ryan, J. P., Ascani, F., Wilson, S. T., Bombar, D., Marin, R.,
27 Richards, K. J., Karl, D. M., Scholin, C. A. and Zehr, J. P.: Ecogenomic sensor reveals
28 controls on N2-fixing microorganisms in the North Pacific Ocean, ISME J, 8(6), 1175–1185,
29 doi:10.1038/ismej.2013.244, 2014.

30 Sañudo-Wilhelmy, S. A., Kustka, A. B., Gobler, C. J., Hutchins, D. A., Yang, M., Lwiza, K.,
31 Burns, J., Capone, D. G., Raven, J. A. and Carpenter, E. J.: Phosphorus limitation of nitrogen
32 fixation by Trichodesmium in the central Atlantic Ocean, Nature, 411(6833), 66–69,
33 doi:10.1038/35075041, 2001.

34 Sargent, E. C., Hitchcock, A., Johansson, S. A., Langlois, R., Moore, C. M., LaRoche, J.,
35 Poulton, A. J. and Bibby, T. S.: Evidence for polyploidy in the globally important diazotroph
36 Trichodesmium, FEMS Microbiology Letters, 363(21), fnw244, doi:10.1093/femsle/fnw244,
37 2016.

38 Shiozaki, T., Kodama, T. and Furuya, K.: Large-scale impact of the island mass effect through
39 nitrogen fixation in the western South Pacific Ocean, Geophys. Res. Lett., 41(8),
40 2014GL059835, doi:10.1002/2014GL059835, 2014.

41 Short, S. M., Jenkins, B. D. and Zehr, J. P.: Spatial and Temporal Distribution of Two
42 Diazotrophic Bacteria in the Chesapeake Bay, Appl. Environ. Microbiol., 70(4), 2186–2192,
43 doi:10.1128/AEM.70.4.2186-2192.2004, 2004.

1 Stewart, W. D. P.: Biological and Ecological Aspects of Nitrogen Fixation by Free-Living
2 Micro-Organisms, Proceedings of the Royal Society of London. Series B, Biological
3 Sciences, 172(1029), 367–388, 1969.

4 Subramaniam, A., Yager, P. L., Carpenter, E. J., Mahaffey, C., Björkman, K., Cooley, S.,
5 Kustka, A. B., Montoya, J. P., Sañudo-Wilhelmy, S. A., Shipe, R. and Capone, D. G.:
6 Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic
7 Ocean, PNAS, 105(30), 10460–10465, doi:10.1073/pnas.0710279105, 2008.

8 Sukenik, A., Kaplan-Levy, R. N., Welch, J. M. and Post, A. F.: Massive multiplication of
9 genome and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum
10 (Cyanobacteria), ISME J, 6(3), 670–679, doi:10.1038/ismej.2011.128, 2012.

11 Taniuchi, Y., Chen, Y. L., Chen, H.-Y., Tsai, M.-L. and Ohki, K.: Isolation and
12 characterization of the unicellular diazotrophic cyanobacterium Group C TW3 from the
13 tropical western Pacific Ocean, Environmental Microbiology, 14(3), 641–654,
14 doi:10.1111/j.1462-2920.2011.02606.x, 2012.

15 Thompson, A., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F. and Zehr, J. P.: Genetic
16 diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte
17 host, Environ Microbiol, 16(10), 3238–3249, doi:10.1111/1462-2920.12490, 2014.

18 Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., Kuypers, M.
19 M. M. and Zehr, J. P.: Unicellular Cyanobacterium Symbiotic with a Single-Celled
20 Eukaryotic Alga, Science, 337(6101), 1546–1550, doi:10.1126/science.1222700, 2012.

21 Tripp, H. J., Bench, S. R., Turk, K. A., Foster, R. A., Desany, B. A., Niazi, F., Affourtit, J. P.
22 and Zehr, J. P.: Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium,
23 Nature, 464(7285), 90–94, doi:10.1038/nature08786, 2010.

24 Turk-Kubo, K. A., Frank, I. E., Hogan, M. E., Desnues, A., Bonnet, S. and Zehr, J. P.:
25 Diazotroph community succession during the VAHINE mesocosm experiment (New
26 Caledonia lagoon), Biogeosciences, 12(24), 7435–7452, doi:10.5194/bg-12-7435-2015, 2015.

27 Van Den Broeck, N., Moutin, T., Rodier, M. and Bouteiller, A. L.: Seasonal variations of
28 phosphate availability in the SW Pacific Ocean near New Caledonia, Mar Ecol Prog Ser, 268,
29 1–12, doi:10.3354/meps268001, 2004.

30 Webb, E. A., Ehrenreich, I. M., Brown, S. L., Valois, F. W. and Waterbury, J. B.: Phenotypic
31 and genotypic characterization of multiple strains of the diazotrophic cyanobacterium,
32 Crocosphaera watsonii, isolated from the open ocean, Environmental Microbiology, 11(2),
33 338–348, doi:10.1111/j.1462-2920.2008.01771.x, 2009.

