UNIVERSITY OF CALIFORNIA
SANTA CRUZ

# INSIGHTS INTO MARINE UNICELLULAR CYANOBACTERIAL AND NON-CYANOBACTERIAL DIAZOTROPHS THROUGH SINGLE-CELL ANALYSES

dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

OCEAN SCIENCES

by

**Katie J. Harding**

June 2021

The Dissertation of Katie J. Harding is approved:

_____
Professor Jonathan P. Zehr, Chair

_____
Dr. Xavier Mayali

_____
Professor Chad Saltikov

_____
Dr. Matthew Mills

_____
Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

**Table of Contents**

## LIST OF TABLES AND FIGURES

**Chapter 2: Symbiotic Unicellular Cyanobacteria Fix Nitrogen in the Arctic Ocean**



**Chapter 3: Direct cell-specific measurements show $N_2$ fixation by particle-attached non-cyanobacterial diazotrophs in the North Pacific Subtropical Gyre**



**Chapter 4: Visualization of two uncultivated marine diazotrophs, the non-cyanobacterium 'Gamma-A' and the symbiotic cyanobacterium UCYN-A, using fluorescent in situ hybridization of the *nifH* gene (geneFISH).**



**Appendix 1: Supplementary information for Chapter 2**

**Appendix 2: Supplementary information for Chapter 3**



**Appendix 3: Supplementary information for Chapter 4**

## Abstract

Katie J. Harding

**Insights into Marine Unicellular Cyanobacterial and Non-Cyanobacterial Diazotrophs through single-cell analyses**

Nutrient availability affects primary productivity, and nitrogen (N) is the limiting nutrient over large regions of the ocean. Biological N fixation (BNF) is the conversion of atmospheric dinitrogen ($N_2$) to bioavailable forms of N (ammonium and amino acids). A select group of Bacteria and Archaea, known as diazotrophs, are the only organisms capable of this conversion. BNF is a critical source of N in nutrient-depleted (oligotrophic) marine waters. As such, BNF studies are a fundamental component for better understanding primary productivity and carbon cycling, as well as parameterizing predictive future global ocean models. Unicellular diazotrophs are widespread in marine systems and, in some cases, important contributors of BNF. This dissertation investigates unicellular diazotrophs at the single-cell level, including UCYN-A, a cyanobacterial diazotroph that lives in symbiosis with an algal host, non-cyanobacterial diazotrophs (NCDs), which are putative heterotrophs, and a specific NCD known as Gamma A, which is a commonly occurring organism throughout the tropics and subtropics that belongs to the class Gammaproteobacteria. The single-cell investigations used in these studies can provide high-resolution insights that are not possible from community-level analyses of BNF, such as the $N_2$ fixation rate of individual cells and cell morphology. Single-cell nanoscale secondary ion mass spectrometry (nanoSIMS) analyses were used in chapters 2 and 3 to investigate the cell-specific $N_2$ fixation rates of UCYN-A in the

cold waters of the Bering and Arctic Seas where $N_2$ fixation was not previously believed to be possible (chapter 2), and NCDs in the North Pacific Subtropical Gyre where we demonstrated the first direct evidence of NCD $N_2$ fixation on open ocean particles (chapter 3). Chapter 4 focused on a specific uncultivated NCD, Gamma A, using geneFISH to visualize single cells in natural seawater samples. We successfully identified Gamma A cells by developing a fluorescent *in situ* hybridization (FISH) probe targeting the diazotrophic functional marker gene *nifH*, revealing Gamma A cells are most often attached to particles. Together these studies provide new insights into the distributions, activities, and physiologies of geographically widespread, unicellular diazotrophs, thus improving our understanding of their roles as sources of fixed N in the surface ocean.

## Acknowledgments

This dissertation would not have been possible without the wonderful people who shared this journey with me. First and foremost, thank you to my advisor Jonathan Zehr. As an undergraduate, Jon inspired me to study marine microbiology, and I feel extremely grateful, he allowed me to pursue that goal in his lab as a graduate student. Jon offered a place to find and follow my interests with guidance, support, and understanding-- especially through challenging projects. As part of Jon's group, I was able to take part in once-in-a-lifetime opportunities I will forever be grateful for. A huge thank you to my technical advisor Xavier Mayali at Lawrence Livermore National Laboratory (LLNL), who encouraged me to apply to study at LLNL. I have learned so much working with him and the group there. Xavier was always encouraging and patient in 'try, and see if it works' situations, and he was always available to chat, give advice and provide feedback. I am very grateful to my committee member Matthew Mills who was an enormous asset in my orientation to nanoSIMS and always provided supportive feedback in committee meetings. Thank you to Chad Saltikov for being part of my committee and for providing engaging questions and ideas towards my project development and completion.

I want to thank past Zehr Lab members who helped me find my feet in the world of $N_2$ fixation, especially Mary Hogan, Hanna Farnelid, and Irina Shilova. I am extremely grateful to Kendra Turk-Kubo, and Rosie Gradoville in the Zehr Lab group for always offering emotional and science support even when extremely busy. Thank you to all the Zehr Lab members, past and present, for making it such an enjoyable place to work. A special thank you to Esther Mak for her always sharing her optimism and going out of her way to be

helpful no matter what she has going on. Additionally, I would like to thank my cohort for their friendship, support, and creating great memories.

Finally, the culmination of this degree and everything that led to this point is due to my parents' unending love and support. They have made so many sacrifices over the years, which allowed me the time and space to follow my curiosities into a career. And lastly, thank you to my wonderful husband, who deserves an honorary degree for all the papers, presentations, and stress he has had to patiently endure over the years.

The author wrote this dissertation in support of requirements for the degree Doctor of Philosophy in Ocean Science at the University of California, Santa Cruz. The research is funded in part by the LLNL Graduate Scholars Program and is not a deliverable for any United States government agency. The views and opinions expressed are those of the author and do not state or reflect those of the United States government or Lawrence Livermore National Security, LLC.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-TH-823869.

The text of this dissertation includes reprints of the following previously published material and submitted manuscript:

1.) Harding, K., Turk-Kubo, K. A., Sipler, R. E., Mills, M. M. and Bronk, D. A., Zehr, J.P.: Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean, PNAS, 115(52), doi:10.1073/pnas.1813658115, 2018.

2.) Harding, K. Turk-Kubo, K. A., Mak, E.W.K., Weber, P.K., Mayali, X., Zehr, J.P.: Direct cell-specific measurements show $N_2$ fixation by particle-attached non-cyanobacterial diazotrophs in the North Pacific Subtropical Gyre. In Review.

J.P. Zehr (a co-author listed on both manuscripts) directed and supervised research, which forms the basis for the dissertation.

# Chapter 1: Introduction

The oceans cover approximately ¾ of the Earth's surface and are responsible for about half of biological productivity. The productivity of the oceans is important for marine food webs but also the cycling of carbon and trace gases between the atmosphere and the oceans. Thus, marine ecosystems play critical roles in food webs and greenhouse gas formation and consumption. Although life in the ocean includes the easily visible whales and fish, the less visible microscopic microorganisms are much more abundant and play critical roles in primary productivity and biogeochemical cycles. Photosynthetic unicellular eukaryotes and cyanobacteria, or phytoplankton, convert carbon dioxide ($CO_2$) into organic carbon (C), known as primary productivity. Primary productivity forms the base of the food chain in both marine and terrestrial ecosystems. In marine systems, the organic C produced by phytoplankton is transferred up the food web into larger organisms such as zooplankton, fish, and whales. However, fixed C is ultimately recycled back to $CO_2$ by heterotrophic bacteria. The productivity of marine food webs is controlled by the availability of nutrients, just as it is in terrestrial ecosystems. One of the most important nutrients that are often present in limiting amounts is nitrogen (N), even though $N_2$ gas is abundant in Earth's atmosphere. Only certain microorganisms can draw upon this pool of nitrogen, called $N_2$-fixers or diazotrophs, which are important in marine and terrestrial ecosystems. Marine $N_2$-fixers are the subject of this dissertation.

Marine primary producers remove as much $CO_2$ from the atmosphere as terrestrial plants accounting for half of the oxygen produced globally (Falkowski et al., 1998; Field et al., 1998). Phytoplankton in the surface ocean are responsible for most of the primary productivity in the oceans. Marine primary production removes $CO_2$, a greenhouse gas, from the atmosphere, generating organic C, which forms the basis of the marine food web. Organic C can then be incorporated into large cells or fecal pellets, which sink out of the euphotic zone (> 200 m), effectively removing C and nitrogen (N) from atmospheric circulation by sequestration in deep water and sediments. Up to 1/3 of phytoplankton fixed C is sequestered in the deep sea facilitated by biological incorporation and sinking, causing C transfer against the concentration gradient known as the biological carbon pump (Falkowski et al., 1998; Karl et al., 2002; Siegenthaler and Sarmiento, 1971). Primary productivity is intrinsically linked with biogeochemical cycles that dictate nutrient availability, all of which include microbial processes (Falkowski, 2015; Falkowski et al., 2008). The projects summarized in this dissertation focus on the microbial process of biological nitrogen fixation (BNF), an important component of the N cycle that provides N in nutrient-limited surface waters and fuels primary productivity and C sequestration.

*Nitrogen Budget*

Primary productivity is limited by the availability of nutrients such as N, phosphorus, and iron (Dugdale and Goering, 1967; Falkowski et al., 1998). In vast areas of the global ocean, primary productivity is limited by N (Capone et al., 1997; Dugdale and Goering, 1967; Falkowski et al., 1998; Lipschultz and Owens, 1996). N

is essential for life as it is used in cells for nucleic acids and proteins. There is a vast supply of dinitrogen ($N_2$) in the atmosphere, but most organisms cannot use it. The select group of organisms that can use $N_2$ are known as diazotrophs and can break the stable triple bond that makes the molecule nearly inert to all other organisms (Howard and Rees, 1996). Nitrate ($NO_3^-$), ammonium ($NH_4^+$), and urea are more biologically accessible forms of N. $NH_4^+$ and urea are products of degraded organic matter such as dead cells and are rapidly recycled within the euphotic zone (Ward, 2012; Zehr and Kudela, 2011). $NO_3^-$ primarily originates from terrestrial sources, including weathering of rocks and riverine input. In the open ocean, away from terrestrial sources, $NO_3^-$ is mixed into the surface layers from nutrient-rich deep water. Deepwater is nutrient-rich due to the mineralization of organic matter sinking below the euphotic zone, which produces $NH_4^+$ (Ward, 2012; Zehr and Kudela, 2011). Below the photic zone, $NH_4^+$ is oxidized to $NO_3^-$ through nitrification, resulting in increasing concentrations of $NO_3^-$ as the water ages (Caffrey et al., 2007; Wuchter et al., 2006). However, stratification and quiescent waters limit vertical mixing in the subtropics and tropics, resulting in $NO_3^-$ depleted surface waters (Karl et al., 2002). $NH_4^+$ and $NO_3^-$ are involved in biogeochemical cycles apart from acting solely as N sources. As mentioned, $NH_4^+$ serves as an electron and energy source in ammonia oxidation (Caffrey et al., 2007; Wuchter et al., 2006) but also in anaerobic ammonia oxidation known as anammox (Van de Graaf et al., 1995). $NO_3^-$ is used as an alternative electron acceptor ($NO_3^-$) in anaerobic denitrification (Ward, 2012; Zumft, 1997). Primary productivity depends on bioavailable forms of N; therefore, knowing

the processes involved and balances between N sources and sinks provides a better understanding of primary productivity rates and limitations.

