Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia Lagoon)

Kendra A. Turk-Kubo[1]*, Ildiko E. Frank[1], Mary E. Hogan[1], Anne Desnues[2], Sophie Bonnet[2], Jonathan P. Zehr[1]

[1] Ocean Sciences Department, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA, 95064 USA

[2] Mediterranean Institute of Oceanography (MIO) – IRD/CNRS/Aix-Marseille University, IRD Noumea, 101 Promenade R. Laroque, BPA5, 98848 Noumea Cedex

Correspondence to: K.A. Turk-Kubo (kturk@ucsc.edu)

Abstract

The VAHINE mesocosm experiment, conducted in the low-nutrient low-chlorophyll waters of the Noumea Lagoon (coastal New Caledonia) was designed to trace the incorporation of nitrogen (N) fixed by diazotrophs into the food web, using large volume (50 m^3) mesocosms. This experiment provided a unique opportunity to study the succession of different N_2-fixing microorganisms (diazotrophs) and calculate *in situ* net growth and mortality rates in response to fertilization with dissolved inorganic phosphate (DIP) over a 23-day period, using quantitative polymerase chain reaction (qPCR) assays targeting widely distributed marine diazotroph lineages. Inside the mesocosms, the most abundant diazotroph was the heterocyst-forming *Richelia* associated with *Rhizosolenia* (Het-1) in the first half of the experiment, while unicellular cyanobacterial Group C (UCYN-C) became abundant during the second half of the experiment. Decreasing DIP concentrations following the fertilization event and increasing temperatures were significantly correlated with increasing abundances of UCYN-C. Maximum net growth rates for UCYN-C were calculated to range between 1.23 ± 0.07 and 2.16 ± 0.07 d^{-1} in the mesocosms, which are among the highest growth rates reported for diazotrophs. Outside the mesocosms in the New Caledonia lagoon, UCYN-C abundances remained low, despite increasing temperatures, suggesting that the microbial community response to the DIP fertilization created conditions favorable for UCYN-C growth inside the

mesocosms. Diazotroph community composition analysis using PCR targeting a component of the nitrogenase gene (*nifH*) verified that diazotrophs targeted in qPCR assays were collectively among the major lineages in the lagoon and mesocosm samples, with the exception of *Crocosphaera*-like phylotypes, where sequence types not typically seen in the oligotrophic ocean grew in the mesocosms. Maximum net growth and mortality rates for nine diazotroph phylotypes throughout the 23-day experiment were variable between mesocosms, and repeated fluctuations between periods of net growth and mortality were commonly observed. The field population of diazotrophs in the New Caledonian lagoon waters appeared to be dominated by Het-1 over the course of the study period. However, results from both qPCR and PCR analysis indicated a diverse field population of diazotrophs was present in the lagoon at the time of sampling. Two ecotypes of the *Braarudosphaera bigelowii* symbiont unicellular group A (UCYN-A) were present simultaneously in the lagoon, with the recently described *B. bigelowii*/UCYN-A2 association present at higher abundances than the *B. bigelowii*/UCYN-A1 association.

1 Introduction

Biological nitrogen fixation (BNF), the microbially-mediated conversion of dinitrogen (N_2) gas into bioavailable nitrogen (N), is a significant source of new N in oligotrophic oceanic regions where primary productivity is N limited (Gruber and Sarmiento, 1997;Karl et al., 1997), and has the potential to directly impact carbon sequestration (Karl et al., 2012;Karl and Letelier, 2008). BNF has historically been considered an important process in the oligotrophic ocean gyres, but primary productivity in oligotrophic tropical coastal regions can also be N limited (Torréton et al., 2010) and such environments have the potential to play a significant role in export production in the worlds oceans due to the transfer of carbon from tidal and wind-generated currents (Hyndes et al., 2014;Gattuso et al., 1998).

The New Caledonian (Noumea) coral lagoon, located off the southwestern coast of New Caledonia (South Western Pacific), is a tropical low-nutrient low-chlorophyll (LNLC) system and is bounded by one of the world's largest barrier reefs. Oligotrophic ocean water enters the lagoon from the south over the open shelf, then is driven north by

63 the trade winds and tidal forces and exits through several deep inlets in the intertidal
64 barrier reef that forms the western boundary of the lagoon (Ouillon et al., 2010). Primary
65 productivity is N limited throughout the year (Torréton et al., 2010), giving
66 microorganisms able to fix N_2 gas into bioavailable nitrogen (diazotrophs) a competitive
67 edge over non-diazotrophic organisms. High rates of BNF during the austral summer
68 have been reported, in both large size fractions (>10 μm) and small size fractions (<10
69 μm; Garcia et al., 2007;Biegala and Raimbault, 2008). Garcia et al., (2007) also reported
70 that the percent of N_2 fixation measured in the large size fraction had high temporal
71 variability. Large blooms of the most conspicuous and well-studied diazotroph,
72 *Trichodesmium*, have been repeatedly detected in this region using both indirect (via
73 satellite observation; Dupouy et al., 2011;Dupouy et al., 2000) and direct measurements
74 (Renaud et al., 2005;Masotti et al., 2007;Rodier and Le Borgne, 2008, 2010). Both free-
75 living and organism- and particle-associated unicellular picocyanobacterial diazotrophs
76 are also suspected to be significant contributors to BNF in the lagoon (Garcia et al.,
77 2007;Biegala and Raimbault, 2008), yet the phylogenetic identity of these
78 picocyanobacteria has yet to be determined. Despite evidence that diverse diazotroph
79 communities exist in the Noumea lagoon, there is very little quantitative diazotroph
80 distribution data from this region (Luo et al., 2012), especially for diazotrophs other than
81 *Trichodesmium*.
82 Marine unicellular cyanobacterial diazotrophs are phylogenetically divided into
83 three groups: the uncultivated unicellular group A (UCYN-A; Zehr et al., 2001;Zehr et
84 al., 2008;Tripp et al., 2010) which live in association with strains of the prymnesiophyte
85 *Braarudosphaera bigelowii* (Thompson et al., 2014;Thompson et al., 2012;Hagino et al.,
86 2013); the free-living *Crocosphaera* sp. (also referred to as unicellular group B or
87 UCYN-B); and the presumably free-living unicellular group C (UCYN-C), which
88 contains several cultivated cyanobacteria including *Cyanothece* sp. strain ATCC 51142
89 (Reddy et al., 1993) and group C TW3 (Taniuchi et al., 2012).
90 The marine filamentous cyanobacterial diazotrophs include the colonial, non-
91 heterocyst-forming *Trichodesmium*, and the heterocyst-forming symbionts associated
92 with diatoms (DDAs). DDAs form between different strains of *Richelia* sp. associated
93 with diatoms of the genera *Rhizosolenia* (Het-1) and *Hemiaulus* (Het-2) (Villareal, 1990,

1992). Het-2 is in an obligate symbiont of *Hemiaulus*, as evidenced by its reduced genome size (Hilton et al., 2013). Although the genome of Het-1 shows evidence of some genome reduction, it remains unclear whether Het-1 is also an obligate symbiont of *Rhizosolenia* (Villareal, 1992;Hilton et al., 2013). The heterocyst-forming *Calothrix* (Het-3) has long been observed living as an epiphyte with *Chaetoceros* (Carpenter and Foster, 2002) but can also grow free from its host (Foster et al., 2010), and has a non-streamlined genome (Hilton et al., 2013).

Although *nifH* genes are regularly recovered from diverse non-cyanobacterial diazotrophs in oligotrophic ocean waters (Halm et al., 2012;Farnelid et al., 2011;Riemann et al., 2010;Langlois et al., 2005;Hewson et al., 2007;Bird and Wyman, 2012;Bonnet et al., 2013;Fong et al., 2008;Turk-Kubo et al., 2014), their activity and relative significance to BNF remains poorly understood (Turk-Kubo et al., 2014). The most widely studied non-cyanobacterial diazotroph, γ-24774A11, is an uncultivated putative gamma proteobacteria most closely related to *Pseudomonas stutzeri* that has been hypothesized to be a potentially important contributor to overall BNF in the South Pacific (Moisander et al., 2014).

Different diazotrophs have potentially different fates in the marine environment. For example, *Trichodesmium* is rarely recovered in sediment traps (Walsby, 1992), but *Trichodesmium*-derived N is efficiently transferred to non-diazotrophic plankton (mainly diatoms and bacteria) at short time scales (48 h) in the surface ocean (Bonnet et al. *in revision*). In contrast, blooms of DDAs fuel an important summer export flux at the ALOHA station (Karl et al., 2012). This highlights the importance of characterizing the diazotroph community composition when performing biogeochemical studies on the fate of N_2 fixation in the ocean. Such studies are commonly performed on oceanographic cruises, where discrete samples are taken from multiple stations along a cruise track, which passes over many different water masses. This approach is of critical importance to describe the biogeographical distribution of diazotrophs with respect to environmental parameters over large oceanic provinces. However, in order to understand how diazotroph assemblages shift in response to rapid environmental perturbations inside the same water mass, and to track the incorporation of their newly fixed N into the food web,

high frequency sampling of a single water body is required, which is rarely accomplished.

We report here data from the VAHINE mesocosm experiment, detailed by Bonnet et al. (2015a), which was a large, multi-institute collaborative project conducted to determine which components of the food web were directly supported by newly fixed N. To answer this question, three large (50 m^3) mesocosms were deployed in the New Caledonian (Noumea) lagoon to isolate a part of the water column from physical dispersion without disturbing light penetration and temperature. The same water masses were monitored for 23 days during austral summer conditions. In order to create conditions favorable for diazotrophs, the mesocosms were fertilized with dissolved inorganic phosphorus (DIP) on day 4 of the incubation. This experiment provided unique opportunities to: 1) track rapid diazotroph assemblage shifts using quantitative techniques (quantitative PCR; qPCR) targeting known major marine diazotroph lineages for a long period of time (23 days) in a single water mass; 2) calculate *in situ* net growth and mortality rates for targeted diazotroph phylotypes in a complex community; 3) determine the abundances of targeted diazotrophs in a coastal LNLC environment, the Noumea lagoon, during the experimental period; and 4) characterize shifts in the diazotroph community composition in both the mesocosms and in the Noumea Lagoon during the experimental period using next generation sequencing of *nifH* gene amplicons.

2 Methods

2.1 Sampling

Three large volume mesocosms (50 m^3), based on the design described in Guieu et al. (2010, 2014), were deployed at the exit of the Noumea lagoon (22°29.073 S - 166°26.905 E), 28 km off the coast of New Caledonia on January 13, 2013. Detailed descriptions of the mesocosm site selection, deployment and sampling strategy are provided in Bonnet et al. (2015a). Large volume samples (50 L) were collected from 1, 6, and 12 m depths from each mesocosm and outside the mesocosms (hereafter called Noumea lagoon) once per day at 7 a.m. using a Teflon® PFA pump and PVC tubing. This daily sample was immediately transferred back to laboratories aboard the R/V Alis, and subsampled for a suite of stock and flux measurements. Samples for DNA analysis

were immediately filtered onto 25 mm 0.2 μm Supor® filters (Millipore, Billarica, CA), using gentle peristaltic pumping. All filters were flash frozen in liquid nitrogen, and stored at -80°C until shipment on dry ice from New Caledonia to the University of California, Santa Cruz.

2.2 Determining targeted diazotroph abundances and net growth rates using the quantitative polymerase chain reaction (qPCR)

DNA was extracted using a Qiagen DNeasy Plant kit (Valencia, CA), with modifications to the protocol optimized to recover high quality DNA from cyanobacteria, including additional cell lysis steps of freeze/thaw cycles, agitation using a bead beater, as well as a proteinase K digestion (Moisander et al., 2008). The quality of DNA extracts was evaluated using a NanoDrop (Thermo Scientific, Waltham, MA), and concentrations were determined using a PicoGreen® dsDNA Quantitation Kit (Molecular Probes, Eugene, OR, USA), according to manufacturer's guidelines.

Nine diazotrophic phylotypes were quantified in samples from the mesocosms and the Noumea lagoon, using quantitative PCR (qPCR), using Taqman® assays widely used in the marine environment. This study targeted representatives from known major marine diazotroph lineages: two unicellular diazotrophic symbionts of different *Braarudosphaera bigelowii* strains, UCYN-A1 (Church et al., 2005a), UCYN-A2 (Thompson et al., 2014); two free-living unicellular diazotroph cyanobacterial phylotypes UCYN-B (*Crocosphaera sp.*; Moisander et al., 2010), and UCYN-C (Foster et al., 2007); filamentous, colonial, non-heterocyst forming *Trichodesmium* spp. (Church et al., 2005a); three diatom-diazotroph associations (DDAs), *Richelia* associated with *Rhizosolenia* (Het-1; Church et al., 2005b), *Richelia* associated with *Hemiaulus* (Het-2; Foster et al., 2007), and *Calothrix* associated with *Chaetoceros* (Het-3; Foster et al., 2007); and a widespread γ-proteobacterial phylotype γ-24774A11 (Moisander et al., 2008).