34 Wells, M. L., Vallis, G. K. and Silver, E. A.: Tectonic processes in Papua New Guinea and
35 past productivity in the eastern equatorial Pacific Ocean, Nature, 398(6728), 601–604,
36 doi:10.1038/19281, 1999.

37 Yoshikawa, C., Nakatsuka, T. and Kawahata, H.: Transition of low-salinity water in the
38 Western Pacific Warm Pool recorded in the nitrogen isotopic ratios of settling particles,
39 Geophys. Res. Lett., 32(14), L14615, doi:10.1029/2005GL023103, 2005.

40 Young, J. P. W.: The Phylogeny and Evolution of Nitrogenases, in Genomes and Genomics of
41 Nitrogen-fixing Organisms, edited by R. Palacios and W. E. Newton, pp. 221–241, Springer
42 Netherlands. [online] Available from: http://link.springer.com/chapter/10.1007/1-4020-3054-
43 1_14 (Accessed 29 November 2016), 2005.

1 Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F.,
2 Hansen, A. and Karl, D. M.: Unicellular cyanobacteria fix N2 in the subtropical North Pacific
3 Ocean, Nature, 412(6847), 635–638, doi:10.1038/35088063, 2001.

4 Zehr, J. P., Methe, B. and Foster, R.: New Nitrogen-Fixing Microorganisms from the Oceans:
5 Biological Aspects and Global Implications, in Biological Nitrogen Fixation, Sustainable
6 Agriculture and the Environment, pp. 361–365, Springer Netherlands. [online] Available
7 from: http://link.springer.com/chapter/10.1007/1-4020-3570-5_87 (Accessed 29 November
8 2016), 2005.

9 Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., Tripp, H. J. and Affourtit,
10 J. P.: Globally Distributed Uncultivated Oceanic N2-Fixing Cyanobacteria Lack Oxygenic
11 Photosystem II, Science, 322(5904), 1110–1112, doi:10.1126/science.1165340, 2008.
12

1 Table 1. Summary of environmental conditions in the surface along the cruise transect.

Region	Stations	surface* DIN‡ (µM)	surface* DIP (µM)	surface* DiSi (µM)	surface* salinity (PSU)	surface* temp. (˚C)
Melanesian archipelago (MA) 160˚ E-178˚ E 170 ˚W - 175 ˚W	SD1-12 LDA LDB	0.02 ± 0.01	0.03 ± 0.02	0.55 ± 0.10	35.13 ± 0.27	29.33 ± 0.45
Subtropical gyre (SG) 160 ˚W- 169˚W	SD13-15 LDC	0.01 ± 0.01	0.18 ± 0.07	0.79 ± 0.04	35.12 ± 0.10	29.34 ± 0.18

2 *5m depth, ‡NO_2+NO

Figure 1. Sampling locations and the horizontal and vertical distributions of diazotrophs and the UCYN-A1 and UCYN-A2 hosts in the study area. Sampling depths are indicated as black dots (white for LD stations) and the abundances are the log *nifH* gene copies L^{-1} for the diazotrophs and 18S rRNA gene copies L^{-1} for the UCYN-A host lineages. The boundary between the Melanesian archipelago and subtropical gyre in the transect map is marked with a dotted line.

Figure 2 a-b. LOG10 transformed mean abundances for 4 of the diazotrophs across the transect: het-1 (blue), *Trichodesmium* (green), UCYN-A1 (yellow) and UCYN-B (purple). The mean *nifH* abundance values are shown as a function of (a) percent (%) of surface irradiance and corresponding depth (m) on the secondary y-axis and (b) at each station. The corresponding depths for the percent of surface irradiance varied little, hence values were pooled and the mean is plotted. The solid black line in (b) designates the transition between the Melanesian archipelago (MA) and the subtropical gyre (SG).

Deleted: across the transect

Deleted: indicates

Figure 3. Hierarchical clustering heat map of Spearman's Rho results. The histogram shows negative (blue) and positive (green) values of correlation strength between parameters. Stars within cells mark significant correlations ($p<0.05$). In brackets are the two distinct groups in the WTSP.

Figure 4 a-b. Multivariate RDA bi-plot (a), depicts the variance explained by the environmental parameters and qPCR data in several dimensions, (b). A majority of the variance (y-axis: 0-4) in the dataset is explained by the RDA1 and RDA2 axes meaning that most of the variance observed is explained by the environmental parameters. The arrows are the constrained explanatory vectors with the dots representing the superimposed unconstrained response variables. PAR and nutrients (DIP and DIN) were omitted due to limited data points.

Deleted: which also

Deleted: included

Deleted: , where

Deleted: RDA1 and RDA2 are the most important

Deleted: As can be seen, a

Deleted: included

1 **Figure 5 a-d.** Meta-analysis bar graphs based on Suppl. Table 6 for a) *Trichodesmium*, b)
2 UCYN-B, c) UCYN-A, and d) het-1. Only the significant ($p<0.05$) parameters are included
3 and arranged as the strongest effect to the left and weakest to the right (either positive or
4 negative). Each parameter is color coded, and the cyanobacterial diazotrophs are shades of
5 orange and brown. Red=temperature, blue=salinity, black=depth, green=chlorophyll *a*,
6 yellow=DIN and purple=DIP.