The overall productivity of a system depends on the balance between sources of fixed N and sinks of fixed N. In marine systems, the balance is defined by denitrification and anammox as losses of fixed N and deepwater mixing, atmospheric deposition, and BNF as sources of fixed N (Codispoti, 2007; Galloway et al., 2004; Gruber and Sarmiento, 1997; Duce et al., 2008). Denitrification is an anaerobic process that removes bio-available $NO_3^-$ from ecosystems through sequential reduction to $N_2$ gas ($NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$) mediated primarily by heterotrophic bacteria that use oxidized inorganic N compounds for respiration in the absence of $O_2$ (Ward, Capone, & Zehr, 2007, Gruber 2008). The other main loss of $NO_3^-$ is through anammox. Anammox is an alternative microbial respiration strategy that uses $NO_2^-$ and $NH_4^+$ for anaerobic autotrophic growth-producing $N_2$ (Van de Graaf et al., 1995; Kartal et al., 2012; Ward et al., 2007). Both denitrification and anammox occur in low $O_2$ and anaerobic environments such as sediments and oxygen minimum zones. Although fixed N is lost strictly in low $O_2$ environments, it affects the productivity of the larger system as the availability of fixed N depends on the balance of losses and gains.

An evaluation of the global marine N budget revealed an imbalance: N sinks such as denitrification and anammox were much larger than N sources such as BNF. The missing source of N was thought to be an underestimation of BNF in the surface ocean (Codispoti, 2007; Gruber and Sarmiento, 1997). BNF represents one of the

largest uncertainties of the global N budget (Galloway et al., 2004), and global marine estimates vary widely from 100 to 200 Tg N fixed annually (Karl et al., 2002; Wang et al., 2019). The considerable variation and potential underestimation of BNF are due to many factors (Zehr and Capone 2021). $N_2$ fixation rate measurements are relatively sparse in the expansive ocean, and there may be $N_2$-fixers that have not yet unidentified (Ward et al., 2007). In addition, there are methodological problems for directly measuring rates. For example, recent studies showed BNF rates determined using calculated $^{15}N_2$ concentrations were overestimated, creating a significant underestimation of the BNF rates (Mohr et al., 2010). Additionally, the unicellular cyanobacteria UCYN-A and its widespread distribution were only acknowledged as a globally significant contributor to BNF within the last two decades (Farnelid et al., 2016; Martínez-Pérez et al., 2016; Moisander et al., 2010; Montoya et al., 2004). Even more recently, previously unidentified NCDs within *Planctomycetes* and *Micavibrionaceae* were found to have the potential to fix $N_2$, which have never previously been linked with BNF (Delmont et al., 2018; Poff et al., 2021). Although whether *Planctomycetes* fix $N_2$ in the surface ocean has not been demonstrated, their recent identification indicates that the known diazotrophic diversity is continuing to expand. The projects described in this dissertation endeavor to provide detailed information on BNF rates of UCYN-A in unexpected locations (cold northern latitudes), measure BNF rates of largely ignored NCDs in the surface ocean, and gain insight into the functional lifestyle of the commonly occurring NCD Gamma A, all of which will help parameterize global BNF estimates.

*Biological Nitrogen Fixation*

In the open ocean, away from terrestrial sources of N, BNF is an essential source of bioavailable (fixed) N. BNF is one of the main sources of N responsible for maintaining the balance of fixed N in the euphotic zone due to N loss via sinking organic matter, these additions of fixed N allow for C sequestration (Galloway et al., 2004, 2008; Luo et al., 2012; Sohm et al., 2011). N fixed by BNF can fuel up to 50% of the primary productivity in oligotrophic regions (Karl et al., 2002). Diazotrophs are a select but diverse group of Bacteria and Archaea and the only organisms capable of BNF, which allows organisms to use atmospheric $N_2$ as a N source to produce ammonia which is then assimilated into amino and nucleic acids (Capone et al., 2005; Karl et al., 2008; Zehr and Kudela, 2011). $N_2$ fixation is carried out by the highly conserved nitrogenase enzyme (Howard and Rees, 1996), a two-component metalloenzyme consisting of an iron (Fe) protein and a molybdenum iron (MoFe) protein. Alternative metal-containing nitrogenases exist, but the Fe/MoFe enzyme is the predominant enzyme, especially in the surface ocean (Farnelid et al., 2011; Zehr et al., 2003; Zehr and Capone 2021; Zhang et al., 2016). Interestingly, the nitrogenase enzyme is irreversibly damaged by $O_2$, so diazotrophs have developed various $O_2$ protection strategies to carry out the reaction (Berman-Frank et al., 2003; Inomura et al., 2017; Paerl et al., 2018). The conversion of $N_2$ into ammonia is energetically expensive due to the stability of the N triple bond. The reaction requires 16 ATP and eight electrons per molecule $N_2$ reduced according to the equation: $N_2 + 8H^+ + 8e^- + 16MgATP \rightarrow 2NH_3 + H_2 + 16MgADP + 16Pi$ (Howard and Rees, 1996). Aside from

the ATP required for the $N_2$ fixation, the reaction also loses energy in the form of hydrogen ($H_2$) in a 1 to 1 stoichiometry with $N_2$. However, much of the $H_2$ is directly recycled back to the cell through efficient uptake hydrogenases recuperating some of the energy lost in BNF (Wilson et al., 2010a, 2010b). The rate of this energetically costly process can alleviate N limitation in ecosystems, thus fueling biological productivity.

Diazotrophs are diverse, with $N_2$ fixation capabilities scattered throughout the phylogenetic tree of prokaryotes. The diversity can be grouped into four general functional categories of open ocean diazotrophs. The first is filamentous cyanobacteria, such as *Trichodesmium* (Breitbarth et al., 2007; Capone et al., 1997), which can form large blooms visible to the unaided eye (Carpenter and Capone, 1992; McKinna, 2015). *Trichodesmium* is easily identified by microscopy as multi-filament colonies known as puffs and tuffs and occurs as individual filaments of ~100 cells known as trichomes (Gradoville et al., 2017; Karl et al., 2002). The second group is heterocyst-forming cyanobacteria, which form short filaments with differentiated heterocyst cells dedicated to $N_2$ fixation (Villareal, 1990, 1994). Heterocyst-forming diazotrophs such as *Richelia* are often found in symbioses with diatoms and, due to their heterocyst-containing morphology, can be identified by microscopy (Foster and Zehr, 2006). The third category is unicellular cyanobacteria, such as *Crocosphaera* and UCYN-A, which are single-cell cyanobacteria capable of $N_2$ fixation (Rippka and Waterbury, 1977; Thompson et al., 2012; Webb et al., 2009; Zehr et al., 2001). *Crocosphaera* are common in warm oligotrophic waters and can be free-living or

aggregated depending on the strain (Bench et al., 2016). Some unicellular cyanobacteria such as *Crocosphaera* can be visually identified by their pigmentation, which contains phycoerythrin (Webb et al., 2009; Zehr et al., 2001). UCYN-A is an interesting unicellular cyanobacterium; it lives in symbiosis with a haptophyte host *Braarudosphaera bigelowii* or closely related species and has a greatly reduced genome lacking photosystem II, Rubisco, and the entire TCA cycle. UCYN-A also lacks the genes for most pigments (Tripp et al., 2010; Zehr et al., 2008), and therefore cannot be identified visually with traditional procedures based on cyanobacterial pigment fluorescence. There are several sublineages of UCYN-A known as A1 to A6 (Turk-Kubo et al., 2017). Based on genome sequencing and visualization of UCYN-A1 and A2, the sublineages live in obligate symbioses with different but related haptophyte hosts (Bombar et al., 2014; Thompson et al., 2014; Tripp et al., 2010; Zehr et al., 2008), have different morphologies (Cabello et al., 2015; Cornejo-Castillo et al., 2016), and occupy different ecological niches (Farnelid et al., 2016; Turk-Kubo et al., 2017). The last category is non-cyanobacterial diazotrophs (NCDs), a broad, diverse group of Bacteria or Archaea that have the genetic potential to fix $N_2$ but are not cyanobacteria. NCDs in the surface ocean are putative heterotrophs or photoheterotrophs only identified by their *nifH* gene sequence (Farnelid et al., 2011), although metagenomic assemblies exist for some NCDs (Delmont et al., 2018; Salazar et al., 2019). NCDs have no known distinct morphology or pigmentation and therefore have not been visually identified. Very few surface ocean NCDs have been successfully cultured, and very little is known about these organisms, including

whether they contribute to BNF in the open ocean (Bombar et al., 2016; Moisander et al., 2017; Riemann et al., 2010). The projects included in this dissertation focus on the unicellular diazotrophs: UCYN-A, NCDs as a general group, and a specific, commonly occurring NCD known as Gamma A.

$O_2$ irreversibly damages the nitrogenase enzyme, yet diazotrophs manage to fix $N_2$ in the oxygenated surface ocean (Berman-Frank et al., 2003; Karl et al., 2002). Cyanobacterial diazotrophs, which evolve $O_2$ in oxygenic photosynthesis, have developed various strategies to mitigate nitrogenase inhibition by $O_2$. As a general simplification, cyanobacterial diazotrophs separate $O_2$ evolution from photosynthesis and BNF spatially or temporally. In N-limited environments, heterocyst-forming diazotrophs such as *Richelia* have a differentiated cell (heterocyst) among a chain of oxygenic photosynthesizing vegetative cells. The heterocyst fixes $N_2$ for the photosynthesizing cells but does not perform oxygenic photosynthesis, keeping the cellular $O_2$ concentration within the $N_2$-fixing cell low, protecting the nitrogenase enzyme (Foster and Zehr, 2006; Stewart, 1973; Villareal, 1990, 1994). However, there are exceptions (such as UCYN-A and *Trichodesmium*) where the temporal or spatial separation does not follow these general patterns (Berman-Frank et al., 2007; Cornejo-Castillo and Zehr, 2019).

Unicellular cyanobacteria such as *Crocosphaera* separate activities temporally, photosynthesizing and producing $O_2$ in the daylight hours and fixing $N_2$ at night. Temporal separation requires that the organic C gained through photosynthesis is stored until it is respired at night, fueling BNF (Dron et al., 2012;

Schneegurt et al., 1994; Shi et al., 2010; Toepel et al., 2008). UCYN-A is an interesting exception in unicellular cyanobacteria; as mentioned, it has a greatly reduced genome that lacks genes encoding photosystem II, among many others (Tripp et al., 2010; Zehr et al., 2008) and is a symbiont of an oxygen-evolving phototrophic haptophyte alga. Therefore, it does not produce $O_2$ and surprisingly fixes $N_2$ during the day (Church et al., 2005a; Munoz-Marin Del Carmen et al., 2019). Even though UCYN-A does not produce $O_2$, it still needs to protect the nitrogenase from $O_2$ produced from its symbiotic haptophyte host cell (Thompson et al., 2012). Bioinformatic analysis indicates UCYN-A may have a hopanoid-rich membrane to minimize $O_2$ diffusion (Cornejo-Castillo and Zehr, 2019) from its photosynthetic host.

Although NCDs do not produce $O_2$ through photosynthesis, the organisms still need to protect the nitrogenase from $O_2$ diffusion from the surrounding seawater. How NCDs mitigate $O_2$ diffusion in the surface ocean remains a mystery (Bombar et al., 2016). Some terrestrial NCDs, such as *Azotobacter vinelandii*, sustain high respiration rates to maintain a reduced intracellular environment, thereby protecting the nitrogenase (Inomura et al., 2017; Sabra et al., 2000). In a different strategy, a NCD isolated from the Baltic Sea (*Pseudomonas strain*) formed aggregates with exopolysaccharides under aerobic conditions (Bentzon-Tilia et al., 2015a; Paerl et al., 2018). However, both strategies are unlikely as they would be energetically challenging (Inomura et al., 2017) in the low-C waters of the oligotrophic ocean. Alternatively, evidence suggests that NCDs may live in the guts of invertebrates such

as copepods (Scavotto et al., 2015) or on particles (Farnelid et al., 2019; Pedersen et al., 2018), both microhabitats could harbor interior microaerobic zones (> 1 mm diameter) that would protect the nitrogenase from $O_2$ (Klawonn et al., 2015a; Paerl and Prufert, 1987; Ploug, 2001). Additionally, as NCDs are likely heterotrophic, the high concentration of organic C present in invertebrate guts and on particles (Bombar et al., 2016; Bonnet et al., 2013; Moisander et al., 2012) could create an ideal niche for heterotrophically fueled $N_2$-fixation.