Recombinant plasmids with the targeted organism *nifH* fragment were used as qPCR standards, and each 96 well plate was run with a serial dilution (10^0-10^7 *nifH* copies reaction^{-1}) of the appropriate standard. All qPCR reactions were set up with 1 μL of DNA extract in a 15 μL volume with the following reagents and final concentrations: 1X Taqman® Master Mix (Applied Biosystems, Carlsbad, CA, USA), 0.4 μM each

forward and reverse primers, and 0.2 μM probe (5'-FAM and 3'-TAMRA labeled). Thermocycle parameters were as described in Goebel et al. (2010) for all assays, with the exception of using a 64°C annealing temperature for UCYN-A2. The qPCR reaction efficiencies were as follows: UCYN-A1 – 96.7 ± 2.8%; UCYN-A2 – 94.9 ± 3.4%; UCYN-B – 101.3 ± 2.2%; UCYN-C – 95.4 ± 2.9%; γ-24774A11 – 96.2 ± 3.2%; *Trichodesmium* – 93.6 ± 1.4%; Het-1 – 96.1 ± 4.7%; Het-2 – 98.3 ± 2.0%; and Het-3 – 97.5 ± 0.9%. Based on the differences in sample volumes, the limits of detection (LOD) and quantification (LOQ) for all qPCR assays ranged between 11-56 *nifH* copies L^{-1}, and 87-444 *nifH* copies L^{-1}, respectively. Samples were determined to be 'detected, not quantified' (DNQ) when calculated abundances were greater than the LOD, but less than the LOQ. Abundances are reported as *nifH* copies L^{-1}, rather then cells L^{-1} because there is currently little information about the number of *nifH* copies per genome in these diazotroph targets.

Preliminary analyses were conducted from three-depth profiles on days 19 and 20, and very little vertical stratification was observed for most of the targeted diazotrophs, with abundances for most diazotrophs measured at the same order of magnitude throughout the 15 m mesocosm (Supplement Fig. S1). *Trichodesmium* and Het-3 were the only exceptions to this observation. Based on these findings, qPCR analyses focused on samples taken from the middle of the mesocosm (6 m), collected on odd days from inside the mesocosms, and even days from outside the mesocosms.

Growth and mortality rates were calculated for individual diazotrophs inside the mesocosms when abundances were higher than the LOQ on two consecutive sampling days, as described in Moisander et al. (2012), using the following formula: $k = 2.303*(\log_{10}(N_{t2}-N_{t1}))/(t_2-t_1)$, where N_x=abundance at time x. This assumes that the organisms were growing exponentially during the experiment, which cannot be easily verified in field populations. These rates implicitly include grazing, mortality and viral lysis, and thus are net growth and mortality rates.

2.3 Determination of diazotroph community composition using high throughput sequencing of *nifH* amplicons

In order to evaluate whether the diazotrophs targeted via qPCR assays were representative of the phylotypes present in the lagoon and mesocosms, partial *nifH* fragments (ca. 360 base pairs) were amplified using a nested PCR assay and universal *nifH* primers nifH1-4, as described in Turk-Kubo et al., (2014). Two lagoon samples (day 1 and day 22), and three samples from each mesocosm (day 3, day 23, and the day where UCYN-C abundances were beginning to increase, i.e. days 11-15, see discussion below) were chosen for analysis. For each sample, triplicate PCR reactions were pooled. Internal primers were modified 5' common sequence (CS) linkers (CS1_nifH1F: 5'-ACACTGACGACATGGTTCTACATGYGAYCCNAARGCNGA, CS2_nifH2R: 5'-TACGGTAGCAGAGACTTGGTCTADNGCCATCATYTCNCC) to facilitate library preparation at the DNA Services (DNAS) Facility at the University of Illinois, Chicago, using the targeted amplicons sequencing (TAS) approach described in Green et al. (2015). These libraries were pooled with other libraries to achieve a target depth of ca. 40,000 sequences per sample. Sequencing of paired end reads was performed at the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign using Illumina MiSeq technology. De-multiplexed raw paired end reads were merged in CLC Genomics workbench, and merged reads between 300-400 base pairs in length were selected. Quality filtering was performed in QIIME (Caporaso et al., 2010) using the usearch quality filter (usearch_qf) pipeline script, which includes steps for denoising, de novo chimera removal using UCHIME (Edgar et al., 2011) and operational taxonomic unit (OTU) determination using usearch6.1 at 97% nucleotide identity (Edgar 2010). Representative nucleotide sequences from OTUs with greater than 100 reads (277 out of 2325 OTUs, representing 92% of all sequences that passed the usearch quality filter) were imported into ARB (Ludwig et al., 2004), translated into protein sequences, where non-*nifH* OTUs or those with frameshifts were discarded. QIIME script exclude_seqs_by_blast.py was used to check for sequences with >92% amino acid identity to known contaminants; none were found. OTUs targeted by each qPCR assay was determined in silico for each group of diazotrophs in ARB by identifying representative sequences that had 0-2 mismatches in either primer or the probe binding region, without exceeding a total of 4 mismatches in all three regions (see Supplement Table S4).

For the characterization of the overall diazotroph community composition in the lagoon and mesocosms, representative sequences from the most highly recovered OTUs (109 OTUs representing 85% of all post quality-filtering sequences) were considered. Translated amino acid sequences were aligned to the existing amino acid alignment in the curated database. Maximum likelihood trees were calculated using translated amino acid sequences from representative sequences and their closest relatives (determined via blastp) in MEGA 6.06 (Tamura et al. 2013), using the JTT matrix based model and bootstrapped with 1000 replicate trees. Distribution of read data across samples for each representative sequence was visualized in the Interactive Tree of Life online tool (Letunic and Bork, 2006).

Raw reads (fastq files) were deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject ID PRJNA300416 and BioSample accessions SAMN04202524-SAMN04202534.

2.4 PCR amplification of *Cyanothece*-like organisms

Partial *nifH* gene fragments (233 base pairs) from *Cyanothece*-like organisms were PCR amplified using primers designed as part of this study to broadly target the UCYN-C group, including cultivated *Cyanothece* spp., *Gleoecapsa* spp., the cyanobionts of diatoms *E. turgida*, and *R. gibba*, as well as many closely related uncultivated *Cyanothece*-like ecotypes. The oligonucleotide primers *Cyanothece*_nifH_F (5'-CTT AGC AGC AGA ACG GGG AA-3') and *Cyanothece*_nifH_R (5'-GCA TTG CGA AAC CAC CAC AA-3') were designed using NCBI's Primer BLAST, screened in silico for cross reactivity to non-target *nifH* phylotypes in a curated *nifH* ARB database (Heller et al., 2014) and synthesized by Sigma Oligos (St. Louis, MO, USA).

Duplicate PCR reactions were carried out in 20 μL volumes with 1X Platinum® Taq PCR buffer (Invitrogen, Carlsbad, CA), 3.0 mM $MgCl_2$, 400 μm dNTP mix, 0.2 μM of each forward and reverse primers, 1 U Platinum® Taq polymerase (Invitrogen), and 2 μL of DNA extract. Thermocycle parameters were as follows: the initial denaturation step at 94°C for 5 min was followed by 30 cycles of 94°C for 30 sec, 58°C for 30 sec, and 72°C for 30 sec and a final elongation step at 72°C for 10 min. PCR amplicons were cleaned using the QiaQuick Gel Extraction Kit (Qiagen), and sequenced directly using

Sanger technology at the UC Berkeley DNA Sequencing Center using Cyanothece_nifH_R to prime the sequencing reaction.

Raw sequences were processed using Sequencher 5.2.4 (Gene Codes Corporation, Ann Arbor, MI) and phylogenetic analyses were conducted in ARB (Ludwig et al., 2004), using a curated database of all *nifH* sequences available in Genbank (Heller et al., 2014) and in MEGA 6.06 (Tamura et al., 2011). Sequences were submitted to NCBI's Genbank database under accession numbers KU160532-KU160537.

2.5 Statistical analyses

Pearson correlation coefficients were calculated for each paired variable, after transformation to assure normality, and samples collected from the incubations and the Noumea lagoon were treated independently. Supplement Tables S1 and S2 detail linear associations and transformations required for each variable inside and outside the mesocosms, respectively. Diazotroph abundance and physico-chemical environmental data was analyzed in R (Team, 2012). Biplot form principal component analysis was used to examine the temporal and spatial variation in environmental parameters. Correspondence analysis of the abundance data in each mesocosm separately was performed with the ca package (Nenadic and Greenacre, 2007). Biplots of sampling days (rows) and organisms (columns) were compared to examine reproducibility across mesocosms. Least squares regression models were calculated to fit and predict the effect of temperature, salinity, and PO_4 on diazotroph abundance.

3 Results and Discussion

3.1 Diazotroph community structure in VAHINE mesocosms experiment and in the Noumea Lagoon

In order to 1) characterize the diazotroph community composition in the mesocosms and the lagoon using a qualitative approach that is complementary to qPCR and 2) evaluate whether the qPCR assays used in this study, which are widely used in studies of the oligotrophic ocean, target the major lineages of diazotrophs in the Noumea Lagoon, fragments of the *nifH* gene were amplified from a subset of samples using a

well-established nested PCR approach (Zehr and Turner, 2001), and sequenced using Illumina MiSeq technology.

In total, 636,848 paired end reads were recovered from 11 samples, and 544,209 passed the quality filtering steps described above (Supplemental Table S4). Clustering at 97% nucleotide identity yielded 2,325 OTUs with greater than 4 reads, and 334 OTUs with greater than 100 reads. 277 of these OTUs (representing 479,402 reads or 88% of all reads that passed quality filtering; Supplemental Table S4) remained after removing non-*nifH* reads or those with frameshifts, and were used in downstream analyses. A majority of these 277 OTUs (152 OTUs, representing 311,550 reads, or 65% of the reads selected for analysis) were affiliated with *nifH* cluster 1B Cyanobacteria. Reads affiliated with *nifH* cluster 1G, which is composed primarily of γ-proteobacterial phylotypes, was the second most highly recovered group (88 OTUs, representing 120,586 reads, or 25.2% of the reads selected for analysis). Reads that were closely related to *nifH* cluster 1J/1K, comprised primarily of α- and β-proteobacteria, were also recovered, but only comprised 8.4% of the reads selected for analysis (19 OTUs, 40,032 reads). Cluster 3 and cluster 1O affiliated reads were recovered, but together accounted for less than 2% of the reads selected for analysis (Table 1).

The two OTUs with the highest relative abundance (OTU1890 and OTU2317), accounted for 31% of the reads selected for analysis and were closely related to the prymnesiophyte symbiont UCYN-A2 (Supplemental Table S5). Both OTUs were present in the lagoon in day 1 and day 22 samples, and were recovered at high relative abundances from all three mesocosms throughout the experiment (Fig. 1). The third most highly recovered OTU (OTU1) was a γ-proteobacteria closely related to γ-24774A11, a heterotrophic diazotroph with widespread occurrence (Moisander et al., 2014; Langlois et al., 2015), that is also preferentially amplified by the *nifH* primers used (Turk et al., 2011). This sequence type was present in the lagoon samples, and had high relative abundances in all three mesocosms in midpoint samplings (days 11, 13, and 15 for M1, M2, and M3, respectively; Fig. 1). OTU2280 (cluster 1J/1K) was the OTU with the fourth highest relative abundance. It does not have high sequence similarity to any uncultivated or cultivated organisms, with the closest relative, an uncultivated rhizosphere isolate (Genbank accession no. KC667160), sharing only 86% nucleotide

sequence similarity. This is true for all but one of the 1J/1K OTUs, OTU119, which is closely related (98% nucleotide identity) to an environmental sequence recovered from Heron Reef (Genbank accession no. EF175779).

Also among OTUs that had high recovery were UCYN-A1 (OTU2008, OTU1754, and OTU2), other UCYN-A2 OTUs (OTU2325, OTU1548, and OTU664), *Trichodesmium* (OTU5 and OTU35), UCYN-C (OTU12) and two OTUs affiliated with cluster 1G (OTU2218 and OTU2199) (Fig.1 and Supplemental Table S5). 1G OTUs are recovered from the lagoon sample at day 22, and have high relative abundances in all three mesocosms by the end of the experiment. These sequence types are not closely related to γ-24774A11, thus are not quantified using qPCR assays in the study, and their quantitative importance in the mesocosm environment cannot be determined based on this qualitative measure. It is important to note that high relative abundances in PCR-based libraries is not indicative of high abundances, as often these sequence types dominate PCR libraries yet are present at low abundances in the environment (Hewson et al., 2007, Bonnet et al., 2013, Turk-Kubo et al., 2014, Shizoaki et al., 2014, Bentzon-Tilia et al., 2015), presumably as a result of preferential amplification (Turk et al., 2011).

