UNIVERSITY OF CALIFORNIA

SANTA CRUZ

DISTRIBUTIONS AND ABUNDANCES OF SUBLINEAGES OF THE NITROGEN-FIXING CYANOBACTERIUM *CANDIDATUS ATELOCYANOBACTERIUM THALASSA* (UCYN-A) IN SOUTH PACIFIC WATERS

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF SCIENCE

in

OCEAN SCIENCES

by

Britt A. Henke

September 2017

The thesis of Britt A. Henke
is approved:

_____
Professor Jonathan Zehr,
Chair

_____
Professor Raphael Kudela

_____
Mathew Mills, PhD

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

ABSTRACT

DISTRIBUTIONS AND ABUNDANCES OF SUBLINEAGES OF THE NITROGEN-FIXING CYANOBACTERIUM *CANDIDATUS ATELOCYANOBACTERIUM THALASSA* (UCYN-A) IN SOUTH PACIFIC WATERS

by

BRITT A. HENKE

Nitrogen (N_2) fixation is a major source of nitrogen that supports primary production in the vast oligotrophic areas of the world's oceans. The Southwestern Pacific Ocean has been identified as a potential hotspot for N_2 fixation based on empirical and modeled data. In the southwest Lagoon of New Caledonia (SLNC), high abundances of the unicellular N_2-fixing cyanobacteria group A (UCYN-A), coupled with daytime N_2 fixation rates associated with the < 10 µm size fraction, suggest UCYN-A may be an important diazotroph (N_2-fixer) in this region. However, little is known about the seasonal variability and diversity of UCYN-A in this region. To assess this, surface waters from a 12 km transect from the mouth of the Dumbéa River to the Dumbéa Pass were sampled monthly between July, 2012 and March, 2014. UCYN-A abundances for two of the defined sublineages, UCYN-A1 and UCYN-A2, were quantified using qPCR targeting the *nifH* gene, and the *nifH*-based diversity of UCYN-A was characterized using oligotyping. Total UCYN-A abundances were dominated by the UCYN-A1 sublineage, peaked in September and October and could be predicted by a suite of nine environmental parameters. At the sublineage level,

UCYN-A1 abundances could be predicted based on lower temperatures (<23°C), nitrate concentrations, precipitation, and wind speed, while UCYN-A2 abundances could be predicted based on silicate, chlorophyll *a* concentrations, wind direction, precipitation and wind speed. Using UCYN-A *nifH* oligotyping, new UCYN-A sublineages were discovered, and similar environmental variables explained the relative abundances of sublineages and their associated oligotypes, with the notable exception of a UCYN-A2 oligotype (oligo43) which had relative abundance patterns distinct from the dominant UCYN-A2 oligotype (oligo3). The results support the emerging picture that UCYN-A is comprised of a diverse group of strains, with sublineages that likely have different ecological niches. By identifying environmental factors that influence the composition and abundance of UCYN-A sublineages, this study helps to explain global UCYN-A abundance patterns, and is important for understanding the significance of N_2 fixation at local and global scales.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Jonathan Zehr, for giving me the chance to do research in his lab. Jon welcomed me into his lab despite my lack of experience and I'm appreciative for his bravery and patience in converting me from a student of the humanities to a scholar of marine microbial ecology. Jon's larger perspective on both science and life make him a great mentor. Thank you Jon! It's an honor to work with you. And thank you to the rest of the Zehr Lab, especially the postdocs, Hanna Farnelid and Mari Munoz Marin, for teaching me so much.

I owe a debt of gratitude to Kendra Turk-Kubo for making the technical aspects of this work possible and enjoyable. From help with experimental design, to cruise packing and set-up, to sequence processing and oligotyping analysis, my Master's progress would have come to an abrupt stop if not for Kendra's guidance and encouragement. Thank you Kendra! Your passion for the science and devotion to the research continue to be an inspiration.

Mary Hogan was also fundamental in the smooth sailing of both field and lab work. As lab mother, Mary is both nurturing and hilariously sharp-tonged, depending on what is needed to support students and keep the lab functioning. I cannot recall all the times that Mary came to my aid, with technical, graduate, and life support. Thank you Mary, it's a pleasure to work with you.

I would like to thank my committee: Jonathan Zehr (chair), Raphael Kudela, and Mathew Mills, for thoughtful comments and edits on this document. I would also like to thank my honorary fourth committee member, Pete Raimondi, for his sage statistical advice. Without Pete's wisdom and patience, much of the statistical analysis in this thesis would not have been possible. Pete received question after question with the dedication only a true teacher could muster. Thank you Pete, I hope to continue learning from you!

This project would not have been possible without our collaborators, Sophie Bonnet of IRD/MIO (Marseille/ Nouméa) for sample collection and IRD-IMAGO-LAMA (Laboratoire des moyens analytiques) for core parameters. I also want to thank METEO FRANCE, for the meterological parameters.

I'm grateful to the Simons Collaboration on Ocean Sciences and Policy (SCOPE) for funding my research and giving me the opportunity to be part this important work.

These years would not have been the same without my wonderful cohort. I'm also grateful to my friend and former College of Marin Professor, Fernando Agudelo-Silva, for supporting me all along the way, including connecting me with a life-changing internship. Finally, I want to acknowledge my parents. Thank you for loving me unconditionally and empowering me to pursue my passions. I love you: Momma, Poppa, Robert, and Judy!

CHAPTER 1

Introduction

The oceans cover nearly three quarters of the earth's surface and are a fundamental link between land and atmospheric processes. Marine microorganisms play a critical role in ocean processes by recycling nutrients (e.g. nitrogen), fueling the food chain, and influencing greenhouse gas production and consumption. Marine phytoplankton (unicellular eukaryotic algae and cyanobacteria) are primary producers at the base of the food chain, and are responsible for approximately one-half of global primary production (Falkowski, 1994). Marine phytoplankton fix as much carbon dioxide (CO_2) as do terrestrial plants (Lalli and Parsons 1997; Field 1998). This Master's thesis focuses on N_2-fixing microorganisms that play a fundamental role in marine ecosystems by alleviating nitrogen limitation, and therefore fuel primary production and carbon sequestration.

Nitrogen is fundamental to life

Nitrogen (N) is a primary element in the biomolecules DNA and RNA (nucleic acids), as well as proteins (amino acids). Over much of the ocean, marine primary production, is limited by availability of nitrogen (N) (Dugdale and Goering 1967; Howarth 1988; Capone 2000), as the bioavailable forms of N that phytoplankton can use are often present in lower concentrations than other necessary nutrients relative to cellular needs (Moore et al. 2013). N-limitation is alleviated by diazotrophs, which catalyze the conversion of dinitrogen gas (N_2) to bioavailable forms of N. The

process mediated by diazotrophs is called N_2 fixation. If not for N_2 fixation, N losses to the atmosphere would dominate and life as we know it would become N-starved (Postgate, 1998).

The N cycle

The N cycle is comprised of various chemical reactions, many that are the result of microbial metabolisms (Figure 1). These reactions transform N from one form to another. The major bioavailable forms of N in most environments include ammonium (NH_4^+), nitrite (NO_2^-), nitrate (NO_3^-), and urea ($CO(NH_2)_2$). Thus, most bioavailable forms of N are inorganic, with the exception of urea. Fully reduced, ammonium requires the least energy expenditure to assimilate and therefore is readily incorporated by bacteria and phytoplankton (Zehr and Ward 2002). In the oceans, nitrate (oxidized form) is the most abundant form of N; however, concentrations remain low at the surface where primary production occurs. Dinitrogen gas (N_2) gas comprises 78% of our atmosphere and is by far the most abundant form of N, although it is unavailable to most organisms.

Biological N_2 fixation is the microbially-mediated reduction of N_2 gas to ammonia (NH_3) (ammonium at the pH of sea water) by diazotrophs.

$$N_2 + 8H^+ + 8e^- + 16ATP \rightarrow 2NH_3 + H_2 + 16ADP + 16P_i$$

Diazotrophs are a diverse group of organisms, including Bacteria from a variety of taxonomic groups, as well as some Archaea. N_2 fixation is catalyzed by the nitrogenase enzyme complex, which is composed of two metalloproteins, each with multi-subunits (Seefeldt et al., 2009). Component I (dinitrogenase) is a molybdenum-iron (MoFe) protein and component II (nitrogenase reductase) is the nitrogenase iron (Fe) protein (Seefeldt et al., 2009). N_2 fixation requires 8 electrons and 16 molecules of adenosine triphosphate (ATP) per molecule, making it an energetically expensive process (Postgate 1998; Zehr and Paerl 2008). For this reason, N_2 fixation is highly regulated.

Bioavailable N undergoes assimilation and remineralization (Figure 1). Newly fixed ammonium enters the food web via the death of diazotrophs, or alternatively, it may be leaked directly into the environment. The amount of N that is leaked into the environment by diazotrophs is controversial, but one study estimated that an average of 52% of fixed N_2 was rapidly released by *Trichodesmium* in the Gulf of Mexico (Mulholland et al., 2006). Equipped with enzymes that reduce oxidized N into ammonium, and subsequently, amino acids, both prokaryotic and Eukaryotic microorganism can assimilate simple inorganic N (nitrate, nitrite, ammonium), as well as urea into biomass or organic N. When microbes die, the majority of the remaining organic N is converted back to ammonium through remineralization by bacteria or fungi.

Nitrifying bacteria convert ammonium to nitrate in a process called nitrification (Ward et al., 2000) (Figure 1). These bacteria preform nitrification to glean energy and electrons from this oxidation. Nitrification occurs in two stages; bacteria such as the *Nitrosomonas* species convert ammonia to nitrite in ammonia oxidation and the *Nitrobacter* species of bacteria oxidize nitrite to nitrate in nitrite oxidation (Ward, 1996). In the ocean, nitrification occurs throughout the water column (Ward et al., 1989; Ward, 2005).

Microbes also mediate the reduction of nitrite and nitrate to N_2 gas, completing the N-cycle (Figure 1). With a reduction potential nearly equal to oxygen (O_2), oxidized N compounds are excellent electron acceptors for biological reactions. Reduction products include nitrite, in a process called dissimilatory nitrate reduction; N_2 gas and nitrous oxide (N_2O), in denitrification and; and ammonia, in dissimilarity nitrate reduction to ammonia (DNRA) (Gardner et al., 2006). The anaerobic oxidation of ammonium (anammox) is another microbially-mediated process that results in N_2 gas formation (Jetten et al., 2009) (Figure 2). Anammox is the conversion of equivalent amounts of nitrite and ammonia (ammonium at the pH of seawater) to N_2 gas and it is a major form of nitrogen conversion in the oceans (Kuypers et al. 2003; Thamdrup and Dalsgaard 2002). Denitrification broadly encompasses anaerobic respiration combined with oxidation of organic matter, and results in N_2 gas formation and the removal of bioavailable N. Denitrification occurs in anoxic or low oxygen environments when nitrate is present.

N cycle in the oceans

In the ocean light is attenuated with depth, limiting primary production to the euphotic zone which extends from the surface to a depth where light intensity is 1% of that at the surface. In this zone, bioavailable N is used up by microbes in the formation of particulate organic nitrogen (PON, Figure 2). Microbial production is categorized as either new (mostly nitrate-based) or regenerated (mostly ammonium-based) (Dugdale and Goering, 1967). N_2 fixation, which supports new ammonium-based production, is an exception to this categorization. Surface-ocean microbial growth is largely supported by regenerated production (Dugdale and Goering, 1967), as PON is consumed, digested, assimilated, and finally released and/or remineralized back to bioavailable N during grazing and decomposition, only to be rapidly assimilated to form new PON. Defined as being able to pass through a filter (< 0.22 μm), dissolved organic nitrogen (DON) is also produced in the euphotic zone (Jiao et al., 2010). DON is exported to depth via mixing and downwelling where it undergoes remineralization. Additionally, a fraction of the PON sinks and is remineralized at depth. Due to accumulation of remineralized N, the deep ocean is a reservoir of bioavailable fixed inorganic N.

In marine systems, the predominate sources of new bioavailable N are riverine and atmospheric deposition and N_2 fixation. Marine N_2 fixation can be a significant source of N in oligotrophic waters, including the oceanic gyres, which are 40% of the

planet's surface area (Karl and Church, 2014). In the tropical Pacific and Atlantic Ocean, N_2 fixation accounts for approximately 36-50% of new production (Karl et al. 1997; Carpenter et al., 1999; Dore et al., 2002). It is estimated that N_2 fixation is equivalent to 50-180% of the flux of nitrate into the euphotic zone (Karl et al., 1997b; Capone et al., 2005). Consequently, in much of the world's oceans, N_2 fixation partially dictates oceanic primary production.

Although geochemists often posit that phosphorus (P) is the limiting nutrient because diazotrophs can counteract N-depletion, biologists generally posit that N limitation ultimately controls oceanic primary production (Tyrrell, 1999). Timescales are central to this debate, and P may ultimately be limiting as its availability is tied to the gradual weathering of rock. Nevertheless, experiments show that N addition alone stimulates productivity and phytoplankton biomass in oligotrophic waters (Graziano et al., 1996; Moutin et al., 2007; Moore et al., 2008) and iron stimulates growth in high-nitrate low-chlorophyll waters (Martin and Fitzwater, 1988; Boyd et al., 2007). Some marine organisms can use non-P lipids in response to P scarcity (Mooy et al., 2009). Although there are strategies for metabolic maintenance at low N levels (Grzymski and Dussaq, 2011), cellular N requirements are more restricted. Additionally, at 7% of marine microbe cellular mass, N accounts for 7 times more cellular mass in marine microbes as compared to P (Geider and La Roche, 2002).

Redfield et al. (1934) was the first to document the composition of carbon (C), N, P in marine phytoplankton, giving rise to the Redfield ratio, which describes the molar ratio of C:N:P as 106:16:1. This proportionality is also reflected in the composition of the deep ocean organic matter NO_3^-:PO_4^{3-} (Redfield 1958). Assuming Redfield proportions, it is possible to identify elemental imbalances and therefore areas of N_2 fixation and denitrification. Accordingly, negative (N-deficit) divergences from the canonical Redfield ratio are sites of N loss (denitrification or anammox) and positive (N-excess) divergences, are sites of N-gain (N_2 fixation) (Sarmiento and Gruber 1997; Michaels et al. 1996). As much of the ocean is in N-deficit, the assumption that follows from canonical Redfield ratio is that much of the ocean is N-limited and would select for N_2 fixation (Mahaffey et al., 2005; Codispoti, 2007). Although still helpful on a first-order basis, it is now known that Redfield proportions vary according to phytoplankton taxa (Quigg et al., 2003) and growth rate (Geider and Roche, 2016), complicating the identification of elemental imbalances and corresponding sites of expected denitrification and N_2 fixation (Mills and Arrigo, 2010).

The balance of modern oceanic N budget is controversial, with some claiming that global oceanic denitrification and anammox rates exceed global N_2 fixation rates (Codispoti et al. 2001; Galloway et al. 2004). This has caused some to argue that global N_2 fixation rates are underestimated (Mahaffey et al., 2005; Codispoti 2007). More recent analyses suggest that N-sinks and N-sources are in equilibrium (Gruber,

2016). Understanding how these processes relate is important for deducing past atmospheric carbon dioxide levels as well as forecasting how current anthropogenic perturbations in the C to N balance will affect biogeochemical cycling (Galloway et al., 2004).

N_2 fixation and diazotroph distributions

It is difficult to quantify the total contribution of marine N_2 fixation to the global N budget (Mahaffey et al., 2005). A major aspect of this challenge is the incomplete understanding of the factors that regulate N_2 fixation for predicting spatial and temporal distributions of N_2-fixers and their activities. Enhanced understanding of nitrogenase expression will help explain global patterns of diazotroph distributions and abundances, and therefore aid regional and global N_2 fixation estimates.

The availability of trace metals such as iron (Fe) has also been suggested as a principal factor controlling N_2 fixation rates. Some argue that the decoupling between N-sinks and N-sources, which has resulted in global denitrification (and anammox) rates exceeding global N_2 fixation rates could have resulted from the reduced availability of trace-medals required for N_2 fixation (Moore and Doney, 2007).

Temperature is another control on N_2 fixation (Luo et al., 2014) and especially for cyanobacterial diazotrophs (Church et al. 2008; Moisander et al. 2010; Cabello et al.

2015). Well-known diazotrophs (*Trichodesmium* and diatom symbionts), once thought to be primarily responsible for marine N_2 fixation (Capone 1997; Villareal and Carpenter 1989; Villareal 1990), flourish in tropical regions. But it is now known that other cyanobacterial diazotrophs (including UCYN-A, the focus of this research) have been shown to have a lower temperature range (19°C - 24°C) (Church et al. 2008; Langlois, Hümmer, and LaRoche 2008), and have been documented to be actively fixing N_2 in cold waters, including the Bering Sea (Shiozaki et al., 2017), as well as the Danish Strait (Bentzon-Tilia et al. 2015). Heterotrophic diazotrophs are also present and active at lower temperatures (Farnelid et al. 2013; Bentzon-Tilia et al. 2015). Thus, although still helpful in predicting diazotroph distributions (Capone 1997; Breitbarth, Oschlies, and La Roche 2007) and consequently N_2 fixation rates (Mahaffey, Michaels, and Capone 2005; Goebel et al. 2010), the relationship between diazotroph abundances and temperature is more specific than once thought (Moisander et al., 2010).

Irradiance is yet another factor that affects diazotrophic activity and has been questioned by recent findings. It was generally believed that the majority of marine N_2 fixation was due to cyanobacterial diazotrophs, which obtain energy through photosynthesis, restricting their distributions to sunlit surface waters. Heterotrophic diazotrophs have been shown to be actively fixing N_2 below the chemocline (200 m) in Baltic Sea waters, which is below the depth to which light penetrates (Farnelid et al., 2013).

The challenge of estimating total global N_2 fixation is complicated by the dynamic nature of diazotroph activity, lack of data, and issues extrapolating basin-scale or global-scale rates from existing data. Measured N_2 fixation can be both spatially (Fong et al., 2008) and temporally (Church et al., 2009; Karl et al., 2012) patchy. As the factors controlling N_2 fixation are still not completely understood, predicting areas of high N_2 fixation can be challenging. The lack of diazotroph abundance data and N_2 fixation rates from vast areas of the ocean, including the southwestern Pacific, South Atlantic, and Indian Oceans (Luo et al. 2012 and 2014) also hinders global N_2 fixation estimates. Finally, the tendency to study diazotroph activity where N_2 fixation rates are high, such as during summer export pulses in the North Pacific Gyre (Karl et al. 2012), further complicates efforts to extrapolate rates across basins and globally. The distribution of N_2-fixing organisms and rates of N_2 fixation are non-random and can be predicted based on ecological characteristics, but the nuances associated with this relationship are not fully understood.

Studying diazotrophs

The advent of molecular analysis in marine microbial ecology has revealed the tremendous diversity of N_2-fixing bacteria. N_2 fixation is catalyzed by nitrogenase which is encoded by the nifHDK operon. Encoding the (Fe)-protein of the nitrogenase enzyme, the *nifH* gene has been widely used to assess the diversity and abundance of diazotrophs (Zehr et al. 1998; Zehr et al. 2003). The *nifH* gene is used

as a phylogenetic marker for the functional group of N_2-fixing microorganisms because the *nifH* and 16S rRNA genes largely correspond (Zehr et al. 2003). Using this information, four clusters (I-IV) of N_2-fixing microorganisms have been defined (Chien et al., 1996). Cluster I includes Cyanobacteria and Proteobacteria. Cluster II includes archaea and microorganism with alternative iron-only nitrogenases. Cluster III includes anaerobic bacteria such as Firmicutes, the Chlorobi, and δ-Proteobacteria. Cluster IV includes organisms with homologues of *nifH* presumably used for functions other than N_2 fixation, although recent studies have called this assumption into question (Zheng et al., 2016).

As discussed above, prior to sequencing the *nifH* gene, the dogma was that *Trichodesmium* and diatom symbionts were largely responsible for the majority of marine N_2 fixation (Capone 1997; Villareal and Carpenter 1989; Villareal 1990). The use of nested primers for polymerase chain reaction (PCR) however, has allowed the identification of low abundance *nifH* genes (Zehr and Turner 2001) in marine systems and revealed a diverse range of novel marine diazotrophs (Zehr et al., 1998; Zehr et al., 2000). Included in this group of previously unrecognized marine N_2-fixers are the unicellular cyanobacteria group A (UCYN-A) (Zehr et al. 2008).

The subject of this study, UCYN-A is now recognized to contribute significant amounts of fixed N_2 to marine systems at certain times and places (Montoya, Holl, and Zehr 2004; Goebel et al. 2007; Church et al., 2009; Turk et al., 2011; Martínez-

Pérez et al., 2016). UCYN-A has at times been observed at equal or greater abundances than other cyanobacterial diazotrophs in tropical or subtropical oceans (Foster et al., 2007; Kong et al., 2011). Further, with a broader distribution than *Trichodesmium*, UCYN-A has expanded the traditional N_2 fixation domain (Short and Zehr 2007; Moisander et al. 2010; Bentzon-Tilia et al., 2015; Messer et al. 2016). For these reasons, UCYN-A has the potential to contribute substantial amounts of N to marine ecosystems.

BIBLIOGRAPHY

Bentzon-Tilia, M., Traving, S. J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J. L., Markager, S., et al. (2015). Significant N2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. *ISME J.* 9, 273–285. doi:10.1038/ismej.2014.119.

Boyd, P. W., Jickells, T., Law, C. S., Blain, S., Boyle, E. a, Buesseler, K. O., et al. (2007). Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science* 315, 612–617. doi:10.1126/science.1131669.

Breitbarth, E., Oschlies, A., and Laroche, J. (2007). Physiological constraints on the global distribution of Trichodesmium – effect of temperature on diazotrophy. 53–61.

Cabello, A. M., Cornejo-Castillo, F. M., Raho, N., Blasco, D., Vidal, M., Audic, S., et al. (2015). Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. *ISME J.* doi:10.1038/ismej.2015.147.

Capone, D. G. (1997). Trichodesmium, a Globally Significant Marine Cyanobacterium. *Science.* 276, 1221–1229. doi:10.1126/science.276.5316.1221.

Capone, D. G. (2000). The marine microbial nitrogen cycle. *Microb. Ecol. Ocean.*, 455–493.

Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., et al. (2005). Nitrogen fixation by Trichodesmium spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochem. Cycles* 19, 1–17. doi:10.1029/2004GB002331.

Carpenter, E. J., Montoya, J. P., Burns, J., Mulholland, M. R., Subramaniam, A., and Capone, D. G. (1999). Extensive bloom of a N2-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean. *Mar. Ecol. Prog. Ser.* 185, 273–283. doi:10.3354/meps185273.

Chien, Y., Zinder, S. H., Microbiology, S., Hall, W., and York, N. (1996). Cloning, Functional Organization, Transcript Studies, and Phylogenetic Analysis of the Complete Nitrogenase Structural Genes (nifHDK2) and Associated Genes in the Archaeon Methanosarcina barkeri 227. 178, 143–148.