BNF in the surface ocean has traditionally been thought to occur in the nutrient-poor, warm surface waters of the tropical and subtropical ocean by cyanobacteria (Gruber and Sarmiento, 1997; Karl et al., 2002). Due to the costly energetics of BNF, the process was only expected to occur when low N availability selects for diazotrophs providing a competitive fitness when less expensive N uptake strategies are not available (Gruber and Sarmiento, 1997; Karl et al., 2002). Additionally, early studies on diazotrophs relied on microscopy, and the visually distinct $N_2$-fixing cyanobacteria, *Trichodesmium,* have the highest abundances in warm (>20$^o$C), oligotrophic waters (Breitbarth et al., 2007; Carpenter and Capone, 1992). However, even after diazotroph identification was not solely reliant on microscopy, with the advent of DNA sequencing, studies continued to focus on cyanobacteria in warm oligotrophic water (Zehr and Bombar, 2015). This is likely because BNF is reliably found in these areas (Zehr and Bombar, 2015), and several well-studied diazotrophs, such as *Trichodesmium* (>20$^o$C) and *Crocosphaera* (>25$^o$C), have narrow temperature ranges limiting their distribution to warm waters

(Brauer et al., 2013; Stal, 2009). Within the last decade, the perception of where $N_2$ fixation can occur is shifting as increased studies investigate habitats outside the tropics and subtropics. Recently, diazotrophs have been found in unconventional locations, such as nutrient-replete coastal waters (Bentzon-Tilia et al., 2015b; Hamersley et al., 2011; Mills et al., 2020; Mulholland et al., 2012), polar regions (Blais et al., 2012; Shiozaki et al., 2017, 2020), and in the aphotic zone (Benavides et al., 2018a; Bonnet et al., 2013; Hewson et al., 2007; Poff et al., 2021; Rahav et al., 2013). As the geographic distribution of diazotrophs is found to be more global, it is clear there is no simple answer as to which factors control diazotroph distribution.

As the diversity of diazotrophs expands with sequencing approaches, so does understanding what factors influence BNF and diazotroph distribution. Light, temperature, and nutrients are the major factors that affect surface ocean diazotrophs, with the additional factor of C sources in the case of NCD. Light is essential for cyanobacterial diazotrophs, which provides the advantage of a reliable source of organic C through photosynthesis. Light is sufficient in the surface waters of the subtropics and tropics but would become a limiting factor if cyanobacterial diazotrophs sink out of the euphotic zone (Benavides et al., 2018a; Poff et al., 2021). Additionally, cyanobacteria at high latitudes would experience seasonal light limitation (Blais et al., 2012; Shiozaki et al., 2017). Interestingly, the NCD Gamma A abundances were found to have a significant positive correlation with light (Langlois et al., 2015; Moisander et al., 2014). This may imply the presence of light-absorbing pigments (Gómez-Consarnau et al., 2019), which could help meet the energy

demands needed for BNF, but there is not yet evidence to support this hypothesis (Inomura et al., 2017; Langlois et al., 2015; Moisander et al., 2017; Zehr and Capone, 2020). Additionally, NCDs in the surface ocean would depend on photosynthetically derived organic C and may therefore be indirectly influenced by light (Halm et al., 2012).

Open ocean BNF has historically been considered a cyanobacterial warm water process that was unlikely to occur at cold temperatures (Stal, 2009). As temperatures decrease, the theoretical cost of fixing $N_2$ increases, making an already energetically expensive process seem unlikely in cold climates. In laboratory studies on the unicellular cyanobacteria *Cyanothece,* $N_2$-fixation becomes inefficient below 18$^o$C because respiration rates slow to a point where intracellular $O_2$ levels exceed the minimum threshold for the nitrogenase enzyme to function (Brauer et al., 2013). Similarly, *Trichodesmium* is generally considered to have a temperature minimum of 20$^o$C (Breitbarth et al., 2007). However, diazotrophic activity has been identified in ice-covered (0$^o$C) Antarctic lakes (Olson et al., 1998; Singh and Elster, 2007), deep-sea sediments (4$^o$C) (Dekas et al., 2018), and cold waters (10$^o$C) at northern latitudes (Blais et al., 2012; Shiozaki et al., 2017) indicating not all diazotrophic activity is constrained by temperature.

Diazotrophs have an unlimited supply of N but can be limited by other nutrients such as phosphorus (P), iron (Fe), and for NCDs, organic C. Many studies hypothesize Fe is likely a limiting nutrient for diazotrophs for two reasons: 1) biologically available forms of Fe have very low concentrations in the open ocean,

and 2) the nitrogenase enzyme requires Fe (Berman-Frank et al., 2007; Paerl et al., 1987; Sohm et al., 2011). Therefore, diazotrophs have a higher Fe requirement than non-diazotrophic cells. The other nutrient that may limit BNF and diazotroph distributions is P (Deutsch et al., 2007; Sañudo-Wilhelmy et al., 2001; Wu et al., 2000). Teasing apart factors that control BNF is useful in understanding the constraints that affect diazotroph distributions and $N_2$ fixation rates, yet determining which nutrient is limiting at any given time is dependent on the diazotrophic species and the surrounding environment (Mills et al., 2004; Moisander et al., 2012; Severin et al., 2015). For example, some studies have found cyanobacterial diazotroph distribution is controlled by temperature, light, or nutrient availability (Moisander et al., 2010; Monteiro et al., 2011; Moore et al., 2009; Sohm et al., 2011; Stal, 2009), while other studies reveal a low correlation with such factors (Church et al., 2009; Ward et al., 2007).

*Methods to Study Biological Nitrogen Fixation*

Many diazotrophs do not have cultured representatives, including UCYN-A and Gamma A, and therefore cannot be studied in a controlled laboratory setting. For uncultured diazotrophs, the advent of molecular techniques revealed a vast diversity of diazotrophs that was invisible through culturing and microscopic observations. The known diversity continues to expand with new sequencing efforts (Delmont et al., 2018; Farnelid et al., 2011; Moisander et al., 2010; Poff et al., 2021; Zehr et al., 1998). Uncultivated diazotrophs are primarily studied using the diazotrophic functional marker gene *nifH,* which encodes a subunit of the Fe protein of the

nitrogenase enzyme complex (Zehr and Mcreynolds, 1989). Although other *nif* genes have been used (*nifK* and *nifD*), the *nifH* gene database is one of the largest databases outside of 16S rRNA genes (Zehr et al., 2003). Sequencing of the *nifH* gene can provide an overview of the relative abundance of diazotrophs present, while quantitative PCR (qPCR) can provide information on absolute abundances of groups using specifically designed primers and probes (Church et al., 2005b; Langlois et al., 2005; Zehr and Bombar, 2015). The same techniques can be used to determine the activity level of *nifH* gene transcription (Church et al., 2005a; Gradoville et al., 2020; Short and Zehr, 2007). However, the presence of *nifH* genes and transcripts do not necessarily imply active nitrogenase proteins or $N_2$ fixation activity due to high transcriptional (Chen et al., 1999) and post-translational regulation (Ohki et al., 1992). A molecular technique to visualize uncultured diazotrophs is fluorescent *in situ* hybridization (FISH), which can be used if the 16S rRNA gene sequences are known (Amann et al., 1995). Unfortunately, most NCDs are only identified by the *nifH* gene making 16S rRNA hybridization-based visualization impossible.

Community BNF rates (rates measured in bulk water samples) are the standard way to quantify the amount of $N_2$ fixation at a given location and represent the $N_2$ fixed by the whole diazotrophic community. The current gold standard for community $N_2$ fixation rate measurements is with stable isotope incubations using $^{15}N_2$ gas. $^{15}N_2$ gas is added to seawater samples using either pre-enriched $^{15}N_2$ seawater that is added to the environmental sample or by adding a bubble of $^{15}N_2$ that is allowed to equilibrate before releasing the bubble prior to the start of the incubation

(Mohr et al., 2010; White et al., 2020). The isotope amended bottles are incubated for a given period before particulate N is collected by filtration, then the $^{15}N$ content of the particulate nitrogen is measured by mass spectrometry. The isotope incubations and subsequent measurements provide a quantitative value of how much $^{15}N_2$ was incorporated into particulate organic matter (cells), which could only result from BNF (Montoya et al., 1996, 2004). Community $N_2$ fixation rates are very informative but cannot provide information on how much N different diazotrophic groups are fixing or if specific groups are fixing $N_2$ at all (Bombar et al., 2016; Moisander et al., 2017). Many studies compare community BNF rates and *nifH* sequencing data to determine which diazotrophic groups are likely responsible for the BNF rates measured (Gradoville et al., 2017; Shiozaki et al., 2020; Turk-Kubo et al., 2014). However, indirect inferences as *nifH* gene or transcript presence do not necessarily represent $N_2$ fixation activity.

Most studies use community BNF rates, but the amount of $N_2$ fixed by NCDs has never been directly quantified (Bombar et al., 2016, Moisander 2014, Reimann 2010), even though NCD *nifH* genes have been noted throughout the tropics and subtropics (Langlois et al., 2015; Riemann et al., 2010; Shiozaki et al., 2017) for over two decades (Zehr et al., 1998). Due to the community level analysis and the presence of cyanobacteria in these regions (Bombar et al., 2016; Moisander et al., 2017), the contribution of $N_2$ fixed by NCDs is obscured. As NCDs have yet to be shown capable of fixing $N_2$ in the open ocean (Bombar et al., 2016), the community-wide $N_2$ fixation rates are generally attributed solely to the cyanobacterial community.

Recently, studies have been able to characterize the metabolism and activity of environmental populations of microorganisms at the single-cell level using nanoscale secondary ion mass spectrometry (nanoSIMS). NanoSIMS combined with isotope addition incubations, such as $^{15}N_2$ for the study of BNF, allows for cell-specific isotope incorporation measurements. NanoSIMS uses a high-energy primary ion current ($Cs^+$ for BNF studies) to ionize a small target area, sending secondary ions released from the sample into a mass spectrometer for quantification (Pett-Ridge and Weber, 2012). An increasing number of studies have employed nanoSIMS to analyze BNF rates of uncultured diazotrophs at the single-cell level (Dekas, Poretsky, and Orphan, 2009; Woebken *et al.*, 2012, Thompson 2012, Martines-Perez, Gradoville 2019). Before nanoSIMS analysis, uncultured diazotrophs in environmental samples must be identified and located, which is commonly done with FISH or similar techniques as long as the 16S rRNA of the organisms is known (Behrens et al., 2008; Dekas et al., 2009; Thompson et al., 2012). NanoSIMS has allowed for a valuable and unprecedented look at the cell-specific BNF rates of uncultured diazotrophs (Dekas et al., 2009; Martínez-Pérez et al., 2016; Thompson et al., 2012; Woebken et al., 2012); however, nanoSIMS analysis is low throughput, time-intensive, and relies on known 16S rRNA genes.