A majority of the OTUs with high relative abundances (159 out of 277, representing 380,556 out of 479,402 reads) affiliated with the following lineages targeted by qPCR assays used in this study: UCYN-A1, UCYN-A2, UCYN-B, UCYN-C, *Trichodesmium*, and *Richelia* associated with *Rhizosolenia* (Het-1), and γ-24774A11. To determine whether the qPCR assays used would target the diazotrophs present, representative sequences were identified that contained between 0-2 mismatches in the primer and probe binding regions, without exceeding a total of 4 mismatches for all three regions. For UCYN-A, *Trichodesmium*, Het-1, and γ-24774A11 lineages, nearly all of the sequence types present met these criteria (between 97-100%), thus would be quantified in the qPCR assays (Table 1). Only 26% of the recovered *Cyanothece*-like sequences would successfully be targeted by the UCYN-C assay, however, this coverage increases to 85% when including two of the most highly recovered OTUs that have a third mismatch at the 5' end of the probe binding region, thus are still likely to be quantified. Together with the results of UCYN-C specific PCRs (see section 3.3 below),

these results indicate that qPCR assays targeting UCYN-C likely quantified most of the phylotypes present, but absolute abundances may be underestimated.

A group of 32 OTUs clustered within the *Chroococcales* and were closely related to both *Crocosphaera watsonii* 8501 (93-97% amino acid identity) and *Cyanothece* sp. WH8904 (93-97% amino acid identity). These OTUs were not targeted by the UCYN-B qPCR assay, with 4-8 total mismatches in the primer/probe binding regions. So despite being considered putative *Crocosphaera* sp. for the purpose of this study, it is possible that they are actually from a *Cyanothece*-like organism (not targeted by the UCYN-C qPCR assay) or another member of the *Chroococcales*. These OTUs were most highly recovered at the final sampling day (day 23) in M2 and M3, but were not recovered in the lagoon libraries.

The heterocyst-forming symbionts of diatoms are conspicuously underrepresented in these libraries, with only two Het-1 OTUs recovered, despite being one of the most abundant diazotrophs measured in the lagoon and mesocosms (see discussion below). Furthermore, there were no *Richelia* associated with *Hemiaulus* (Het-2) or *Calothrix* associated with *Chaetoceros* (Het-3) sequences recovered despite being detected using qPCR assays. These findings lend further support to previous reports that this degenerate PCR assay does not work well for these phylotypes (Foster et al., 2006). This underscores a major limitation of characterizing community composition using solely qualitative PCR techniques.

3.2 Most abundant diazotrophs in VAHINE mesocosms experiment shift from *Richelia* to *Cyanothece*-like ecotypes

During the VAHINE mesocosm experiment, three periods could be roughly defined based on the identity of the most abundant diazotroph and biogeochemical parameters (described in detail in Berthelot et al., 2015). A diazotroph community similar to the Noumea lagoon field community and low DIP concentrations characterized the period prior to the DIP spike (hereafter referred to as P0). The second period, days 5-14 (hereafter referred as P1), was characterized by high abundances of Het-1, high DIP availability, as well as moderate N_2 fixation rates (10.1±1.3 nmol N L^{-1} d^{-1}; Bonnet et al., 2015b). In the second half of the experiment, days 15-23 (hereafter referred as P2), the

UCYN-C population grew to high abundances, and was characterized by low DIP availability, and high N_2 fixation rates (27.3±1.0 nmol N L^{-1} d^{-1}; Bonnet et al., 2015b), with rates reaching > 60 nmol N L^{-1} d^{-1}, ranking among the highest reported in marine waters (Luo et al., 2012).

At the initiation of the mesocosm experiment, the most abundant diazotrophs in the Noumea lagoon were Het-1 (3.1 x10^4 *nifH* copies L^{-1}), and *Richelia* associated with *Hemiaulus* (Het-2; 1.2x10^4 *nifH* copies L^{-1}), as well as the *B. bigelowii*-associated UCYN-A2 (1.5 x 10^4 *nifH* copies L^{-1}) and UCYN-A1 (5.6 x 10^3 *nifH* copies L^{-1}). *Trichodesmium* and the uncultivated unicellular cyanobacterial group C (UCYN-C) were present but at lower abundances, 2.8x10^3 *nifH* copies L^{-1} and 7.8x10^2 *nifH* copies L^{-1}, respectively. γ-24774A11 and *Calothrix* associated with *Chaetoceros* (Het-3) were also present but at abundances too low to quantify (DNQ), and no *Crocosphaera* sp. (UCYN-B) could be detected (Supplement Table S3).

During P0, evidence of shifts from the original Noumea lagoon diazotroph assemblage could be observed in all three mesocosms (Fig. 2b-e). In M1, M2 and M3, Het-1 remained the most abundant diazotroph, and the relative proportions of Het-2 and UCYN-A1 were similar to those measured outside the mesocosms. However, UCYN-A2 abundances decreased and *Trichodesmium* abundances increased over this three-day period with respect to their abundances in the lagoon. These changes were most pronounced in M3. As in the lagoon, UCYN-C and γ-24774A11 were present at low abundances (ca. 10^2 *nifH* copies L^{-1}) in the day 3 sampling.

Diazotroph assemblages within each mesocosm remained relatively consistent throughout the first thirteen days of the experiment, with Het-1 the most abundant diazotroph (Fig. 2b-d). However, a significant shift in community composition began about halfway through the 23-day experiment, evidenced by the increasing abundances of UCYN-C and Het-3, and decreasing abundances of Het-1 and UCYN-A1 (Fig. 2b-d and 3). Increasing UCYN-C abundances were first seen in M1 (day 11), then M2 (day 13), followed by M3 (day 15), which reflected the time in each mesocosm when the DIP turnover time dropped below 1 d^{-1}, signaling DIP limitation (Berthelot et al., 2015;Moutin et al., 2005). In all three mesocosms, which acted as biological replicates, UCYN-C abundances were not only strongly correlated with the decreasing DIP

concentrations (p=0.004, r=-0.51), but also increasing temperatures in the mesocosms (p=0.000, r=+0.86) and could be weakly correlated with increasing salinity (p=0.000,r=+0.56), decreasing Het-1 (p=0.002,r=-0.54), Het-2 (p=0.01, r=-0.37) and UCYN-A1 (p=0.000, r=-0.64) abundances. A similar correlation between UCYN-C abundance and both increasing salinity and temperature was observed outside in the Noumea Lagoon (although UCYN-C abundances in the lagoon remained low; see section 3.4). Salinity in each of the three mesocosms and in the lagoon did increase gradually during the experimental period, but were elevated inside the mesocosms in P2, likely due to evaporative loss (Bonnet et al., 2015a). Together, this data suggests that temperatures below 25.6^{o}C are not optimal temperatures for this group, and that it may tolerate slightly elevated salinities better than other diazotrophs present. This is consistent with the warmer temperatures required for growth for the UCYN-C isolate TW3 (Taniuchi et al., 2012), and the occurrence of the UCYN-C group in low latitude, warm waters (> 28^{o}C) in the Tropical North Atlantic (Langlois et al., 2008). However, changes in temperature and salinity alone cannot explain the high abundances of UCYN-C in these mesocosms, as similar increases were recorded in the lagoon without a corresponding increase in UCYN-C abundances, and other diazotrophs also can grow optimally at these temperatures (e.g. *Crocosphaera* sp. (Webb et al., 2009) and *Trichodesmium* sp. (Breitbarth et al., 2007)).

Het-3 was undetected in all mesocosms until day 7, and then remained at low abundances (usually below quantitation limits) until day fifteen of the incubation. After day 15, Het-3 abundances were of similar order of magnitude to their abundances in the Noumea Lagoon (ca. 10^2-10^3 *nifH* copies L^{-1}). Increases in Het-3 abundances after day 15 could be strongly correlated to increases in temperature (p=0.000, r=+0.82) as well as salinity (p=0.000, r=+0.78), total chl a (p=0.000, r=+0.71), and UCYN-C abundances (p=0.000, r=+0.77). Het-3 abundances could also be weakly correlated to increases in bulk N_2 fixation rates (p=0.01, r=+0.49), and decreases in PO_4 concentrations (p=0.003, r=-0.54), UCYN-A1 (p=0.000, r=-0.64) and Het-1 (p=0.03, r=-0.38) abundances. Het-3 abundances were never greater than 5.8 x 10^3 *nifH* copies L^{-1}, however their distribution throughout the water column (Supplement Fig. S1) and their recovery in the sediment traps (Bonnet et al., 2015b) suggests that these diazotrophs are sinking out of the water

462 column, and could possibly play a role in supplying fixed N to sediments in this shallow
463 coastal system.
464 Both UCYN-A2 and UCYN-A1 abundances peaked early in the mesocosms, at
465 days 5 and 7, respectively. UCYN-A2 abundances were not strongly correlated to any
466 environmental factors, and were only weakly correlated to bulk N_2 fixation rates
467 (p=0.008, r=+0.49) and N_2 fixation rates associated with the <10 μm size fraction
468 (p=0.045, r=+0.38). However, decreasing UCYN-A1 abundances could be strongly
469 correlated to temperature increases (p=0.000, r=-0.75), and decreasing PO_4
470 concentrations (p=0.000, r=+0.69). This implies that the *B. bigelowii*/UCYN-A1
471 association benefitted either directly or indirectly from the day 4 introduction of PO_4 in
472 the early part of the experiment. There is evidence that UCYN-A1 *nifH* transcription (a
473 proxy for active N_2 fixation) is limited by the availability of inorganic P (Turk-Kubo et
474 al., 2012). UCYN-A1 is unable to directly utilize components of the dissolved organic
475 phosphate (DOP) pool, such as phosphoesters and phosphonates (Tripp et al., 2010), but
476 nothing is known about the capability of the *B. bigelowii* host to utilize DOP substrates.
477 The correlation between UCYN-A1 abundances (thus assumed correlation between the
478 symbiosis as a whole) and inorganic P concentrations in the VAHINE mesocosms
479 provides further evidence that inorganic P may be the preferred P source for the *B.*
480 *bigelowii* host.
481 Multiple regression models underscore the importance of a small number of
482 environmental factors in the abundances of three of the diazotrophs targeted. The log
483 abundances of UCYN-A1, UCYN-C and Het-3 in all three mesocosms and in the lagoon
484 can be modeled well using only the temperature, salinity, and PO_4 data, with R^2 values
485 between 0.77-0.85, and R^2-cv between 0.60-0.76 (Table 2), when the sample location
486 (M1-M3 and outside) are included as a variable. Although the three-predictor model
487 provides the highest quality fit to the data, the goodness of prediction values (R^2-cv) are
488 comparable between models that use all three predictors versus models using just T alone
489 in the case of UCYN-A1 and UCYN-C, and just PO_4 alone for all three diazotrophs
490 (Table 2). These results from the linear regression model imply that the most important
491 environmental factor best correlated with the dramatic increase in UCYN-C and Het-3
492 abundances was the decreasing concentration of PO_4.

493 Considering that UCYN-C maintained low population abundances in the Noumea
494 Lagoon during the time of the mesocosm experiment, despite changes in temperature and
495 salinity that mirror changes in the mesocosms (Bonnet et al., 2015a), it follows that
496 UCYN-C may have benefitted either indirectly from the DIP fertilization, or directly
497 from a physical aspect of the mesocosms themselves. A biofilm had accumulated on the
498 sides of the mesocosm bags, and although this was not sampled for molecular analysis, it
499 is possible that the UCYN-C ecotype was a component of this biofilm community, thus
500 dependent on a physical environment not representative of water column conditions.
501 Another explanation may be that the decreased turbulence in the mesocosm environment
502 created favorable conditions for this ecotype. Finally, it is possible that the inverse
503 correlation between UYCN-C abundances and DIP concentrations may be a result of
504 UCYN-C being able to outcompete other diazotrophs for organic P substrates, under low
505 DIP conditions. *Cyanothece* sp. strains PCC 8801 and 8802 have genes used in
506 phosphonate metabolism (Bandyopadhyay et al., 2011), which strongly implies that some
507 strains are able to use this organic P substrate to meet cellular P requirements. The strain
508 most closely related to the UCYN-C ecotype, *Cyanothece* sp. CCY0100, also has genes
509 for phosphonate metabolism and transport (JGI website). It is evident that some
510 component of the microbial community in the mesocosms was utilizing DOP, as DOP
511 stocks were drawn down to levels seen in the lagoon (0.105 ± 0.011 μmol L^{-1}) in the
512 second half of the experiment (Berthelot et al., 2015), however, UCYN-C abundances
513 could not be correlated to this drawdown (data not shown). *Trichodesmium erythraeum*
514 IMS101 and *Calothrix rhizosoleniae* SC01 (Het-3) are the only two other diazotrophs
515 targeted in this study known to possess the metabolic capability to utilize phosphonates
516 (Dyhrman et al., 2006;Hilton et al., 2013). *Richelia* associated with *Hemiaulus* (Het-2) do
517 not have any genes for the metabolism of phosphonates (Hilton et al., 2013), but it
518 remains unclear whether the symbiont of *Rhizosolenia* (Het-1) is able to use
519 phosphonates, as four genes related to phosphonate metabolism were identified in the
520 genome (Hilton, 2014). However, as with UCYN-C, no correlation between DOP
521 concentration and *Trichodesmium*, Het-1 or Het-3 abundances were seen (data not
522 shown). This may be in part due to shortcomings in our understanding of the chemical

composition of the DOP pool (Dyhrman et al., 2007), and which organic P compounds are bioavailable to which organisms.