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

Church, M. J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., and Karl, D. M. (2009). Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochem. Cycles* 23. doi:10.1029/2008GB003418.

Codispoti, L. a. (2007). An oceanic fixed nitrogen sink exceeding 400 Tg N a−1 vs the concept of homeostasis in the fixed-nitrogen inventory. *Biogeosciences*

Discuss. 3, 1203–1246. doi:10.5194/bgd-3-1203-2006.

Codispoti, L. A., Brandes, J. A., Christensen, J. P., Devol, A. H., Naqvi, S. W. A., Paerl, H. W., et al. (2001). The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Sci. Mar.* 65, 85–105. doi:10.3989/scimar.2001.65s285.

Dore, J. E., Brum, J. R., Tupas, L., and Karl, D. M. (2002). Seasonal and interannual variability in sources of nitrogen supporting export in the oligotrophic subtropical North Pacific Ocean . *Limnol. Ocean.* 47, 1595–1607.

Dugdale, R. C., and Goering, J. J. (1967). Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnol. Oceanogr.* 12, 196–206. doi:10.4319/lo.1967.12.2.0196.

Falkowski, P. G. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles *. 235–258.

Farnelid, H., Bentzon-Tilia, M., Andersson, A. F., Bertilsson, S., Jost, G., Labrenz, M., et al. (2013). Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. *ISME J.* 7, 1413–23. doi:10.1038/ismej.2013.26.

Field, C. B. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science.* 281, 237–240. doi:10.1126/science.281.5374.237.

Fong, A. A., Karl, D. M., Lukas, R., Letelier, R. M., Zehr, J. P., and Church, M. J. (2008). Nitrogen fixation in an anticyclonic eddy in the oligotrophic North Pacific Ocean. 663–676. doi:10.1038/ismej.2008.22.

Foster, R. A., Mahaffey, C., Carpenter, E. J., and Angeles, L. (2007). Influence of the Amazon River plume on distributions of free-living and symbiotic cyanobacteria in the western tropical north Atlantic Ocean. 52, 517–532.

Galloway, J. N., Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., et al. (2004). *Nitrogen cycles: Past, present, and future*. doi:10.1007/s10533-004-0370-0.

Gardner, W. S., Mccarthy, M. J., An, S., Sobolev, D., Sell, K. S., Brock, D., et al. (2006). Nitrogen fixation and dissimilatory nitrate reduction to ammonium (DNRA) support nitrogen dynamics in Texas estuaries. 51, 558–568.

Geider, R. J., and La Roche, J. (2002). Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. 1–17. doi:10.1017?S0967026201003456.

Geider, R., and Roche, J. La (2016). Redfield revisited : variability of C : N : P in marine microalgae and its biochemical basis Redfield revisited : variability of C : N : P in marine microalgae and its biochemical basis. 262, 1–17.

doi:10.1017/S0967026201003456.

Goebel, N. L., Edwards, C. A., Church, M. J., and Zehr, J. P. (2007). Modeled contributions of three types of diazotrophs to nitrogen fixation at Station ALOHA. 606–619. doi:10.1038/ismej.2007.80.

Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., et al. (2010). Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N2 fixation in the tropical Atlantic Ocean. *Environ. Microbiol.* 12, 3272–3289. doi:10.1111/j.1462-2920.2010.02303.x.

Graziano, L. M., Geider, R. J., Li, W. K. W., and Olaizola, M. (1996). Nitrogen limitation of North Atlantic phytoplankton: Analysis of physiological condition in nutrient enrichment experiments. *Aquat. Microb. Ecol.* 11, 53–64. doi:10.3354/ame011053.

Gruber, N. (2016). Elusive marine nitrogen fixation. 113, 4246–4248. doi:10.1073/pnas.1603646113.

Gruber, N., and Sarmiento, J. L. (1997). and With Some Recent Suggestions That , Pelagic Nitrogen Fixat , Ion May Be. *Expedition* 11, 235–266.

Grzymski, J. J., and Dussaq, A. M. (2011). The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J.* 6, 71–80. doi:10.1038/ismej.2011.72.

Howarth, R. W. (1988). ECOSYSTEMS. 89–110.

Jetten, M. S. M., Niftrik, L. van, Strous, M., Kartal, B., Keltjens, J. T., and Op den Camp, H. J. M. (2009). Biochemistry and molecular biology of anammox bacteria. *Crit. Rev. Biochem. Mol. Biol.* 44, 65–84.

Jiao, N., Herndl, G. J., Hansell, D. a, Benner, R., Kattner, G., Wilhelm, S. W., et al. (2010). Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* 8, 593–599. doi:10.1038/nrmicro2386.

Karl, D., Church, M. J., Dore, J. E., Letelier, R. M., and Mahaffey, C. (2012). Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Pnas* 109, 1842–1849. doi:10.1073/pnas.1120312109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1120312109.

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

Karl, D. M., and Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* 12, 699–713. doi:10.1038/nrmicro3333.

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

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

Kuypers, M. M. M., Sliekers, A. O., Lavik, G., Schmid, M., Jørgensen, B. B., Kuenen, J. G., et al. (2003). Anaerobic Ammonium Oxidation by Anammox Bacteria in the Black Sea. *Nature* 422, 608–611. doi:10.1038/nature01526.1.

Lalli, C., and Parsons, T. R. (1997). *Biological oceanography: an introduction*. Butterworth-Heinemann.

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

Luo, Y. W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., et al. (2012). 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.

Luo, Y. W., Lima, I. D., Karl, D. M., Deutsch, C. A., and Doney, S. C. (2014). Data-based assessment of environmental controls on global marine nitrogen fixation. *Biogeosciences* 11, 691–708. doi:10.5194/bg-11-691-2014.

Mahaffey, C., Michaels, A. F., and Capone, D. G. (2005). The conundrum of marine N2 fixation. *Am. J. Sci.* 305, 546–595. doi:10.2475/ajs.305.6-8.546.

Martin, J. H., and Fitzwater, S. E. (1988). Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. *Nature* 331, 341–343.

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., et al. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat. Microbiol.* 1, 16163. doi:10.1038/nmicrobiol.2016.163.

Messer, L., Mahaffey, C., M Robinson, C., Jeffries, T. C., Baker, K. G., Bibiloni Isaksson, J., et al. (2016). High levels of heterogeneity in diazotroph diversity and activity within a putative hotspot for marine nitrogen fixation. *ISME J.* 10, 1–15. doi:10.1038/ismej.2015.205.

Michaels, a. F., Olson, D., Sarmiento, J. L., Ammerman, J. W., Fanning, K., Jahnke, R., et al. (1996). Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. *Biogeochemistry* 35, 181–226. doi:10.1007/BF02179827.

Mills, M. M., and Arrigo, K. R. (2010). Magnitude of oceanic nitrogen fixation influenced by the nutrient uptake ratio of phytoplankton. *Nat. Geosci.* 3, 412–

416. doi:10.1038/ngeo856.

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., et al. (2010). Unicellular cyanobacterial distributions broaden the oceanic N2 fixation domain. *Science* 327, 1512–1514. doi:10.1126/science.1185468.

Montoya, J. P., Holl, C. M., and Zehr, J. P. (2004). High rates of N2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. 430, 1027–1031. doi:10.1038/nature02744.1.

Moore, C. M., Mills, M. M., Langlois, R., Milne, A., Achterberg, E. P., La Roche, J., et al. (2008). Relative influence of nitrogen and phosphorus availability on phytoplankton physiology and productivity in the oligotrophic sub-tropical North Atlantic Ocean. *Limnol. Oceanogr.* 53, 291–305. doi:10.4319/lo.2008.53.1.0291.

Moore, J. K., and Doney, S. C. (2007). Iron availability limits the ocean nitrogen inventory stabilizing feedbacks between marine denitrification and nitrogen fixation. *Global Biogeochem. Cycles* 21, 1–12. doi:10.1029/2006GB002762.

Mooy, B. A. S. Van, Fredricks, H. F., Pedler, B. E., Dyhrman, S. T., Karl, D. M., Lomas, M. W., et al. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. 458, 69–72. doi:10.1038/nature07659.

Moutin, T., Karl, D. M., Duhamel, S., Rimmelin, P., Raimbault, P., Van Mooy, B. A. S., et al. (2007). Phosphate availability and the ultimate control of new nitrogen input by nitrogen fixation in the tropical Pacific Ocean. *Biogeosciences Discuss.* 4, 2407–2440. doi:10.5194/bgd-4-2407-2007.

Mulholland, M. R., Bernhardt, P. W., Heil, C. a., Bronk, D. a., and Neil, J. M. O. (2006). Nitrogen fixation and release of fixed nitrogen by Trichodesmium spp. in the Gulf of Mexico. *Limnol. Oceanogr.* 51, 1762–1776. doi:10.4319/lo.2006.51.4.1762.

Postgate, J. (1998). *Nitrogen Fixation*. 3rd editio. Cambridge: Cambridge University Press.

Quigg, A., Finkel, Z. V, Irwin, A. J., Rosenthal, Y., Ho, T. Y., Reinfelder, J. R., et al. (2003). Plastid inheritance of elemental stoichiometry in phytoplankton and its imprint on the geological record. *Nature* 425, 291–294.

Redfield, A. (1958). The biological control of chemical factors in the enviroment. *Am. Sci.* 46, 205–221.

Redfield, A. C. (1934). On the proportions of organic derivatives in a sea water and their relation to the composition of plankton. *James Johnstone Meml. Vol.*, 177–192.

Seefeldt, L. C., Hoffman, B. M., and Dean, D. R. (2009). Mechanism of Mo-dependent nitrogenase. *Annu. Rev. Biochem.* 78, 701–722.

Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., et al. (2017). Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. *Global Biogeochem. Cycles*. doi:10.1002/2017GB005681.

Short, S. M., and Zehr, J. P. (2007). Nitrogenase gene expression in the Chesapeake Bay Estuary. *Environ. Microbiol.* 9, 1591–1596. doi:10.1111/j.1462-2920.2007.01258.x.

Thamdrup, B., and Dalsgaard, T. (2002). Production of N2 through Anaerobic Ammonium Oxidation Coupled to Nitrate Reduction in Marine Sediments. *Appl. Environ. Microbiol.* 68, 1312–1318. doi:10.1128/AEM.68.3.1312.

Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B., and Zehr, J. P. (2017). Distinct ecological niches of marine symbiotic N2-fixing cyanobacterium Candidatus Atelocyanobacterium thalassa sublineages. *J. Phycol.* 53, 451–461. doi:10.1111/jpy.12505.

Turk, K. A., Rees, A. P., Zehr, J. P., Pereira, N., Swift, P., Shelley, R., et al. (2011). Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. *ISME J.* 5, 1201–1212. doi:10.1038/ismej.2010.205.

Tyrrell, T. (1999). The relative influences of nitrogen and phosohorus on oceanic primary production. *Nature* 400, 525–531. doi:10.1038/22941.

Villareal, T. A. (1990). Laboratory culture and preliminary characterization of the nitrogen-fixing *Rhizosolenia-Richelia* symbiosis. *Mar. Ecol.* 11, 117–132. doi:10.1111/j.1439-0485.1990.tb00233.x.

Villareal, T. A., and Carpenter, E. J. (1989). Nitrogen fixation, suspension characteristics, and chemical composition of Rhizosolenia mats in the central North Pacific gyre. *Biol. Oceanogr.* 6, 327–346. doi:10.1080/01965581.1988.10749535.

Ward, B. B. (1996). Nitrification and denitrification: probing the nitrogen cycle in aquatic environments. *Microb. Ecol.* 32, 247–261.

Ward, B. B. (2005). Temporal variability in nitrification rates and related biogeochemical factors in Monterey Bay, California, USA. *Mar. Ecol. Prog. Ser.* 292, 97–109. doi:10.3354/meps292097.

Ward, B. B., Kilpatrick, K. a., Renger, E. H., and Eppley, R. W. (1989). Biological nitrogen cycling in the nitracline. *Limnol. Oceanogr.* 34, 493–513.

Ward, B. B., Martino, D. P., Diaz, M. C., and Al, W. E. T. (2000). Analysis of Ammonia-Oxidizing Bacteria from Hypersaline Mono Lake , California , on the Basis of 16S rRNA Sequences. 66, 2873–2881.

Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., et al. (2008). Globally Distributed Uncultivated. 322, 1110–1113.

Zehr, J. P., Carpenter, E. J., and Villareal, T. A. (2000). New perspectives on nitrogen-fixing microorganisms in tropical and subtropical oceans. *Trends Microbiol.* 8, 68–73. doi:10.1016/S0966-842X(99)01670-4.

Zehr, J. P., and Hiorns, W. D. (1998). "Molecular Approaches to Studies of the Activities of Marine Organisms," in *Molecular Approaches to the Study of the Ocean*, ed. K. E. Cooksey (Dordrecht: Springer Netherlands), 91–111. doi:10.1007/978-94-011-4928-0_3.

Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Env. Microbiol* 5, 539–554. doi:10.1046/j.1462-2920.2003.00451.x.

Zehr, J. P., Mellon, M. T., Zani, S., and York, N. (1998). New Nitrogen-Fixing Microorganisms Detected in Oligotrophic Oceans by Amplification of Nitrogenase (nifH) Genes. 64, 3444–3450.

Zehr, J. P., and Paerl, H. W. (2008). "Molecular Ecological Aspects of Nitrogen Fixation in the Marine Environment," in *Microbial Ecology of the Oceans*, ed. D. L. Kirchman (New Jersey: Wiley-Blackwell), 481–525. doi:10.1002/9780470281840.ch13.

Zehr, J. P., and Ward, B. B. (2002). Nitrogen Cycling in the Ocean: New Perspectives on Processes and Paradigms MINIREVIEW Nitrogen Cycling in the Ocean: New Perspectives on Processes and Paradigms. *Appl. Environ. Microbiol.* 68, 1015–1024. doi:10.1128/AEM.68.3.1015.

Zheng, H., Dietrich, C., Radek, R., and Brune, A. (2016). Endomicrobium proavitum, the first isolate of Endomicrobia class. nov. (phylum Elusimicrobia) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. 18, 191–204. doi:10.1111/1462-2920.12960.

CHAPTER 2

DISTRIBUTIONS AND ABUNDANCES OF SUBLINEAGES OF THE NITROGEN-FIXING CYANOBACTERIUM *CANDIDATUS ATELOCYANOBACTERIUM THALASSA* (UCYN-A) IN SOUTH PACIFIC WATERS

1.0 Introduction

Biological nitrogen fixation (BNF) is the conversion of dinitrogen (N_2) gas into bioavailable nitrogen (N) by certain microorganisms (diazotrophs). BNF supports carbon and nitrogen losses from the surface ocean through vertical export (Karl et al. 2012) by providing a significant source of new N in oligotrophic regions where primary production is N limited (Karl et al. 1997). The Western Tropical South Pacific (WTSP) is a global hot spot of N_2 fixation with rates averaging 570 µmol N m^{-2} d^{-1} in a vast area covering 5 x 10^{12} m^2 (Bonnet et al., 2017). Oligotrophic coastal regions from the WTSP can also be N-limited (Torréton et al., 2010), but the diazotroph communities in these areas are far less studied than in open ocean regions.

The 23,400 km^2 Southwestern lagoon of New Caledonia (SLNC) is located off the southwestern coast of New Caledonia in the Coral Sea of WTSP (Figure 3). Bounded by one of the world's largest barrier reefs, the SLNC is a tropical low-nutrient low-chlorophyll (LNLC) system. It has average depth of 17.5 m, along with many deeper canyons (~60 m) and reef islands. The lagoon runs roughly 100 km from north to south and is 5 km wide in the north and 40 km wide in the south (Ouillon et al., 2010).

The SLNC is characterized by seasonal weather patterns. The southeasterly trade winds dominate year-round but are most consistent from October – May. Southeasterly trade winds (110°) drive the oligotrophic South Equatorial Current (SEC) into the lagoon between the reef and Ouen Island at the southern end, exiting through one of two main passes north of this entrance, with Dumbéa Pass being the more northerly (Ouillon et al., 2010). Cooler sea and air temperatures prevail from June - September. The transition to warmer temperatures can cause thermal stratification during the dry warm months of October – December. Haline stratification is most likely during the warm-wet months of January – May. With a drainage of 233 km^2, the Dumbéa river provides a major freshwater input to the lagoon, especially during the wet months, which enhances haline stratification near the river mouth.

Primary production in the SLNC is N-limited throughout the year (Torréton et al., 2010) providing a selective advantage for diazotrophs. High rates of BNF have been measured in the large (>10 µm, filamentous diazotrophs) and small (<10 µm, unicellular pico- and nanoplanktonic bacteria and cyanobacteria) size-fractions in the SLNC during the austral summer (November – May) (Garcia, Raimbault, and Sandroni 2007; Biegala and Raimbault 2008). Large blooms of *Trichodesmium* spp., a colony-forming, filamentous diazotroph have been reported via satellite imagery (Dupouy et al. 2010; Ganachaud et al. 2010) and direct measurements (Renaud,

Pringault, and Rochelle-Newall 2005; Rodier and Le Borgne 2008, 2010). *Trichodesmium* blooms have been correlated with warm sea surface temperatures (SST) (≥24°) (Rodier and Le Borgne, 2008) and are frequently documented during November – May when SST is maximum (Garcia, Raimbault, and Sandroni 2007; Dupouy et al. 2010; Ganachaud et al. 2010). The community composition of diazotrophs in the <10 µm size-fraction is less well-characterized (Luo et al. 2012). Biegala and Raimbault (2008) found UCYN-A in the SLNC. UCYN-A however, is poorly studied in this area, in particular during the winter season, and along the seasonal cycle.

It is importaint to understand which diazotrophs are active in the lagoon because diazotrophs differ in size and ecology with implications for the fate of fixed N_2 (Bronk and Glibert 1994; Mulholland et al. 2012; Foster et al. 2011; Karl et al. 2012; Benavides et al. 2013). This is illustrated in the VAHINE microcosm experiment, which was initially dominated by diatom diazotroph associations (DDAs) and then later by *Cyanothece*-like (UCYN-C) populations in a shift that correlated with near tripling of the N_2 fixation rates and a more moderate increase in primary production (Leblanc et al., 2016). Characterization of the composition of active N_2-fixing assemblages, in conjunction with size-fractionated N_2 fixation rates, is therefore essential for understanding the importance of N_2 fixation in a given ecosystem.

The unicellular cyanobacteria group A (UCYN-A) is a picoplankton diazotroph that has been reported in high abundances in many regions (Cabello et al. 2015; Farnelid et al. 2016; Turk-Kubo et al. 2017) and is believed to be a significant contributor to BNF in the SLNC (Garcia, Raimbault, and Sandroni 2007; Biegala and Raimbault 2008). High rates of daytime N_2 fixation have been recorded in the small size-fraction (<10 µm) (Biegala and Raimbault 2008), which is assumed typical for UCYN-A based on expression patterns of the *nifH* gene (Church et al., 2005). However, little is known about the seasonality and extent of UCYN-A abundances, or the genetic diversity of these picocyanobacteria in the SLNC.

UCYN-A lives symbiotically with a single-celled eukaryotic alga in the prymnesiophyte (haptophyte) group (Thompson et al., 2012). Nutrient exchange is believed to be the basis for the symbiosis; UCYN-A provides fixed N_2 in return for fixed carbon (Thompson et al., 2012). The UCYN-A genome has been sequenced, showing that UCYN-A lacks oxygenic photosynthesis, carbon fixation and many other basic metabolic pathway genes (Tripp et al., 2010). The large degree of genome reduction (Tripp et al. 2010; Bombar et al. 2014) and tight coupling of nutrient exchanges suggests an obligate symbiosis (Thompson et al., 2012). Thompson et al. (2014) defined three distinct sublineages: UCYN-A1, UCYN-A2 and UCYN-A3. Recently, a 4th (UCYN-A4) sublineage was identified (Farnelid et al., 2016), followed by a 5th (UCYN-A5) and a 6th (UCYN-A6) (Turk-Kubo et al. 2017). With the highest relative abundances (Turk-Kubo et al., 2017) and largest geographic range (Cabello et

al. 2015), UCYN-A1 and UCYN-A2 are the dominant UCYN-A sublineages and will be the focus of much of this analysis.

UCYN-A1 and UCYN-A2 differ in morphology and they may inhabit different niches. UCYN-A1 and UCYN-A2 have genetically distinct prymesiophyte hosts (Thompson et al., 2014). CARD-FISH probes specific to UCYN-A 16S rRNA (Cornejo-castillo et al., 2016), show that UCYN-A1 and its host are much smaller (1.5 μm and 2.5 μm diameter, respectively) than UCYN-A2 and its host (4 μm and 8 μm diameter, respectively). Furthermore, molecular studies indicate that sublineages may have different patterns of abundances, with the UCYN-A1/prymesiophyte association having higher abundances in open ocean waters (Thompson et al. 2014; Farnelid et al. 2016) and the UCYN-A2/prymnesiophyte association possibly more commonly found at high abundances in coastal regions (Thompson et al. 2014; Farnelid et al. 2016; Turk-Kubo et al. 2017). This theory is still being investigated. Using cell-specific rates, UCYN-A2 was recently reported to have much higher per cell N_2 fixation rates than UCYN-A1 (Martínez-Pérez et al., 2016). However, the probes used in that study (Cornejo-castillo et al., 2016) have subsequently been shown to amplify the 16S rRNA genes for both UCYN-A2 and UCYN-A3, and so it is possible that the UCYN-A2 reported in Martínez-Pérez et al. (2016) is UCYN-A3 and not UCYN-A2 (Turk-Kubo et al. 2016; Farnelid et al. 2016). Regardless, it is clear that UCYN-A1 and UCYN-A2 differ, with implications for their physiology, nutrient requirements, ecology, and global distribution patterns. Much remains

unknown about the other sublineages, including their host identities and ecological niches.

The objectives of the present study were to (1) quantify abundances, and document the spatial-temporal distributions, of UCYN-A1 and UCYN-A2; (2) determine which environmental factors contributed most to overall abundances and distributions of UCYN-A; (3) differentiate the environmental factors associated with UCYN-A1 vs. UCYN-A2 abundances; (4) determine the diversity of UCYN-A in the SLNC; and (5) determine if there is a relationship between UCYN-A community composition (diversity) and specific environmental conditions. To do so, samples from a 20-month time-series, collected at four stations within the SLNC, were analyzed for UCYN-A abundances, diversity and biogeography. In addition, hydrological and meteorological variables were also analyzed with the goal of better understanding environmental controls on UCYN-A abundance and distributions. The four-stations were along a 12 km transect between Dumbéa Pass and Dumbéa Bay (Figure 3), following a gradient from oligotrophic to relatively greater coastal influences. UCYN-A-specific assays targeting the *nifH* gene were used for quantitative (qPCR) and qualitative (next-generation sequencing) assessment. *nifH* sequences were analyzed using an oligotyping-method, in which oligotypes are defined based on nucleotide positions with high variability or Shannon entropy (Eren et al., 2013). This study provides new data describing the extent of UCYN-A abundances, seasonality, and genetic diversity in the WTSP, where such data is lacking.