*Summary*

This dissertation aimed to better understand the distribution, activity, and lifestyle of the unicellular diazotrophs, UCYN-A, NCDs in general, and the commonly occurring NCD known as Gamma A at the single-cell level. Single-cell

investigations are a direct means of analyzing cells, providing more detailed information than available with community-level analysis. The projects described here address questions regarding BNF that could only be answered through single-cell investigations. Chapter 1 addresses whether UCYN-A is present and fixing $N_2$ in the cold water (>10$^{o}$C) of the Bering Strait and the Arctic Ocean. We used community $N_2$ fixation rates from $^{15}N_2$ gas incubations, *nifH* sequencing, and UCYN-A specific nanoSIMS measurements of $^{15}N$ enrichment in cells to determine the presence of UCYN-A in the Arctic and estimate their $N_2$ fixation contribution. Our data shows that the same strains of UCYN-A1 and UCYN-A2 found in subtropical waters are present and fixing $N_2$ in cold Northern waters. Chapter 2 addresses whether NCDs are capable of $N_2$ fixation in the surface ocean of the North Pacific Subtropical Gyre. To answer this question, we used dual-isotope ($^{15}N_2$ and $^{13}C$-bicarbonate) incubations and an untargeted nanoSIMS approach to scan thousands of cells looking for cells enriched in $^{15}N$ but not $^{13}C$, indicating that they are $N_2$-fixers but are not photosynthetic. Our results show NCD are fixing $N_2$ in the surface ocean while attached to particles and contribute to the fixed N of the system. Chapter 3 addresses whether *nifH* geneFISH can be used to visualize uncultivated diazotrophs, such as Gamma A, in environmental samples and gain insights into their lifestyles. This project involved designing and optimizing *nifH* geneFISH probes for UCYN-A and Gamma A to visualize uncultivated cells via fluorescence microscopy. This method-based study concluded that geneFISH is a viable option to visualize diazotrophs for which only the *nifH* gene sequence is known. Additionally, successful

geneFISH identification of Gamma A showed that the majority of cells are attached to particles in the surface ocean, providing information on their lifestyle, despite being uncultivated. Single-cell analyses such as these offer high-resolution insight into uncultured diazotrophs which can help us better understand their lifestyles and $N_2$ fixation capability. As N limitation is directly linked to primary productivity and carbon cycling, gaining a better knowledge of BNF and microorganisms involved, particularly the less well-known uncultivated diazotrophs, can help to understand oceanic $N_2$ fixation, as well as constrain and predict $N_2$ fixation rates, which are critical for understanding predicting the effects of global climate change.

## References


Amann, R. I., Ludwig, W., & Schleifer, K.H. (1995). Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. *Microbiological Reviews*, *59*(1), 143–169.

Behrens, S., Lösekann, T., Pett-Ridge, J., Weber, P. K., Ng, W. O., Stevenson, B. S., Spormann, A. M. (2008). Linking microbial phylogeny to metabolic activity at the single-cell level by using enhanced element labeling-catalyzed reporter deposition fluorescence in situ hybridization (EL-FISH) and NanoSIMS. *Applied and Environmental Microbiology*, *74*(10), 3143–3150. https://doi.org/10.1128/AEM.00191-08

Benavides, M., Shoemaker, K. M., Moisander, P. H., Niggemann, J., Dittmar, T., Duhamel, S., Bonnet, S. (2018). Aphotic $N_2$ fixation along an oligotrophic to ultraoligotrophic transect in the western tropical South Pacific Ocean. *Biogeosciences*. https://doi.org/10.5194/bg-15-3107-2018

Bench, S. R., Frank, I., Robidart, J., & Zehr, J. P. (2016). Two subpopulations of *Crocosphaera watsonii* have distinct distributions in the North and South Pacific. *Environmental Microbiology*, *18*(2), 514–524. https://doi.org/10.1111/1462-2920.13180

Bentzon-Tilia, M., Severin, I., & Hansen, L. H. (2015). Genomics and Ecophysiology of Heterotrophic Nitrogen-Fixing Bacteria Isolated from Estuarine Surface Water. *MBio*, *6*(4), 1–11. https://doi.org/10.1128/mBio.00929-15.Editor

Bentzon-Tilia, M., Traving, S. J., Mantikci, M., Knudsen-leerbeck, H., Hansen, J. L. S., Markager, S., & Riemann, L. (2015). Significant $N_2$ fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries, 273–285. https://doi.org/10.1038/ismej.2014.119

Berman-Frank, I., Lundgren, P., & Falkowski, P. (2003). Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Research in Microbiology*, *154*(3), 157–164. https://doi.org/10.1016/S0923-2508(03)00029-9

Berman-Frank, I., Quigg, A., Finkel, Z. V., Irwin, A. J., & Haramaty, L. (2007). Nitrogen-fixation strategies and Fe requirements in cyanobacteria. *Limnology and Oceanography*, *52*(5), 2260–2269. https://doi.org/10.4319/lo.2007.52.5.2260

Blais, M., Tremblay, J., Jungblut, A. D., Gagnon, J., Martin, J., Thaler, M., & Lovejoy, C. (2012). Nitrogen fixation and identification of potential diazotrophs in the Canadian Arctic. *Global Biogeochemical Cycles*, *26*(3). https://doi.org/10.1029/2011GB004096

Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J., & Zehr, J. P. (2014). Comparative genomics reveals surprising divergence of two closely related

strains of uncultivated UCYN-A cyanobacteria. *The ISME Journal*, *8*(12), 2530–2542. https://doi.org/10.1038/ismej.2014.167

Bombar, D., Paerl, R. W., & Riemann, L. (2016). Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. *Trends in Microbiology*, *24*(11), 16–927. https://doi.org/10.1016/j.tim.2016.07.002

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K. A., Moutin, T., Hamersley, R. M., Grosso, O., Capone, D. G. (2013). Aphotic $N_2$ Fixation in the Eastern Tropical South Pacific Ocean. *PLoS ONE*, *8*(12), e81265. https://doi.org/10.1371/journal.pone.0081265

Brauer, V. S., Stomp, M., Rosso, C., Beusekom, S. A. M. Van, Emmerich, B., Stal, L. J., & Huisman, J. (2013). Low temperature delays timing and enhances the cost of nitrogen fixation in the unicellular cyanobacterium *Cyanothece*. *Isme Journal*, *7*(11), 2105–2115. https://doi.org/10.1038/ismej.2013.103

Breitbarth, E., Oschlies, A., & LaRoche, J. (2007). Physiological constraints on the global distribution of *Trichodesmium*- Effect of temperature on diazotrophy. *Biogeosciences*, *4*(1), 53–61. https://doi.org/10.5194/bg-4-53-2007

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

Caffrey, J. M., Bano, N., Kalanetra, K., & Hollibaugh, J. T. (2007). Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. *ISME Journal*, *1*(7), 660–662. https://doi.org/10.1038/ismej.2007.79

Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., & Carpenter, E. J. (1997). *Trichodesmium*, a Globally Significant Marine Cyanobacterium. *Science*, *276*(5316), 1221–1229. https://doi.org/10.1126/science.276.5316.1221

Capone, Douglas G, Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., … Carpenter, E. J. (2005). Nitrogen fixation by *Trichodesmium* spp .: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, *19*. https://doi.org/10.1029/2004GB002331

Carpenter, E., & Capone, D. G. (1992). Nitrogen Fixation in *Trichodesmium* Blooms. In *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs* (pp. 1–12).

Chen, Y. B., Dominic, B., Zani, S., Mellon, M. T., & Zehr, J. P. (1999). Expression of photosynthesis genes in relation to nitrogen fixation in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. IMS 101.

*Plant Molecular Biology*.

Church, M J, Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Vertical distributions of nitrogen-fixing phylotypes at Station ALOHA in the oligotrophic North Pacific Ocean. *Aquatic Microbial Ecology*, *38*, 3–14. https://doi.org/10.3354/ame038003

Church, Matthew J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., & Karl, D. M. (2009). Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochemical Cycles*, *23*(2), n/a-n/a. https://doi.org/10.1029/2008GB003418

Church, Matthew J, Short, C. M., Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH* ) expression in the oligotrophic North Pacific Ocean. *Applied and Environmental Microbiology*, *71*(9), 5362–5370. https://doi.org/10.1128/AEM.71.9.5362

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*, *4*(2), 233–253. https://doi.org/10.5194/bg-4-233-2007

Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sa, P., Lima-mendez, G., Hingamp, P., … Zehr, J. P. (2016). Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nature Communications*, (March). https://doi.org/10.1038/ncomms11071

Cornejo-Castillo, F. M., & Zehr, J. P. (2019). Hopanoid lipids may facilitate aerobic nitrogen fixation in the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(37), 18269–18271. https://doi.org/10.1073/pnas.1908165116

Dekas, A. E., Fike, D. A., Chadwick, G. L., Green-Saxena, A., Fortney, J., Connon, S. A., … Orphan, V. J. (2018). Widespread nitrogen fixation in sediments from diverse deep-sea sites of elevated carbon loading, *20*, 4281–4296. https://doi.org/10.1111/1462-2920.14342

Dekas, A. E., Poretsky, R. S., & Orphan, V. J. (2009). Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science*, *326*(October), 422–426.

Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T. M., Rappé, M. S., … Eren, A. M. (2018). Nitrogen-fixing populations of *Planctomycetes* and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, *3*(July). https://doi.org/10.1038/s41564-018-0176-9

Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N., & Dunne, J. P. (2007). Spatial coupling of nitrogen inputs and losses in the ocean. *Nature*, *445*(7124),

163–167. https://doi.org/10.1038/nature05392

Dron, A., Rabouille, S., Claquin, P., Le Roy, B., Talec, A., & Sciandra, A. (2012). Light-dark (12:12) cycle of carbon and nitrogen metabolism in *Crocosphaera watsonii* WH8501: Relation to the cell cycle. *Environmental Microbiology*, *14*(4), 967–981. https://doi.org/10.1111/j.1462-2920.2011.02675.x

Duce, R. A., LaRoche, J., Altieri, K., Arrigo, K. R., Baker, A. R., Capone, D. G., … Zamora, L. (2008). Impacts of atmospheric anthropogenic nitrogen on the open ocean. *Science*, *320*(5878), 893–897. https://doi.org/10.1126/science.1150369

Dugdale, R. C., & Goering, J. J. (1967). Uptake of New and Regenerated Forms of Nitrogen in Primary Productivity. *Limnology and Oceanography*, *12*(2), 196–206. https://doi.org/10.4319/lo.1967.12.2.0196

Falkowski, P. G. (2015). *Life's Engines*. Princeton University Press. Retrieved from http://www.jstor.org/stable/j.ctvc776vb

Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science*, *281*(5374), 200–206. https://doi.org/10.1126/science.281.5374.200

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Microbial Ecology*, (3), 1034–1040.

Farnelid, H., Andersson, A. F., Bertilsson, S., Al-Soud, W. A., Hansen, L. H., Sørensen, S., … Riemann, L. (2011). Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0019223

Farnelid, H., Helle, K. T., Justin, P., James, E. O., Mooy, B. A. S. Van, & Zehr, J. P. (2019). Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *The ISME Journal*, 170–182. https://doi.org/10.1038/s41396-018-0259-x

Farnelid, H., Turk-Kubo, K., Munoz-Marin Del Carmen, M., & Zehr, J. P. (2016). New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquatic Microbial Ecology*, *77*(3), 128–138. https://doi.org/10.3354/ame01794

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

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

Galloway, J N, Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., … Vo, C. J. (2004). *Nitrogen cycles: past, present, and future*.