3.3 Phylogenetic identity of diazotrophs targeted with UCYN-C qPCR assay

The UCYN-C qPCR assay used in this study was designed to target a *nifH* sequence type recovered from Amazon-influenced waters in the Tropical North Pacific (Foster et al., 2007). In addition to uncultivated sequences of marine origin, this assay also targets cultivated members of the UCYN-C group including *Cyanothece* sp. strains AT51142, AT51472, CCY 0110, as well as some freshwater cyanobacterial symbionts of diatoms (Fig. 4). Due to the importance of this group in the mesocosm experiment, and the uncertainty about exactly which organism(s) were targeted with the qPCR assay, PCR amplification of a partial *nifH* fragment was used to specifically characterize the identity of the *Cyanothece*-like organisms. Two closely related sequence types were recovered from 6 m samples on days 15 and 20, represented by 60341CB and 60343CB (Fig. 4), which shared 95% nucleotide similarity. The *Cyanothece*-like sequences recovered from the mesocosms were most closely related to *Cyanothece* sp. CCY0100 (92-93% nucleotide identity; Fig. 4), which was isolated from coastal waters in Chwaka Bay, Zanzibar. They shared only 90-91% nucleotide identity to the sequences used to design the oceanic UCYN-C primer of Foster et al. (2007), and the UCYN-C isolate TW3 (Taniuchi et al., 2012). Thus the ecotypes that reached such high abundances in the mesocosms were more closely related to an ecotype reported in a coastal environment than those recovered from open water regimes.

It is important to note that the sequences amplified using this *Cyanothece*-specific PCR are targeted by the UCYN-C qPCR assay, and the phylotypes with 3 mismatches in the probe binding region were not recovered with this assay. This may be due to differences primer efficiency, the number of amplification steps, and/or depth of sequencing, but these results lend further support that important UCYN-C phylotypes were quantified with qPCR assays.

3.4 *In situ* net growth and mortality rates

553 The VAHINE mesocosm experiment provided a rare opportunity to repeatedly
554 sample the same water mass for an extended period of time, thus the ability to empirically
555 determine *in situ* net growth or mortality rates for individual diazotroph phylotypes,
556 based on the change in *nifH* gene copies L^{-1} between sampling days. Growth (and
557 mortality) rates are critical input parameters for mathematical models of oceanic N
558 budgets. Culture-based rates or estimates are often employed because there have been so
559 few measurements of *in situ* rates from natural populations of diazotrophs, and some
560 species such as UCYN-A remain uncultivated. However, due to the lack of competition
561 and grazing in culture, these rates may be overestimated compared to *in situ* rates.
562 Surprisingly, maximum net growth rates for each diazotroph phylotype were among
563 the highest reported for oceanic diazotrophic organisms (even when considering culture-
564 based growth rates), yet showed substantial variability between mesocosms, both in terms
565 of the absolute rates and the patterns of growth and mortality across time (Table 3). For
566 example, maximum growth rates for UCYN-A2 ranged between 0.52 and 1.71 d^{-1}, but
567 occurred early in the experiment in M3 (day 5), in the middle of the experiment in M2
568 (day 11) and at the end of the experiment in M1 (day 20). The two phylotypes with
569 consistent timing of maximum growth rates across mesocosms were UCYN-C and Het-3.
570 UCYN-C had the highest maximum growth rates of all phylotypes, which ranged
571 between 1.23 - 2.16 d^{-1} and occurred within a four-day period (day 11-15) in all
572 mesocosms. Het-3, which was virtually absent for the first half of the experiment in all
573 three mesocosms, had maximum net growth rates between 0.57 - 1.09 day^{-1}, that occurred
574 within a five day period (day 15 – 20) in all mesocosms.
575 Moisander et al. (2012) reported maximum net growth rates from nutrient amendment
576 experiments conducted in the South Western Pacific close to New Caledonia for UCYN-
577 A1 (0.19 d^{-1}), UCYN-B (0.61 d^{-1}) and γ-24774A11 (0.52 d^{-1}). The maximum net growth
578 rates calculated for these phylotypes during the VAHINE project were considerably
579 higher, at 0.73 d^{-1}, 1.38 d^{-1}, and 1.07 d^{-1} for UCYN-A1, UCYN-B and γ-24774A11,
580 respectively. These results are unexpected considering that the rates determined by
581 Moisander et al. (2012) were from a series of nutrient amendment incubations in 4.5-L
582 bottles, where presumably favorable conditions for the growth of diazotrophs were
583 present.

584 The maximum net growth rates determined *in situ* for UCYN-A2 and UCYN-C were
585 among the highest measured at 1.71 d^{-1}, and 2.16 d^{-1}, respectively, and represent the only
586 reported growth rates for these uncultivated diazotrophs. Very little is known about the
587 newly described uncultivated *B. bigelowii*/UCYN-A2 association, but the difference
588 between UCYN-A1 and UCYN-A2 net growth rates and their patterns within the same
589 mesocosm indicate that the growth rates of each association are likely dependent upon
590 different environmental variables. It also seems plausible, due to the difference in size
591 between the two *B. bigelowii* hosts (Thompson et al., 2014), that they are grazed by
592 different protists.
593 Experimental growth rates as high as 1.92 d^{-1} have been reported for batch cultures of
594 *Cyanothece* sp. ATCC 51142 (Vu et al., 2012), and Taniuchi et al. (2012) measured rates
595 up to 0.85 d^{-1} in the presumably oligotrophic UCYN-C TW3 isolated from the Kuroshio
596 Current. Thus, rates calculated for UCYN-C in the VAHINE mesocosms experiment are
597 similar in magnitude. However, such high rates in culture are a direct result of the lack of
598 competition with other organisms for nutrients important for growth and the lack of
599 grazing pressure. Despite high growth rates in culture, TW3 is found at low abundances
600 in the Kuroshio Current, presumably due to these factors (Taniuchi et al., 2012).
601 Although the factors behind such high growth rates for UCYN-C in the mesocosms are
602 not clear, it is possible that UCYN-C was free from significant grazing in these
603 experiments.
604 Maximum net growth rates for the filamentous diazotrophs were also generally higher
605 than previously reported growth rates. *Trichodesmium* sp. maximum net growth rates for
606 were as high as 1.46 ± 0.05 (M1) and 1.55 ± 0.02 d^{-1} (M3), and microscopic analyses
607 indicated (data not shown) that both *T. erythraeum and T. thiebautti* were present in the
608 mesocosms, the former being the dominant *Trichodesmium* species. These calculated
609 rates are much higher than the specific growth rate previously reported for cultures of *T.*
610 *erythraeum* (0.29 ± 0.04 d^{-1}; Hutchins et al., 2007) and net growth rates for *T. thiebautii*
611 populations in the Noumea lagoon (0.11 - 0.38 d^{-1}; Rodier and Le Borgne, 2008).
612 Maximum growth rates for the DDAs Het-1, Het-2, and Het-3 were 1.28 d^{-1}, 2.24 d^{-1}, and
613 1.09 d^{-1}, respectively. Although these are the first reported growth rates for these DDAs
614 determined using *in situ* cellular abundances, Foster et al. (2011) determined maximum

net growth rates to be much lower when using nanoSIMS-based techniques to quantify ^{15}N incorporation in biomass, which is to be expected if the DDAs are using other N sources (eg. dissolved organic N and/or recycled N) in addition to fixing N.

For many diazotrophs, the pattern of net growth and mortality rates indicated a very dynamic process that appeared specific to each mesocosm. Most diazotrophs experienced multiple transitions between net growth and net mortality within a single mesocosm throughout the 23-day incubation (Table 3). For example, in M1, net growth rates for *Trichodesmium* were observed approximately every other sampling period (on days 9, 13, 15,19, and 23), and in some intervening periods experienced mortality rates similar in magnitude to growth rates (e.g. day 7 and 11). In contrast, net growth occurred on days 7, 11, 13, 20, and 23 in M3, and one of the largest net mortality rates was measured at day 9 (-2.16 +/- 0.18 d^{-1}). There are no aspects of the experimental design that can be invoked to explain this variability; in fact biogeochemical parameters and picoplankton population dynamics were well replicated in all three mesocosms (Bonnet et al., 2015a). Therefore the dynamic nature of diazotroph growth and mortality rates in each mesocosm most likely results from a combination of grazing pressure and viral lysis, which can be expected to reflect natural variations in the grazers and virus present.

3.5 Symbiotic *Richelia* and UCYN-A ecotypes are abundant in the Noumea lagoon during the VAHINE mesocosms experiment

The VAHINE mesocosm experiment provided an additional opportunity to characterize the abundances of targeted diazotrophs in the Noumea lagoon using quantitative techniques. As described above, the diazotroph assemblage at the onset of the VAHINE mesocosms experiment was dominated by Het-1, Het-2, UCYN-A2 and UCYN-A1 (see section 3.1). Het-1 remained the most abundant of the targeted diazotrophs for most of the sampling days ($3.1x10^4$ – $5.5x10^5$ *nifH* copies L^{-1}) during the period of study (Fig. 2e, Supplement Table S3). Het-2 abundances were consistently lower than Het-1, and variable throughout the period of study, ranging between $4.0x10^3$ – $1.6x10^4$ *nifH* copies L^{-1}, and reaching peak abundances on day 10 and again on day 18 (Supplement Table S3). Bonnet et al. (2015a) discusses the BNF rates in detail, however,

it should be noted that in the lagoon only Het-2 abundances had significantly strong positive correlation with bulk N_2 fixation rates (p=0.01, r=+0.78).

Although *Richelia* living in association with *Rhizosolenia* (Het-1) and *Hemiaulus* (Het-2) are known to be widely distributed and potentially significant N_2-fixers in oligotrophic oceans (Carpenter et al., 1999;Subramaniam et al., 2008;Karl et al., 2012), reports of their presence in coastal waters in the Southwest Pacific are rare, and previous studies during the austral summer in the Noumea Lagoon reported very low abundances of heterocystous symbionts (Rodier and Le Borgne, 2010;Biegala and Raimbault, 2008). This is the first report on quantitative abundances of *Richelia* in the Noumea lagoon, which are comparable to abundances reported from the Tropical North Atlantic (Foster et al., 2007;Goebel et al., 2010) and the North Pacific Subtropical Gyre (Church et al., 2008;Foster and Zehr, 2006), where the DDAs are thought to have a significant impact on C sequestration due to their productivity and rapid sinking (Subramaniam et al., 2008;Karl et al., 2012;Villareal et al., 2012), and may play a role in supporting the benthic community (Houlbreque and Ferrier‐Pagès, 2009).

UCYN-A2 was the second most abundant member of the targeted diazotrophic assemblage in the lagoon for the first 10 days of the experiment (Fig. 2e), and was present at abundances as high as $1.1x10^5$ *nifH* copies L^{-1} (day 6; Supplement Table S3). UCYN-A2 abundances declined steadily throughout the period of the 23-day experiment (p=0.04, r=-0.63), and had a significantly strong negative correlation to sea temperature (p=0.003, r=-0.82) and salinity (p=0.03, r=-0.89), both of which were increasing in Noumea lagoon throughout the VAHINE deployment (Bonnet et al., 2015b). UCYN-A1 abundances were consistently lower than those of UCYN-A2, ranging between $2.8x10^3$ – $6.4x10^4$ *nifH* copies L^{-1} with peaks in abundance at day 6 and again at day 22 (Supplement Table S3). Interestingly, UCYN-A1 abundances had significant positive correlation to total chl *a* in the > 10 µm size fraction (p=0.03, r=+0.72). The significant inverse relationship between UCYN-A2 abundances and water temperatures in the Noumea lagoon, suggests that the *B. bigelowii*/UCYN-A2 association may thrive at lower temperatures than other diazotrophs (e.g. *Trichodesmium*), similar to the *B. bigelowii*/UCYN-A1 association (Moisander et al., 2010).