2.0 Methods

2.1 Sampling

Using a trace metal clean Teflon pump connected to a polyethylene tube, surface seawater (3 m depth) was collected between July 4, 2012 - April 3, 2014 in the SLNC along a transect running from the Dumbéa Pass to the mouth of Dumbéa Bay (Figure 3). The transect included four sampling sites. From the North East to the South West, the stations were: D39 (Dumbéa Bay): 22°13.229' S - 166°22.430' E, M99: 22°14.798' S - 166°18.954' E, M09: 22°18.970' S - 166°17.700' E, and L2 (Dumbéa Pass): 22°20.963' S - 166°15.763' E.

Samples for DNA analysis were immediately filtered through 25 mm diameter 0.2 μm pore-size Supor® filters (Millipore, Billarica, CA), using gentle peristaltic pumping. Sample volume varied (930 ml - 2350 ml) depending on the amount of water that could be filtered within a reasonable amount of time. For this reason, abundances are calculated separately for each sample. All filters were flash-frozen in liquid N_2, and stored at -80°C until shipment on dry ice from New Caledonia to the University of California, Santa Cruz.

2.2 Temperature, Nutrient, and Chlorophyll Measurements

Surface temperature and fluorescence were recorded using a Seabird 911 plus conductivity-temperature-depth (CTD) profiler. Surface seawater samples (3 m

depth) for nutrient analyses were collected in acid-washed polyethylene bottles, poisoned with $HgCl_2$ (10 μg L^{-1} final concentration) and stored in the dark at 4°C until analysis.

Soluble reactive phosphorus (SRP) and NOx [nitrite ($NO2^-$)+nitrate ($NO3^-$)] concentrations were quantified using standard colorimetric techniques (Aminot and Kérouel, 2007) on a Bran Luebbe AA3 autoanalyzer. Detection limits for the procedures were 0.01 μM for SRP and 0.03 μM for NOx. The acidification protocol was used for chlorophyll *a* analysis (chl *a*). For this, seawater samples were collected in 0.55 L flasks, filtered onto GF/F Whatman filters, and extracted with 95% methanol after 30 minutes of incubation in the dark at ambient temperature (Herbland et al., 1985). Using a Turner Design fluorometer, the first fluorescence measurement was made, the sample was acidified (20 μl HCL 0.3 mol/l), and the second measurement was made. The fluorometer used was outfitted with chl *a* extracted acidification module (module a 7200–040) calibrated with pure chl *a* standard (Sigma). For each sampling, a linear least squares regression was used to align the *in situ* CTD fluorescence measurements with the extracted chl *a* concentrations.

2.3 Meteorological Data

Meteorological data was acquired from the Météo-France station at Faubourg Blanchot (Nouméa) 22°16.30 S - 166°27.06 E. Daily values were obtained for wind direction, wind velocity, and precipitation. These values were averaged to derive the

three weekly average (mean of daily values the week prior to sampling) and three monthly average (mean of daily values the month prior to sampling) covariate terms. These six terms make up the meteorological covariates used in this analysis.

2.4 DNA extraction

DNA was extracted using a Qiagen DNeasy Plant kit (Valencia, CA), with protocol modifications 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 protenase K digestion (Moisander et al., 2008). The purity and quantity of DNA extracts was determined using a NanoDrop (Thermo Scientific, Waltham, Ma), according to the manufacturer's guidelines.

2.5 Determining UCYN-A abundances using quantitative polymerase chain reaction (qPCR)

UCYN-A was quantified using qPCR with Taqman® assays. This study targeted UCYN-A1 (Church et al. 2005a) and UCYN-A2 (Thompson et al., 2014). The UCYN-A2 assay is now known to also amplify the UCYN-A2, UCYN-A3, and UCYN-A4 sublineages (Farnelid et al., 2016). For this reason, it is currently necessary to perform *nifH* sequence analysis to confirm the presence/absence of particular sublineage. In keeping with the literature, all organisms targeted by the Thompson et al. (2014) UCYN-A2 primer are referred to as UCYN-A2 throughout this study even though UCYN-A3 or UCYN-A4 were also present.

Recombinant plasmids containing cloned *nifH* fragments from the target diazotrophs were used as qPCR standards, and each 96 well plate was run with a serial dilution ($10^0 – 10^7$ *nifH* copies reaction) of the appropriate standard. All qPCR reactions had 2 µL of DNA extract in a 10 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 primer, and 0.2 µM probe (5'-FAM and 3'-TAMRA labeled). Thermocycling parameters were as described in Goebel et al. (2010) for all assays, except the annealing temperature for UCYN-A2 was 64°C.

The qPCR reaction efficiencies were as follows: 105 ± 3% for UCYN-A1 and 93 ± 3% for UCYN-A2. Based on the differences in sample volumes, the limits of detection (LOD) and quantification (LOQ) for all qPCR assays ranged between 17 and 145 *nifH* copies L^{-1}, and between 133 and 1156 *nifH* copies L^{-1}, respectively (Goebel et al., 2010). LOD and LOQ ranges are large because sample volumes varied. Samples were determined to be "detected, not quantified" (DNQ) when calculated abundances were greater than the LOD, but less than the LOQ. While it is known that UCYN-A1 and UCYN-A2 have one *nifH* copy per genome, except during replication, it is not known when genomes are undergoing replication. Therefore, *nifH* copies L^{-1} rather than cells L^{-1} is used as a proxy for relative abundance.

2.6 UCYN-A community composition determined by high-throughput sequencing of *nifH* amplicons

Samples from each season were sequenced to evaluate if the diazotrophs targeted via qPCR detected all the UCYN-A phylotypes present in the lagoon as well as to provide more detail about the seasonal ecology of the UCYN-A phylotypes. Twenty-eight lagoon samples (all four stations from the following dates: 1/31/13, 3/27/13, 4/3/14, 6/12/13, 8/13/13, 9/3/13, and 12/17/13) were chosen for analysis. For each sample, triplicate PCR reactions were pooled.

A nested PCR approach was used involving the universal *nifH* primers nifH3/nifH4 (Zehr and Turner 2001) for the first amplification and UCYN-A-specific *nifH* primers with 5' common sequence linkers, as per Turk-Kubo et al. (2017) for the second amplification. Thermo-cycle parameters were as described in Turk-Kubo et al. (2017). UCYN-A *nifH* fragments were successfully amplified from all 28 samples. A dual PCR approach was used in library preparation (Green et al., 2015). Sequencing was conducted at the DNA Service Facility at the University of Chicago, Illinois, using the Illumina MiSeq platform.

From the 28 samples sequenced, 981,274 raw paired-end UCYN-A *nifH* reads were obtained. Sequences were processed according to a pipeline described in Turk-Kubo et al. (2017). This involved merging paired ends using Paired-End reAd mergeR (PEAR) software, quality filtering with Quantitative Insights Into Microbial Ecology

(QIIME) scripts, as well as additional filtering in ARB to remove stop codons and non-*nifH* sequences. The majority of sequences (569,021 out of 981,274) passed quality filtering and were subsequently trimmed of primer regions, aligned to a reference alignment for UCYN-A sequences in a *nifH* database, and prepared for oligotyping analysis as per Turk-Kubo et al. (2017). Shannon entropy analysis and oligotyping was conducted according to Turk-Kubo et al. (2017), involving oligotyping pipeline version 2.2 (Eren et al., 2013). This analysis defined 60 unique oligotypes, which represented 100.00% of the sequences submitted for analysis, with a total purity score of 0.62.

2.7 Statistical Analysis of qPCR Data

Relationship between predictor variables and total UCYN-A abundance

To assess the relationships between potential predictor variables and *nifH* copies L^{-1}, a general linear modeling approach was used. Both categorical and continuous predictor variables were included. Categorical variables were station and sublineage (UCYN-A1 and UCYN-A2) while continuous predictor variables included temperature, chl *a*, NOx, SRP, weekly average wind direction, monthly average wind direction, weekly average wind speed, monthly average wind speed, weekly average precipitation, monthly average precipitation.

Variance inflation factor (VIF) scores for continuous variables were much less than 10 indicating the lack of co-linearity at a level that would compromise the modeling.

The approach was to start with a full model, containing all terms including pairwise interactions between categorical and continuous predictor variables, then to sequentially drop variables not contributing to model fit using Bayesian information criterion (BIC) standards, culminating in a reduced model containing only informative variables or interactions between variables.

Environmental characteristics of stations

To determine the factors that explain variability of UCYN-A across stations, a discriminant function analysis (DFA) model in conjunction with a MANOVA approached was developed using a quadratic model fitting routine. The MANOVA informs how variability in measured parameters varied as a function of station, while the DFA assesses if station is predictable based on a set of those measured parameters. There was significant and predictable variation among stations. Moreover, in the DFA, the resulting model was highly predictive, correctly predicting, based on the measured parameters, the true station in 96% of instances. Here a stepwise approach was used.

Ratio of UCYN-A1 to UCYN-A2 across samples

The ratio UCYN-A1 to UCYN-A2 was quite variable, and to explore if the ratio of UCYN-A1 to UCYN-A2 could be predicted based on measured variables, another DFA was constructed. This involved converting the data from continuous abundances to categorical states by distinguishing between two community states: (1)

disproportionate numbers of A1 or (2) disproportionate numbers of A2. These two states were based on whether a sample ratio of A1-A2 (Table 2) was greater than the long-term average value. A forward stepwise DFA approach was used for model selection. Terms were added if they increased the overall discrimination ability without compromising the overall fit of the model (using Pillai Trace estimation).

2.8 Statistical Analysis of Sequence Data

Spatial-temporal relationship among UCYN-A oligotype distributions

A multi-dimensional scaling (MDS) plot based on Bray-Curtis (dis)similarities was constructed to assess the way in which UCYN-A oligotypes cluster together. This analysis was intended to determine if the genetic clustering of oligotypes from the same sublineage indicated anything about relative abundance patterns of oligotypes in nature. MDS plots are an ordination tool, used to create graphical representation of relationships among oligotypes based on the relative abundance of the oligotypes, using station and date as replicates. MDS operates on the similarity matrix, using the ecological distances between samples, to construct a map of the samples in a specified number of dimensions. The axes in the MDS plots are for orientation only, with no meaning beyond this, and if it is included, scaling in MDS plots is arbitrary. A stress value ranging from 0 to 1.0 is used to measure the reliability of the ordination with zero signifying a perfect fit – all rank orders correctly signified by the relative distance between all pairs of points in the graph – and with values > 0.3

corresponding to near arbitrary placement of points in the graph (Clarke and Warwick, 2001).

Relationship between predictor variables and oligotype distributions

The relationship between corresponding oligotypes and the environmental patterns was assessed using a matrix-matching permutation-based test (BIO-ENV method implemented in PRIMER). Two matrices were initially developed: (1) a similarity (or dis-similarity) matrix for pairwise comparisons of samples based on oligotype relative abundances, (2) a distance matrix for pairwise comparisons of samples based on environmental data (e.g. chlorophyll *a*). Usually biological similarity is assessed using a Bray-Curtis approach, and the environmental matrix is assessed using normalized data (to standardize units of measurement) and Euclidean distances (Clarke and Ainsworth, 1993). In order to identify the environmental metrics most associated with the biological matrix the elements of the two corresponding similarity matrices are ranked, and the two matching sets of ranks are compared by calculating a Spearman correlation coefficient (ρ_s). This is repeated after sequential generation of new environmental matrices that differ in the inclusion of environmental metrics (e.g. all environmental metric except temperature). The value of an environmental metric to the overall association of environmental and biotic matrices can be assessed based on the change of fit to the overall model when the metric is omitted from the generation of the environmental matrix. The significance level of the matches can be calculated by comparing the value of ρ_s from any match to a null distribution of ρ_s

values generated by iteratively (usually 1000 runs) randomizing the pairing of matrix elements in the Spearman Rank correlation.

Nonmetric multidimensional scaling (NMDS) plots were employed to create graphical representations of the relationship among samples based on the relative abundances of the various oligotypes and to illustrate temporal and special patterns of community structure. These Non-metric NMDS plots are just as described above; however, in this instance stations and dates were ordinated using oligotyes as replicates, whereas above the oligotypes were ordinated using stations and dates as replicates. Another difference is that Non-metric NMDS operates on the rank order of the elements in the similarity matrix, rather than on the matrix itself (as in the MDS plots described above), to construct a map of the samples in a specified number of dimensions.

3.0 Results

Environmental Conditions

Over the course of the study, temperatures ranged from 21.5°C to 28.1°C, and were fairly consistent across stations (Figure 4a and 4b, Table 1). Peak temperatures were obtained between the end of December and the end of February while minimum temperatures were between August and September. Precipitation was maximum between January and July. During the wet months, average precipitation was 4 mm per day, as compared to a daily average of 1.5 mm during the dry months. From

November until May, mean monthly average wind direction ranged between northeast and southeast (45°-135°) with a consistently strong ($\geq$ 9.5 m s^{-1}) mean monthly wind speed. From June – October, mean monthly average wind speed was more southerly (>135°) with a weaker (< 9.5 m s^{-1}) mean monthly average wind speed. Although more variable, a similar pattern was observed with mean daily wind speed and direction. Chl *a* concentrations ranged from 0.13 μg $L^{-1}$to 0.99 μg L^{-1}. NOx and SRP concentrations ranged from DL to 0.36 μM and DL to 0.37 μM, respectively. There was no discernable seasonal pattern associated with chl *a*, NOx, SRP, or with any other nutrient concentrations across stations.

3.1 UCYN-A abundances (qPCR)

Seasonal abundances of UCYN-A were variable (Figure 4a and 4a, Table 2). UCYN-A1 had higher mean qPCR-based abundances than UCYN-A2 across the entire dataset ($2.0 \times 10^4 \pm 3.3 \times 10^3$ *nifH* copies per L^{-1}, including non-detects), but was detected sporadically (below the limit of detection 18 times). Conversely, UCYN-A2 had an order of magnitude lower mean abundances ($2.0 \times 10^3 \pm 3.4 \times 10^2$ *nifH* copies per L^{-1}, including non-detects), but was more consistently detected (below the limit of detection only 5 times). UCYN-A1 and UCYN-A2 population oscillations coincided with each other, rising in the early austral winter (June - August) and peaking in the late austral winter (September and October), but with lower and more variable abundances in the austral summer (December - May). Generally, the pattern of

seasonal abundances varied inversely with annual temperature fluctuations (Figure 4a and 4b).

Abundances of UCYN-A1 differed with respect to station. Maximum abundances of UCYN-A1 were observed at the outer reef station (L2) and the difference in abundance across station was significant ($p<0.001$). UCYN-A1 was always detectable at station L2, where abundances ranged seasonally from $10^3 – 10^6$ *nifH* copies L^{-1}. Detection of UCYN-A1 was more sporadic at the inner stations, especially the inner-most station (D39). Nevertheless, UCYN-A1 mean log abundances were fairly consistent across the three inner stations (D39, M99, and M09). Although maximum abundances of UCYN-A2 were observed at station D39 and declined towards the reef passage, log abundances did not differ significantly across stations.

Relationships between predictor variables and total UCYN-A abundances

A reduced model developed through the Bayesian information criteria (BIC) process, to determine which predictor variables were related to abundance, was very explanatory ($p<0.001$), with an adjusted r^2 of 0.57. The reduced model contained nine predictor variables, all individually significant ($p<0.05$) or highly significant ($p<0.001$), suggesting the importance of a suite of factors in predicting abundances (Table 3).

The natural log (ln) of temperature (ln(temp)) was the most significant term in the model (Table 3). The next most significant term was sublineage (UCYN-A1 or UCYN-A2), followed by weekly average precipitation. Monthly average wind speed was the next most important term. Station (D39, M99, M09, and L2) was next, followed by the ln(NOx) and then the ln(chl *a*). The last two terms in the model were interactions, sublineage*ln(temp) and sublineage*station. These last two terms indicate that depending on the sublineage (UCYN-A1 or UCYN-A2), temperature and station have different effects on abundances. The second term (sublineage) reinforces the discriminant effect that identification as UCYN-A1 vs. UCYN-A2 has on abundance. Model findings are strongly driven by UCYN-A1, because UCYN-A1 dominated total UCYN-A qPCR abundances.

Environmental differences among stations

As ratio of UCYN-A1 and UCYN-A2 was variable across stations, a DFA was constructed using quadratic model fitting to assess if that variability was due to differences in environmental characteristics among station (Figure 7, Table 4). The resulting DFA station model was accurate, correctly predicted station 96% of the time, and revealed that the stations formed a gradient of environmental factors. A 95% confidence level ellipse was plotted for each mean. The confidence level of stations L2, M09, and M99 all fell on the negative side of canonical axis 1 of scoring coefficient and overlap somewhat with each other, but not with station D39, which fell on the positive side of canonical axis 1. This indicates the distinct nature of

station D39. Pillai's Trace test was used to assess the model's overall significance ($p < .001$).

Using a stepwise approach, all the covariates contributed to the model's accuracy and were thus retained. However, the lower value (<0.5) of the canonical structure coefficients associated with the meteorological data (precipitation and wind), temperature, SRP, NOx, and nstar, indicate that these terms explain little of the variance (Table 4). This was especially true of the meteorological data (canonical structure coefficients <0.1), which was acquired from the Météo-France station at Faubourg Blanchot (Nouméa), nonetheless retention of these terms improved the model's accuracy. The top covariates that discriminate for station D39, on the positive side of canonical axis 1, are ln(silicate) and ln(chl *a*) (Table 4).

Ratio of UCYN-A1 to UCYN-A2 across samples

The results of the DFA indicated that the ratio of UCYN-A1 to UCYN-A2 could be predicted with great accuracy. The DFA model predicted 97% of events where the ratio of A1 to A2 was less than predicted and 81% of samples where the ratio was greater (Table 5). Pillai's Trace test was used to assess the model's overall significance ($p < 0.009$).

Using the stepwise model, incorporating both scoring coefficients and canonical structure, discrimination of UCYN-A2 from A1 (positive values of scoring

coefficient canonical axis 1) is driven by increasing ln(silicate), ln(temp), ln(chl *a*), monthly average wind speed, weekly average precipitation, and weekly average wind direction, in order of relative importance. Discrimination of UCYN-A1 vs A2 (negative values of scoring coefficient canonical axis 1) is driven by monthly precipitation average, ln(NOx), and weekly average wind speed, in order of relative importance. Based on the value of the canonical structure coefficients (≥0.5), the terms that explain most of the variance in this model are ln(temp), ln(silicate), and ln(chl *a*), all of which are associated with the discrimination of UCYN-A2 over UCYN-A1 (Table 5).

The most predictive term in the reduced model was ln(temp), and after other variables have been modeled the overall effect of temperature on UCYN-A abundance is negative ($p<0.001$, Figure 5, Table 3). The negative correlation is driven by the dominance of UCYN-A1, as when the relationship between ln(temp) and UCYN-A2 abundance is analyzed on its own a positive relationship is observed (Figure 6). In the reduced model, total UCYN-A abundance is negatively impacted by increasing ln(chl *a*) concentration and positively impacted by ln(NOx) concentration after accounting for all other factors (Table 3). The DFA also highlights the polarizing effect of ln(chl *a*) and ln(NOx) on the abundance of UCYN-A as indicated by the correlation between UCYN-A2 abundance and ln(chl *a*), as well as UCYN-A1 abundance and ln(NOx) (Table 5).

3.2 UCYN-A genetic diversity within and outside of major sublineages (UCYN-A1 and UCYN-A2)

To assess the genetic diversity of UCYN-A in the SLNC, oligotyping analysis was performed on amplified UCYN-A partial *nifH* gene sequences. The resulting UCYN-A *nifH* amplicon dataset was comprised of 44 previously identified oligotypes and 16 novel oligotypes, that represented all known sublineages of UCYN-A, as well as some potentially new sublineages. Twelve of these – in order of descending total relative abundance: oligo3, oligo1, oligo43, oligo4, oligo45, oligo46, oligo2, oligo40, oligo13, oligo30, oligo34, oligo27 – accounted for 96.7% of all sequences recovered (Table 6). Four major oligotypes – oligo3, oligo1, oligo43, and oligo4 – accounted for 93.1% of all sequences recovered. The remainder of the dataset was comprised of minor oligotypes, present at low relative abundances across the dataset. Of the 16 newly defined oligotypes, several fell to the UCYN-A3 sublineage and many others fell to the UCYN-A2 sublineage. Other newly defined oligotypes were genetically distinct from previously defined oligotypes, and may be new sublineages. These are referred to here as UCYN-A7 and UCYN-A8.

Oligo3, which is within the UCYN-A2 sublineage (Turk-Kubo et al., 2016), dominated the UCYN-A *nifH* amplicon dataset. Oligo3 accounted for 45% of total sequences. In 13 out of the 28 total samples analyzed, oligo3 accounted for over 50% of the sequences. Oligo1, which is affiliated with the UCYN-A1 sublineage (Zehr, et al. 2008; Tripp, et al. 2010), was the second most abundant oligotype in this dataset,

accounting for 31% of total recovered sequences. In 8 out of the 28 total samples analyzed, oligo1 accounted for over 50% of the sequences. The third most abundant oligotype, oligo43, clusters with the UCYN-A2 sublineage, and represented 14% of total sequences. Oligo43 has been found exclusively in the Nouméa Lagoon of New Caledonia (Turk-Kubo et al., 2017). Oligo4, the fourth most abundant oligotype, clusters with the recently defined UCYN-A4 sublineage (Farnelid, et al. 2016). Oligo4 accounted for 2% of total sequences. Also of note is oligo2, the seventh most abundant, which clusters in the UCYN-A3 sublineage defined by Thompson, et al. (2014). Little is known about this sublineage. Roughly equal numbers (15/36 and 16/36) of the minor oligotypes were phylogenetically affiliated with UCYN-A2 and UCYN-A1 sublineages respectively.

Spatial-temporal relationships among oligotype distributions

In the MDS plot using Bray-Curtis based dissimilarities, oligotypes within the same sublineage largely co-occur (Figure 8). There are a few exceptions to this, with oligo 46 (A2), oligo40 (A2) and oligo99 (A2) clustering with the UCYN-A1 sublineage. As is typical with multidimensional ordination techniques, objects oriented closer to one another are more similar than those oriented further away. In this case, (dis)similarity is defined by relative abundance and frequency of occurrence by the Bray-Curtis-based dissimilarity matrix. As the majority of variation [59.7%] is contained in axis 1, points distributed along this axis are more dissimilar than those that may be similarly distributed along axis 2. Beyond the exceptions noted above,

along axis 1 there is a clear divide between the oligotypes associated with UCYN-A1 and those associated with UCYN-A2. Other sublineages co-occur. For instance, UCYN-A4 (Oligo4) clusters with the UCYN-A2 oligotypes, close to oligo43. The newly defined UCYN-A7 sublineage (oligo80, oligo100, and oligo98) and UCYN-A8 sublineage (oligo85) also cluster with UCYN-A2. The UCYN-A3 sublineage (oligo2, and newly defined oligo86, oligo77, and oligo101) cluster with UCYN-A1. The UCYN-A6 (oligo24) and UCYN-A5 sublineages (oligo15) also cluster with UCYN-A1. The overall stress of the MDS is 0.15, indicating that the MDS is a fairly good indicator of the relationship among oligotypes. Thus, the overall pattern is UCYN-A1 and UCYN-A3 clustering together and UCYN-A2 and UCYN-A4 clustering together.