Galloway, James N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., … Sutton, M. A. (2008). Transformation of the Nitrogen Cycle : *Science*, *320*(May), 889–892. https://doi.org/10.1126/science.1136674

Gómez-Consarnau, L., Needham, D. M., Weber, P. K., Fuhrman, J. A., & Mayali, X. (2019). Influence of light on particulate organic matter utilization by attached and free-living marine bacteria. *Frontiers in Microbiology*, *10*(JUN). https://doi.org/10.3389/fmicb.2019.01204

Gradoville, M. R., Bombar, D., Crump, B. C., Letelier, R. M., Zehr, J. P., & White, A. E. (2017). Diversity and activity of nitrogen-fixing communities across ocean basins. *Limnology and Oceanography*, *62*, 1895–1909. https://doi.org/10.1002/lno.10542

Gradoville, M. R., Farnelid, H., White, A. E., Turk-Kubo, K. A., Stewart, B., Ribalet, F., … Zehr, J. P. (2020). Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnology and Oceanography*, *65*, 1858–1875. https://doi.org/10.1002/lno.11423

Gruber, N., & Sarmiento, J. L. (1997). Global patterns of marine nitrogen fixation and denitrification. *Biogeochemical Cycles*, *11*(2), 235–266.

Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., … Kuypers, M. M. M. (2012). Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *The ISME Journal*, *6*(6), 1238–1249. https://doi.org/10.1038/ismej.2011.182

Hamersley, M. R., Turk, K. A., Leinweber, A., Gruber, N., Zehr, J. P., Gunderson, T., & Capone, D. G. (2011). Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight. *Aquatic Microbial Ecology*, *63*(2), 193–205. https://doi.org/10.3354/ame01494

Hewson, I., Moisander, P. H., Achilles, K. M., Carlson, C. A., Jenkins, B. D., Mondragon, E. A., … Zehr, J. P. (2007). Characteristics of diazotrophs in surface to abyssopelagic waters of the Sargasso Sea. *Aquatic Microbial Ecology*, *46*(1), 15–30. https://doi.org/10.3354/ame046015

Howard, J. B., & Rees, D. C. (1996). Structural Basis of Biological Nitrogen FIxation. *Chemical Reviews*, *96*(7), 2965–2982. https://doi.org/10.1021/cr9500545

Inomura, K., Bragg, J., & Follows, M. J. (2017). A quantitative analysis of the direct and indirect costs of nitrogen fixation: a model based on *Azotobacter vinelandii*. *ISME Journal*, 166–175. https://doi.org/10.1038/ismej.2016.97

Karl, D M, Michaels, A., Bergman, B., & Capone, D. (2002). Dinitrogen fixation in the world's oceans. *Biochemistry*, *57*(58), 47–98. https://doi.org/10.1023/A:1015798105851

Karl, David M, Bidigare, R. R., Church, M. J., Dore, J. E., Letelier, R. M., Mahaffey, C., & Zehr, J. P. (2008). *The Nitrogen Cycle in the North Pacific Trades Biome : An Evolving Paradigm*. https://doi.org/10.1016/B978-0-12-372522-6.00016-5

Kartal, B., van Niftrik, L., Keltjens, J. T., Opden Camp, H. J. M., & Jetten, M. S. M. (2012). Anammox-Growth Physiology, Cell Biology, and Metabolism. In R. K. B. T.-A. in M. P. Poole (Ed.), *Advances in Microbial Physiology* (Vol. 60, pp. 211–262). Academic Press. https://doi.org/10.1016/B978-0-12-398264-3.00003-6

Klawonn, I., Bonaglia, S., Bruchert, V., & Ploug, H. (2015). Aerobic and anaerobic nitrogen transformation processes in $N_2$-fixing cyanobacterial aggregates. *The ISME Journal*, *9*(15), 1456–1466. https://doi.org/10.1038/ismej.2014.232

Langlois, R., Großkopf, T., Mills, M., Takeda, S., & LaRoche, J. (2015). Widespread Distribution and Expression of Gamma A (UMB), an Uncultured, Diazotrophic, γ-Proteobacterial *nifH* Phylotype. *Plos One*, *10*(6), e0128912. https://doi.org/10.1371/journal.pone.0128912

Langlois, R. J., LaRoche, J., & Raab, P. a. (2005). Diazotrophic Diversity and Distribution in the Tropical and Subtropical Atlantic Ocean Diazotrophic Diversity and Distribution in the Tropical and Subtropical Atlantic Ocean. *Applied and Environmental Microbiology*, *71*(12), 7910–7919. https://doi.org/10.1128/AEM.71.12.7910

Lipschultz, F., & Owens, N. J. P. (1996). An assessment of nitrogen fixation as a source of nitrogen to the North Atlantic Ocean. *Biogeochemistry*, *35*(1), 261–274. https://doi.org/10.1007/BF02179830

Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., … Zehr, J. P. (2012). Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth System Science Data*, *4*(1), 47–73. https://doi.org/10.5194/essd-4-47-2012

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., … Kuypers, M. M. M. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, *1*(11). https://doi.org/10.1038/nmicrobiol.2016.163

McKinna, L. I. W. (2015). Three decades of ocean-color remote-sensing *Trichodesmium* spp. in the World's oceans: A review. *Progress in Oceanography*, *131*(December 2014), 177–199. https://doi.org/10.1016/j.pocean.2014.12.013

Mills, M. M., Ridame, C., Davey, M., La Roche, J., & Geider, R. J. (2004). Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature*, *435*(7039), 232. https://doi.org/10.1038/nature03632

Mills, M. M., Wilson, S. T., Arrigo, K. R., Zehr, J. P., Henke, B. A., & Harding, K. (2020). Unusual marine cyanobacteria/haptophyte symbiosis relies on $N_2$ fixation even in N-rich environments. *The ISME Journal*, (3), 2395–2406. https://doi.org/10.1038/s41396-020-0691-6

Mohr, W., Großkopf, T., Wallace, D. W. R., & Laroche, J. (2010). Methodological Underestimation of Oceanic Nitrogen Fixation Rates. *PLoS ONE*, *5*(9). https://doi.org/10.1371/journal.pone.0012583

Moisander, P. H., Beinart, R. a, Hewson, I., White, A. E., Johnson, K. S., Carlson, C. a, … Zehr, J. P. (2010). Unicellular cyanobacterial distributions broaden the oceanic $N_2$ fixation domain. *Science (New York, N.Y.)*, *327*(5972), 1512–1514. https://doi.org/10.1126/science.1185468

Moisander, P. H., Benavides, M., Bonnet, S., & Berman-frank, I. (2017). Chasing after Non-cyanobacterial Nitrogen Fixation in Marine Pelagic Environments. *Frontiers in Microbiology*, *8*(September). https://doi.org/10.3389/fmicb.2017.01736

Moisander, P. H., Serros, T., Paerl, R. W., Beinart, R. a, & Zehr, J. P. (2014). Gammaproteobacterial diazotrophs and *nifH* gene expression in surface waters of the South Pacific Ocean. *The ISME Journal*, *8*(10), 1962–1973. https://doi.org/10.1038/ismej.2014.49

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

Monteiro, F. M., Dutkiewicz, S., & Follows, M. J. (2011). Biogeographical controls on the marine nitrogen fixers. *Global Biogeochemical Cycles*, *25*(June 2010), 1–8. https://doi.org/10.1029/2010GB003902

Montoya, J. P., Holl, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., & Capone, D. G. (2004). High rates of $N_2$ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature*, *430*(August), 1027–1031. https://doi.org/10.1038/nature02744.1.

Montoya, J. P., Voss, M., Kahler, P., & Capone, D. G. (1996). A Simple, High-Precision, High-Sensitivity Tracer Assay for $N_2$ Fixation. *Applied and Environmental Microbiology*, *62*(3), 986–993. https://doi.org/10.1128/AEM.62.3.986.993

Moore, C. M., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M.

I., … Woodward, E. M. S. (2009). Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nature Geoscience*, *2*(12), 867–871. https://doi.org/10.1038/ngeo667

Mulholland, M. R., Bernhardt, P. W., Blanco-Garcia, J. L., Mannino, a., Hyde, K., Mondragon, E., … Zehr, J. P. (2012). Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank. *Limnology and Oceanography*, *57*(4), 1067–1083. https://doi.org/10.4319/lo.2012.57.4.1067

Munoz-Marin Del Carmen, M., Shilova, I. N., Shi, T., Farnelid, H., Cabello, A. M., & Zehr, J. P. (2019). The Transcriptional Cycle Is Suited to Daytime $N_2$ Fixation in the Unicellular Cyanobacterium Candidatus Atelocyanobacterium thalassa (UCYN-A). *MBio*, *10*(1), 1–17.

Ohki, K., Zehr, J. P., & Fujita, Y. (1992). Regulation of nitrogenase activity in relation to the light-dark regime in the filamentous non-heterocyctous cyanobacterium *Trichodesmium* Sp. NIBB 1067. *Journal of General Microbiology*, *138*, 2679–2685.

Olson, J. B., Steppe, T. F., Litaker, R. W., & Paerl, H. W. (1998). $N_2$-fixing microbial consortia associated with the ice cover of Lake Bonney, Antarctica. *Microbial Ecology*, *36*(3), 231–238. https://doi.org/10.1007/s002489900110

Paerl, H. W., Crocker, K. M., & Prufert, L. E. (1987). Limitation of $N_2$ fixation in coastal marine waters: Relative importance of molybdenum, iron, phosphorus, and organic matter availability. *Limnology and Oceanography*, *32*(3), 525–536. https://doi.org/10.4319/lo.1987.32.3.0525

Paerl, H. W., & Prufert, L. E. (1987). Oxygen-Poor Microzones as Potential Sites of Microbial $N_2$ Fixation in Nitrogen-Depleted Aerobic Marine Waters. *Applied and Environmental Microbiology*, *53*(5), 1078–1087. https://doi.org/10.1128/AEM.53.5.1087.1987

Paerl, R. W., Hansen, T. N. G., Henriksen, N. N. S. E., & Olesen, A. K. (2018). $N_2$ fixation and related $O_2$ constraints on model marine diazotroph Pseudomonas stutzeri BAL361. *Aquatic Microbial Ecology*, *81*, 125–136. https://doi.org/10.3354/ame01867

Pedersen, J. N., Bombar, D., Paerl, R. W., Riemann, L., & Seymour, J. R. (2018). Diazotrophs and $N_2$-Fixation Associated With Particles in Coastal Estuarine Waters. *Frontiers in Microbiology*, *9*(November), 1–11. https://doi.org/10.3389/fmicb.2018.02759

Pett-Ridge, J., & Weber, P. K. (2012). NanoSIP: NanoSIMS applications for microbial biology. *Methods in Molecular Biology*, *881*, 375–408. https://doi.org/10.1007/978-1-61779-827-6_13

Ploug, H. (2001). Small-scale oxygen fluxes and remineralization in sinking aggregates. *Limnology and Oceanography*, *46*(7), 1624–1631. https://doi.org/10.4319/lo.2001.46.7.1624

Poff, K. E., Leu, A. O., Eppley, J. M., Karl, D. M., & Delong, E. F. (2021). Microbial dynamics of elevated carbon flux in the open ocean's abyss. *PNAS*, *118*(4), 1–11. https://doi.org/10.1073/pnas.2018269118

Rahav, E., Bar-Zeev, E., Ohayon, S., Elifantz, H., Belkin, N., Herut, B., … Berman-Frank, I. (2013). Dinitrogen fixation in aphotic oxygenated marine environments. *Frontiers in Microbiology*, *4*(August), 1–11. https://doi.org/10.3389/fmicb.2013.00227

Riemann, L., Farnelid, H., & Steward, G. F. (2010). Nitrogenase genes in non-cyanobacterial plankton : prevalence, diversity, and regulation in marine waters. *Aquatic Microbial Ecology*, (December). https://doi.org/10.3354/ame01431

Rippka, R., & Waterbury, J. B. (1977). The Synthesis of Nitrogenase by Non-Heterocyctous Cyanobacteria. *FEMS Microbiology Letters*, *2*, 83–86. https://doi.org/10.1017/CBO9781107415324.004