675 The coexistence of the *B. bigelowii*/UCYN-A1 and *B. bigelowii*/UCYN-A2
676 ecotypes in the Noumea Lagoon, both present at reasonably high abundances, indicates
677 that the ecological niches of these cryptic symbioses overlap. Very little information
678 currently exists about the ratio of UCYN-A1 to UCYN-A2 in tropical oligotroph oceanic
679 regimes, but UYCN-A1 is typically found at higher relative abundances (as inferred from
680 clone-library based studies; Thompson et al., 2014) in oligotrophic regions. The Noumea
681 Lagoon is the first location where the co-occurrence of each ecotype has been verified
682 using quantitative techniques, but it must be noted that the difference in abundances may
683 be a result of the number of UCYN-A2 cells associated with each *B. bigelowii* host,
684 which has been reported to be as high as 11:1 (Thompson et al., 2014).
685 *Trichodesmium* spp. have been routinely observed in the Noumea Lagoon using
686 satellite observations (e.g. Dupouy et al., 2000;Dupouy et al., 2011) and direct field
687 measurements (Renaud et al., 2005;Masotti et al., 2007;Rodier and Le Borgne, 2010,
688 2008). Conspicuous blooms that occur in the austral spring and summer months have
689 been associated with warm sea surface temperatures (>26^{o}C) and recent NO_3^- and soluble
690 reactive P (SRP) enrichments in the lagoon (Rodier and Le Borgne, 2008). Although
691 *Trichodesmium* was present at relatively low abundances during the first 8 days of the
692 experiment ($3.4x10^2 – 6.5x10^3$ *nifH* copies L^{-1}; Supplement Table S3), abundances
693 increased steadily throughout the experiment (p=0.01, r=+0.73). *Trichodesmium* reached
694 peak abundances of 1.4×10^5 *nifH* copies L^{-1} on the final sampling day (day 22; Fig. 2e).
695 The only environmental parameter with which changes in *Trichodesmium* abundances
696 could be strongly correlated was NH_4^+ (p=0.02, r=+0.81).
697 The UCYN-C group comprised a minor part of the Noumea lagoon diazotroph
698 community, and was detected at abundances between $3.2x10^2$ - $4.8x10^3$ *nifH* copies L^{-1}
699 (Supplement Table S3) during the experimental period. However, UCYN-C abundances
700 did show a linear increase over the duration of the experiment (p=0.01, r=+0.74), and
701 significant positive correlations to changes in salinity (p=0.02, r=+0.80), and NH_4^+
702 (p=0.03, r=+0.76). The unicellular UCYN-C group was first described when *nifH*
703 sequences with 90% nucleotide identity to the cultivated *Cyanothece* sp. PCC 51142
704 were recovered from surface waters in the Tropical North Atlantic (Langlois et al.,
705 2005;Foster et al., 2007). This phylotype has rarely been quantified in the oligotrophic

706 ocean, and when present, abundances are low (Foster et al., 2007;Langlois et al.,
707 2008;Needoba et al., 2007;Goebel et al., 2010;Ratten et al., 2014), therefore very little is
708 known about its distribution or importance. The qPCR assay used in this study (Foster et
709 al., 2007) amplifies not only the uncultivated Atlantic phylotype, but also many
710 cultivated *Cyanothece*- and *Gloeothece*-like isolates, which are expected to be present in
711 an environment like the Noumea Lagoon, as *Cyanothece*-and *Gloeothece*-like organisms
712 have been reported in shallow marine sediments (Hunter et al., 2006;Bauer et al., 2008)
713 and intertidal sands (Reddy et al., 1993;Ohki et al., 2008).
714 γ-24774A11 was also consistently detected at low abundances (ca. 10^2-10^3 *nifH*
715 copies L^{-1}; Supplement Table S3), which increased throughout the duration of the
716 experiment (p=0.03, r=+0.65). Despite being the most widely studied marine
717 heterotrophic diazotroph, very little is known about this phylotype. Recent studies
718 suggest that this γ-proteobacterial diazotroph is likely to be free-living and may be able to
719 sustain in a broad range of environmental conditions, as evidenced by low but uniform
720 abundances throughout the photic zone in the South Pacific (Moisander et al., 2014).
721 The epiphytic *Chaetoceros* symbiont, *Calothrix* (Het-3), was also consistently
722 detected at low abundances (ca. 10^2-10^3 *nifH* copies L^{-1}; Supplement Table S3).
723 *Chaetoceros* can often be found as part of the neritic diatom assemblage, therefore this
724 association is generally observed in coastal or coastal transition zone regions (Gómez et
725 al., 2005). This is the first report of Het-3 in the coastal oligotrophic waters surrounding
726 New Caledonia.
727 *Crocosphaera* sp. (UCYN-B), previously reported to be members of the
728 unicellular diazotroph community in the Noumea Lagoon (Biegala and Raimbault, 2008),
729 were not initially detected, but were present at low abundances that increased over the 23
730 day period (p=0.04, r=+0.63) to abundances as high as 2.7×10^3 *nifH* copies L^{-1}, and had a
731 significant strong correlation with total chl *a* (p=0.02, r=+0.81). It must be noted that the
732 *Crocosphaera*-like sequences that were recovered from the mesocosm libraries would not
733 likely amplify efficiently with the qPCR assay used. Therefore, it is not clear whether
734 these abundances included a *Crocosphaera* population not recovered in the PCR-based
735 *nifH* libraries, or were in part the result of cross-reactivity with the *Crocosphaera*-like

736 population that were amplified from mesocosm samples during P1 and P2, and possibly
737 present in the lagoon as well.
738
739 ## 4 Conclusions
740 The VAHINE mesocosm experiment was conducted to trace the incorporation of N_2
741 fixed by diazotrophs into the food web (Bonnet et al., 2015a). Despite lack of sampling
742 replication within each mesocosm, consistent patterns in both the relative abundances of
743 targeted diazotrophic phylotypes, as well as shifts in diazotrophic community
744 composition, were reproducibly seen in all three mesocosms. Although the timing of the
745 increases were specific to an individual mesocosm, UCYN-C and Het-3 abundances
746 increased over time, while UCYN-A1 and Het-1 abundances decreased over time in all
747 three mesocosms (Fig, 3). The experimental conditions selected for the growth of UCYN-
748 C during P2, an ecotype never before quantified at high abundances in the marine water
749 column, which enabled the calculation of growth rates of this uncultivated ecotype, and
750 provided insight into its dynamics with respect to environmental parameters. Although
751 the data strongly suggests that the drawdown of DIP provided an environment favorable
752 for high UCYN-C growth rates, further studies are required to better understand the
753 environmental conditions that stimulated this bloom, and whether such blooms are seen
754 in the Noumea Lagoon itself.
755 The experimental set up provided a rare opportunity to calculate *in situ* net growth
756 rates for natural populations of diazotrophs, including the uncultivated UCYN-A. This
757 study provided the first growth rates for the UCYN-C phylotype and for UCYN-A2, both
758 of which were surprisingly high, implying not only favorable conditions, but also a lack
759 of grazing pressure. Maximum net growth rates were high for all diazotroph ecotypes, but
760 most also experienced intermittent periods of growth and mortality within the 23-day
761 experiment, which was also an unexpected finding. Along with net rates recently reported
762 by Moisander et al. (2012), we anticipate that this data will be important for future
763 modeling efforts.
764 The analysis of the diazotroph assemblage outside the mesocosms represents both the
765 first quantitative data on targeted diazotroph phylotypes, as well as the first *nifH*-based
766 diversity libraries on the populations in the Noumea Lagoon. Not previously considered

767 to be a significant diazotrophs in this region, DDAs must now be considered a potentially
768 important contributor to BNF in these waters, especially in the austral summer. Although
769 the presence of UCYN-A in the lagoon has long been suspected due to the relative
770 importance of daytime nitrogen fixation in the small size fraction (Biegala and
771 Raimbault, 2008), we report the first quantitative data on UCYN-A abundances in the
772 Noumea lagoon. Furthermore, the co-occurrence of two UCYN-A ecotypes revealed in
773 this study, provides important insight into the overlap in environmental niches for these
774 two ecotypes.
775

776 **Author contributions**

777 SB designed and executed the experiments, and SB and AD sampled the experiments for
778 molecular analyses. MH extracted DNA samples, and KT conducted all qPCR and PCR
779 analysis, and analyzed the data. IF performed statistical analyses. KT and JZ prepared
780 the manuscript with input from all co-authors.
781

782 **Acknowledgements**

783 Funding for this research was provided by the Agence Nationale de la Recherche (ANR
784 starting grant VAHINE ANR-13-JS06-0002), INSU-LEFE-CYBER program, GOPS and
785 IRD. The authors thank the captain and crew of the R/V Alis. We also thank the SEOH
786 divers service from Noumea, as well as the technical support and the divers of the IRD
787 research center of Noumea. We gratefully acknowledge C. Guieu, J-M. Grisoni and F.
788 Louis for the mesocosms design and their useful advice. We would also like to thank five
789 anonymous reviewers for their time and critiques. The molecular component of this work
790 was funded by a Gordon and Betty Moore Foundation Marine Investigator award (JPZ)
791 and the National Science Foundation Science Center for Microbial Oceanography
792 Research and Education (C-MORE, Grant #EF-0424599).
793

794 **References**

795 Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stöckel, J., Liberton, M., Min, H., Sherman, L. A.,
796 and Pakrasi, H. B.: Novel metabolic attributes of the genus *Cyanothece*, comprising a group of
797 unicellular nitrogen-fixing cyanobacteria, MBio, 2, e00214-00211, 2011.
798

Bauer, K., Díez, B., Lugomela, C., Seppälä, S., Borg, A. J., and Bergman, B.: Variability in benthic diazotrophy and cyanobacterial diversity in a tropical intertidal lagoon, FEMS Microbiol Ecol, 63, 205-221, 2008.

Berthelot, H., Moutin, T., L'Helguen, S., Leblanc, K., Hélias, S., Grosso, O., Leblond, N., Charrière, B., and Bonnet, S.: N_2 fixation and DON as significant N sources for primary and export productions during the VAHINE mesocosms experiment (New Caledonia lagoon), Biogeosciences, 2015.

Bentzon-Tilia, M., Traving, S. J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J. L., Markager, S., and Riemann, L.: Significant N_2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries, The ISME journal, 9(2), 273-285, 2015.

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

Bird, C., and Wyman, M.: Transcriptionally active heterotrophic diazotrophs are widespread in the upper water column of the Arabian Sea, FEMS Microbiol Ecol, 84, 189-200, 10.1111/1574-6941.12049, 2012.

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K. A., Moutin, T., Hamersley, R. M., Grosso, O., Zehr, J. P., and Capone, D. G.: Aphotic N_2 Fixation in the Eastern Tropical South Pacific Ocean, PLoS ONE, 8, e81265, doi:10.1371/journal.pone.0081265, 2013.

Bonnet, S.: Introduction to the project VAHINE: VAriability of vertical and tropHIc transfer of fixed N_2 in the south wEst Pacific, Biogeosciences, in preparation, 2015a.

Bonnet, S., Berthelot, H., Turk-Kubo, K. Fawcet, F. Rahav, E., Berman-Frank, I., and l'Helguen, S. Dynamics of N_2 fixation and fate of diazotroph-derived nitrogen during the VAHINE mesocosm experiment (New Caledonia), in preparation 2015b.

Bonnet, S., Berthelot, H., Turk-Kubo, K., Cornet-Bartaux, V., Fawcett, S. E., Berman-Frank, I., Barani, A., Dekaezemacker, J., Benavides, M., Charriere, B., and Capone, D. G.: Diazotroph derived nitrogen supports diatoms growth in the South West Pacific: a quantitative study using nanoSIMS, Limnology and Oceanography, *under revision*.

Breitbarth, E., Oschlies, A., and LaRoche, J.: Physiological constraints on the global distribution of *Trichodesmium* - effect of temperature on diazotrophy, Biogeosciences, 4, 53-61, doi:10.5194/bg-4-53-2007, 2007.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., and Gordon, J. I.: QIIME allows analysis of high-throughput community sequencing data, Nature Methods, 7, 335-336, doi:10.1038/nmeth.f.303, 2010.

Carpenter, E. J., Montoya, J. P., Burns, J., Mulholland, M. R., Subramaniam, A., and Capone, D. G.: Extensive bloom of a N_2-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean, Mar Ecol Prog Ser, 185, 273-283, 1999.

Carpenter, E. J., and Foster, R. A.: Marine cyanobacterial symbioses, in: Cyanobacteria in Symbiosis, edited by: Rai, A. N., Bergman, B., and Rasmussen, U., Kluwer Academic Publishers, Netherlands, 11-17, 2002.

Church, M., Jenkins, B., Karl, D., and Zehr, J.: Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific, Aquat Microb Ecol, doi:10.3354/ame038003, 2005a.

Church, M., Short, C., Jenkins, B., Karl, D., and Zehr, J.: Temporal Patterns of Nitrogenase Gene (*nifH*) Expression in the Oligotrophic North Pacific Ocean, Appl Environ Microbiol, 2005b.

Church, M. J., Bjorkman, 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, 63-77, 2008.

Dupouy, C., Neveux, J., Subramaniam, A., Mulholland, M. R., Montoya, J. P., Campbell, L., Carpenter, E. J., and Capone, D. G.: Satellite captures *Trichodesmium* blooms in the southwestern tropical Pacific, EOS, Transactions American Geophysical Union, 81, 13-16, 2000.