Relationship between predictor variables and oligotype relative abundances

The BIO-ENV procedure showed that there was a strong correlation between oligotype community composition and environmental conditions. Using all 17 environmental variables, the Spearman Rank correlation was 0.193 with a p-value of $p<0.022$ (Table 7). The solution that maximized the Spearman rank correlation between the two resemblance matrices was a four-variable combination – ln(temp), chl *a*, NOx, and weekly average wind direction ($\rho_s = 0.358$, $p<0.001$, Figure 9). The other four-variable combination, which substitutes temp for ln(temp) is equivalent. The five-variable combinations performed slightly better; however, due to the

collinearity of these combinations they are excluded on the basis that the collinear variable did not change the interpretation.

Temperature [ln(temp)] was the top determinant of community composition and the only one-variable solution proposed by the BIO-ENV analysis. Samples extending from the positive direction of the temp vector correspond with higher temperatures (Figure 10). Correlating samples in in these NMDS plots with SST data (Figure 2a and 2b), it is possible to confirm that in general, the austral summer water temperatures (12/17/2013 and 1/31/2013) are the highest, austral winter (6/12/2013, 8/13/13, and 9/3/2013) water temperatures are the lowest and austral fall water temperatures (3/27/2013 and 4/3/2014) are intermediate. Chlorophyll *a* concentration is inversely related to temperature, as depicted in the NMDS plots (Figure 10) by the perfectly opposite orientation of these two vectors. NOx is almost 90°from the chl *a* and ln(temp) vector, suggesting they have independent effects. Conversely, the effects imposed by weekly average wind direction and NOx may occur in conjunction, based on their similar orientation.

Weekly average wind direction was another strong predictor of community composition. In the NMDS plots, Oligo1 (A1) had higher relative abundances in samples corresponding to the positive orientation of the wind direction vector (Figure 11c). Conversely, Oligo3 (A2), oligo43 (A2), and oligo4 (A4) all had higher relative abundances associated with samples on the negative side of the wind direction vector

(Figure 11a, 11b, and 11d). The wind data was binned into four categories: north (340°-20°), east (80°-120°), southeast (121°-160°), and south (161°-110°), illustrating a major distinction in relative abundance between oligo1 and ologo3 (Figure 12). When the average wind direction was from the north or east, oligo3 comprised 50-60% of relative abundance sequence reads. When the average wind direction was from the south, oligo1 comprised 60% of relative abundance sequence reads. When the wind was from the southeast, oligo1 made up 40% of relative abundance sequence reads while oligo3 comprised 35% of relative abundance. Oligo43 exhibited a similar pattern as oligo3, demonstrating higher relative abundances when the average wind was from the north, east, and southeast in order of preference. Oligo4 had higher relative abundances associated with east and southeast wind directions.

A similar pattern was observed in the abundances of UCYN-A1 (Figure 13); maximum UCYN-A1 abundances correspond with southern or southeastern wind. Interestingly, maximum UCYN-A2 qPCR abundance also correspond with southern or southeastern wind and there is no difference between UCYN-A1 and UCYN-A2 abundances when the wind is from the north or east. Temperature also appears to be related to wind direction, with colder SST associated with southerly or southeasterly wind, which is when peak UCYN-A abundances are observed (Figure 13**)**. Binning the recorded temperatures into temperature categories [high ($26°C \leq x$), mid ($23°C \leq x < 26°C$), low ($x < 23°C$)], the same relationship is observed; colder water

temperatures are observed in association with southeasterly wind (~140°) and this is when peak qPCR abundances occurred.

Oligo3 and Oligo43, which both cluster with the UCYN-A2 sublineage, had somewhat different patterns of relative abundances (Figure 11a and 11b). The relative abundance of oligo43 appears to correspond to ln(temp), as noted in the disproportionately large diameter of the circles in Fig. 11b associated with austral summer sampling dates (12/17/2013 and 1/31/2013) when SST is warmest. Further, of the three dates when temperatures are generally the coldest (6/12/2013, 8/13/13, and 9/3/2013), oligo43 is virtually absent from the first two of these dates and only present at the inner stations on the third date. Using temperature categories [high ($26°C \leq x$), mid ($23°C \leq x < 26°C$), low ($x<23°C$)], oligo3 did not have a consistent pattern with temperature, while oligo43 has higher relative abundances in the high temperature category (Figure 14). Although there appears to be some correspondence between oligo43 and chl *a*, as demonstrated by the low relative abundance of oligo43 on 4/3/14, when chlorophyll *a* declined abruptly, and strong relative abundance on 3/27/13, when chlorophyll *a* peaked (Table 1), this relationship is not maintained by an analysis of variance nor bar-chart exploration. Similarly, no relationship was established between oligo3 relative abundance and chl *a* concentrations. NOx also did not appear to correspond with the relative abundance of these oligotypes. Oligo43 and oligo3 only differ genetically by one nucleotide; it is interesting that

such distinct patterns of relative abundance could be associated with such closely related organisms.

The oligo1 NMDS plot (Figure 11c) suggests that the relative abundance of oligo1 is negatively correlated with temperature. This is demonstrated by the disproportionately large diameter of the circles, representing higher oligo1 relative abundances, in Figure 11c, associated with austral winter and early spring (6/12/2013, 8/13/13, and 9/3/2013) sampling dates when water temperatures are the lowest. Using an analysis of variance, there was a significant inverse relationship between oligo1 relative abundance and temperature ($p<0.001$). The higher relative abundance of oligo1 in lower temperatures ($x<23°C$) is also demonstrated in the bar-charts (Figure 14). Based on the relative abundance of oligo1 in relation to the NOx vector, the NMDS plot also suggests that Oligo1 relative abundance relates to NOx concentration. The 6/12/2013 station L2 sample, may serve as an example of this correspondence, as this sample had maximum NOx concentrations (0.36 µM, Table 1) as well as a high oligo1 relative abundance (Table 6). When the NOx data was separated into categories [high ($0.075\ \mu M \leq x$), mid ($0.03\ \mu M \leq x < 0.075\ \mu M$), low ($x<0.03\ \mu M$)], oligo1 had a higher relative abundance in the mid-NOx category and high variability in the high-NOx category (Figure 15). However, NOx concentrations were low across the study, with more than half of all measurements falling below DL. Chlorophyll *a* did not appear to have any relationship with oligo1 relative abundance.

Oligo4 (A4) has an unusual pattern of relative abundance, with presence largely limited to the two inner most stations (D39 and M99). Oligo4 displayed high relative abundance on 12/17/2013 at stations D39 and M99, which both had maximum temperatures (27.2°C and 26.5°C). The relationship between oligo4 relative abundances and temperatures is further supported in bar-chart form, where oligo4 relative abundance is noticeably higher in the high temperature category ($26°C \leq x$) (Figure 14). Consistent with this pattern, the two inner stations (D39 and M99) had slightly higher temperatures relative to the other stations throughout the study (Figure 4a and 4b). No clear relationship between oligo4 relative abundance and chl *a* nor NOx were observed.

There was more variability in environmental conditions across time than across locations. As temperature is the only covariate to display a strong seasonal pattern, peaking in the austral summer and declining in the austral winter (Figure 4a and 4b), it is representative of date (or time) in this study. Stations (or locations) also reflected a gradient of environmental conditions and was correlated with overall UCYN-A abundance. Nevertheless, UCYN-A relative abundance is more related to date (or temperature) than station, as indicated in the way samples from similar dates tend to cluster together (Figure 10). Consequently, there was more variation in UCYN-A composition across time than across locations.

4.0 Discussion

4.1 UCYN-A1 is present at higher abundances than UCYN-A2 in the SLNC

The high UCYN-A1 abundances (10^3 and 10^6 *nifH* copies L^{-1}) reported in this study are in agreement with previous WTSP studies (Bonnet et al. 2015; Moisander et al. 2010; Turk-Kubo et al. 2015). On an open-ocean transect in the Coral Sea between Australia and New Caledonia and around Fiji, maximum UCYN-A1 abundances (10^6 *nifH* copies L^{-1}) were measured at the stations nearest New Caledonia (Moisander et al., 2010). Similarly, Bonnet et al. (2015) sampled along a transect off the West coast of New Caledonia, extending Southwest between 155 °E and 165 °E, and found UCYN-A1 surface abundances between $10^3 – 10^5$ *nifH* copies L^{-1} with peak abundances right off New Caledonia, just north of Dumbéa Pass. In a *nifH* sequence analysis from the Coral Sea (off the Northeast corner of Australia), UCYN-A1 comprised 42% of total *nifH* sequences (Messer et al. 2015). Stenegren et al. (2017) found patchy UCYN-A1 abundances that peaked around 10^4 *nifH* copies L^{-1} around New Caledonia and south of the Fijian Islands (~180 °W). In a SLNC study that occurred south of this study location, UCYN-A1 abundances ranging from $10^3 – 10^4$ *nifH* copies L^{-1} (Turk-Kubo et al., 2015). Thus, the results presented here are consistent with previous studies of high (10^3 - 10^6 *nifH* copies L^{-1}) UCYN-A1 abundances in the WTSP around New Caledonia and the Fijian Islands, as well as in the SLNC.

UCYN-A2 is less well known in the WTSP. Bonnet et al. (2015) identified peak UCYN-A2 abundances of 10^3 *nifH* copies L^{-1} off the Southwest corner of New Caledonia, however, the UCYN-A2 Taqman® assay targets UCYN-A2, UCYN-A3

and UCYN-A4 sublineages, and the sublineage present in the Bonnet et al., (2015) samples was later revealed to be UCYN-A3 (Turk-Kubo et al., 2017). In the Northern part of the Coral Sea, Messer et al. (2015) identified low relative abundances of UCYN-A2 from *nifH* sequence data. Patchy UCYN-A2 abundances were observed by Stenegren et al. (2017), with a peak (10^4 *nifH* copies L^{-1}) around New Caledonia and the Fijian Islands (~180°W) and many non-detects in the Subtropical Gyre. Referenced above, the only other SLNC study where quantitative data has been reported for both UCYN-A1 and UCYN-A2 sublineages, identified higher UCYN-A2 (10^4 *nifH* copies L^{-1}) than UCYN-A1 (10^3 *nifH* copies L^{-1}) abundances at a SLNC site south of where samples for this study were collected (Turk-Kubo et al., 2015). There is a lack of quantitative UCYN-A2 data in the WTSP, but existing data indicates that UCYN-A2 abundances may be moderate ($\leq 10^4$ *nifH* copies L^{-1}) in the WTSP, and limited to coastal areas (around New Caledonia and the Fijian Islands), as well as in the SLNC (Turk-Kubo et al., 2015; Stenegren et al., 2017).

4.2 Total UCYN-A abundances are negatively correlated with temperature

This is the first documentation of year-round UCYN-A abundances in the SLNC with a September-October peak ($10^4 – 10^6$ *nifH* copies L^{-1}) and lower, more-variable numbers from November to May. UCYN-A1 dictated this pattern of seasonal abundance which fluctuated inversely with temperature. Biegala and Raimbault (2008) also identified high abundances (up to 130 cells ml^{-1}) of picoplankton

diazotrophs from August to October using TSA-FISH. Bonnet et al., (2015) identified high UCYN-A qPCR abundances around New Caledonia in September. In another October study, 42% of *nifH* sequences recovered from the Northern part of the Coral Sea were UCYN-A (Messer et al. 2016). The pattern of UCYN-A abundance determined in this study is offset from that of *Trichodesmium*, which has maximum abundances from November - April (Garcia et al., 2007; Dupouy et al., 2011; Ganachaud et al. 2010). Staggering peak abundances would likely give UCYN-A a competitive advantage. This temporal separation of niches contrasts with the special separation that has been observed in the open ocean, where *Trichodesmium* is often at higher abundances in surface waters, and UCYN-A often at higher abundances deeper in the water column (Moisander et al. 2010; Stenegren et al. 2017). These observations may be related to the warm temperature (≥~25°C) generally thought to be required for *Trichodesmium* blooms (Capone et al., 1997; Montoya et al., 2004). Given the lagoon's year-round N-limitation (Torréton et al., 2010), these findings indicate that diazotrophs play a year-round role in the ecology of SLNC.

The results indicate that the two major UCYN-A sublineages are found in a range of conditions within the SLNC. At the outer stations, model studies suggest that hydrodynamically driven mixing and rapid uptake of nutrients by phytoplankton are sufficient to dissipate river inputs and maintain oligotrophic conditions under trade wind forcing (Pinazo et al., 2004; Torréton et al., 2007; Faure et al., 2010). The

Dumbéa Bay however, is protected by the Nouméa peninsula and nourished by the Dumbéa River, fostering a more mesotrophic environment (Pinazo et al., 2004). Station D39, appears to be at the cusp of this protected zone, as average concentrations of chl *a*, NOx or SRP were not appreciably different across stations. Nevertheless, corresponding peak in weekly average precipitation, chl *a,* and silicate at station D39 in May, 2013, suggest that it is influenced by significant runoff events (Table 1). Further, the higher average silicate levels (+ 0.16 µM) observed along the coast (station D39) are likely a result of silicate and iron-rich runoff (Ouillon et al., 2010). Runoff may also be a source of ammonium and SRP (Le Borgne et al., 2010), possibly explaining the correlation between station D39 and chl *a* concentrations (Jacquet et al., 2006; Torréton et al., 2010). Less oligotrophic conditions, as well as more variation in nutrient concentration and biomass, are typical of costal zones in the SLNC (Biegala and Raimbault, 2008; Torréton et al., 2010). The variation of environments and corresponding variation of UCYN-A abundances across station indicated that, UCYN-A1 and UCYN-A2 correlate with different environmental characteristics despite having some overlap in niches.

Temperature, which is often linked to cyanobacterial diazotroph distribution (Church et al. 2008; Moisander et al. 2010; Cabello et al. 2015), was the top predictor of UCYN-A abundance. Consistent with previous studies, which note correspondence between high UCYN-A abundances and the 19°C - 24°C temperature range (Church et al. 2008; Langlois, Hümmer, and LaRoche 2008), a decline in UCYN-A

abundances was observed in association with rising SST (x > 24°C, Figure 4a and b). Seasonal variation in irradiance, which is maximal from December to January (10–62 E m^2 d^{-1}) and minimal from June to July (5– 25 E m^2 d^{-1}), contributes to seasonal temperature fluctuations (Torréton et al., 2010). Cooler SST may also be produced by shifts in wind direction. More frequent from May to October, intermittent westerlies may bring cool SEC water directly through the Dumbéa Pass, bypassing the southern part of the SLNC (Le Borgne et al., 2010; Torréton et al., 2010). Representing less than 12% of yearly wind occurrences (Blaize and Lacoste 1995; Torréton et al., 2010), westerly wind is nonetheless significant as cyclones and tropical depressions tend to generate winds from that direction (Ouillon et al., 2010). Additionally, because of the western orientation of the SLNC, water residence time can be longer under western wind, which could give UCYN-A time to bloom (Pinazo et al., 2004; Le Borgne et al., 2010). It's not clear what changes associated with colder SST create better conditions for UCYN-A. Although temperature is not the only controller (Table 3), it does appear to be the primary control on UCYN-A abundance.

4.3 UCYN-A1 and UCYN-A2 abundances are correlated with distinct environmental characteristics

UCYN-A1 and UCYN-A2 had different environmental predictors. UCYN-A2 abundances were positively linked with silicate, temperature, and chl *a*, while UCYN-A1 abundances were positively correlated with inverse conditions. Wind speed was

associated with abundances of both major sublineages. Interestingly, SRP which has been speculated to control N_2 fixation in the WTSP, was not correlated with UCYN-A1 nor UCYN-A2 abundances (Van Den Broeck et al., 2004). The inverse relationship between UCYN-A1 abundance and temperature reported in this study is consistent with previous findings (Moisander et al. 2010; Bonnet et al. 2015). Messer et al., (2015) also identified the same temperature and silicate relationships in an Australian estuary, where higher UCYN-A2 abundance were found in warmer, higher silicate waters and higher UCYN-A1 abundance in colder, lower silicate waters (Messer et al. 2015). Another Messer et al. (2016) study also found a negative correlation between UCYN-A1 *nifH* relative abundances and silicate as well as phosphate. As suggested by Messer et al. (2016), UCYN-A1 analysis implies an affinity for oligotrophy.

UCYN-A1 abundances were also connected with nitrate (NOx), although NOx concentrations were moderate in this study. The observed relationship between NOx and UCYN-A1 abundance was also noted in another WTSP study, where the correlation was especially prominent in waters where nutrients had been recently entrained (Moisander et al., 2010). While it is unusual for diazotroph abundances to be associated with nitrate, it has been suggested that the higher observed nitrate concentrations are related to shoaling isopycnal surfaces (Moisander et al., 2010). This explanation may also be applicable in this study, given documentation of upwelling outside the lagoon, as well as and the sporadic peaks in NOx

concentrations documented at the outermost Dumbéa Pass station (L2) (Hénin and Cresswell 2005; Neveux et al. 2010; Ganachaud et al. 2014). The relationship between UCYN-A1 and NOx may also be driven by the prymesiophyte host.

Other studies have identified seemingly contradictory correlations between UCYN-A abundances and environmental characteristics, reinforcing the intricacies implicit in microbial ecology. In a 23-day SLNC study with daily sampling, Turk-Kubo et al. (2015) found UCYN-A1 abundances positively correlated with chl *a* and UCYN-A2 abundances negatively correlated with temperature. However, the dissimilarity in study length and sampling frequency, and corresponding smaller range in measurements observed in that study (chl *a*: 0.11 µg L^{-1} to 0.65 µg L^{-1}, temp: 25.30°C to 26.24°C), as compared to this study (chl *a*: 0.13 µg L^{-1} to 0.99 µg L^{-1}, temp: 21.5°C to 28.1°C), make comparisons challenging. However, UCYN-A2 abundances have been associated with colder waters in other studies including Bentzon-Tilia et al. (2015) and Turk-Kubo et al. (2017), which identified a predominance of UCYN-A2 sequences in the cold Danish Strait waters. Clearly, the factors driving UCYN-A abundances are complex and often intertwined. Further research will be needed to assess the correlations identified in this study as well as those identified elsewhere, to see which factors continue to prove useful predictors of UCYN-A abundances. Finer-scale genetic diversity that is not-detected using commonly applied qPCR approaches may help explain seemingly contradictory findings, as although oligotypes within the same sublineage tended to respond similarly to environmental stimuli, some

differences were detected (discussed below) and these may be critical in understanding global patterns of UCYN-A distributions.

4.4 UCYN-A genetic diversity in the SLNC

The oligotyping method is an emerging tool that is helping to define UCYN-A community composition and interpret geographic distribution (Turk-Kubo et al., 2017). At the time of this study, 44 UCYN-A oligotypes had been defined using a global survey of UCYN-A sequences that included UCYN-A sequences from previous studies (Bentzon-Tilia et al. 2015; Turk-Kubo et al. 2015; Messer et al. 2015; Messer et al. 2016). Turk-Kubo et al. (2017) identified four dominant oligotypes, that affiliated with different sublineages, oligo1(UCYN-A1), oligo2 (UCYN-A3), oligo3 (UCYN-A2) and oligo4 (UCYN-A4). Like the global survey, 93% of sequences were associated with only four oligotypes and three of these top four ologotypes [oligo1 (A1), oligo3 (A2), and oligo4(A4)] were the same across studies. An interesting difference between the results of these studies is the presence of oligo43 (A2), which was present at high relative abundances in our study. Oligo43 was defined in Turk-Kubo et al. (2017) and has only been identified in the SLNC. Somewhat surprising, given the qPCR-based dominance of UCYN-A1, oligo3 (A2) was the dominant oligotype in this study, accounting for 45% of total sequences. Turk-Kubo et al. (2017) also identified oligo3 as the dominant oligotype in the SLNC (Noumea Lagoon dataset), accounting for 59.3% of sequences. Turk-Kubo et al. (2017) identified oligo43 as the second most abundant oligotype in the SLNC, not

oligo1, as determined here. In disagreement with previous reports of UCYN-A2 dominating the genetic diversity of UCYN-A in the SLNC (Turk-Kubo et al., 2017), this study found roughly equal proportions of minor oligotypes affiliated with the UCYN-A1 vs. UCYN-A2 sublineages. The identification of 16 new oligotypes likely indicates that there is significant UCYN-A diversity yet to be discovered. Findings may also be an indication of the high endemism known in New Caledonia (Morat 1993; Veillon et al. 2013).

The ordination of oligotypes by Bray-Curtis (dis)similarities indicates that some sublineages are found together owing to similar patterns of relative abundance across the dataset. The separation between oligotypes affiliated with UCYN-A1 and UCYN-A3 vs. those associated with UCYN-A2 and UCYN-A4 was also observed in Turk-Kubo et al. (2017). Turk-Kubo et al. (2017) noted that this division appears to be driven by the co-occurrence of the UCYN-A1 and UCYN-A3 in oligotrophic samples, which contrasts with the appearance of UCYN-A2 and UCYN-A4 in coastal samples. The variation in environmental character and abundances and across station, with significantly higher UCYN-A1 abundances at the more oligotrophic, open-ocean station (L2) ($p<0.001$), and higher UCYN-A2 abundances at the less oligotrophic, coastal station (D39), is consistent with the Bray-Curtis ordination of oligotypes. Results from both the quantitative (qPCR) and qualitative (sequence analysis) parts of this study suggest that UCYN-A1 and UCYN-A2 are distinct ecotypes.

4.5 Oligotype distributions are predictable from environmental characteristics

As indicated by the co-occurrence of oligotypes from the same sublineage, the point mutations and subsequent genetic variability that give rise to distinct oligotypes unlikely confer any selective advantage. The top predictors of oligotype relative abundance are temperature, chl *a*, NOx, and wind direction. Consistent with the findings of the ratio of UCYN-A1 to UCYN-A2 DFA, oligo1 (A1) correlates with lower temperature (<23°C) and moderate NOx concentrations ($0.03\mu M - 0.075\mu M$), while oligo43 (A3) correlates with higher temperatures (≥26°C). Thus, UCYN-A1 and UCYN-A2 are comprised of genetically distinct subgroups, but sublineage predictors are a reasonable first-order hypothesis of oligotype predictors. However, there are exceptions, with some oligotypes from the same sublineage having high relative abundances in different environmental conditions, despite having high genetic similarity.

For instance, oligo3 (A2) and oligo43 (A2) diverged with respect to SST. Oligo3 abundances did not have a consistent relationship with temperature, conversely oligo43 abundances correlated with high temperatures (≥26°C). Three other UCYN-A2 oligotypes (Oligo40, oligo46, and oligo99) are also inversely correlated with temperature, which is why they cluster with the UCYN-A1 oligotypes in the Bray-Curtis MDS ordination. Diversity at the oligotype-level may also help explain the discrepancy between this study's overall finding of UCYN-A2 abundances

correlating with higher temperatures, despite the fact that UCYN-A2 is the dominant sublineage in the cold Danish Strait waters (Turk-Kubo et al., 2017). Oligo43 has only been identified in the SLNC. Conversely, oligo3 dominated Danish Strait samples. This variation is note-worthy given the high genetic similarity of these two UCYN-A2 oligotypes, which only differed genetically by one nucleotide base pair. These findings strongly suggest that fine-scale genetic variation is likely important in understanding global UCYN-A distributions and that environmental predictors are not always consistent for all oligotypes within the same sublineage.