Sabra, W., Zeng, A. P., Lunsdorf, H., & Deckwer, W. D. (2000). Effect of oxygen on formation and structure of *Azotobacter vinelandii* alginate and its role in protecting nitrogenase. *Applied and Environmental Microbiology*, *66*(9), 4037–4044. https://doi.org/10.1128/AEM.66.9.4037-4044.2000

Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., … Dimier, C. (2019). Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. Cell, 179(Nov), 1068–1093. https://doi.org/https://doi.org/10.1016/j.cell.2019.10.014

Sañudo-Wilhelmy, S. A., Kustka, A. B., Gobler, C. J., Hutchins, D. A., Yang, M., Lwiza, K., … Carpenter, E. J. (2001). Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature*, *411*(6833), 66–69. https://doi.org/10.1038/35075041

Scavotto, R. E., Dziallas, C., Bentzon-Tilia, M., Riemann, L., & Moisander, P. H. (2015). Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. *Environmental Microbiology*, *17*, 3754–3765. https://doi.org/10.1111/1462-2920.12777

Schneegurt, M. A., Sherman, D. M., Nayar, S., & Sherman, L. A. (1994). Oscillating behavior of carbohydrate granule formation and dinitrogen fixation in the cyanobacterium *Cyanothece* sp. strain ATCC 51142. *Journal of Bacteriology*, *176*(6), 1586–1597. https://doi.org/10.1128/jb.176.6.1586-1597.1994

Severin, I., Bentzon-Tilia, M., Moisander, P. H., & Riemann, L. (2015). Nitrogenase expression in estuarine bacterioplankton influenced by organic carbon and

availability of oxygen. *FEMS Microbiology Letters*, *326*, 1–7. https://doi.org/10.1093/femsle/fnv105

Shi, T., Ilikchyan, I., Rabouille, S., & Zehr, J. P. (2010). Genome-wide analysis of diel gene expression in the unicellular $N_2$-fixing cyanobacterium *Crocosphaera watsonii* WH 8501. *ISME Journal*, *4*(5), 621–632. https://doi.org/10.1038/ismej.2009.148

Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., … Furuya, K. (2017). Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. *AGU Publications*, 996–1009. https://doi.org/10.1002/2017GB005681

Shiozaki, T., Fujiwara, A., Inomura, K., Hirose, Y., Hashihama, F., & Harada, N. (2020). Biological nitrogen fixation detected under Antarctic sea ice. *Nature Geoscience*. https://doi.org/10.1038/s41561-020-00651-7

Short, S. M., & Zehr, J. P. (2007). Brief report Nitrogenase gene expression in the Chesapeake Bay Estuary. *Environmental Microbiology*, *9*, 1591–1596. https://doi.org/10.1111/j.1462-2920.2007.01258.x

Siegenthaler, U., & Sarmiento, J. L. (1971). Atmospheric carbon dioxide and the biosphere. *Nature*, *1*(4), 249–261. https://doi.org/10.1016/0013-9327(71)90019-X

Singh, S. M., & Elster, J. (2007). Cyanobacteria in Antarctic Lake Environments BT - Algae and Cyanobacteria in Extreme Environments. In J. Seckbach (Ed.) (pp. 303–320). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6112-7_16

Sohm, J. A., Webb, E. A., & Capone, D. G. (2011). Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, *9*(7), 499–508. https://doi.org/10.1038/nrmicro2594

Stal, L. J. (2009). Is the distribution of nitrogen-fixing cyanobacteria in the oceans related to temperature?: Minireview. *Environmental Microbiology*, *11*(7), 1632–1645. https://doi.org/10.1111/j.1758-2229.2009.00016.x

Stewart, W. D. P. (1973). Nitrogen Fixation by photosynthetic Microorganisms. *Annual Review of Microbiology*, *2*(1), 1–19. https://doi.org/10.1146/annurev.biochem.64.1.721

Thompson, A. W., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F., & Zehr, J. P. (2014). Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environmental Microbiology*, *16*, doi:10.1111/1462-2920.12490. https://doi.org/10.1111/1462-2920.12490

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., …

Zehr, J. P. (2012). Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. *Science (New York, NY)*, *337*(September), 1546–1550. https://doi.org/10.1126/science.1222700

Toepel, J., Welsh, E., Summerfield, T. C., Pakrasi, H. B., & Sherman, L. A. (2008). Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp. strain ATCC 51142 during light-dark and continuous-light growth. *Journal of Bacteriology*, *190*(11), 3904–3913. https://doi.org/10.1128/JB.00206-08

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

Turk-Kubo, K. A., Farnelid, H., Shilova, I. N., Henke, B., & Zehr, J. P. (2017). Distinct ecological niches of marine symbiotic $N_2$-fixing cyanobacterium Candidatus Atelocyanobacterium thalassa sublineages. *Journal of Phycology*, *53*(2), 451–461. https://doi.org/10.1111/jpy.12505

Turk-Kubo, K. A., Karamchandani, M., Capone, D. G., & Zehr, J. P. (2014). The paradox of marine heterotrophic nitrogen fixation: Abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. *Environmental Microbiology*, *16*(10), 3095–3114. https://doi.org/10.1111/1462-2920.12346

Van de Graaf, A. A., Mulder, A., De Bruijn, P., Jetten, M. S. M., Robertson, L. A., & Kuenen, J. G. (1995). Anaerobic oxidation of ammonium is a biologically mediated process. *Applied and Environmental Microbiology*, *61*(4), 1246–1251. https://doi.org/10.1128/aem.61.4.1246-1251.1995

Villareal, T. A. (1990). Laboratory Culture and Preliminary Characterization of the Nitrogen-Fixing *Rhizosolenia-Richelia* Symbiosis. *Marine Ecology*, *11*(2), 117–132. https://doi.org/https://doi.org/10.1111/j.1439-0485.1990.tb00233.x

Villareal, T. A. (1994). Widespread Occurrence of the Hemiaulus-Cyanobacterial Symbiosis in the Southwest North Atlantic Ocean. *Bulletin of Marine Science*, *54*(1), 1–7.

Wang, W., Moore, J. K., Martiny, A. C., & François, W. (2019). Convergent estimates of marine nitrogen fixation. *Nature*, 205–213. https://doi.org/10.1038/s41586-019-0911-2

Ward, B. (2012). *Fundamentals of Geobiology*. *Fundamentals of Geobiology*. https://doi.org/10.1002/9781118280874

Ward, B. B., Capone, D. G., & Zehr, J. P. (2007). What's New in the Nitrogen Cycle? *Oceanography*, *20*(June), 101–109. https://doi.org/http://dx.doi.org/10.5670/oceanog.2007.53

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

White, A. E., Granger, J., Selden, C., Gradoville, M. R., Potts, L., Bourbonnais, A., … Chang, B. X. (2020). A critical review of the $^{15}N_2$ tracer method to measure diazotrophic production in pelagic ecosystems. *Limnology and Oceanography*, 129–147. https://doi.org/10.1002/lom3.10353

Wilson, S. T., Foster, R. A., Zehr, J. P., & Karl, D. M. (2010). Hydrogen production by *Trichodesmium erythraeum Cyanothece* sp. and *Crocosphaera watsonii*. *Aquatic Microbial Ecology*, *59*(2), 197–206. https://doi.org/10.3354/ame01407

Wilson, S. T., Tozzi, S., Foster, R. A., Ilikchyan, I., Kolber, Z. S., Zehr, J. P., & Karl, D. M. (2010). Hydrogen cycling by the unicellular marine diazotroph crocosphaera watsonii strain WH8501. *Applied and Environmental Microbiology*, *76*(20), 6797–6803. https://doi.org/10.1128/AEM.01202-10

Woebken, D., Burow, L. C., Prufert-bebout, L., Bebout, B. M., Hoehler, T. M., Pett-ridge, J., … Singer, S. W. (2012). Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *The ISME Journal*, *6*, 1427–1439. https://doi.org/10.1038/ismej.2011.200

Wu, J., Sunda, W., Boyle, E. A., & Karl, D. M. (2000). Phosphate depletion in the Western North Atlantic Ocean. *Science*, *289*(5480), 759–762. https://doi.org/10.1126/science.289.5480.759

Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., Van Bleijswijk, J., Timmers, P., … Damsté, J. S. S. (2006). Archaeal nitrification in the ocean. *PNAS*, *103*(33), 12317–12322. https://doi.org/10.1073/pnas.0600756103

Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., … Affourtit, J. P. (2008). Globally Distributed Uncultivated Oceanis $N_2$-Fixing Cyanobacteria Lack Oxygenic Photosystem II. *Science*, *322*(November), 1110–1112. https://doi.org/10.1126/science.1165340

Zehr, J. P., & Bombar, D. (2015). Cyanobacteria and Archaea Marine Nitrogen Fixation : Organisms, Significance, Enigmas, and Future Directions. In *Biological Nitrogen Fixation* (Vol. 2, pp. 857–871).

Zehr, J. P., & Capone, D. G. (2021). Marine Nitrogen Fixation. Springer.

Zehr, J. P., & Capone, D. G. (2020). Changing perspectives in marine nitrogen fixation. *Science*, *9514*, 1–9. https://doi.org/10.1126/science.aay9514

Zehr, J. P., Jenkins, B. D., Short, S. M., & Steward, G. F. (2003). Nitrogenase gene

diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology*, *5*(7), 539–554. https://doi.org/10.1046/j.1462-2920.2003.00451.x

Zehr, J. P., & Kudela, R. M. (2011). Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. *Annual Review of Marine Science*. https://doi.org/10.1146/annurev-marine-120709-142819

Zehr, J. P., & Mcreynolds, L. A. (1989). Use of Degenerate Oligonucleotides for Amplification of the *nifH* Gene from the Marine Cyanobacterium *Trichodesmium thiebautii*. *Applied and Environmental Microbiology*, 2522–2526. https://doi.org/10.1128/AEM.55.10.2522-2526.1989

Zehr, J. P., Mellon, M. T., Zani, S., & York, N. (1998). New Nitrogen-Fixing Microorganisms Detected in Oligotrophic Oceans by Amplification of Nitrogenase (*nifH*) Genes. *Applied and Environmental Microbiology*, *64*(9), 3444–3450. https://doi.org/10.1128/AEM.64.9.3444-3450.1998

Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., … Karl, D. M. (2001). Unicellular cyanobacteria fix $N_2$ in the subtropical North Pacific Ocean. *Nature*, *412*(6847), 635–638. https://doi.org/10.1038/35088063

Zhang, X., McRose, D. L., Darnajoux, R., Bellenger, J. P., Morel, F. M. M., & Kraepiel, A. M. L. (2016). Alternative nitrogenase activity in the environment and nitrogen cycle implications. *Biogeochemistry*, *127*(2–3), 189–198. https://doi.org/10.1007/s10533-016-0188-6

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews : MMBR*, *61*(4), 533–616. https://doi.org/10.1128/.61.4.533-616.1997

# Chapter 2

## Symbiotic Unicellular Cyanobacteria Fix Nitrogen in the Arctic Ocean

Harding, K., Turk-kubo, K. A., Sipler, R. E., Mills, M. M. and Bronk, D. A., Zehr, J.P. (2018). PNAS, 115(52) 13371-13375

**Abstract**

Biological dinitrogen ($N_2$) fixation is an important source of nitrogen (N) in low-latitude open oceans. The unusual $N_2$-fixing unicellular cyanobacteria (UCYN-A)/haptophyte symbiosis has been found in an increasing number of unexpected environments, including northern waters of the Danish Straight Bering and Chukchi Seas. We used nanoscale secondary ion mass spectrometry (nanoSIMS) to measure $^{15}N_2$ uptake into UCYN-A/haptophyte symbiosis and found that UCYN-A strains identical to low-latitude strains are fixing $N_2$ in the Bering and Chukchi Seas, at rates comparable to subtropical waters. These results show definitively that cyanobacterial $N_2$ fixation is not constrained to subtropical waters, challenging paradigms and models of global $N_2$ fixation. The Arctic is particularly sensitive to climate change, and $N_2$ fixation may increase in Arctic waters under future climate scenarios.