Dupouy, C., Benielli-Gary, D., Neveux, J., Dandonneau, Y., and Westberry, T.: A new algorithm for detecting *Trichodesmium* surface blooms in the South Western Tropical Pacific, Biogeosciences, 8, 1-17, doi:10.5194/bg-8-1-2011, 2011.

Dyhrman, S. T., Chappell, P. D., Haley, S. T., Moffett, J. W., Orchard, E. D., Waterbury, J. B., and Webb, E. A.: Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*, Nature, 439, 68-71, 2006.

Dyhrman, S. T., Ammerman, J. W., and Van Mooy, B. A. S.: Microbes and the Marine Phosphorus Cycle, Oceanography, 20, 110-116, 2007.

Edgar, R. C.: Search and clustering orders of magnitude faster than BLAST, Bioinformatics, 26, 2460-2461,1367-4803, 2010.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R.: UCHIME improves sensitivity and speed of chimera detection, Bioinformatics, 27, 2194-2200, 2011.

Farnelid, H., Andersson, A. F., Bertilsson, S., Al-Soud, W. A., Hansen, L. H., Sorensen, S., Steward, G. F., Hagstrom, A., and Riemann, L.: Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria, PLoS ONE, 6, e19223, 10.1371/journal.pone.0019223, 2011.

Fong, A. A., Karl, D., Lukas, R., Letelier, R. M., Zehr, J. P., and Church, M. J.: Nitrogen fixation in an anticyclonic eddy in the oligotrophic North Pacific Ocean., ISME Journal, 2, 663-676, 2008.

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

Foster, R. A., Subramaniam, A., Mahaffey, C., Carpenter, E. J., Capone, D. G., and Zehr, J. P.: Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical north Atlantic Ocean, Limnol Oceanogr, 52, 517-532, 2007.

Foster, R. A., Goebel, N. L., and Zehr, J. P.: Isolation of *Calothrix rhizosoleniae* (CYANOBACTERIA) strain SC01from *Chaetoceros* (BACILLARIOPHYTA) spp. diatoms of the subtropical north Pacific Ocean, J Phycol, 46, 1028-1037, doi:10.1111/j.1529-8817.2010.00885.x, 2010.

Foster, R. A., Kuypers, M. M. M., Vagner, T., Paerl, R. W., Musat, N., and Zehr, J. P.: Nitrogen fixation and transfer in open ocean diatom–cyanobacterial symbioses, ISME J, 5, 1484-1493, doi:10.1038/ismej.2011.26, 2011.

Garcia, N., Raimbault, P., and Sandroni, V.: Seasonal nitrogen fixation and primary production in the Southwest Pacific: nanoplankton diazotrophy and transfer of nitrogen to picoplankton organisms, Mar Ecol Prog Ser, 343, 25-33, 2007.

Gattuso, J.-P., Frankignoulle, M., and Wollast, R.: Carbon and carbonate metabolism in coastal aquatic ecosystems, Annu Rev Ecol Syst, 405-434, doi:10.1146/annurev.ecolsys.29.1.405, 1998.

Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R. W., Hewson, I., Morrison, A. E., Montoya, J. P., Edwards, C. A., and Zehr, J. P.: Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N_2 fixation in the tropical Atlantic Ocean, Environ. Microbiol., 12, 3272-3289, doi:10.1111/j.1462-2920.2010.02303.x, 2010.

Gómez, F., Furuya, K., and Takeda, S.: Distribution of the cyanobacterium *Richelia intracellularis* as an epiphyte of the diatom *Chaetoceros compressus* in the western Pacific Ocean, J Plankton Res, 27, 323-330, 2005.

Green, S. J., Venkatramanan, R., and Naqib, A.: Deconstructing the Polymerase Chain Reaction: Understanding and Correcting Bias Associated with Primer Degeneracies and Primer-Template Mismatches, PLOS One, doi: 10.1371/journal.pone.0128122, 2015.

Gruber, N., and Sarmiento, J. L.: Global patterns of marine nitrogen fixation and denitrification, Global Biogeochem. Cy., 11, 235-266, 1997.

Hagino, K., Onuma, R., Kawachi, M., and Horiguchi, T.: Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae), PLoS One, 8, e81749, doi:10.1371/journal.pone.0081749, 2013.

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

Heller, P., Tripp, H. J., Turk-Kubo, K., and Zehr, J. P.: ARBitrator: a software pipeline for on-demand retrieval of auto-curated *nifH* sequences from GenBank, Bioinformatics, doi:10.1093/bioinformatics/btu417, 2014.

Hewson, I., Moisander, P. H., Achilles, K. M., Carlson, C. A., Jenkins, B. D., Mondragon, E. A., Morrison, A. E., and Zehr, J. P.: Characteristics of diazotrophs in surface to abyssopelagic waters of the Sargasso Sea, Aquat Microb Ecol, 46, 15-30, 2007.

Hilton, J.: Ecology and evolution of diatom-associated cyanobacteria through genetic analyses, Ph.D., Ocean Sciences Department, University of California, Santa Cruz, Santa Cruz, 2014.

Hilton, J. A., Foster, R. A., Tripp, H. J., Carter, B. J., Zehr, J. P., and Villareal, T. A.: Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont, Nat Commun, 4, 1767, doi:10.1038/ncomms2748, 2013.

Houlbreque, F., and Ferrier−Pagès, C.: Heterotrophy in tropical scleractinian corals, Biological Reviews, 84, 1-17, doi:10.1111/j.1469-185X.2008.00058.x, 2009.

Hunter, E. M., Mills, H. J., and Kostka, J. E.: Microbial community diversity associated with carbon and nitrogen cycling in permeable shelf sediments, Appl Environ Microbiol, 72, 5689-5701, 2006.

Hutchins, D. A., Fu, F. X., Zhang, Y., Warner, M. E., Feng, Y., Portune, K., Bernhardt, P. W., and Mulholland, M. R.: CO_2 control of *Trichodesmium* N_2 fixation, photosynthesis, growth rates, and elemental ratios: Implications for past, present, and future ocean biogeochemistry, Limnol Oceanogr, 52, 1293-1304, 2007.

Hyndes, G. A., Nagelkerken, I., McLeod, R. J., Connolly, R. M., Lavery, P. S., and Vanderklift, M. A.: Mechanisms and ecological role of carbon transfer within coastal seascapes, Biological Reviews, 89, 232-254, 2014.

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, 533-538, 1997.

Karl, D. M., and Letelier, R. M.: Nitrogen fixation-enhanced carbon sequestration in low nitrate, low chlorophyll seascapes, Mar Ecol Prog Ser, 364, 257-268, doi:10.3354/meps07547, 2008.

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, Proceedings of the National Academy of Sciences, 109, 1842-1849, doi:10.1073/pnas.1120312109, 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, 7910-7919, 2005b.

Langlois, R. J., Hummer, D., and LaRoche, J.: Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean, Appl Environ Microbiol, 74, 1922-1931, 2008.

Langlois, R., Großkopf, T., Mills, M., Takeda, S., and LaRoche, J.: Widespread Distribution and Expression of Gamma A (UMB), an Uncultured, Diazotrophic, γ-Proteobacterial nifH Phylotype, PLoS ONE, 10, e0128912, 0121932-0126203, 2015.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüßmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., and Schleifer, K. H.: ARB: a software environment for sequence data, Nucleic Acids Res, 32, 1363-1371, 2004.

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 Jr, 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, 47-73, doi:10.5194/essd-4-47-2012, 2012.

Masotti, I., Ruiz-Pino, D., and Le Bouteiller, A.: Photosynthetic characteristics of *Trichodesmium* in the southwest Pacific Ocean: importance and significance, Marine Ecology-Progress Series, 338, 47-59, 2007.

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, 954-967, 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 N_2 fixation domain, Science, 327, doi:1512-1514, 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, 733-744, 2012.

Moisander, P. H., Serros, T., Paerl, R. W., Beinart, R. A., and Zehr, J. P.: Gammaproteobacterial diazotrophs and *nifH* gene expression in surface waters of the South Pacific Ocean, ISME J, doi:10.1038/ismej.2014.49, 2014.

Moutin, T., Van Den Broeck, N., Beker, B., Dupouy, C., Rimmelin, P., and Le Bouteiller, A.: Phosphate availability controls *Trichodesmium* spp. biomass in the SW Pacific Ocean, Marine Ecology-Progress Series, 297, 15-21, 2005.

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, 1317-1327, 2007.

Nenadic, O., and Greenacre, M.: Correspondence analysis in R, with two-and three-dimensional graphics: The ca package, 2007.

Ohki, K., Kamiya, M., Honda, D., Kumazawa, S., and Ho, K. K.: MORPHOLOGICAL AND PHYLOGENETIC STUDIES ON UNICELLULAR DIAZOTROPHIC CYANOBACTERIA (CYANOPHYTES) ISOLATED FROM THE COASTAL WATERS AROUND SINGAPORE, J Phycol, 44, 142-151, 2008.

Ouillon, S., Douillet, P., Lefebvre, J.-P., Le Gendre, R., Jouon, A., Bonneton, P., Fernandez, J.-M., Chevillon, C., Magand, O., and Lefèvre, J.: Circulation and suspended sediment transport in a coral reef lagoon: The south-west lagoon of New Caledonia, Mar Pollut Bull, 61, 269-296, 2010.

Ratten, J., LaRoche, J., Desai, D. K., Shelley, R., Landing, W. M., Boyle, E. A., Cutter, G. A., and Langlois, R.: Source of iron and phosphate affect the distribution of diazotrophs in the North Atlantic, Deep Sea Res II, doi:10.1016/j.dsr2.2014.11.012, 2014.

Reddy, K., Haskell, J., Sherman, D., and Sherman, L.: Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus *Cyanothece*, J Bacteriol, 175, 1284, 1993.

Renaud, F., Pringault, O., and Rochelle-Newall, E.: Effects of the colonial cyanobacterium *Trichodesmium* spp. on bacterial activity, Aquat Microb Ecol, 41, 261-270, 2005.

Riemann, L., Farnelid, H., and Steward, G. F.: Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters, Aquat Microb Ecol, 61, 225-237, doi:10.3354/Ame01431, 2010.

Rodier, M., and Le Borgne, R.: Population dynamics and environmental conditions affecting *Trichodesmium* spp.(filamentous cyanobacteria) blooms in the south–west lagoon of New Caledonia, J Exp Mar Biol Ecol, 358, 20-32, 2008.

Rodier, M., and Le Borgne, R.: Population and trophic dynamics of *Trichodesmium thiebautii* in the SE lagoon of New Caledonia. Comparison with *T. erythraeum* in the SW lagoon, Mar Pollut Bull, 61, 349-359, 2010.

Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S., and Furuya, K.: Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean, Global Biogeochemical Cycles, 28(10), 1096-1110, 2014.

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

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S.: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol Biol Evol, 28, 2731-2739, doi:10.1093/molbev/msr121, 2011.

Taniuchi, Y., Chen, Y.-l. L., Chen, H.-Y., Tsai, M.-L., and Ohki, K.: Isolation and characterization of the unicellular diazotrophic cyanobacterium Group C TW3 from the tropical western Pacific Ocean, Environ. Microbiol., 14, 641-654, doi:10.1111/j.1462-2920.2011.02606.x, 2012.

Team, R. D. C.: R: A language and environment for statistical computing, Foundation for Statistical Computing, http://www.R-project.org/, 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, 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, 1546-1550, doi:10.1126/science.1222700, 2012.

Torréton, J.-P., Rochelle-Newall, E., Pringault, O., Jacquet, S., Faure, V., and Briand, E.: Variability of primary and bacterial production in a coral reef lagoon (New Caledonia), Mar Pollut Bull, 61, 335-348, 2010.

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, 90-94, doi:10.1038/nature08786, 2010a.

Turk, K., Rees, A. P., Zehr, J. P., Pereira, N., Swift, P., Shelley, R., Lohan, M., Woodward, E. M. S., and Gilbert, J.: Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic, ISME J, 5, 1201-1212, 10.1038/ismej.2010.205, 2011.

Turk-Kubo, K. A., Achilles, K. M., Serros, T. R., Ochiai, M., Montoya, J. P., and Zehr, J. P.: Nitrogenase *(nifH)* gene expression in diazotrophic cyanobacteria in the Tropical North Atlantic in response to nutrient amendments, Frontiers in Aquatic Microbiology, 3, 386, doi:10.3389/fmicb.2012.00386, 2012.

Turk-Kubo, K. A., Karamchandani, M., Capone, D. G., and Zehr, J. P.: The paradox of marine heterotrophic nitrogen fixation: abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific, Environ Microbiol, 16, 3095-3114, doi:10.1111/1462-2920.12346, 2014.

Villareal, T. A.: Laboratory culture and preliminary characterization of the nitrogen-fixing *Rhizosolenia-Richelia* symbiosis, Mar Ecol, 11, 117-132, 1990.