UCYN-A1 abundances are correlated with south and southeast wind

Wind direction and water temperature in the SLNC were linked at the time of this study; colder temperatures and southeast wind occur in conjunction. Given the orientation of New Caledonia (315°), wind from the southeast is known to augment the turn-over of lagoon water, replacing it with cool, oligotrophic SEC water (Torréton et al., 2010). Southerly wind and waves that may accompany this turnover are known to propagate through the reef passages, sending energy to the harbor (Ouillon et al., 2010). Conversely, Northern or even Eastern wind, to which maximum oligo3 relative abundances were correlated, would tend to have less of an effect on the lagoon community (Ouillon et al., 2010), especially at the more protected Dumbéa Bay station (D39). Wind from the south and southeast are correlated with maximum UCYN-A1 qPCR abundances and maximum oligo1 (A1) relative abundances, which suggests that UCYN-A1 may be transported from the

environment outside the lagoon. Previous UCYN-A studies support this distinction, finding higher UCYN-A2 to UCYN-A1 abundances in the lagoon (Turk-Kubo et al., 2015) and high abundances of UCYN-A1 outside the lagoon (Bonnet et al. 2015). An alternative explanation may be that the larger-sized UCYN-A2 sublineage successfully competes with diatoms and other coastal organisms in the lagoon when nutrients are available, but that it is outcompeted by smaller organisms, including UCYN-A1, when southeastern or southern wind brings more oligotrophic water.

The late austral winter peak in UCYN-A abundances documented in this study corresponds with previous reports that showed high day-time N_2 fixation rates in the small (<10 μm, pico- and nanoplankton) size-fractions from August – October. With a few exceptions (Garcia, Raimbault, and Sandroni 2007; Biegala and Raimbault 2008; Bonnet et al. 2015), previous N_2 fixation studies from the SLNC and WTSP are largely focused on the austral summer, which has been considered the main N_2 fixation season in the region (Garcia et al., 2007). Nonetheless, moderate pico- and nanoplankton N_2 fixation rates (0.4 to 0.7 nmol N l^{-2} d^{-1}) have been documented in the lagoon (Garcia et al., 2007) and around New Caledonian (Bonnet et al., 2015a) in the austral spring, at times with 78% of BNF occurring during the day-time (Garcia et al., 2007). Taken together, the literature suggests that "off-season" N_2 fixation may be an important source of new N in the SLNC, and in the context of these reports, the high austral spring UCYN-A abundances reported in this study, suggest that UCYN-A may be a significant contributor to "off-season" N_2 fixation in the WTSP. As primary

production peaks in the austral winter (Lyne et al. 2005; Brewer et al. 2007), this source of added N could have a large effect on primary production.

Conclusions

UCYN-A has the potential to contribute substantial amounts of N to marine systems, due to high rates of N_2 fixation (Montoya et al., 2004; Church et al., 2009; Martínez-Pérez et al., 2016) and global distribution (Cabello et al., 2015; Farnelid et al., 2016). UCYN-A is found in colder waters (19 ˚C - 24˚C) than were previously associated with diazotrophs, allowing it to thrive deeper in the water column (down to ~ 150 m) as well as at higher latitudes (Moisander et al., 2010). High UCYN-A abundances (10^3 - 10^6 nifH copies L^{-1}) have been found in coastal, (Bentzon-Tilia et al., 2015) hypersaline (Messer et al. 2015), N-replete waters (Short and Zehr 2007; Moisander et al. 2010; Mulholland et al. 2012), as well as actively fixing N_2 in sediments (Brown and Jenkins, 2014). The study of the biology and ecology of UCYN-A has greatly broadened the domain of oceanic N_2 fixation (Moisander et al., 2010). UCYN-A has also demonstrated a significant N_2 fixation capacity (>100 fmol N $cell^{-1}$ day^{-1}), equal to that of *Trichodesmium* (Martínez-Pérez et al., 2016), and with growth rates five to ten times higher than those of *Trichodesmium* (Martínez-Pérez et al., 2016). Thus, our improved understanding of UCYN-A is challenging the long-held paradigm that *Trichodesmium* and diatom symbionts are largely responsible for the majority of marine N_2 fixation (Capone 1997; Villareal and Carpenter 1989; Villareal 1990). Continued investigation of UCYN-A distribution, abundance, and associated N_2

fixation rates are thus important to understanding global marine biogeochemical cycles.

This analysis investigated UCYN-A abundances and diversity in relation to environmental characteristics, finding genetically distinct oligotypes that often biologically co-occur with other oligotypes of the same sublineage, and suggests that similar environmental forcing likely governs all oligotypes within a sublineage. Thus, results indicate that sublineage-level distinctions are valuable, especially between UCYN-A1 and UCYN-A2. We know that UCYN-A1 and UCYN-A2 differ significantly, preferring distinctive niches, having different environmental predictors, as well as growth (Turk-Kubo et al. 2015; Martínez-Pérez et al. 2016) and N_2 fixation rates (Martínez-Pérez et al., 2016). Furthermore, given the differing sizes of the UCYN-A1 and UCYN-A2 consortia (Thompson et al., 2014), it is likely they also have different grazing pressures. Nevertheless, genetic diversity at the oligotype level is also important for understanding UCYN-A distributions. In order to further understand marine N budgets, it is critical to continue extrapolating correlations between UCYN-A sublineage abundances and oligotype relative abundances in regard to environmental characteristics. In conclusion, this study demonstrates that UCYN-A is a highly diverse group of diazotrophs, and continuing to investigate these organisms and their pattern of abundance is likely to be essential to our understanding of oceanic primary production and carbon capture as it relates to N_2 fixation

BIOGRAPHY

Aminot, A., and Kérouel, R. (2007). *Dosage automatique des nutriments dans les eaux marines: méthodes en flux continu*. Editions Quae.

Benavides, M., Bronk, D. A., Agawin, N. S. R., Pérez-Hernández, M. D., Hernández-Guerra, A., and Arístegui, J. (2013). Longitudinal variability of size-fractionated N2 fixation and DON release rates along 24.5°N in the subtropical North Atlantic. *J. Geophys. Res. Ocean.* 118, 3406–3415. doi:10.1002/jgrc.20253.

Bentzon-Tilia, M., Traving, S. J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J. L., Markager, S., et al. (2015). Significant N2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. *ISME J.* 9, 273–285. doi:10.1038/ismej.2014.119.

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

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

Bonnet, S., Caffin, M., Berthelot, H., and Moutin, T. (2017). Hot spot of N2 fixation in the western tropical South Pacific pleads for a spatial decoupling between N2 fixation and denitrification. 114, 2800–2801. doi:10.1073/pnas.1619514114.

Bonnet, S., Grand, M. M., Measures, C. I., Hatta, M., Hiscock, W. T., Landing, W. M., et al. (2015a). Global Biogeochemical Cycles. 1–22. doi:10.1002/2014GB004920.Received.

Bonnet, S., Rodier, M., Turk-Kubo, K. A., Germineaud, C., Menkes, C., Ganachaud, A., et al. (2015b). Contrasted geographical distribution of N2 fixation rates and nifH phylotypes in the Coral and Solomon Seas (southwestern Pacific) during austral winter conditions. *Global Biogeochem. Cycles* 29, 1874–1892. doi:10.1002/2015GB005117.

Brewer, D. T., Flynn, a, Skewes, T. D., Corfield, J., Pearson, B., Alowa, J., et al. (2007). Ecosystems of the east marine planning region. . *Rep. to Dep. Environ. Water Resour.*, 150.

Bronk, D. a., and Glibert, P. M. (1994). The fate of the missing 15N differs among marine systems. *Limnol. Oceanogr.* 39, 189–195. doi:10.4319/lo.1994.39.1.0189.

Brown, S. M., and Jenkins, B. D. (2014). Profiling gene expression to distinguish the likely active diazotrophs from a sea of genetic potential in marine sediments. *Environ. Microbiol.* 16, 3128–3142. doi:10.1111/1462-2920.12403.

Cabello, A. M., Cornejo-Castillo, F. M., Raho, N., Blasco, D., Vidal, M., Audic, S., et

al. (2015). Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. *ISME J.* doi:10.1038/ismej.2015.147.

Capone, D. G. (1997). Trichodesmium, a Globally Significant Marine Cyanobacterium. *Science.* 276, 1221–1229. doi:10.1126/science.276.5316.1221.

Capone, D. G., Zehr, J. P. J. P., Paerl, H. W. H. W., Bergman, B., and Carpenter, E. J. E. J. (1997). Trichodesmium, a Globally Significant Marine Cyanobacterium. *Science.* 276, 1221–1229. doi:10.1126/science.276.5316.1221.

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

Church, M. J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., and Karl, D. M. (2009). Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochem. Cycles* 23. doi:10.1029/2008GB003418.

Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., and Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (nifH) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol.* 71, 5362–5370. doi:10.1128/AEM.71.9.5362.

Clarke, K. R., and Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* 92, 205–219. doi:10.3354/meps092205.

Clarke, K. R., and Warwick, R. M. (2001). A further biodiversity index applicable to species lists: Variation in taxonomic distinctness. *Mar. Ecol. Prog. Ser.* 216, 265–278. doi:10.3354/meps216265.

Cornejo-castillo, F. M., Cabello, A. M., Salazar, G., Sa, P., Lima-mendez, G., Hingamp, P., et al. (2016). Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. 1–9. doi:10.1038/ncomms11071.

Dupouy, C., Neveux, J., Ouillon, S., Frouin, R., Murakami, H., Hochard, S., et al. (2010). Inherent optical properties and satellite retrieval of chlorophyll concentration in the lagoon and open ocean waters of New Caledonia. *Mar. Pollut. Bull.* 61, 503–518. doi:10.1016/j.marpolbul.2010.06.039.

Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., et al. (2013). Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* 4, 1111–1119. doi:10.1111/2041-210X.12114.

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

Faure, V., Pinazo, C., Torréton, J. P., and Jacquet, S. (2010). Modelling the spatial and temporal variability of the SW lagoon of New Caledonia I: A new biogeochemical model based on microbial loop recycling. *Mar. Pollut. Bull.* 61, 465–479. doi:10.1016/j.marpolbul.2010.06.041.

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

Ganachaud, A., Cravatte, S., Melet, A., Schiller, A., Holbrook, N. J., Sloyan, B. M., et al. (2014). The Southwest Pacific Ocean circulation and climate experiment (SPICE). *J. Geophys. Res. Ocean.* 119, 7660–7686. doi:10.1002/2013JC009678.Received.

Ganachaud, A., Vega, A., Rodier, M., Dupouy, C., Maes, C., Marchesiello, P., et al. (2010). Observed impact of upwelling events on water properties and biological activity off the southwest coast of New Caledonia. *Mar. Pollut. Bull.* 61, 449–464. doi:10.1016/j.marpolbul.2010.06.042.

Garcia, N., Raimbault, P., and Sandroni, V. (2007). Seasonal nitrogen fixation and primary production in the Southwest Pacific: nanoplankton diazotrophy and transfer of nitrogen to picoplankton organisms. 343, 25–33. doi:10.3354/meps06882.

Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., et al. (2010). Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N2 fixation in the tropical Atlantic Ocean. *Environ. Microbiol.* 12, 3272–3289. doi:10.1111/j.1462-2920.2010.02303.x.

Green, S. J., Venkatramanan, R., and Naqib, A. (2015). Deconstructing the polymerase chain reaction: Understanding and correcting bias associated with primer degeneracies and primer-template mismatches. *PLoS One* 10, 1–21. doi:10.1371/journal.pone.0128122.

Hénin, C., and Cresswell, G. R. (2005). Upwelling along the western barrier reef of New Caledonia. *Mar. Freshw. Res.* 56, 1005–1010. doi:10.1071/MF04266.

Herbland, A., Bouteiller, A. L. E., and Raimbault, P. I. (1985). Size structure of phytoplankton biomass in the equatorial Atlantic Ocean.

Jacquet, S., Delesalle, B., Torreton, J. P., and Blanchot, J. (2006). Response of phytoplankton communities to increased anthropogenic influences (southwestern lagoon, New Caledonia). *Mar. Ecol. Prog. Ser.* 320, 65–78. doi:10.3354/meps320065.

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

/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1120312109.

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

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

Le Borgne, R., Douillet, P., Fichez, R., and Torréton, J. P. (2010). Hydrography and plankton temporal variabilities at different time scales in the southwest lagoon of New Caledonia: A review. *Mar. Pollut. Bull.* 61, 297–308. doi:10.1016/j.marpolbul.2010.06.022.

Leblanc, K., Cornet, V., Caffin, M., Rodier, M., Desnues, A., Berthelot, H., et al. (2016). Phytoplankton community structure in the VAHINE mesocosm experiment. *Biogeosciences* 13, 5205–5219. doi:10.5194/bg-13-5205-2016.

Luo, Y. W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., et al. (2012). 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.

Lyne, V., Hayes, D., Smith, R., Griffiths, B., S. Condie, A., and Hallegraef., G. (2005). Pelagic regionalisation: National marine bioregionalisation integration project. *Hobart CSIRO Mar. Atmos. Res.*

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., et al. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat. Microbiol.* 1, 16163. doi:10.1038/nmicrobiol.2016.163.

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

Messer, L., Mahaffey, C., M Robinson, C., Jeffries, T. C., Baker, K. G., Bibiloni Isaksson, J., et al. (2016). High levels of heterogeneity in diazotroph diversity and activity within a putative hotspot for marine nitrogen fixation. *ISME J.* 10, 1–15. doi:10.1038/ismej.2015.205.

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., et al. (2010). Unicellular cyanobacterial distributions broaden the oceanic N2 fixation domain. *Science* 327, 1512–1514. doi:10.1126/science.1185468.

Moisander, P. H., Beinart, R. A., Voss, M., and Zehr, J. P. (2008). Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *ISME J.* 2, 996–996. doi:10.1038/ismej.2008.84.

Montoya, J. P., Holl, C. M., and Zehr, J. P. (2004). High rates of N2 fixation by

unicellular diazotrophs in the oligotrophic Pacific Ocean. 430, 1027–1031. doi:10.1038/nature02744.1.

Morat, P. (1993). Biodiversity Letters. 1, 72–81.

Mulholland, M. R., Bernhardt, P. W., Blanco-Garcia, J. L., Mannino, A., Hyde, K., Mondragon, E., et al. (2012). Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank. *Limnol. Oceanogr.* 57, 1067–1083. doi:10.4319/lo.2012.57.4.1067.

Neveux, J., Lefebvre, J. P., Le Gendre, R., Dupouy, C., Gallois, F., Courties, C., et al. (2010). Phytoplankton dynamics in the southern New Caledonian lagoon during a southeast trade winds event. *J. Mar. Syst.* 82, 230–244. doi:10.1016/j.jmarsys.2010.05.010.

Ouillon, S., Douillet, P., Lefebvre, J. P., Le Gendre, R., Jouon, A., Bonneton, P., et al. (2010). Circulation and suspended sediment transport in a coral reef lagoon: The south-west lagoon of New Caledonia. *Mar. Pollut. Bull.* 61, 269–296. doi:10.1016/j.marpolbul.2010.06.023.

Pinazo, C., Bujan, S., Douillet, P., Fichez, R., Grenz, C., and Maurin, a. (2004). Impact of wind and freshwater inputs on phytoplankton biomass in the coral reef lagoon of New Caledonia during the summer cyclonic period: a coupled three-dimensional biogeochemical modeling approach. *Coral Reefs* 23, 281–296. doi:10.1007/s00338-004-0378-x.

Renaud, F., Pringault, O., and Rochelle-Newall, E. (2005). Effects of the colonial cyanobacterium Trichodesmium spp. on bacterial activity. *Aquat. Microb. Ecol.* 41, 261–270. doi:10.3354/ame041261.

Rodier, M., and Le Borgne, R. (2008). Population dynamics and environmental conditions affecting Trichodesmium spp. (filamentous cyanobacteria) blooms in the south-west lagoon of New Caledonia. *J. Exp. Mar. Bio. Ecol.* 358, 20–32. doi:10.1016/j.jembe.2008.01.016.

Rodier, M., and Le Borgne, R. (2010). 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. doi:10.1016/j.marpolbul.2010.06.018.

Short, S. M., and Zehr, J. P. (2007). Nitrogenase gene expression in the Chesapeake Bay Estuary. *Environ. Microbiol.* 9, 1591–1596. doi:10.1111/j.1462-2920.2007.01258.x.

Stenegren, M., Caputo, A., Berg, C., Bonnet, S., and Foster, R. A. (2017). Distribution and drivers of symbiotic and free-living diazotrophic cyanobacteria in the Western Tropical South Pacific. *Biosci. Discuss*, 1–47. doi:10.5194/bg-2017-63.

Thompson, A., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F., and Zehr, J. P. (2014). Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ. Microbiol.* 16, 3238–49. doi:10.1111/1462-2920.12490.

Thompson, A., Foster, R. a., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., et al. (2012). Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. *Science (80-.).* 337, 1546–1550. doi:10.1126/science.1222700.

Torréton, J. P., Rochelle-Newall, E., Jouon, A., Faure, V., Jacquet, S., and Douillet, P. (2007). Correspondence between the distribution of hydrodynamic time parameters and the distribution of biological and chemical variables in a semi-enclosed coral reef lagoon. *Estuar. Coast. Shelf Sci.* 74, 667–677. doi:10.1016/j.ecss.2007.05.018.

Torréton, J. P., Rochelle-Newall, E., Pringault, O., Jacquet, S., Faure, V., and Briand, E. (2010). Variability of primary and bacterial production in a coral reef lagoon (New Caledonia). *Mar. Pollut. Bull.* 61, 335–348. doi:10.1016/j.marpolbul.2010.06.019.

Tripp, H. J., Bench, S. R., Turk-Kubo, K. A., Foster, R. a, Desany, B. a, Niazi, F., et al. (2010). Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* 464, 90–94. doi:10.1038/nature08786.

Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B., and Zehr, J. P. (2016). Distinct ecological niches of marine symbiotic N2-fixing cyanobacterium candidatus atelocyanobacterium thalasa sublineages. *J. Phycol.* 38, 42–49. doi:10.1111/jpy.12505-16-180.

Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B., and Zehr, J. P. (2017). Distinct ecological niches of marine symbiotic N2-fixing cyanobacterium Candidatus Atelocyanobacterium thalassa sublineages. *J. Phycol.* 53, 451–461. doi:10.1111/jpy.12505.

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

Van Den Broeck, N., Moutin, T., Rodier, M., and Le Bouteiller, A. (2004). Seasonal variations of phosphate availability in the SW Pacific Ocean near New Caledonia. *Mar. Ecol. Prog. Ser.* 268, 1–12.

Veillon, J., Wulff, A. S., Hollingsworth, P. M., Ahrends, A., Jaffre, T., Huillier, L. L., et al. (2013). Conservation Priorities in a Biodiversity Hotspot: Analysis of Narrow Endemic Plant Species in New Caledonia. 8. doi:10.1371/journal.pone.0073371.

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

doi:10.1111/j.1439-0485.1990.tb00233.x.

Villareal, T. A., and Carpenter, E. J. (1989). Nitrogen fixation, suspension characteristics, and chemical composition of Rhizosolenia mats in the central North Pacific gyre. *Biol. Oceanogr.* 6, 327–346. doi:10.1080/01965581.1988.10749535.

TABLES AND FIGURES

Table 1: Summary of environmental data. Meterological data is from the Météo-France station at Faubourg Blanchot (Nouméa).