*Significance Statement*

Biological dinitrogen ($N_2$) fixation (BNF) is an important source of nitrogen (N) in marine systems. Until recently, it was believed to be primarily limited to subtropical open oceans. Marine BNF is mainly attributed to cyanobacteria; recently, an unusual $N_2$-fixing unicellular cyanobacteria (UCYN-A)/haptophyte symbiosis was reported with a broader temperature range than other $N_2$-fixing cyanobacteria. We report that the UCYN-A symbiosis is present and fixing $N_2$ in the Western Arctic and Bering Seas, further north than any previously reported $N_2$-fixing marine cyanobacteria. Nanoscale secondary ion mass spectrometry (nanoSIMS) enabled us to directly show that the symbiosis was fixing $N_2$. These results show that $N_2$-fixing

cyanobacteria are not constrained to subtropical waters and challenge commonly held ideas about global marine $N_2$ fixation.

## Introduction

Biological $N_2$ fixation, the reduction of atmospheric $N_2$ to biologically available nitrogen, is an important source of N in oligotrophic tropical and subtropical oceans (1). Historically, studies of ocean $N_2$ fixation focused on the well-known cyanobacterium *Trichodesmium* and the diatom symbiont *Richelia*, which were reported primarily from warm (> 20$^o$C) waters (2) with low concentrations of fixed N (nitrate and ammonium) (3). The discovery of a unicellular cyanobacterial symbiont (UCYN-A) of a haptophyte alga (4, 5) expanded the geographic distribution of marine $N_2$-fixers to waters with lower temperatures and higher concentrations of fixed inorganic N (6, 7). These regions include the high latitude waters of the Danish Strait (8) and the Western Arctic (9, 10). The presence of $N_2$-fixers does not confirm $N_2$ fixation activity since $N_2$ fixation is a highly regulated process inhibited by multiple environmental factors. Here we demonstrate that the cyanobacterial symbiont UCYN-A fixes $N_2$ in the cold, high latitude waters of the Western Arctic.

UCYN-A is an unusual cyanobacterium lineage that has lost many of the typical cyanobacterial metabolic pathways, including the ability to fix $CO_2$ and evolve $O_2$ in photosynthesis (5). It is a symbiont with a small planktonic unicellular haptophyte alga, related to *Braarudosphaera bigelowii* (4, 11). UCYN-A is uncultivated and can only be detected by its DNA or through visualization with catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH). UCYN-A is comprised of multiple closely related lineages with distinct haptophyte hosts that can be differentiated by distinct DNA sequences (12). Generally,

haptophytes are geographically widespread, including cold, high latitude waters (13). The symbiosis between a $N_2$-fixing cyanobacterium and a haptophyte may facilitate a unique adaptation to $N_2$ fixation in colder waters, such as the Arctic Ocean, where other $N_2$-fixing marine cyanobacteria have not been found.

Little is known about marine $N_2$ fixation in polar regions, partially because low temperatures are believed to inhibit the growth and activity of $N_2$-fixing cyanobacteria, such as in *Trichodesmium* and *Crocosphaera* (2, 14). *Trichodesmium* has occasionally been reported in high latitudes (62$^o$N) (15) but does not appear to fix $N_2$ when advected into cold waters (16). However, temperature alone does not necessarily preclude $N_2$ fixation since microorganisms can fix $N_2$ in ice-covered Antarctic lakes to near boiling temperatures at hydrothermal vents (17, 18). A few studies have reported low but measurable $N_2$ fixation rates in the Arctic (0.02 to 7.7 nmol N $L^{-1}$ $day^{-1}$) (19, 20). Recent studies in the Bering and Chukchi Seas (9, 10) reported bulk water $N_2$ fixation rates of 2.3 to 3.6 nmol N $L^{-1}d^{-1}$ and detected DNA from UCYN-A and other Bacteria but could not link $N_2$ fixation to specific microorganisms.

**Results and Discussion**

We collected water samples in the Bering, Chukchi, and Beaufort Seas during September 2016 (SI Appendix, Fig. S1 and Materials, and Methods) and identified $N_2$-fixing microorganisms. The *nifH* gene, a widely-used proxy for $N_2$-fixing microorganisms, was amplified by PCR and sequenced. UCYN-A *nifH* sequences were the only marine cyanobacterial *nifH* sequences found and were present in the Bering, Chukchi, and Beaufort Sea samples (SI Appendix, Fig. S2). The abundances

of the two strains identified, UCYN-A1 and UCYN-A2, were estimated using quantitative PCR (qPCR) (Fig. 1, SI Appendix, Fig. S3, and Table S2). In the Bering Sea near Nome, AK, UCYN-A abundances ($10^5$ to $10^6$ *nifH* copies $L^{-1}$) were comparable to those at subtropical latitudes (6, 11, 21) but were considerably lower in the Chukchi Sea, on the north eastern Chukchi shelf, and in the Beaufort Seas (Fig. 1, SI Appendix, Fig. S3, and Table S2). Both UCYN-A lineages were primarily found in surface samples in low salinity ice-melt waters (SI Appendix, Fig. S4, and Table S2). These results verified that two strains of the $N_2$-fixing cyanobacterium UCYN-A are present in polar waters, consistent with recent findings (9, 10).

Our genomic and morphological characterization of the high-latitude UCYN-A symbioses show these strains are indistinguishable from those reported in subtropical oceans. The UCYN-A1 and UCYN-A2 *nifH* sequences from the Arctic samples were identical to the most common sequences reported from subtropical oceans (Fig. 2 and SI Appendix, Fig. S2). Visualization of the different haptophyte hosts and their respective UCYN-A1 and UCYN-A2 cyanobacterial symbionts using lineage-specific CARD-FISH probes (22, 23) (Fig. 3) showed that the symbioses were similar in size and morphology (~3 µm and 5 µm, respectively) to the symbioses in tropical and subtropical oceans (4, 7, 23, 24). Other globally distributed eukaryotic picoplankton have specific cold-adapted strains (25, 26), so the discovery of the subtropical strains of UCYN-A in the Arctic was unexpected.

$^{15}N_2$ uptake experiments coupled with CARD-FISH and nanoscale secondary ion mass spectrometry (nanoSIMS) showed that the UCYN-A symbiosis was actively

fixing $N_2$. This is the first direct evidence of $N_2$ fixation by planktonic marine cyanobacteria (or any marine microorganism) in the Arctic region (72$^o$ N) and with low water temperatures (4$^o$C). The UCYN-A1 and UCYN-A2 symbioses in the Bering Sea had mean cell-specific $N_2$ fixation rates of 7.6 ± 14.5 fmol N cell$^{-1}$d$^{-1}$ (n=6) and 13.0 ± 7.7 fmol N cell$^{-1}$d$^{-1}$(n=8), respectively (Fig. 4). In the Chukchi Sea, the cell-specific $N_2$ fixation rates of the UCYN-A2 symbiosis were considerably lower, but two out of six cells were detectable with an average of 1.1 ± 2.0 fmol N cell$^{-1}$d$^{-1}$. Surprisingly, the UCYN-A1 cell-specific $N_2$ fixation rates measured in the Bering Sea at 10.1$^o$C are similar to rates reported from the warmer waters (>25$^o$C) of the subtropical North Atlantic (0.45 - 12 fmol N cell$^{-1}$d$^{-1}$; 27, 28). UCYN-A $N_2$ fixation in polar waters shows that low temperature does not limit the distribution or activity of $N_2$-fixing cyanobacteria.

$N_2$ fixation by UCYN-A accounted for the total measured $N_2$ fixation rates in the Bering Sea but not in the Chukchi Sea. UCYN-A $N_2$ fixation rates were estimated to be 10.5 ± 18.4 nmol N L$^{-1}$d$^{-1}$ and 0.004 ± 0.007 nmol N L$^{-1}$d$^{-1}$ in the Bering Sea and the Chukchi Sea, respectively, when per cell rates were scaled to volumetric rates (SI Appendix, Materials and Methods, and Table S3). Measured $N_2$ fixation rates in the bulk water sample from the Bering Sea (Station 1) were 6.9 ± 3.8 nmol N L$^{-1}$d$^{-1}$, indicating that $N_2$ fixation by UCYN-A accounted for total bulk rates (within error). In contrast, extrapolated rates from UCYN-A2 cellular $N_2$ fixation in the Chukchi Sea were two orders of magnitude less than the measured bulk $N_2$ fixation of 0.2 ± 0.2 nmol N L$^{-1}$d$^{-1}$ (DNQ). More research is needed to determine the quantitative

significance of $N_2$ fixation by UCYN-A in this region, but the presence of actively $N_2$-fixing unicellular cyanobacteria in Arctic waters is surprising from an ecological perspective and important for mathematical models that predict global $N_2$ fixation.

It is unclear whether the UCYN-A symbioses in the Bering Sea and Western Arctic are advected into the Western Arctic Seas through the Bering Strait or are endemic populations. Microbial community structure in Arctic waters is heavily influenced by the originating water mass (29), and Shiozaki et al. (10) suggest the UCYN-A symbiosis detected by DNA assays originates from Pacific waters transported to the Arctic in the Alaskan Coastal Current, as has been reported for other species (30). However, UCYN-A1, which is widespread in the North Pacific Subtropical Gyre, disappears at the front between the North Pacific Subtropical Gyre and the North Pacific Subarctic Gyre (9, 21), and the UCYN-A2 symbiosis is not commonly found in the oligotrophic North Pacific (12). This suggests that the UCYN-A populations may be maintained throughout the year in the Arctic and may be endemic populations.

Our results provide support for resource ratio theory-based predictions that Bering Sea waters would be favorable for $N_2$-fixers (31) and extend the biogeographical range of active UCYN-A symbioses into the Chukchi and Beaufort Seas. New models are needed for predicting the biogeography of $N_2$-fixing microorganisms and $N_2$ fixation in the world ocean, including other Arctic regions and the Southern Ocean. The results of this study also have implications for global $N_2$ fixation and global environmental change. Arctic ecosystems are rapidly changing.

Predicted effects include increased Pacific inflow, phytoplankton growing season, stratification, nutrient limitation, and sea surface temperatures (32), all of which may select for UCYN-A and other $N_2$-fixing species that are commonly found in warm, oligotrophic waters. The results of this study change the paradigm that $N_2$ fixation and $N_2$-fixing cyanobacteria are common only in warm tropical or subtropical waters, which is critical for understanding and predicting global patterns of $N_2$ fixation.

**Materials and Methods**

Samples were taken in the Bering Sea, Chukchi Sea, on the Chukchi Shelf, and in the Beaufort Sea in September 2016.

*DNA extraction and nifH amplification*

DNA extraction, *nifH* amplification, and quantitative PCR- Samples (2-4L) were filtered by a peristaltic pump onto sequential 3 and 0.2 µm polyphenylene ether filters (0.2 µm 25 mm Supor-200, Pall Life Sciences, Port Washington, NY, USA) in Swinnex filter holders. DNA was extracted using a modified DNeasy Plant Mini Kit (Qiagen, Germantown, MD) protocol, described in detail in (33). PCR amplification of the *nifH* gene used degenerate universal *nifH* primers Yanni/450 and up/down in a nested reaction (34), with the second round primers (up/down) modified to contain common sequence (CS) linkers (35). Library preparation was carried out by the DNA Sequencing Core Facility at the University of Illinois at Chicago (http://rrc.uic.edu/cores/genome-research/sequencing-core/). Amplicons were sequenced using Illumina MiSeq, to a sequencing depth of 40,000 sequences per sample.