Villareal, T. A.: Marine nitrogen-fixing diatom - cyanobacteria symbioses, in: Marine Pelagic Cyanobacteria: *Trichodesmium* and other Diazotrophs, edited by: Carpenter, E. J., Capone, D. G., and Rueter, J. G., Series C: Mathematical and Physical Sciences, Kluwer Academic Publishers, The Netherlands, 163-175, 1992.

Villareal, T. A., Brown, C. G., Brzezinski, M. A., Krause, J. W., and Wilson, C.: Summer diatom blooms in the North Pacific subtropical gyre: 2008–2009, PLoS ONE, 7, e33109, doi:10.1371/journal.pone.0033109, 2012.

Vu, T. T., Stolyar, S. M., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Brown, R. N., Lipton, M. S., Osterman, A., Fredrickson, J. K., and Konopka, A. E.: Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium *Cyanothece* sp. ATCC 51142, Plos Computational Biology, 8, e1002460, doi:10.1371/journal.pcbi.1002460, 2012.

Walsby, A.: The gas vesicles and buoyancy of *Trichodesmium*, in: Marine Pelagic Cyanobacteria: *Trichodesmium* and Other Diazotrophs, Springer, 141-161, 1992.

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, 338-348, 2009.

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 N_2 in the subtropical North Pacific Ocean, Nature, 412, 635-638, 2001.

Zehr, J. P., and Turner, P. J.: Nitrogen fixation: nitrogenase genes and gene expression, in: Methods in Microbiology: Marine Microbiology, edited by: Paul, J. H., Academic Press, Ltd., London, 271-286, 2001.

Zehr, J., Jenkins, B., Short, S., and Steward, G.: Nitrogenase gene diversity and microbial community structure: a cross‐system comparison, Environ. Microbiol., 5, 539-554, 2003.

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 N_2-fixing cyanobacteria lack oxygenic photosystem II, Science, 322, 1110-1112, 2008.

Tables

		targeted by qPCR	
	no. OTUs (no. sequences)	no. OTUs (no. sequences)	% OTUs (% sequences)
1B	**152 (311550)**		
UCYN-A	68 (260221)	60 (258005)	88% (99%)
UCYN-B	32 (15917)	0 (0)	0% (0%)
UCYN-C	18 (11324)	9 (2915)	50% (26%) ⊙
Tricho.	18 (20034)	14 (19386)	78% (97%)
Het-1	2 (1738)	2 (1738)	100% (100%)
Het-2	0 (0)	0 (0)	0% (0%)
Het-3	0 (0)	0 (0)	0% (0%)
other	14 (2316)	na	
1G	**88 (120586)**		
γ-24774A11	22 (51594)	18 (50833)	82% (99%)
other	66 (68992)	na	
1J/1K	**19 (40032)**		
3	**16 (6825)**		
1O	**2 (409)**		

⊙ 61% (85%) when a third mismatch at the 5' end (probe) is allowed

Table 1. In silico *qPCR coverage analysis.* Taxonomic assignment for all *nifH* amplicons reads that passed the quality filtering steps (first column), and the number of OTUs affiliated with each group that are successfully targeted by qPCR assays used in this study. Partial *nifH* sequences were classified according to the convention defined in Zehr et al. (2003). OTU – operational taxonomic unit.

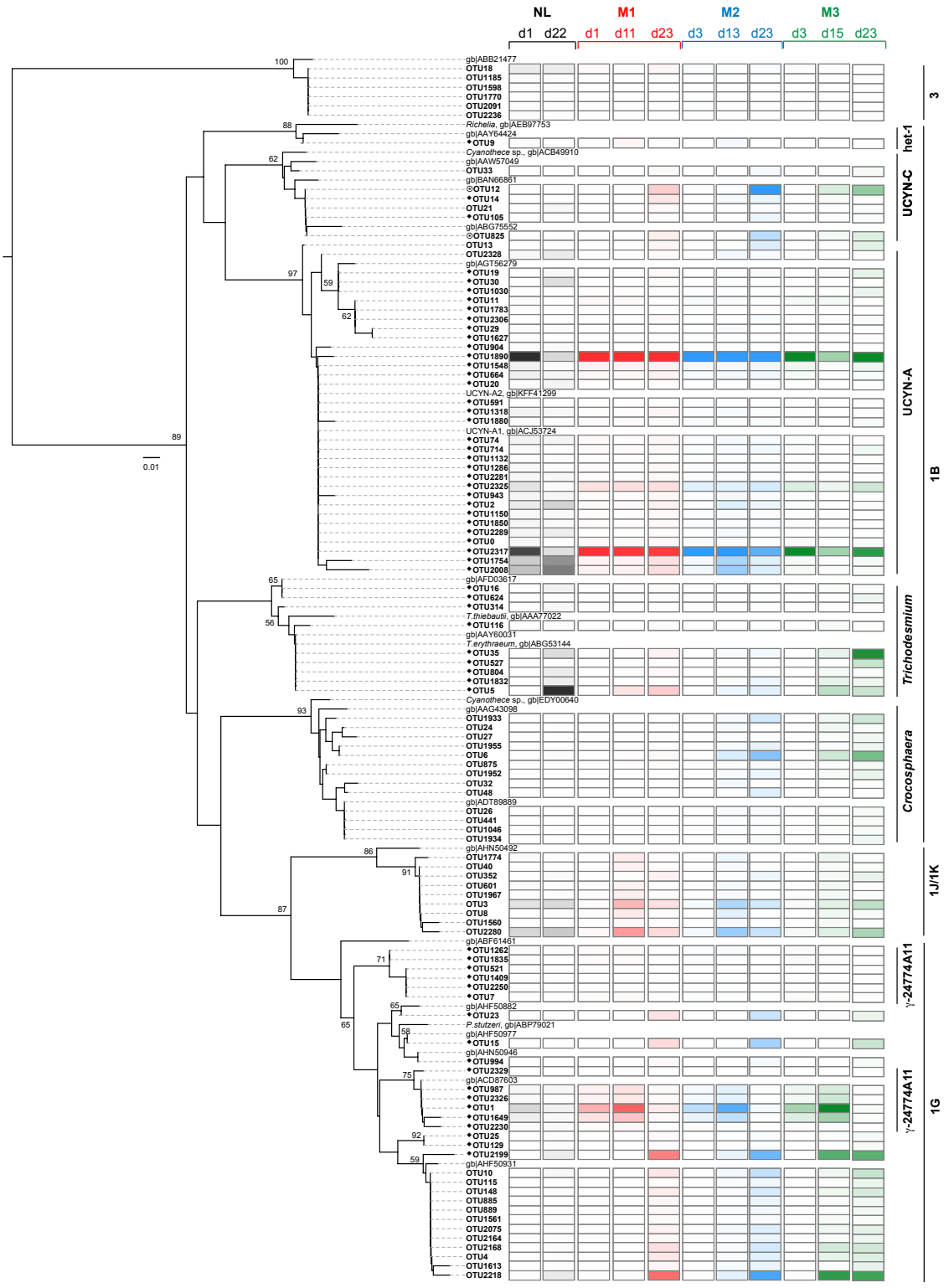
	Temp., Sal. & PO_4		Temp. & PO_4		Temp.		PO_4	
	R^2	R^2-cv	R^2	R^2-cv	R^2	R^2-cv	R^2	R^2-cv
UCYN-A1	0.85	0.76	0.84	0.76	0.70	0.59	0.73	0.63
UCYN-C	0.84	0.72	0.79	0.68	0.71	0.60	0.75	0.66
Het-3	0.77	0.60	0.70	0.57	0.55	0.39	0.70	0.60

Table 2. ***Results from multiple regression models predicting diazotroph abundances from environmental parameters.*** Abundances from both the Noumea lagoon and the VAHINE mesocosms experiments are included in the linear regression models. R^2-cv is the cross-validated fit of the model, and when similar to the R^2 value, indicates a high predictive ability for future datasets.

		Net growth rate / net mortality rate (d^{-1})								
	time	UCYN-A1	UCYN-A2	UCYN-B	UCYN-C	γ-24774A11	*Tricho.*	Het-1	Het-2	Het-3
M1	5	0.54 ± 0.05	0.26 ± 0.04	na	-0.04 ± 0.53	0.27 ± 0.13	0.15 ± 0.04	0.07 ± 0.05	-0.55 ± 0.18	na
	7	0.08 ± 0.07	-0.45 ± 0.04	na	na	-0.08 ± 0.11	-0.94 ± 0.05	0.03 ± 0.03	0.3 ± 0.17	na
	9	-0.81 ± 0.07	-0.48 ± 0.04	-0.41 ± 0.16	na	-0.24 ± 0.06	**1.46 ± 0.05**	**0.34 ± 0.04**	0.55 ± 0.03	na
	11	-0.47 ± 0.01	-0.02 ± 0.11	-0.65 ± 0.16	0.46 ± 0.23	0.19 ± 0.03	-1.12 ± 0.03	-0.15 ± 0.09	0.01 ± 0.04	na
	13	**0.73 ± 0.04**	-0.16 ± 0.15	**0.85 ± 0.03**	**1.23 ± 0.07**	-0.15 ± 0.22	0.39 ± 0.03	-0.23 ± 0.09	-0.21 ± 0.03	na
	15	-0.61 ± 0.04	0.18 ± 0.11	na	0.52 ± 0.03	**0.27 ± 0.22**	0.16 ± 0.04	-1.39 ± 0.04	-0.41 ± 0.02	0.21 ± 0.1
	18	-0.23 ± 0.08	-0.03 ± 0.08	na	0.06 ± 0.02	0.2 ± 0.02	-0.62 ± 0.08	0.7 ± 0.01	-0.03 ± 0.02	0.15 ± 0.16
	19	0.55 ± 0.08	0.31 ± 0.08	0.1 ± 0.11	0.29 ± 0.02	0.03	0.71 ± 0.08	0.7 ± 0.02	0.49 ± 0	-0.13 ± 0.22
	20	0.6 ± 0.05	**0.77 ± 0.04**	0.65 ± 0.13	-0.24 ± 0.07	0.2 ± 0.03	0.01 ± 0.01	-0.35 ± 0.02	-0.37 ± 0.02	**0.57 ± 0.16**
	23	0.47 ± 0.04	0.02 ± 0.06	0.55 ± 0.08	0.22 ± 0.07	-0.8 ± 0.07	0.74 ± 0.02	0.06 ± 0.01	0.01 ± 0.06	-0.1 ± 0.04
M2	5	--	--	--	--	--	--	--	--	--
	7	na	-1.63 ± 0.22	na	na	na	na	na	na	na
	9	-0.06 ± 0.03	**0.52 ± 0.26**	-0.36 ± 0.13	0.51 ± 0.24	-0.91 ± 0.04	0.03 ± 0.01	-0.53 ± 0.01	-0.06 ± 0.01	na
	11	-0.15 ± 0.03	0.29 ± 0.14	na	0.58 ± 0.24	0.31 ± 0.09	0.04 ± 0.02	**0.56 ± 0.07**	0.1 ± 0.01	na
	13	-0.47 ± 0.01	0.38 ± 0.1	na	0.79 ± 0.14	0.18 ± 0.09	-0.5 ± 0.06	-0.26 ± 0.07	0.44 ± 0.03	na
	15	-0.5 ± 0.07	0.33 ± 0.11	0.18 ± 0.07	**2.16 ± 0.07**	0.43 ± 0.06	**0.22 ± 0.05**	-1.29 ± 0.03	-0.63 ± 0.03	na
	18	-0.08 ± 0.11	na	0.3 ± 0.17	-0.27 ± 0.03	0.04 ± 0.07	0.14 ± 0.04	0.35 ± 0.02	-0.04 ± 0.01	**0.76 ± 0.15**
	19	--	--	--	--	--	--	--	--	--
	20	na	-0.46 ± 0.14	na	na	na	na	na	na	na
	23	**0.43 ± 0.05**	na	**0.36 ± 0.07**	0.05 ± 0.09	**1.07 ± 0.07**	0.09 ± 0.04	0.00 ± 0.07	0.04 ± 0.09	0.04 ± 0.09
M3	5	0.13 ± 0.02	**1.71 ± 0.32**	0.7 ± 0.48	0.96 ± 0.43	na	0.17 ± 0.03	0.64 ± 0.03	-0.35 ± 0.04	na
	7	0.29 ± 0.02	-1.64 ± 0.12	-0.42 ± 0.15	-0.38 ± 0.09	-0.39 ± 0.09	0.45 ± 0.06	-0.08 ± 0.05	0.54 ± 0.04	na
	9	-0.38 ± 0.03	-0.03 ± 0.1	-0.14 ± 0.3	0.13	0.29 ± 0.12	-2.16 ± 0.18	-0.63 ± 0.05	-0.72 ± 0.03	na
	11	-0.4 ± 0.03	0.18 ± 0.05	0.68 ± 0.27	0.1 ± 0.15	0.06 ± 0.14	1.2 ± 0.18	1.17 ± 0.06	0.84 ± 0.03	na
	13	-0.43 ± 0.07	0.16 ± 0.1	na	-0.02 ± 0.09	0.5 ± 0.09	0.8 ± 0.05	0.06 ± 0.06	-0.03 ± 0.03	na
	15	-0.48 ± 0.11	0.42 ± 0.11	na	**1.91 ± 0.02**	-0.95 ± 0.1	-0.4 ± 0.01	-1.16 ± 0.03	-0.66 ± 0.05	na
	18	-0.22 ± 0.21	-0.03 ± 0.07	0.21 ± 0.01	0.2 ± 0.04	-0.56 ± 0.11	-0.35 ± 0.01	-0.48 ± 0.01	-0.69 ± 0.07	0.06 ± 0.05
	19	-0.3 ± 0.19	-0.36 ± 0.14	-0.99 ± 0.06	-0.44 ± 0.06	**1.03 ± 0.1**	0.05 ± 0.01	-0.16 ± 0.11	2.24 ± 0.06	**1.09 ± 0.06**
	20	**0.55 ± 0.12**	1.25 ± 0.13	**1.38 ± 0.06**	0.28 ± 0.07	na	**1.55 ± 0.02**	**1.28 ± 0.11**	0.45 ± 0.07	-1.12 ± 0.05
	23	0.45 ± 0.14	-0.04 ± 0.08	0.07 ± 0.02	-0.02 ± 0.06	na	0.72 ± 0.03	-0.36 ± 0.02	-0.14 ± 0.08	0.77 ± 0.11

Table 3. ***Diazotroph net growth and mortality rates (d^{-1}) during VAHINE mesocosms experiment.*** Rates in bold are the maximum rates measured in each mesocosm. '- -' denotes periods where no rates could be calculated due to missing data, and 'na' denotes missing rates due to abundances being 'detected, not quantified' (DNQ) or undetected (UD). Standard error (±) reported for each growth rate is derived from qPCR measurements for replicate amplifications.