Sample	Station	Date	Temp (°C)	Chla (µg L^{-1})	Silicate (µM)	NOx (µM)	SRP (µM)	Monthly ave precip (mm)	Monthly ave wind speed (M/S)	Monthly ave wind direction (°)
64813	L2	7/4/12	23.05	0.25	0.95	0.17	0.03	1.36	7.68	141.03
64814	M09	7/4/12	23.08	0.24	1.02	0.03	0.10	1.36	7.68	141.03
64815	M99	7/4/12	23.06	0.23	1.54	0.03	0.02	1.36	7.68	141.03
64816	D39	7/4/12	22.73	0.42	7.32	0.10	0.02	1.36	7.68	141.03
64817	L2	8/3/12	22.60	0.39	1.22	0.05	0.05	1.94	7.16	159.03
64818	M09	8/3/12	22.60	0.34	1.11	0.03	0.03	1.94	7.16	159.03
64819	M99	8/3/12	22.56	0.30	1.64	0.01	0.04	1.94	7.16	159.03
64820	D39	8/3/12	22.31	0.35	5.53	0.07	0.05	1.94	7.16	159.03
64821	L2	10/2/12	22.35	0.22	1.28	0.13	0.02	1.85	9.47	130.97
64822	M09	10/2/12	22.41	0.19	1.62	0.11	0.20	1.85	9.47	130.97
64823	M99	10/2/12	22.67	0.22	2.31	0.05	0.37	1.85	9.47	130.97
64824	D39	10/2/12	23.14	0.31	2.73	0.10	0.09	1.85	9.47	130.97

64825	L2	11/27/12	24.16	0.38	0.84	0.02	0.07	1.53	9.49	122.33
64826	M09	11/27/12	24.35	0.21	0.98	0.06	0.04	1.53	9.49	122.33
64827	M99	11/27/12	24.59	0.22	2.20	0.03	0.10	1.53	9.49	122.33
64828	D39	11/27/12	24.97	0.16	3.15	0.02	0.06	1.53	9.49	122.33
64829	L2	12/6/12	25.24	0.17	0.99	0.22	0.20	2.62	9.14	118.28
64830	M09	12/6/12	25.95	0.00	1.54	0.12	0.05	2.62	9.14	118.28
64831	M99	12/6/12	25.76	0.00	1.17	0.17	0.05	2.62	9.14	118.28
64832	D39	12/6/12	26.94	0.31	7.95	0.01	0.16	2.62	9.14	118.28
64833	L2	1/31/13	26.47	0.22	1.48	0.12	0.06	0.34	na	na
64834	M09	1/31/13	26.49	0.19	1.77	0.03	0.05	0.34	na	na
64835	M99	1/31/13	26.75	0.22	3.34	0.02	0.09	0.34	na	na
64836	D39	1/31/13	26.96	0.31	5.79	0.04	0.03	0.34	na	na
64837	L2	2/28/13	24.93	0.31	1.64	0.05	0.07	0.19	na	na
64838	M09	2/28/13	25.95	0.35	2.35	0.02	0.05	0.19	na	na
64839	M99	2/28/13	25.81	0.46	3.24	0.02	0.04	0.19	na	na
64840	D39	2/28/13	26.80	0.67	4.60	0.02	0.04	0.19	na	na
64841	L2	3/27/13	25.50	0.38	2.24	0.06	0.02	4.74	9.12	110.00
64842	M09	3/27/13	25.63	0.28	2.03	0.02	0.03	4.74	9.12	110.00

64843	M99	3/27/13	25.94	0.35	3.51	0.03	0.01	4.74	9.12	110.00
64844	D39	3/27/13	26.07	0.52	9.36	0.03	0.02	4.74	9.12	110.00
64845	L2	5/2/13	24.03	0.41	2.09	0.13	0.14	1.87	7.62	149.35
64846	M09	5/2/13	24.48	0.32	4.30	0.05	0.01	1.87	7.62	149.35
64847	M99	5/2/13	23.90	0.38	3.96	0.07	0.02	1.87	7.62	149.35
64848	D39	5/2/13	24.16	0.99	15.01	0.11	0.05	1.87	7.62	149.35
64849	L2	6/12/13	23.00	0.24	1.93	0.36	0.13	2.30	7.92	142.33
64850	M09	6/12/13	22.08	0.25	2.01	0.00	0.07	2.30	7.92	142.33
64851	M99	6/12/13	22.49	0.34	2.93	0.07	0.02	2.30	7.92	142.33
64852	D39	6/12/13	22.39	0.50	5.23	0.03	0.03	2.30	7.92	142.33
64853	L2	8/13/13	22.56	0.23	1.78	0.07	0.04	1.82	7.56	144.19
64854	M09	8/13/13	22.54	0.26	2.01	0.07	0.04	1.82	7.56	144.19
64855	M99	8/13/13	22.69	0.16	1.99	0.04	0.03	1.82	7.56	144.19
64856	D39	8/13/13	22.40	0.37	4.52	0.02	0.03	1.82	7.56	144.19
64857	L2	9/3/13	21.71	0.31	1.54	0.04	0.14	1.49	7.59	165.67
64858	M09	9/3/13	21.48	0.26	1.95	0.04	0.03	1.49	7.59	165.67
64859	M99	9/3/13	21.71	0.21	2.92	0.05	0.27	1.49	7.59	165.67
64860	D39	9/3/13	21.92	0.25	5.13	0.02	0.02	1.49	7.59	165.67

64861	L2	10/25/13	23.88	0.33	1.85	0.08	0.04	1.87	9.20	141.61
64862	M09	10/25/13	23.97	0.50	1.85	0.06	0.03	1.87	9.20	141.61
64863	M99	10/25/13	24.37	0.29	3.56	0.13	0.02	1.87	9.20	141.61
64864	D39	10/25/13	25.06	0.54	7.61	0.06	0.01	1.87	9.20	141.61
64865	L2	11/25/13	25.87	0.13	1.81	0.06	0.03	1.63	8.24	153.00
64866	M09	11/25/13	26.02	0.25	2.12	0.09	0.02	1.63	8.24	153.00
64867	M99	11/25/13	26.18	0.23	2.13	0.15	0.01	1.63	8.24	153.00
64868	D39	11/25/13	26.71	0.67	2.97	0.05	0.01	1.63	8.24	153.00
64869	L2	12/17/13	26.55	0.31	1.88	0.06	0.00	2.52	8.75	123.87
64870	M09	12/17/13	26.75	0.34	2.30	0.08	0.01	2.52	8.75	123.87
64871	M99	12/17/13	26.62	0.26	2.95	0.06	0.01	2.52	8.75	123.87
64872	D39	12/17/13	27.29	0.38	3.96	0.07	0.02	2.52	8.75	123.87
64873	L2	1/22/14	25.80	0.30	1.69	0.12	0.02	5.09	11.26	121.94
64875	M99	1/22/14	26.58	0.36	10.12	0.08	0.02	5.09	11.26	121.94
64876	D39	1/22/14	27.14	0.56	6.83	0.11	0.03	5.09	11.26	121.94
64877	L2	2/22/14	26.86	0.15	2.92	0.00	0.02	4.89	8.66	137.00
64878	M09	2/22/14	26.71	0.22	2.91	0.00	0.02	4.89	8.66	137.00
64879	M99	2/22/14	27.24	0.22	3.50	0.06	0.01	4.89	8.66	137.00

64880	D39	2/22/14	28.08	0.35	5.39	0.00	0.00	4.89	8.66	137.00
64881	L2	4/3/14	25.81	0.00	2.79	0.02	0.03	0.24	7.73	123.00
64882	M09	4/3/14	25.60	0.00	3.36	0.02	0.01	0.24	7.73	123.00
64883	M99	4/3/14	25.86	0.00	4.18	0.05	0.00	0.24	7.73	123.00
64884	D39	4/3/14	26.12	0.00	6.75	0.22	0.01	0.24	7.73	123.00

Table 2: Quantitative (qPCR) abundances of UCYN-A1 and UCYN-A2 and the ratio of UCYN-A1 to UCYN-A2 abundance.

sample	UCYN-A1 [*nifH* copies L-1]		UCYN-A2 [*nifH* copies L-1]		Ratio of A1:A2
	Ave.	***Stdev.***	***Ave.***	***Stdev.***	
64813	5.16E+04	4.56E+03	7.09E+02	1.86E+02	27.59
64814	5.06E+04	1.45E+03	1.23E+03	0.00E+00	27.65
64815	1.33E+03	5.44E+01	4.74E+02	8.66E+00	6.62
64816	2.67E+02	na	6.85E+02	7.36E+01	8.32
64817	6.36E+04	9.46E+03	1.05E+03	4.68E+02	31.49
64818	1.00E+00	na	1.67E+01	na	0.04
64819	1.00E+00	na	1.67E+01	na	0.04
64820	3.42E+03	1.69E+02	1.35E+03	3.43E+02	5.70
64821	5.71E+05	7.36E+04	8.47E+03	4.56E+02	42.61
64822	1.86E+03	6.01E+01	1.11E+03	2.15E+02	6.06
64823	8.75E+03	8.17E+01	1.24E+04	8.79E+02	7.81
64824	3.34E+03	5.69E+02	8.13E+02	3.36E+00	6.90
64825	4.96E+04	6.68E+01	1.87E+03	1.50E+02	33.74
64826	1.17E+05	2.74E+04	4.07E+02	1.72E+01	46.80
64827	1.09E+04	4.03E+03	2.32E+03	4.22E+02	8.65
64828	5.49E+03	6.74E+02	3.40E+01	na	5.63
64829	2.24E+05	1.44E+04	1.83E+03	5.56E+01	26.45
64830	4.83E+03	1.21E+01	2.27E+01	na	5.05
64831	1.09E+05	9.25E+03	1.33E+03	7.58E+01	46.98
64832	1.00E+00	na	2.43E+01	na	0.04
64833	5.04E+02	2.56E+02	2.33E+01	na	12.86
64834	2.17E+01	na	9.76E+02	5.95E+01	0.93
64835	1.00E+00	na	3.13E+01	na	0.05

64836	1.00E+00	na	1.00E+00	na	0.05
64837	1.06E+04	1.59E+03	2.22E+01	na	8.62
64838	2.27E+01	na	2.01E+02	6.04E+01	0.97
64839	1.00E+00	na	3.51E+01	na	0.05
64840	1.00E+00	na	3.92E+01	na	0.05
64841	1.00E+00	na	1.00E+00	na	0.05
64842	1.00E+00	na	2.23E+01	na	0.05
64843	7.66E+02	6.03E+01	4.48E+01	na	19.01
64844	3.57E+01	na	3.57E+01	na	1.34
64845	4.11E+02	9.07E+01	2.50E+01	na	11.51
64846	2.33E+01	na	2.33E+01	na	0.96
64847	1.00E+00	na	1.00E+00	na	0.05
64848	1.00E+00	na	1.00E+00	na	0.05
64849	1.26E+05	2.99E+03	1.26E+03	3.08E+02	22.07
64850	3.04E+02	8.80E+00	1.00E+00	na	8.94
64851	7.41E+03	8.15E+02	2.22E+01	na	6.68
64852	2.20E+03	3.02E+01	2.16E+01	na	5.25
64853	2.12E+05	9.56E+02	2.25E+03	5.90E+01	26.08
64854	1.33E+05	7.06E+03	2.02E+03	4.05E+02	18.97
64855	8.17E+04	6.01E+03	7.01E+03	7.33E+02	36.31
64856	9.97E+03	3.50E+02	1.47E+03	5.07E+02	8.90
64857	4.36E+05	5.54E+04	8.13E+03	1.13E+03	33.28
64858	3.74E+05	1.63E+04	5.71E+03	2.42E+02	30.16
64859	4.84E+04	2.59E+03	2.50E+03	4.19E+02	35.85
64860	3.14E+04	3.13E+03	1.34E+04	1.74E+03	23.61
64861	4.46E+03	1.20E+02	2.13E+01	na	6.29
64862	4.44E+03	1.28E+02	2.13E+01	na	6.48
64863	2.34E+02	5.60E+00	3.07E+02	6.18E+01	7.48

64864	1.00E+00	na	4.32E+02	2.48E+01	0.06
64865	4.67E+03	4.20E+02	2.17E+01	na	5.74
64866	2.51E+03	7.99E+01	2.13E+01	na	5.81
64867	1.92E+03	2.46E+02	2.17E+01	na	4.72
64868	1.99E+02	4.30E+01	2.27E+01	na	6.57
64869	4.05E+02	6.15E+01	4.84E+02	6.47E+00	11.34
64870	1.00E+00	na	2.50E+01	na	0.06
64871	2.66E+01	na	2.66E+01	na	1.06
64872	4.03E+01	na	4.03E+01	na	1.49
64873	6.44E+03	2.82E+02	9.56E+02	1.21E+02	6.13
64875	3.35E+02	6.65E+01	2.70E+01	na	9.54
64876	3.57E+01	na	3.57E+01	na	1.43
64877	2.60E+03	5.38E+01	1.31E+04	7.67E+02	5.49
64878	1.04E+03	5.34E+02	3.63E+03	1.81E+02	23.21
64879	1.00E+00	na	3.21E+01	na	1.00
64880	1.00E+00	na	3.03E+01	na	1.00
64881	1.24E+05	1.16E+04	1.12E+03	2.51E+01	34.16
64882	1.00E+00	na	1.12E+03	2.57E+02	1.00
64883	1.00E+00	na	6.00E+02	1.59E+02	1.00
64884	1.00E+00	na	4.19E+02	5.20E+01	1.00

Table 3: Linear Model to assess the relationship between predictor variables and *nifH* copies L^{-1} – Effect Test. P-values (Prob > F) indicate the fit of individual terms.

Source	**DF**	**Sum of Squares**	**F Ratio**	**Prob > F**
ln(temp)	1	53.492131	50.5696	<.0001
Sublineage	1	20.143497	19.0430	<.0001
Weekly ave precipitation	1	13.198302	12.4772	<.0006
Monthly ave wind speed	1	10.906539	10.3107	0.0018
Station	1	7.686324	7.2664	0.0082
ln(chl *a*)	1	6.346722	6.0000	0.0160
ln(NOx)	1	6.952398	6.5726	0.0118
ln(temp)*Sublineage	1	5.197246	4.9133	0.0289
Sublineage*Station	1	4.996625	4.7236	0.0321

Table 4: Discriminant Function Analysis (DFA) model to predict environmental character of station – Scoring Coefficient and Canonical Structure Table. Values indicate how environmental factors discriminate between stations.

Source	Scoring Coefficients: Canon 1	Total Canonical Structure: Canon 1	Scoring Coefficients: Canon 2	Total Canonical Structure: Canon 2
ln(temp)	-2.01846	0.1924225	2.9775578	-0.080746
ln(chl *a*)	1.4942655	0.4875073	2.8949591	0.508363
ln(silicate)	3.4836746	0.8626549	-1.032289	-0.227086
ln(NOx)	-0.347989	-0.28449	-0.16296	-0.045744
ln(SRP)	0.4431259	-0.138644	1.1749781	0.0923449
weekly ave precipitation	-0.1258	0.0074404	-0.082706	-0.032349
weekly ave wind speed	-0.387207	-0.024881	-0.127863	-0.01403
weekly ave wind direction	0.0117307	0.0125998	-0.003367	-0.057657
monthly ave precipitation	-0.465602	0.0959725	-0.103149	-0.164974
monthly ave wind speed	0.4662937	0.0403427	0.1703088	-0.052843
montly ave wind direction	-0.021061	-0.028406	-0.000887	-0.025671
nstar	0.3682769	0.1322696	0.9841624	0.2354044

Table 5: Discriminant Function Analysis (DFA) model to predict the ratio of UCYN-A1 to UCYN-A2 across samples – Scoring Coefficient and Canonical Structure Table. Values indicate how environmental factors discriminate between two community states: (1) disproportionate numbers of A1 or (2) disproportionate numbers of A2. These two states were based on whether a sample ratio of A1-A2 was greater than the long-term average value.

Source	**Scoring Coefficients: Canon 1**	**Total Canonical Structure: Canon 1**
ln(temp)	5.7434101	0.6212042
ln(chl *a*)	0.4674547	0.482846
ln(silicate)	1.2997311	0.8002822
ln(NOx)	-0.321243	-0.267556
monthly ave precipitation	-0.34702	0.4116781
monthly ave wind speed	0.6159435	0.4301353
weekly ave precipitation	0.068581	0.2982829
weekly ave wind speed	-0.223516	0.2113514
weekly ave wind direction	0.0049191	0.02119

Table 6: Relative abundances of the top 12 oligotypes, listed in order of descending overall relative abundance. Oligotype sublineage is indicated at the top of the column.

	A2	A1	A2	A4	A1	A2	A3	A2	A1	A2	A2	A2
Sample	oligo 3	oligo 1	oligo 43	oligo 4	oligo 45	oligo 46	oligo 2	oligo 40	oligo 13	oligo 30	oligo 34	oligo 37
64833	47%	9%	38%	0%	1%	0%	0%	0%	0%	0%	0%	0%
64834	50%	1%	43%	0%	0%	0%	1%	0%	0%	0%	0%	0%
64835	67%	0%	26%	0%	0%	0%	0%	0%	0%	1%	0%	0%
64836	83%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
64841	58%	17%	17%	0%	2%	1%	0%	0%	1%	0%	0%	0%
64842	81%	3%	10%	0%	0%	0%	0%	0%	0%	0%	3%	0%
64843	49%	18%	18%	4%	1%	1%	0%	0%	0%	0%	0%	1%
64844	60%	1%	36%	0%	0%	0%	0%	0%	0%	1%	0%	0%
64849	16%	68%	0%	0%	1%	2%	6%	1%	0%	0%	0%	0%
64850	77%	10%	6%	0%	1%	1%	0%	0%	0%	0%	0%	0%
64851	41%	44%	1%	0%	2%	2%	3%	1%	1%	0%	0%	0%
64852	32%	56%	0%	0%	2%	2%	0%	1%	0%	0%	0%	0%
64853	9%	88%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
64854	17%	80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

64855	42%	46%	4%	0%	1%	1%	0%	0%	0%	0%	0%	0%
64856	52%	25%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%
64857	34%	54%	1%	0%	3%	4%	0%	1%	1%	0%	0%	0%
64858	33%	52%	4%	0%	3%	3%	0%	1%	1%	0%	0%	0%
64859	28%	39%	20%	5%	2%	1%	0%	0%	1%	0%	0%	0%
64860	39%	21%	29%	5%	1%	1%	0%	0%	0%	0%	0%	0%
64869	33%	19%	44%	0%	1%	0%	0%	0%	0%	0%	0%	0%
64870	41%	5%	50%	0%	0%	0%	0%	0%	0%	1%	0%	0%
64871	30%	17%	36%	14%	0%	0%	0%	0%	0%	0%	0%	0%
64872	35%	19%	20%	24%	0%	0%	0%	0%	0%	0%	0%	0%
64881	8%	78%	0%	0%	1%	1%	4%	0%	0%	0%	0%	0%
64882	92%	0%	4%	0%	0%	0%	0%	0%	0%	0%	0%	0%
64883	97%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
64884	88%	4%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 7: Results of matrix-matching permutation test (the BIO-ENV routine in PRIMER) to assess the degree of similarity between corresponding species and environmental patterns ($\rho_s = 0.193$, $p<0.022$).

Parameters
Correlation method: Spearman rank
Sample statistic (Rho): 0.193
Significance level of sample statistic: 2.2 %
Number of permutations: 999
Number of permuted statistics greater than or equal to Rho: 21

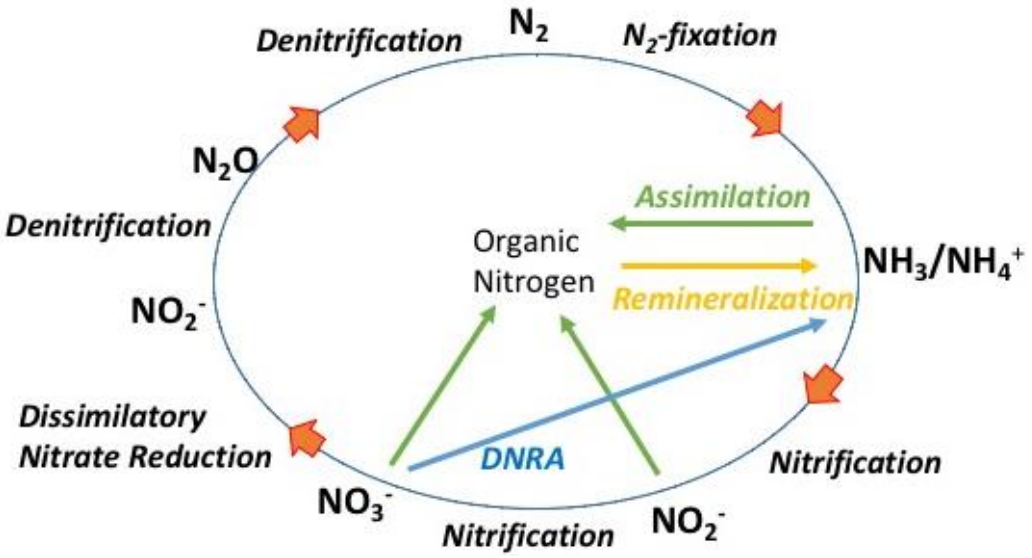

Figure 1: Simplified N Cycle. Anammox not depicted.

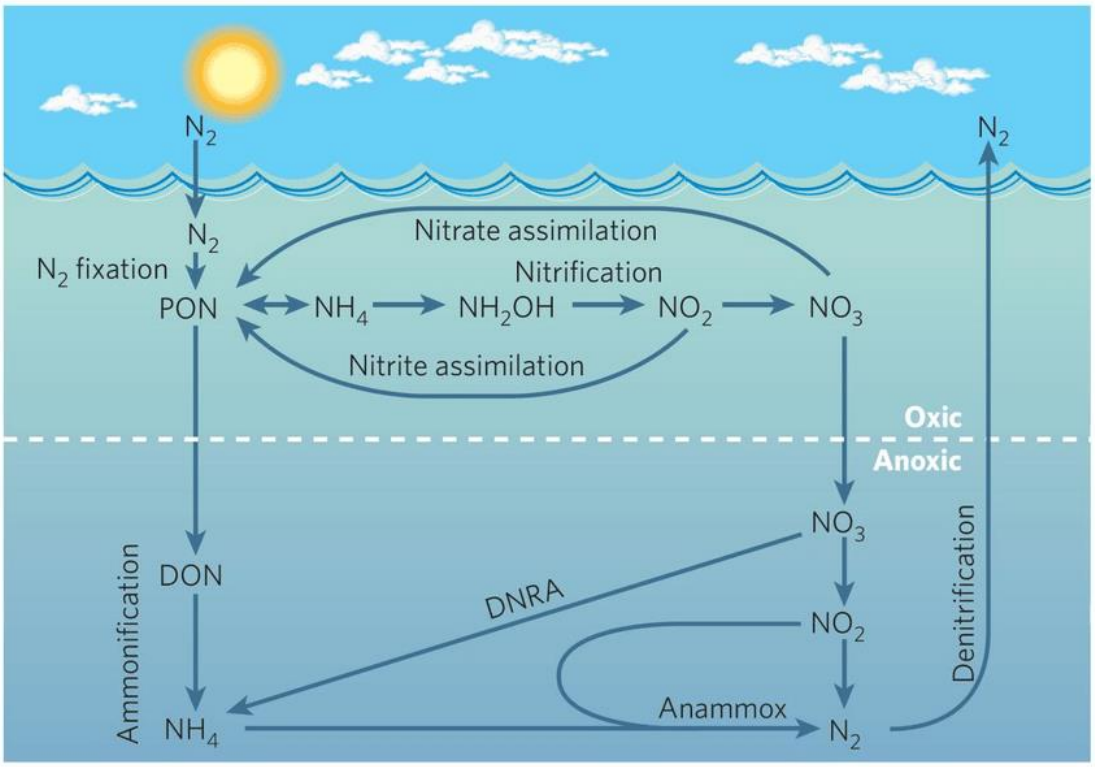

Figure 2: Marine N cycle (Arrigo, 2005). PON, particulate organic nitrogen, including phytoplankton; DON, dissolved organic nitrogen; DNRA, dissimilatory nitrate reductase to ammonium.

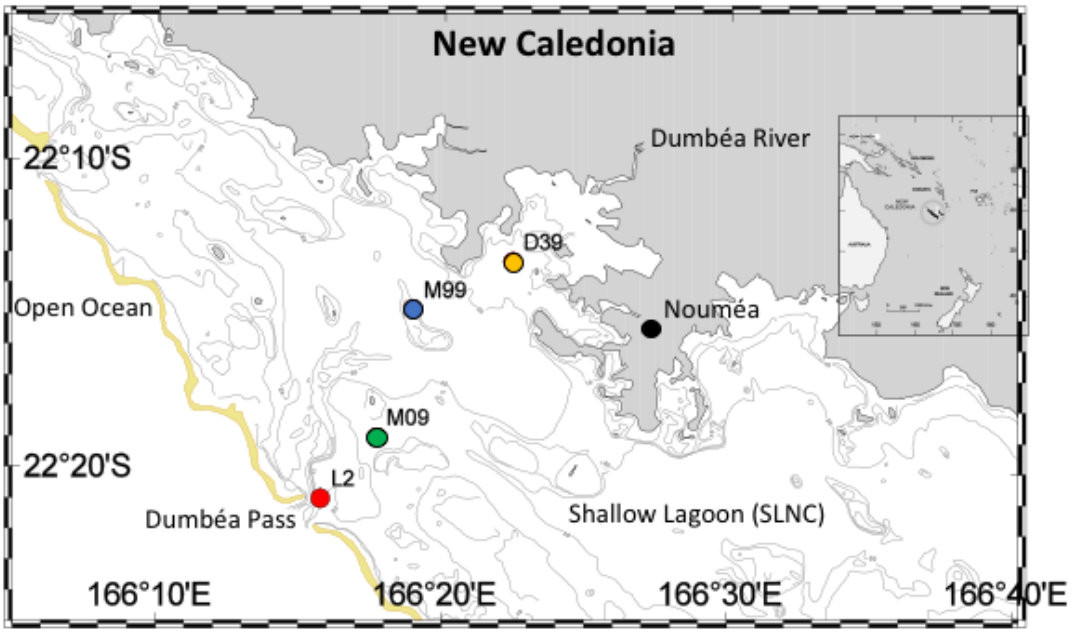

Figure 3: Map of study site, including stations D39, M99, M09, and L2. Study site is located off the southwest corner of New Caledonia in the South Lagoon New Caledonia (SLNC).

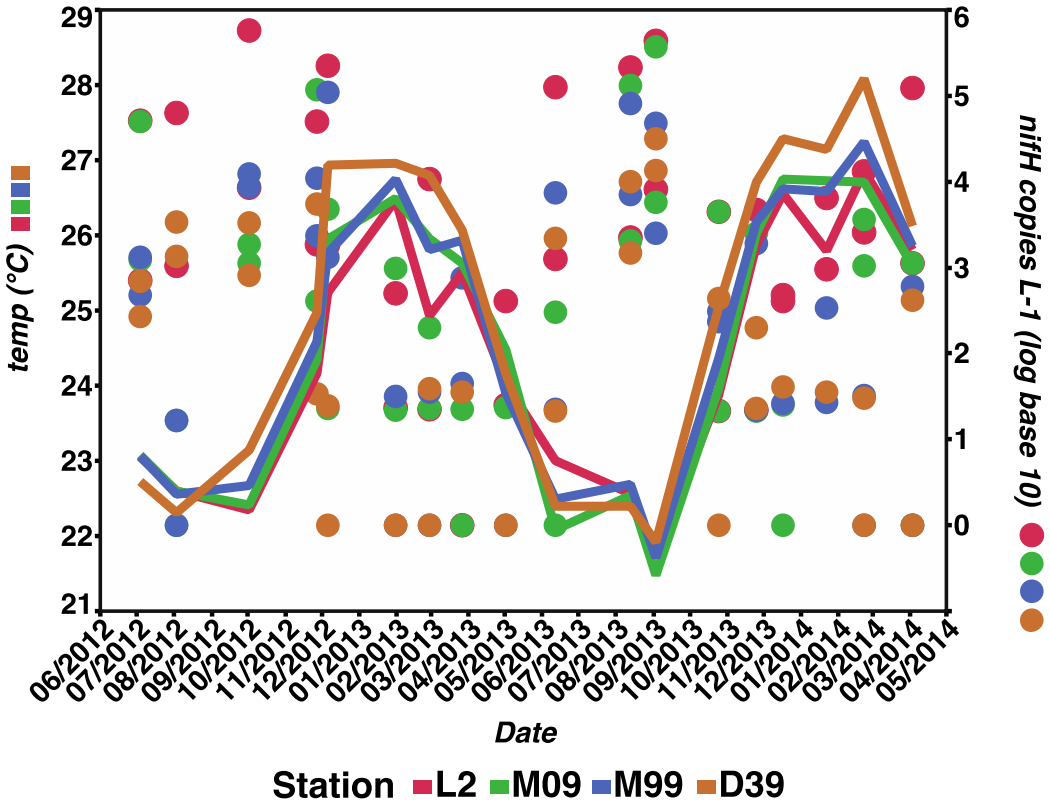

Figure 4a: The relationship between UCYN-A1 *nifH* L^{-1} (log base 10) and mean temperature (°C) across date grouped by station.

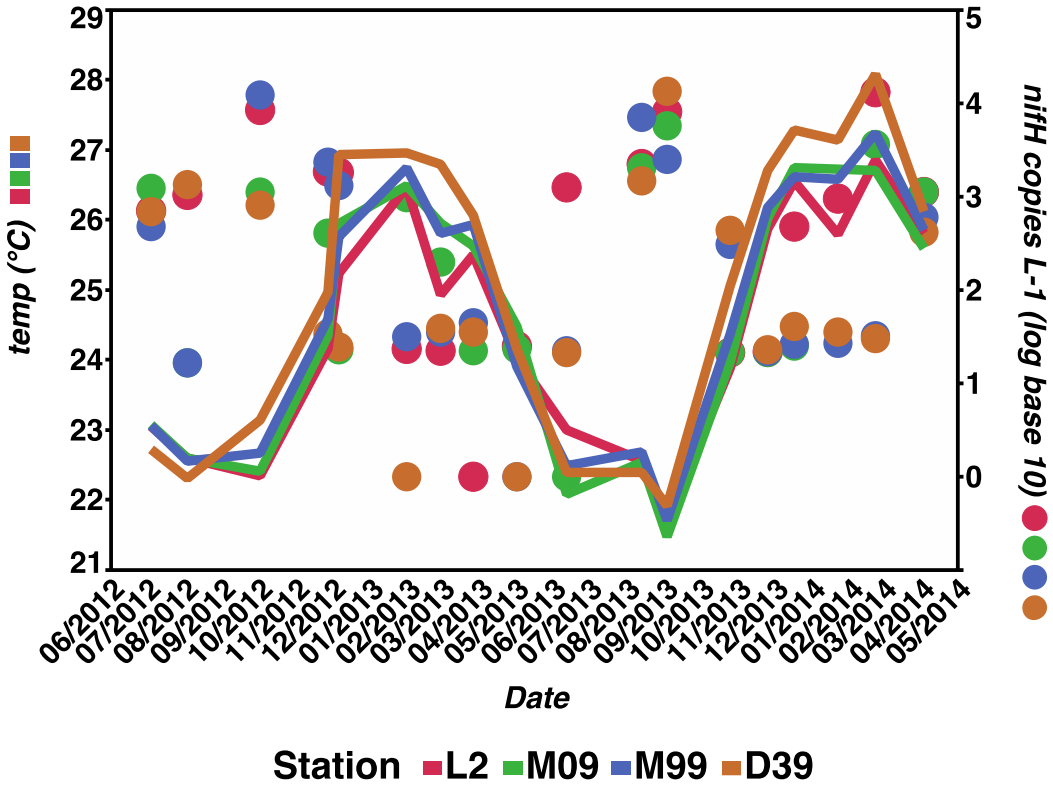

Figure 4b: The relationship between UCYN-A2 *nifH* L^{-1} (log base 10) and mean temperature (°C) across date grouped by station.

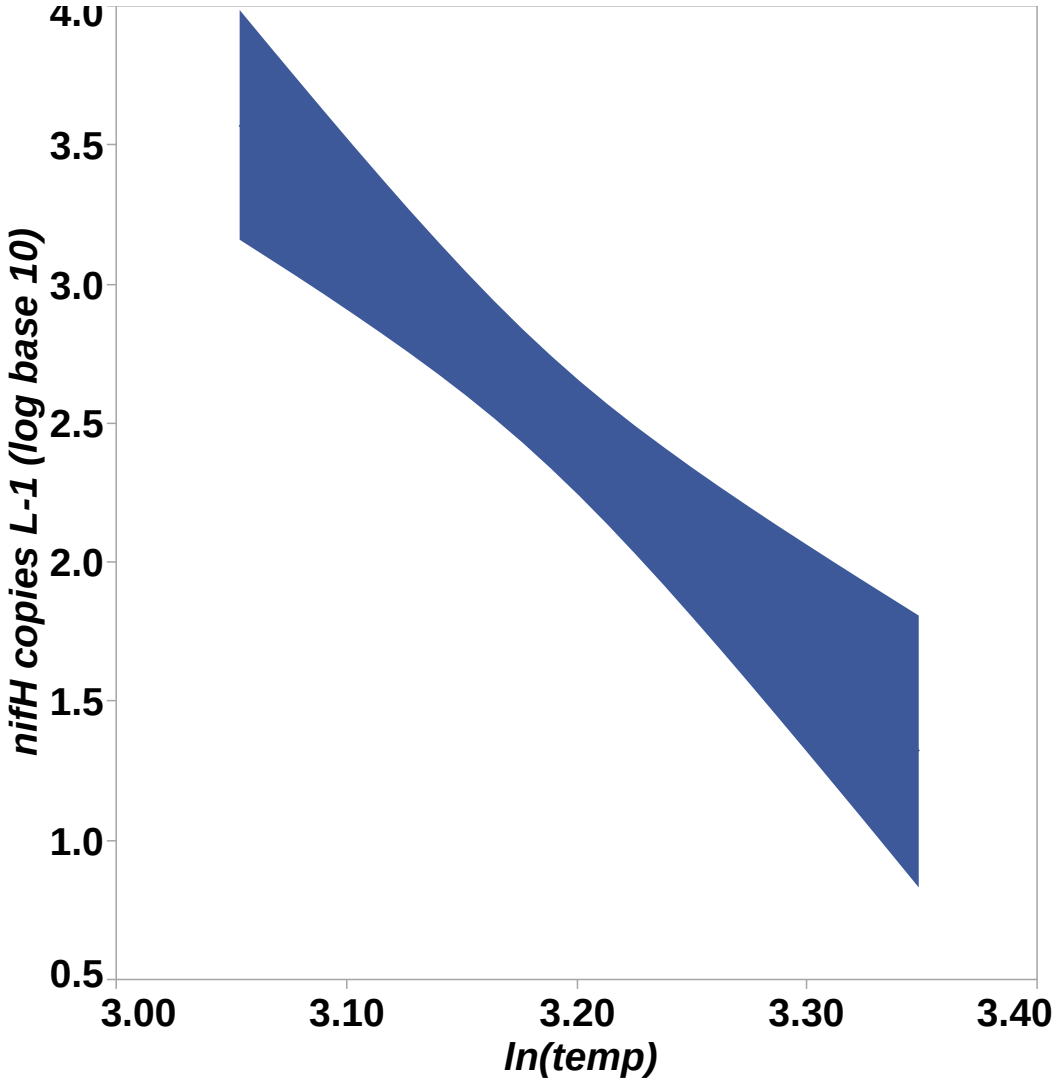

Figure 5: The relationship between *nifH* L^{-1} (log base 10) and ln(temp). (The blue band is the 95% confidence interval of the slope of adjusted means after accounting for other model effects.)

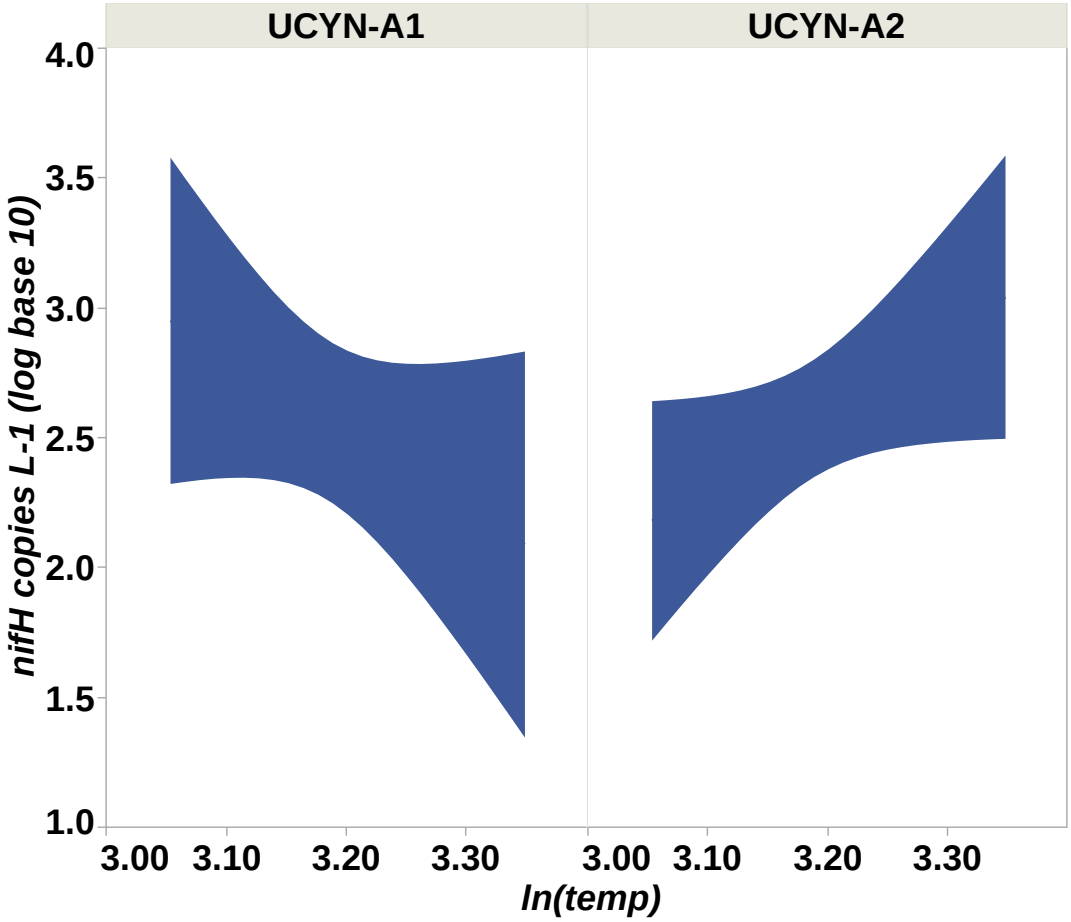

Figure 6: The relationship between *nifH* L^{-1} (log base 10) and ln(temp) separated by target. (The blue band is the 95% confidence interval of the slope of adjusted means after accounting for other model effects.)

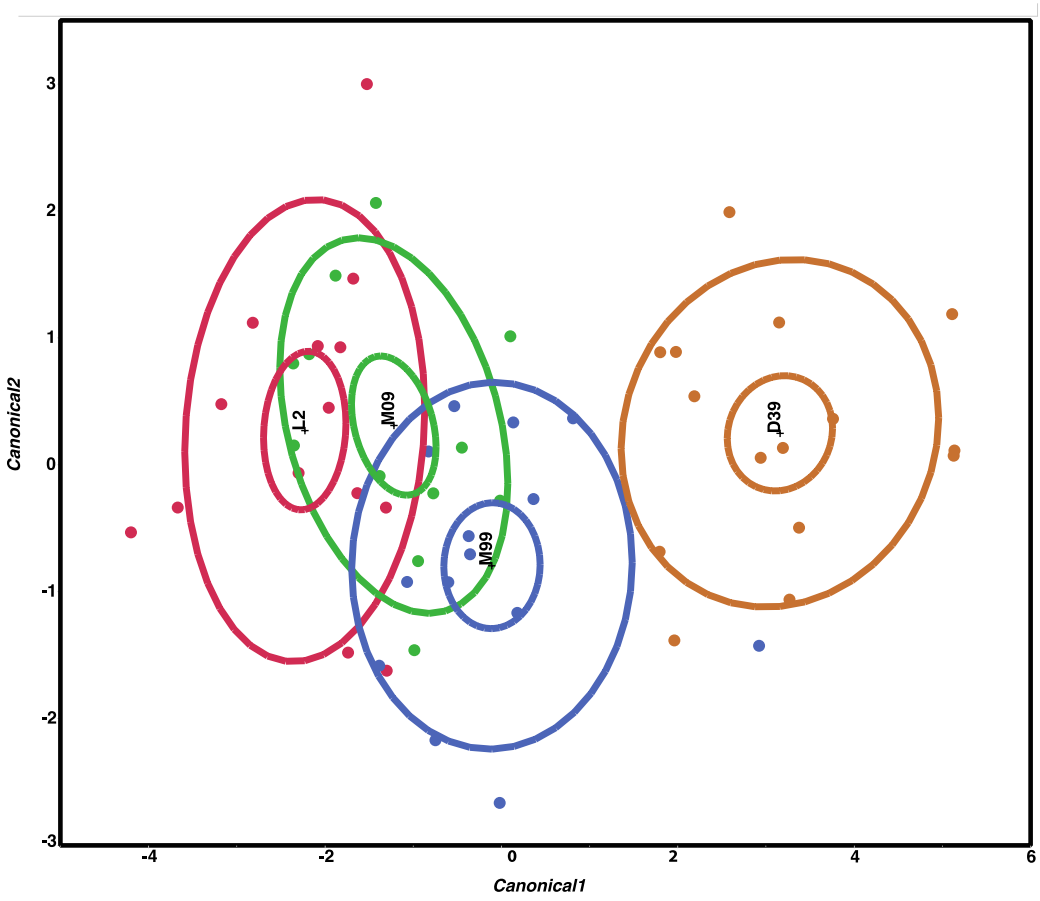

Figure 7: Discriminant Function Analysis (DFA) model to predict environmental character of station – Canonical Plot.

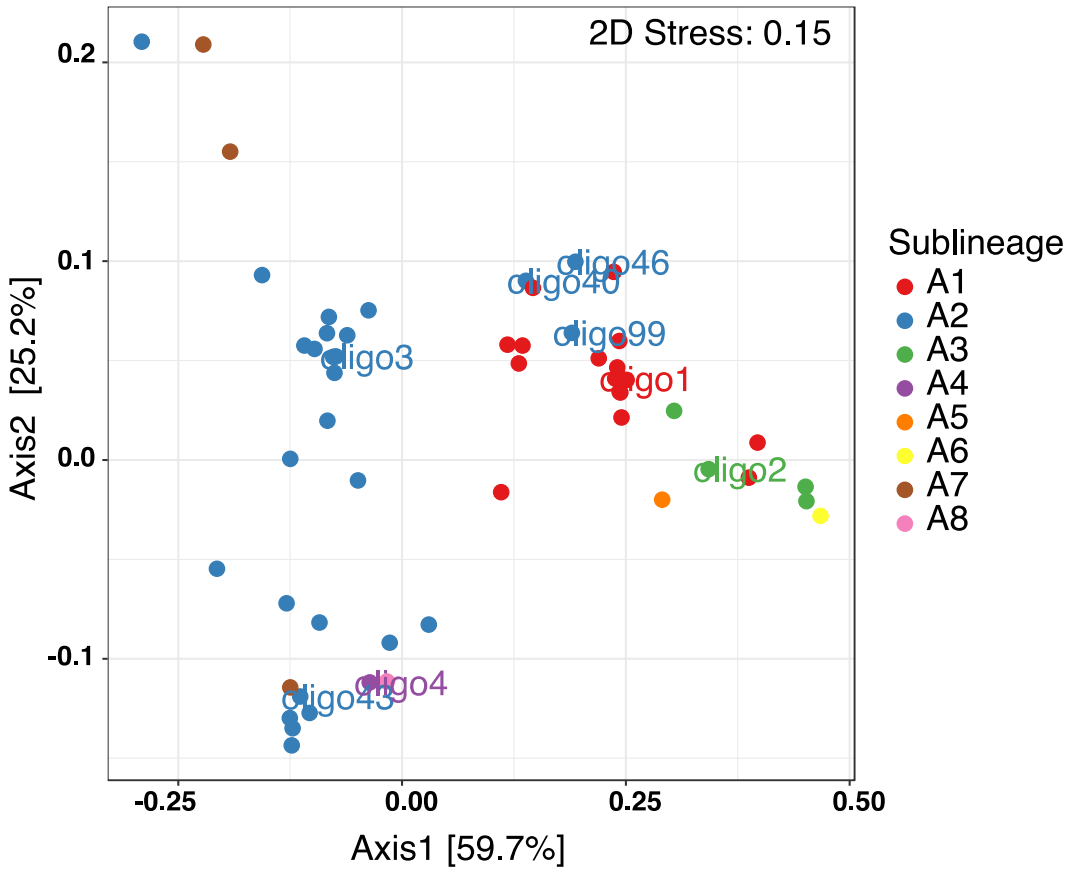

Figure 8: Metric Multidimensional Scaling (MDS) using the Bray-Curtis ecological index to determine dissimilarity between samples based on relative abundance.

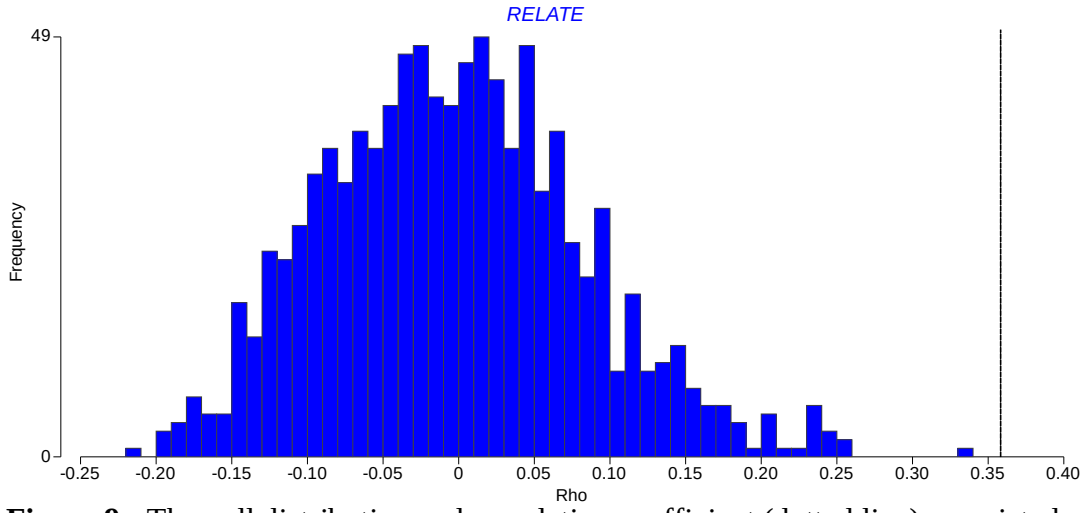

Figure 9: The null distribution and correlation coefficient (dotted line) associated with the four-variable solution ($\rho_s = 0.358$, $p < 0.001$). Null distribution is created via PRIMER BIO-ENV routine by permuting the data.

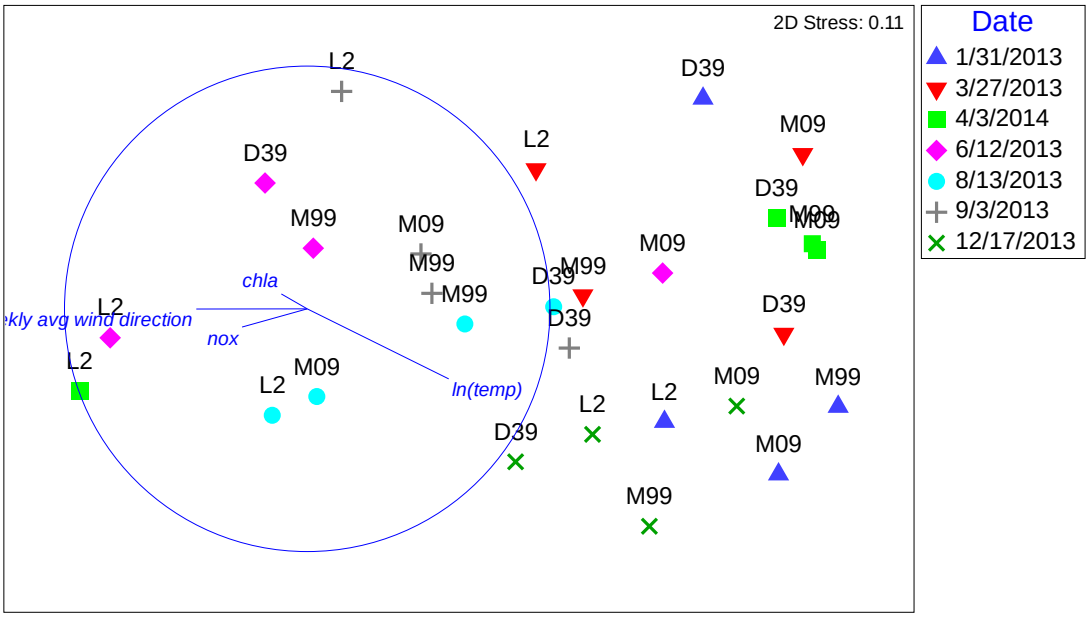

Figure 10: Nonmetric Multidimensional scaling (NMDS): plot of 28 biological samples, coded by date and station with best explanatory environmental variable overlay. Data represent sequence relative abundances; resemblances based on Bray-Curtis similarity. Plots are overlaid with best explanatory environmental variable vectors, pointing in the direction of increasing ln(temp), chla, nox, and weekly average wind direction. The length of the vectors shows the strength of the relationship.