1244 **Figures**



1245

Figure 1. ***Maximum likelihood tree calculated using partial*** **nifH** ***amino acid sequences recovered from the Noumea Lagoon (NL) and mesocosms (M1, M2, and M3).*** Relative abundances of *nifH* reads associated with each operational taxonomic unit (OTU) are indicated for each sample by shaded boxes, with intense shading indicating high relative abundances, and light shading indicating low relative abundances. Trees were bootstrapped using 1000 replicate trees, and nodes with values >50 are displayed. Branch lengths were inferred using the JTT model, and the scale bar indicates the number of substitutions per site. OTUs that are targeted by qPCR assays used in this study are marked with a black diamond (◆), and two UCYN-C sequences that are likely to amplify are marked with a circle (⊙). *nifH* cluster designations according to the convention in Zehr et al. (2003) are notated at the right. d1 – day 1; d11 – day 11, d13 – day 13, d15 – day 15, d22 – day 22, d23 – day 23.

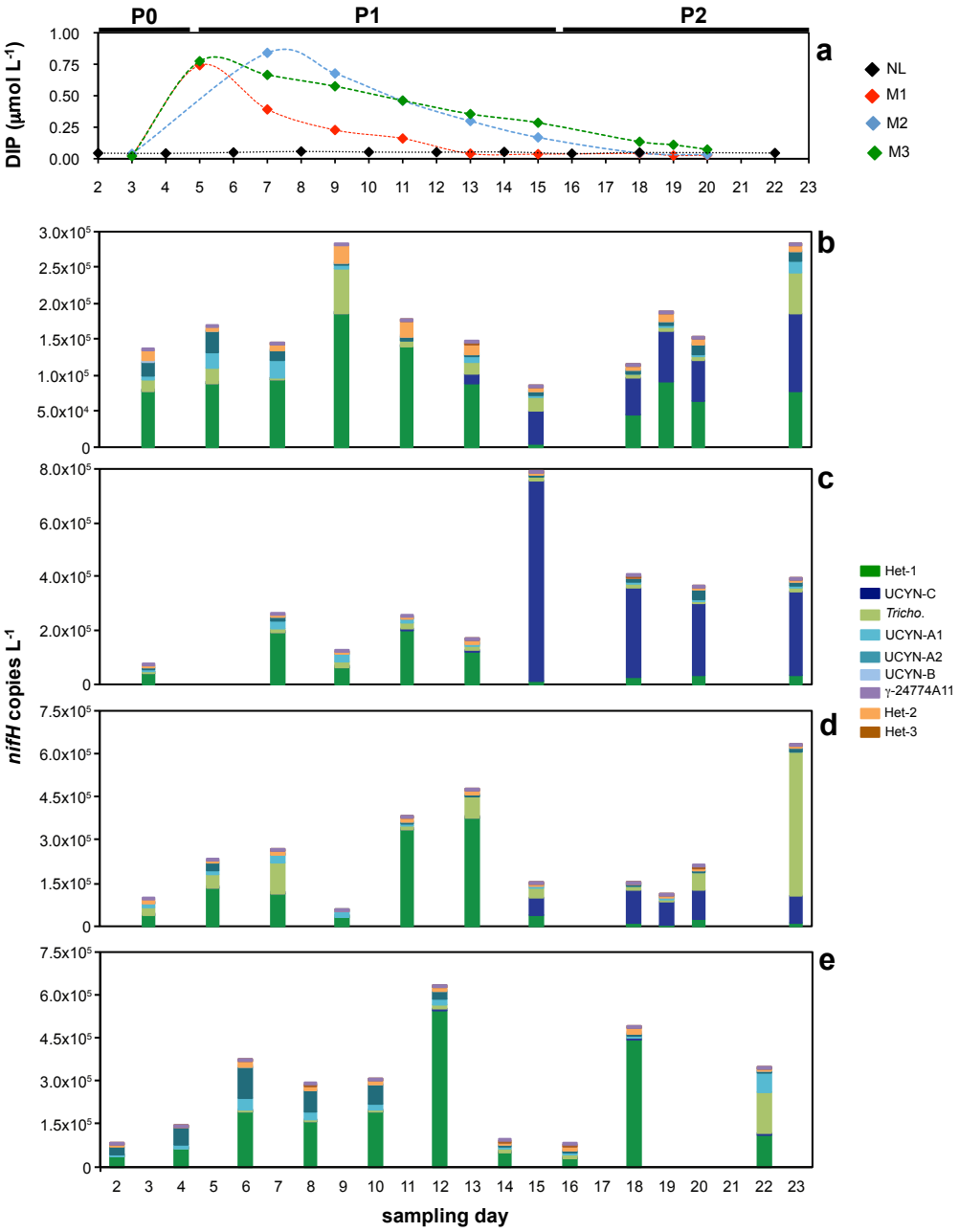


Figure 2. ***Abundances of targeted diazotrophs at the 6m depth during the VAHINE mesocosm experiment and in the Noumea lagoon (NL) during the experimental period.*** Het-1 is the most abundant diazotroph in P0 and P1, while UCYN-C abundances become abundant during P2 in all three mesocosms, M1 (b), M2 (c), and M3 (d). In the NL (e), increasing abundances of UCYN-C during the experimental period is not observed. DIP concentrations (a) decreased steadily following the spike at day 4 (data from in Bonnet et al., 2015a).

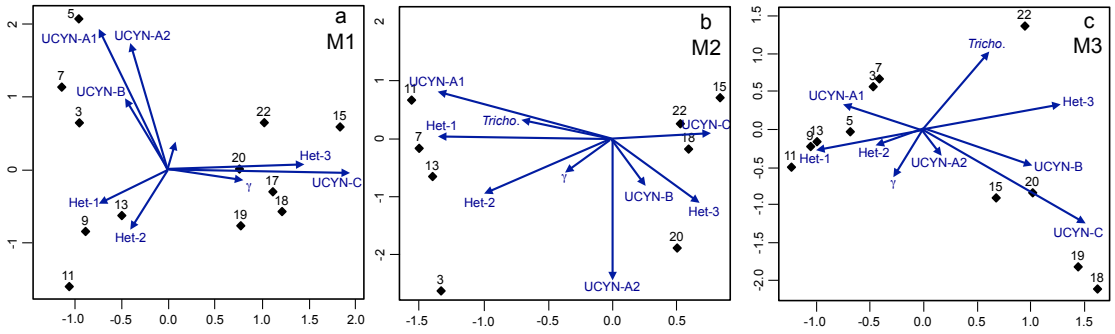


Figure 3. ***Correspondence analysis biplot of diazotroph abundances for each mesocosm.*** The horizontal axis is representative of time, evidenced by the progression of time points projected onto the x-axis. Variances covered by the two axes are 61%+18% in M1, 88%+5% in M2, and 56%+30% in M3.

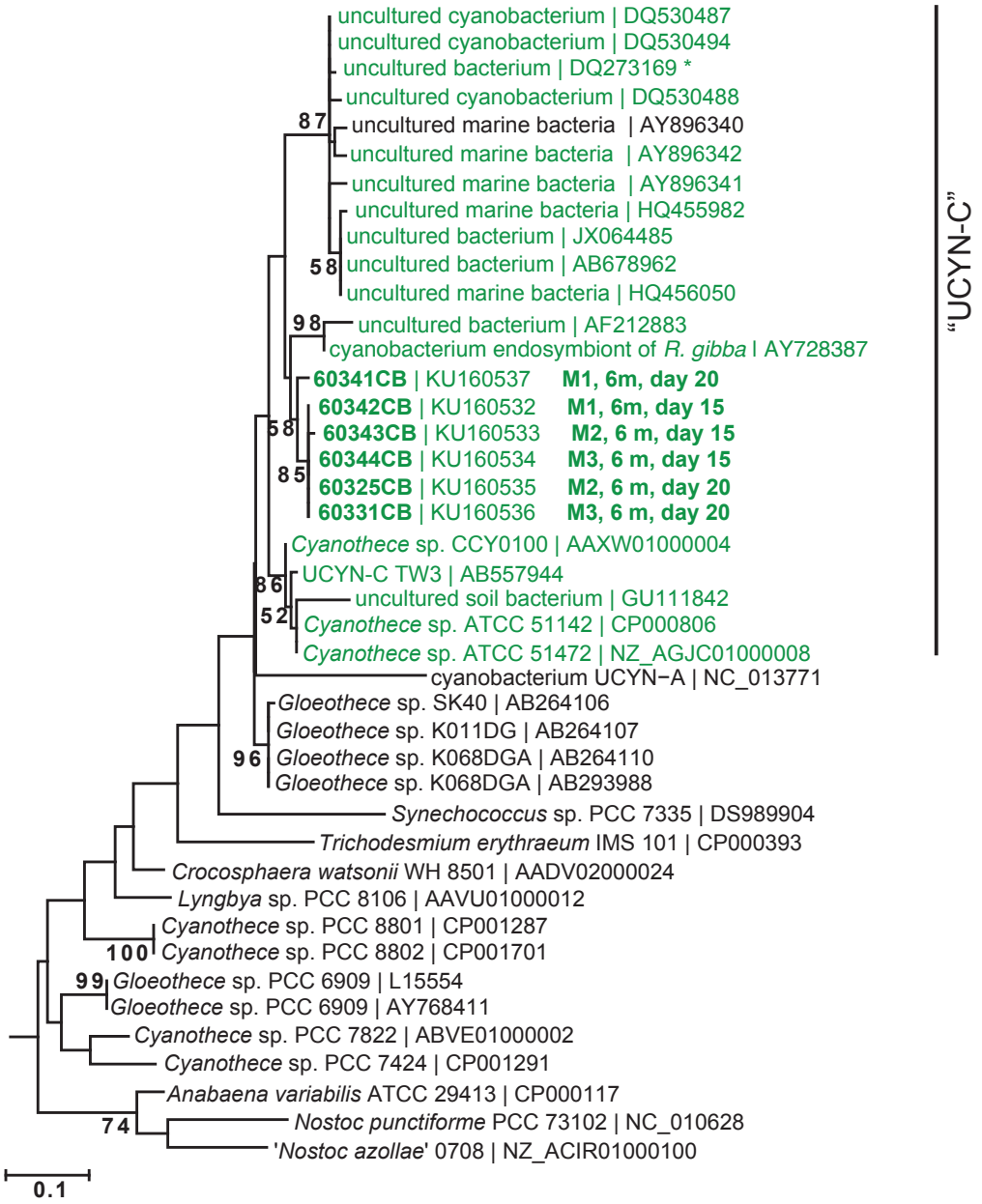


Figure 4. ***Maximum likelihood tree of Cyanothece-like diazotrophs based on partial nifH nucleotide sequences.*** Sequences that are targeted by the UCYN-C qPCR assay (Foster et al., 2007) with no greater than 2 mismatches in each primer and probe sequence are green, and the original sequence used for sequence design is marked with an asterisk (*). *Cyanothece*-like sequences recovered from the VAHINE mesocosms are bold. Bootstrap trees were calculated using 1000 replicate trees, and nodes with values

1287 >50 are displayed. Branch lengths were inferred using the Tamura-Nei model, and the
1288 scale bar indicates the number of substitutions per site.