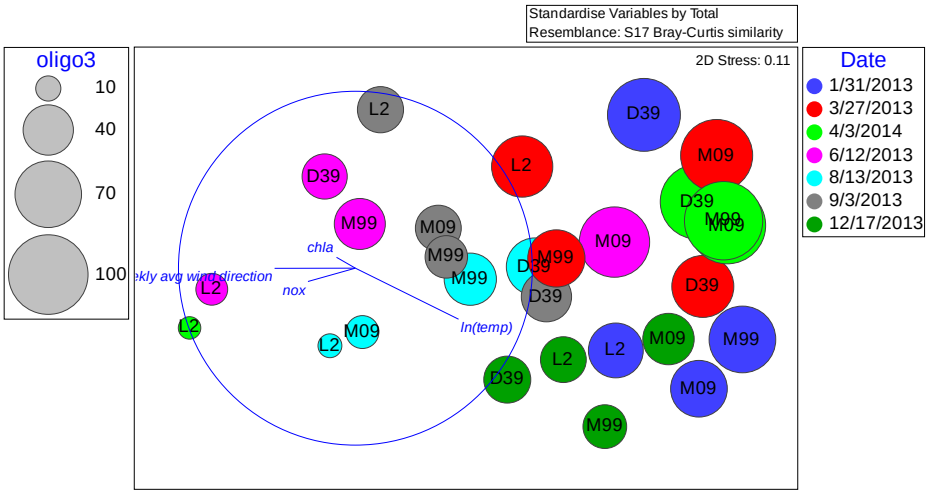

Figure 11a

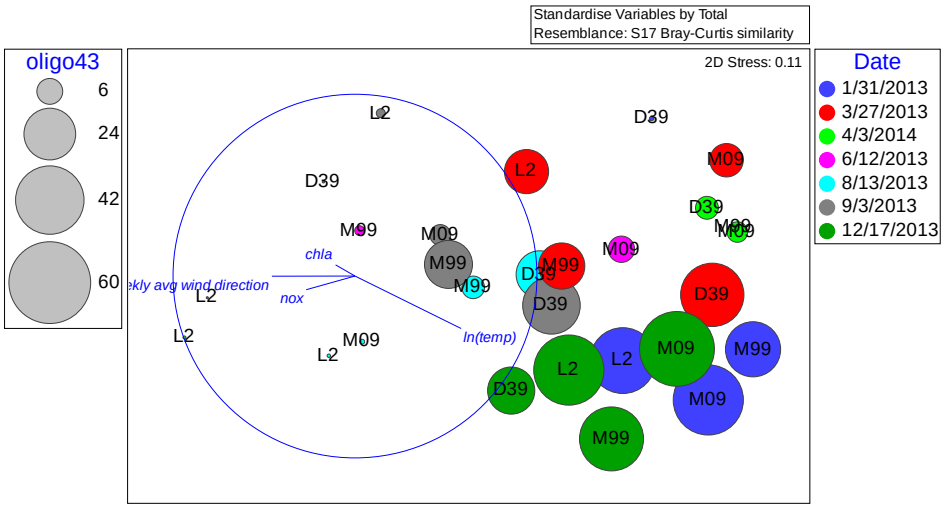

Figure 11b

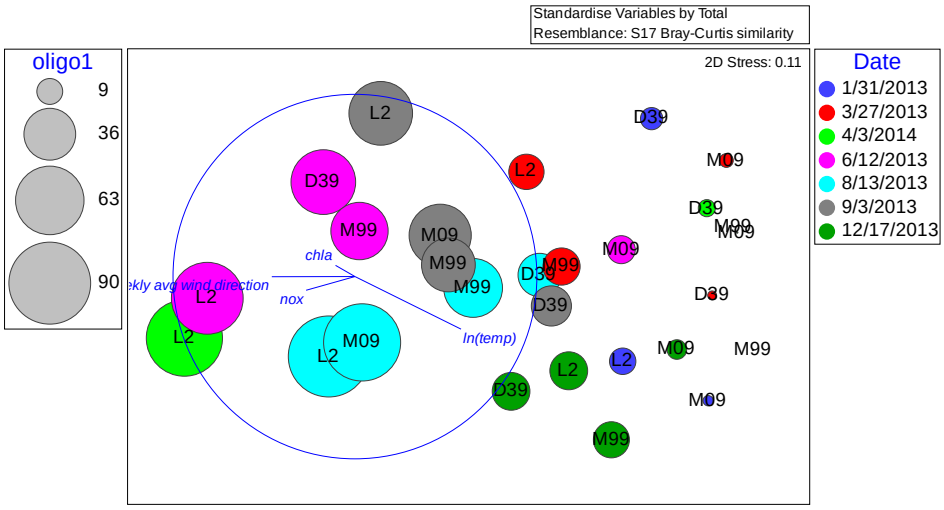

Figure 11c

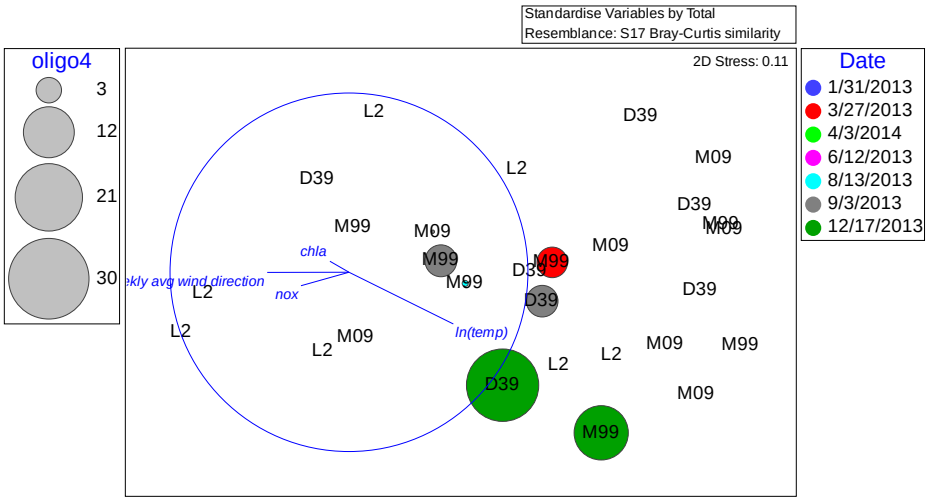

Figure 11d

Figure 11a-11d: Nonmetric Multidimensional scaling (NMDS): plot of 28 biological samples, coded by date and station with best explanatory environmental variable overlay. Data represent sequence relative abundances; resemblances based on Bray-Curtis similarity. Shown from top to bottom are the relative abundances of (a) oligo3, (b) oligo43, (c) oligo1, and (d) oligo4. The oligotype represented in the plot is defined in key at left of MDS plot. The key also correlates circle diameters with relative abundance for each oligotype. Note the variation in relative abundance from one oligotype to the other, with the small circles representing 10 relative abundance sequence reads for oligo3, nine for oligo1, six for oligo43, and three relative abundance sequence reads for oligo4. Colored bubbles offer a way to visualize what may drive relationships between communities.

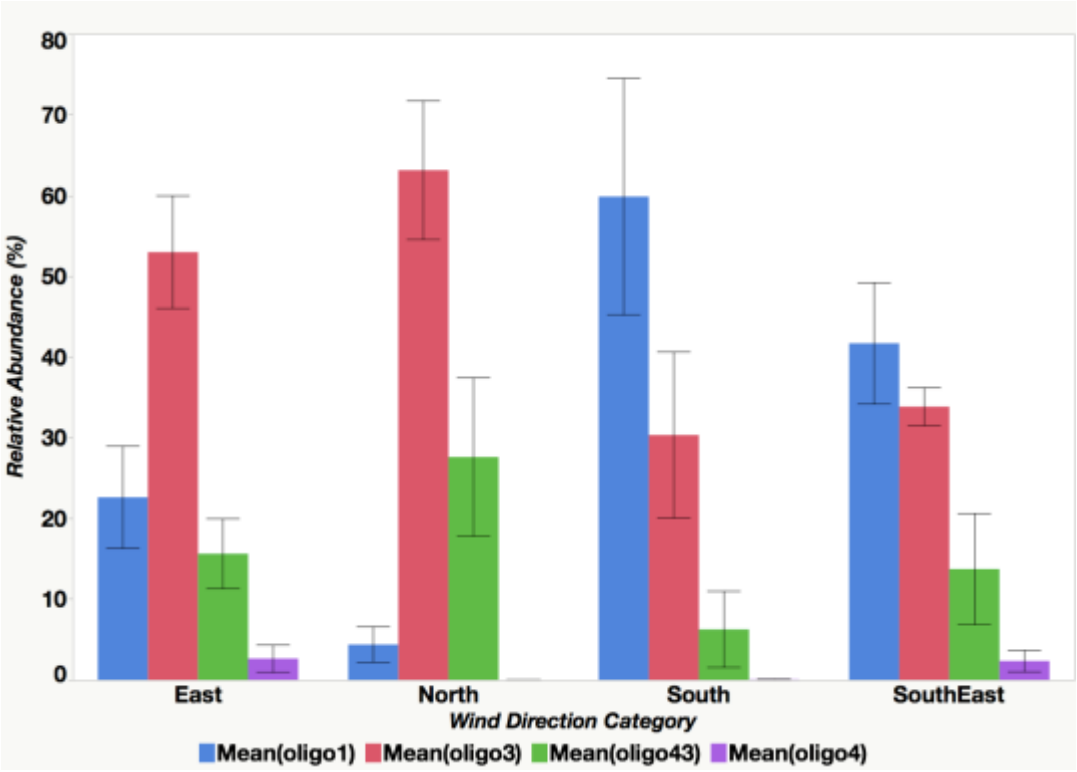

Figure 12: The relationship between mean oligotype relative abundance and weekly average wind direction category. Categories are defined as follows: north (0°-25°), east (90°-115°), southeast (120°-140°), and south (170°-190°).

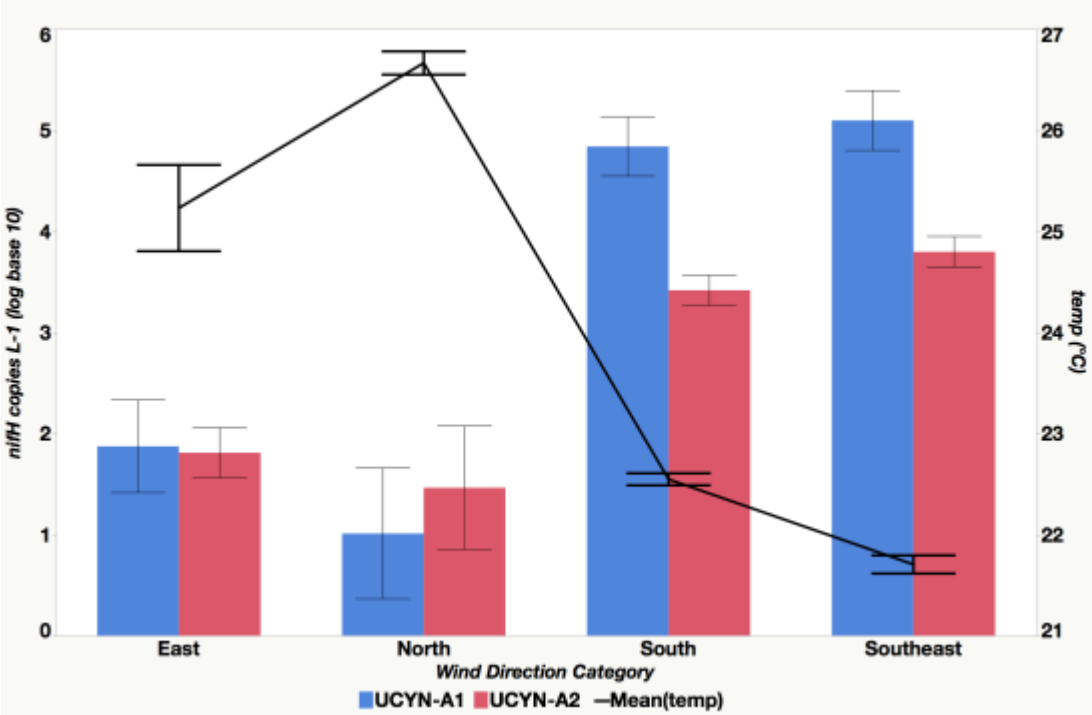

Figure 13: Relationship between mean UCYN-A1 and UCYN-A2 qPCR *nifH* L^{-1} (log base 10) and temperature across weekly average wind direction categories. Categories are defined as follows: north (0°-25°), east (90°-115°), southeast (120°-140°), and south (170°-190°).

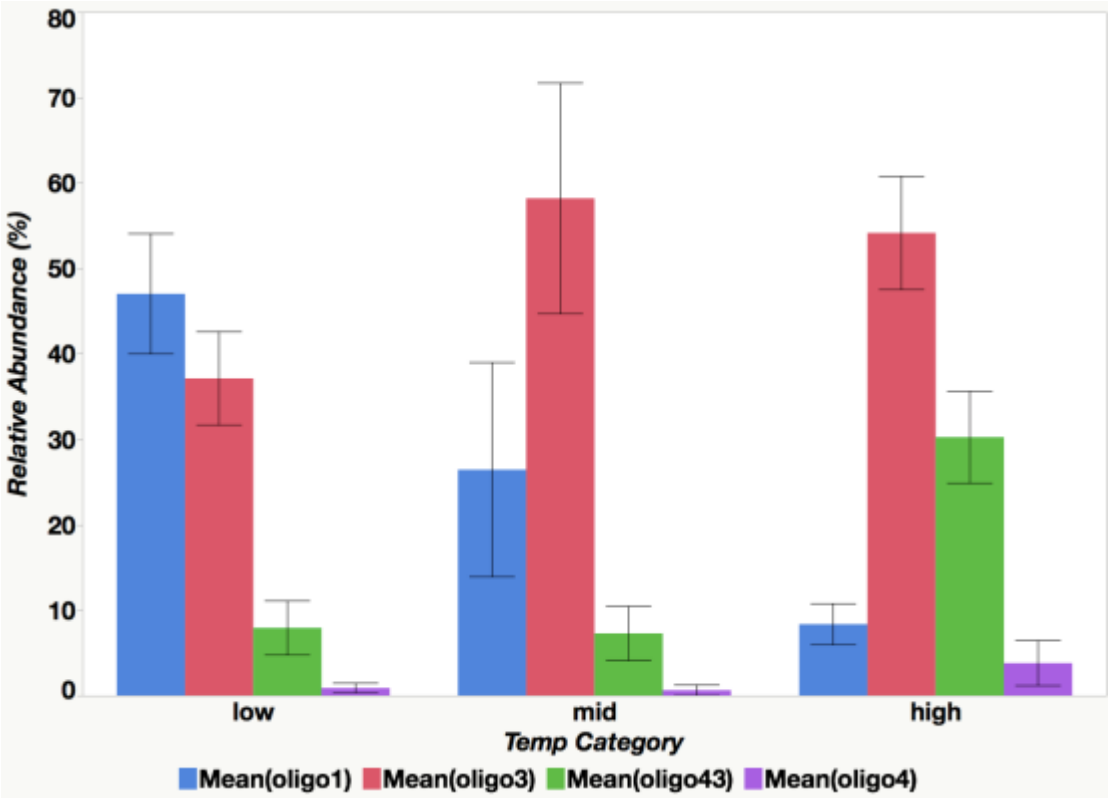

Figure 14: The relationship between mean oligotype relative abundance and temperature category. Categories are defined as follows: high (26°$\boldsymbol{C} \leq$ x), mid (23°$\boldsymbol{C} \leq \boldsymbol{x} < \boldsymbol{26}$°C), low (x<23°C).

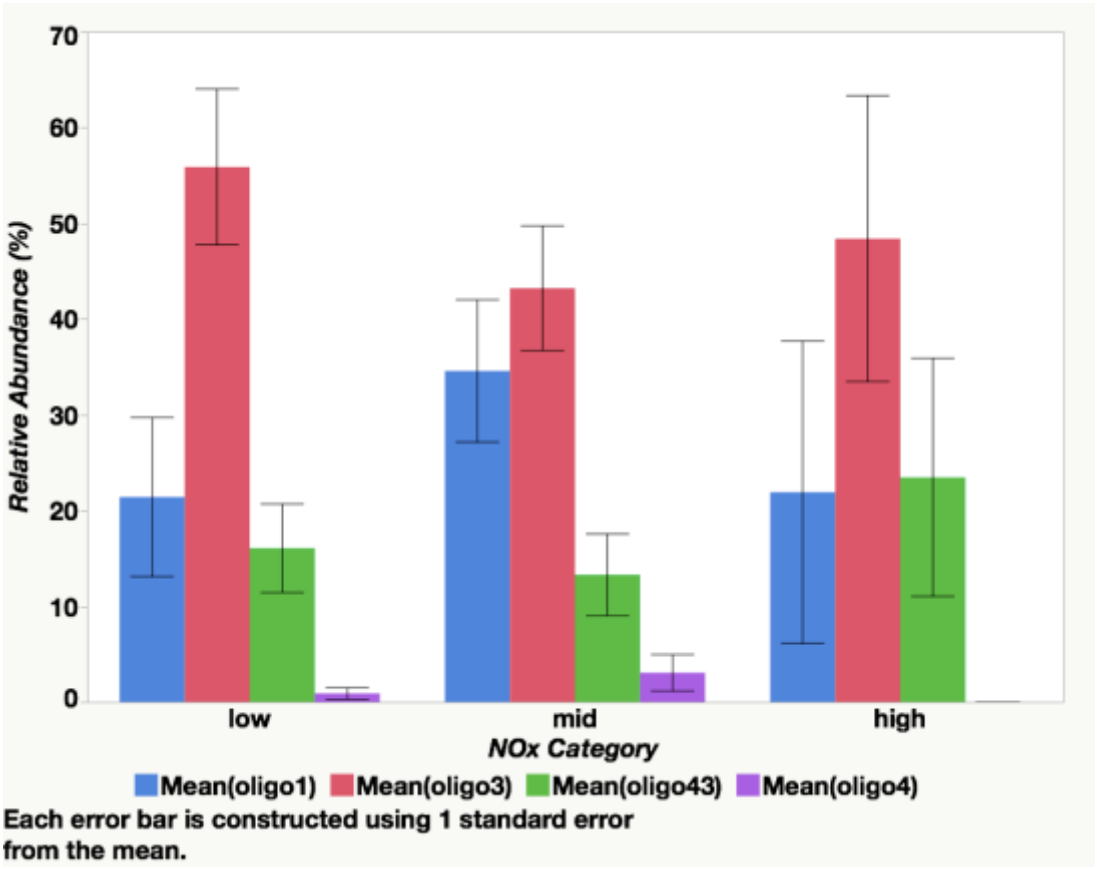

Figure 15: The relationship between mean oligotype relative abundance and NOx category. Categories are defined as follows: high ($0.075\mu M \leq x$), mid ($0.03\mu M \leq x < 0.075\mu M$), low ($x<0.03\mu M$).

CHAPTER 3

Conclusion

UCYN-A was detected year-round in the SLNC. Results indicate that the two major UCYN-A sublineages are found in all major environments of the lagoon. Given the lagoon's year-round N-limitation, findings suggest that UCYN-A plays a role in the ecology of SLNC during all seasons. A follow-up study of UCYN-A activity, including N_2 fixation rates or *nifH* transcripts, will be necessary to assess the total contribution of UCYN-A to lagoon N_2 fixation.

UCYN-A1 dominated qPCR abundance and dictated a seasonal pattern of abundance which peaked in September and October, when temperatures were lowest. Abundance could be predicted by a suite of nine environmental variables, with temperature being the most significant term. To evaluate the usefulness of this model beyond this study, further research is needed to determine if the same set of factors predict UCYN-A1 abundances at other times and places.

UCYN-A1 and UCYN-A2 correlated with different environmental characteristics. UCYN-A1 corresponded to the lower temperature and the low nutrient concentrations characteristic of the outer station, while UCYN-A2 corresponded to the inverse conditions. Findings support the notion that UCYN-A1 and UCYN-A2 are different ecotypes with some overlap in niches. UCYN-A is a globally distributed dizotroph. Different sublineages have distinct patterns of distribution and rates of N_2 fixation. Results from this analysis help to identify the factors driving the divergent biogeography of these sublineages, and consequently help in extrapolating local and global N_2 fixation rates. Subsequent analysis is needed to confirm the theory that UCYN-A1 and UCYN-A2 are different ecotypes. This question could be addressed in a SLNC follow-up study, involving more frequent monitoring as well as additional data collection such as salinity to differentiate water masses, as well as stations outside the lagoon, to distinguish how changes in the lagoon community relate to

those outside the SLNC.

Additional UCYN-A diversity was identified within the major sublineages (UCYN-A1 and UCYN-A2), as well as in other UCYN-A sublineages. Findings indicate that UCYN-A is a diverse group of diazotrophs. Similar environmental variables explained the relative abundances of sublineages as well as their associated oligotypes, with the notable exception being a UCYN-A2 oligotype (oligo43) which had relative abundance patterns distinct from the dominant UCYN-A2 oligotype (oligo3). This surprising finding indicates that fine-scale genetic diversity is also important in understanding UCYN-A distributions, as environmental predictors are not always consistent for all oligotypes within the same sublineage. Results suggest that oligotype level diversity is also important in allowing UCYN-A to thrive globally. Oligotyping analysis will therefore likely be an important aspect of future UCYN-A studies.

UCYN-A is a globally distributed diazotroph consisting of a diverse group of strains. UCYN-A distributions are non-random and can be predicted based on ecological characteristics. Abundance is typically tied to a multitude of factors rather than one or two environmental indicators. Sublineage is a good first-order indicator of environmental preference, and may even signal distinct ecological divisions. However, at present, even sublineage level diversity is frequently confused in UCYN-A studies because the UCYN-A2 primer (Thompson et al., 2014) targets UCYN-A2, UCYN-A3, and UCYN-A4. Further, as indicated in this study, finer-scale oligotype level diversity cannot be ignored, owing to significant divergences between oligotypes within the same sublineage and their response to environmental forcing. The relationship between diversity, abundance, and biogeography is both specific and complex and will require further analysis to improve estimates of marine N_2-fixatation and its role in global N budgets