UCYN-A1 and UCYN-A2 abundances were estimated using TaqMan® quantitative PCR (qPCR) chemistry and primers and probes specific for UCYN-A1 (36) and UCYN-A2 (11) and their respective haptophyte partners, UCYN-A1 host (SI Appendix, Materials, and Methods) and UCYN-A2 host (11), in samples positive for *nifH* amplification.

*$^{15}N_2$ rate measurement incubations*

$N_2$ fixation was assessed using a modified version of the $^{15}N$ bubble method (Montoya *et al.* 1996). Water samples for rate measurement incubations were collected from Niskin bottles into gas-tight 1L glass media bottles (KIMAXTM model # 611001000) capped with black open-top caps with gray butyl septa (model # 240680). The caps and septa were preconditioned in a saltwater brine for 60 days prior to use. The media bottles and caps were acid-washed (10% HCl) and rinsed with copious amounts of high purity water (18.2 MΩ $cm^{-1}$). The glass media bottles were also combusted at 500˚C for 4 hours prior to use.

*Measuring cell-specific $N_2$ fixation rates using NanoSIMS*

To visualize and map both strains and their respective hosts (UCYN-A1/UCYN-A1 host and UCYN-A2/UCYN-A2 host), a double CARD-FISH protocol was used according to the protocols detailed in (22, 23). The full suite of HRP probes, competitor oligonucleotides, and helper probes are given in SI Appendix, Table S1. Prior to nanoSIMS analysis, cells were transferred to a gridded silicon chip (1.2 cm x 1.2 cm with a 1 mm x 1 mm raster, Pelcotec™ SFG12 Finder Grid Substrate) and imaged and mapped under epifluorescence on a Zeiss Axioplan epifluorescence

microscope equipped with digital imaging at UCSC. $^{15}N$ measurements of individual cells were determined by NanoSIMS analyses performed at Stanford Nano Shared Facilities (SNSF; https://snsf.stanford.edu) on a Cameca NanoSIMS 50L at Stanford University, CA, USA**.** Image planes were accumulated after first being aligned. Isotope data was taken as a sum of counts in each plane per pixel. Cell outlines and regions of interest (ROI) were determined as the best fit based on the original CARD-FISH image, electron microscopy image, and accumulated images in $^{12}C^{14}N^-$ and $^{12}C^-$. Cell size was determined based on the ROI of the defined haptophyte or UCYN-A cell. Cell-specific $N_2$ fixation rates were determined by calculating the carbon content per cell based on a spherical cell volume (V) from the measured cell diameter determined by the ROI following the calculations of (*27*). The C:N ratio of 6.3 was measured in UCYN-A from the tropical North Atlantic (*28*) and was used in our calculation to estimate the N content of the cell. The limit of detection was determined to be three times the standard deviation of $^{15}N$ in unenriched samples (0.02 At%), similar to LOD determination in Jayakumar *et al.* (37). More detailed methods and calculations can be found in the SI appendix.

*Bulk $N_2$ fixation rates measurements*

Bottles were filled in triplicate and capped with ambient air bubbles removed. All bottles were immediately placed in mesh bags to mimic the light intensity at collection depth. Different depths received different levels of screening. The bottles were then amended with 1.1 or 2.5 mL of enriched (> 99%) $^{15}N_2$ gas purchased from Cambridge Isotope Laboratories, Inc. (lot # I-199168A). Higher volumes of $^{15}N_2$ gas

were used in all samples after Station 1 to obtain enrichment levels closer to 10% (average $^{15}N_2$ enrichment of 5.8±2.1%). Samples were incubated for 24 hours in flow through incubators on deck (surface samples) or environmental chambers (deep samples) set to 0°C ± 1°C. Prior to use in the incubations, subsamples of the $^{15}N_2$ gas stocks were assessed for $^{15}NH_4^+$, $^{15}NO_3^-$, and $^{15}NO_2^-$ contamination according to methods described in (38). No contamination was measured. Incubations were terminated after 24 hours. A membrane inlet mass spectrometer (MIMS) was used to assess the level of $^{15}N$ enrichment in each sample immediately upon incubation termination. The MIMS data for each individual bottle were used to calculate uptake rates.

Size fractionated bulk $N_2$ fixation rates were determined by filtering in series through 3.0 µm silver filters and then pre-combusted (450˚C for 2 hours) GF-75 filters with a nominal pore size of 0.3 µm. Filters were stored frozen at -20°C in sterile microcentrifuge tubes until analysis. Filters were thawed and dried overnight at 40°C and analyzed on an Integra2 combined Isotope ratio mass spectrometer with an SL autosampler tuned to low mass samples. The mass range of calibration standards was 1-10 µg N (low range) or 5-20 µg N (high range) of Sigma-Aldrich ammonium sulfate salt (0.366022/-0.77), which was calibrated against the NIST USGS40 L-glutamic acid (0.36465/-4.52) with a precision of 0.315 per mil. The limit of detection for the mass of N was 0.51 µg N. The mass range of samples analyzed was 2.66 to 19.91 µg N. The limit of detection (i.e., 3x mass of the $^{15}N$ blank) was 0.103 At% and the average minimum quantifiable rate of the bulk $N_2$ fixation rates was 0.4 ± 0.7

nmol N $L^{-1}$ $day^{-1}$. The limit of detection (LOD) and minimum quantifiable rates (MQR) were calculated according to Montoya, Voss, Kahler, and Capone (39) and Gradoville *et al.* (40) for each size fraction and propagated as error to represent total $N_2$ fixation. Controls (natural abundance) samples were collected from the Niskin in dedicated, acid-washed (10% HCl) high-density polyethylene (HDPE) bottles and filtered in a separate laboratory on a filtration unit designated for no isotope used. Blank natural abundance samples were analyzed on an Integra2 combined Isotope ratio mass spectrometer with an SL autosampler that had not been exposed to enriched samples. The average $\delta^{15}N$ was 7.8 ± 4.2 (0.37 At%) for the >3µm size fraction and 7.2 ± 2.9 (0.37 At%) for the 0.3 – 3 µm size fraction.

**Acknowledgments**

We gratefully acknowledge Mary-Kate Rogener (UGA) for providing MIMS analysis, Quinn Roberts (VIMS) for providing EA-IMRS analysis, Rosie Gradoville (UCSC) for discussions about $N_2$ fixation rate measurements, as well as Laurie Juranek (OSU) and the captain and crew of the R/V Sikuliaq for field logistical support. We also thank Chuck Hitzman at the Stanford Nano Shared facility for nanoSIMS consultation, Lubos Polerecky for look@nanoSIMS consulting, and Stefan Green and his staff at the DNA Services Facility and the University of Illinois, Chicago for sequencing consultation. We greatly appreciate Mick Follows (MIT) and Kevin Arrigo (Stanford University) for helpful discussions. This research was funded by the National Science Foundation (PLR-1503614, J.Z., and PLR-1504307, R.S.), National Science Foundation Awards (number 1241093 and 1559152), and the Simons Foundation (SCOPE Award ID 329108, J.Z.).

*Data and materials availability:*

All data is provided in the manuscript, and SI Appendix with the exception of the raw UCYN-A *nifH* sequences, which were deposited in the Sequence Read Archive at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/sra) under Bioproject ID PRJNA476143. Data will be available on BCO DMO upon the completion of the project.

*Author contributions:*

R. S., K. T-K. D. B., and J. Z. conceived the project. K. H., R.S., and D.B. collected samples and performed molecular analyses. M.M. and K.H. performed single-cell measurements and analysis. R.S. and D. B. measured the bulk rates. K.H., J.Z., K. T.-K wrote the manuscript. All authors were involved in reviewing results and reviewing and editing the manuscript.

**References:**

1. Falkowski PG (1998) Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Proc Natl Acad Sci USA* 281(5374):200-206.

2. Breitbarth E, Oschlies A, & LaRoche J (2007) Physiological constraints on the global distribution of *Trichodesmium* - effect of temperature on diazotrophy. *Biogeosciences* 4(1):53-61.

3. Tyrrell T (1999) The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* 400(6744):525-531.

4. Thompson AW, *et al.* (2012) Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. *Science* 337(6101):1546-1550.

5. Tripp HJ, *et al.* (2010) Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* 464(7285):90-94.

6. Moisander PH, *et al.* (2010) Unicellular cyanobacterial distributions broaden the oceanic $N_2$ fixation domain. *Science* 327(5972):1512-1514.

7. Farnelid H, Turk-Kubo K, Munoz-Marin MD, & Zehr JP (2016) New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquat Microb Ecol* 77(3):125-138.

8. Bentzon-Tilia M, *et al.* (2015) Significant $N_2$ fixation by heterotrophs, photoheterotrophs, and heterocystous cyanobacteria in two temperate estuaries. *ISME J* 9(2):273-285.

9. Shiozaki 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. Cy.* 31(6):996-1009.

10. Shiozaki T, *et al.* (2018) Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnol Oceanogr*.

11. Thompson A, *et al.* (2014) Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ Microbiol* 16(10):3238-3249.

12. Turk-Kubo KA, Farnelid HM, Shilova IN, Henke B, & Zehr JP (2017) Distinct Ecological Niches of Marine Symbiotic $N_2$-Fixing Cyanobacterium *Candidatus Atelocyanobacterium Thalassa* Sublineages. *J Phycol* 53(2):451-461.

13. Comeau AM, Li WKW, Tremblay JE, Carmack EC, & Lovejoy C (2011) Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. *Plos One* 6(11).

14. Brauer VS, *et al.* (2013) Low temperature delays timing and enhances the cost of nitrogen fixation in the unicellular cyanobacterium Cyanothece. *ISME J* 7(11):2105-2115.

15. Rivero-Calle S, *et al.* (2016) Interdecadal *Trichodesmium* variability in cold North Atlantic waters. *Global Biogeochem. Cy.* 30(11):1620-1638.

16. Rees AP, Gilbert JA, & Kelly-Gerreyn BA (2009) Nitrogen fixation in the western English Channel (NE Atlantic Ocean). *Marine Ecology-Progress Series* 374:7-12.

17. Mehta MP & Baross JA (2006) Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon. *Science* 314(5806):1783-1786.

18. Olson J, Steppe T, Litaker R, & Paerl H (1998) $N_2$-Fixing Microbial Consortia Associated with the Ice Cover of Lake Bonney, Antarctica. *Microb Ecol* 36(3):231-238.

19. Sipler RE, *et al.* (2017) Preliminary estimates of the contribution of Arctic nitrogen fixation to the global nitrogen budget. *Limnology and Oceanography Letters* 2(5):159-166.

20. Blais M, *et al.* (2012) Nitrogen fixation and identification of potential diazotrophs in the Canadian Arctic. *Global Biogeochem. Cy.* 26(3):GB3022.

21. Church MJ, Bjorkman KM, Karl DM, Saito MA, & Zehr JP (2008) Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. *Limnol Oceanogr* 53(1):63-77.

22. Cornejo-Castillo FM, *et al.* (2016) Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat Commun* 7:11071.

23. Cabello AM, *et al.* (2016) Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. *ISME J* 10(3):693-706.

24. Krupke A, *et al.* (2014) Distribution of a consortium between unicellular algae and the $N_2$ fixing cyanobacterium UCYN-A in the North Atlantic Ocean. *Environ. Microbiol.* 16(10):3153-3167 %@ 1462-2920.

25. Lovejoy C, *et al.* (2007) Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas. *J Phycol* 43(1):78-89.

26. Paulsen ML, *et al.* (2016) Synechococcus in the Atlantic gateway to the Arctic Ocean. *Frontiers in Marine Science* 3:191.

27. Krupke A, *et al.* (2015) The effect of nutrients on carbon and nitrogen fixation by the UCYN-A-haptophyte symbiosis. *ISME J* 9(7):1635-1647.

28. Martinez-Perez C, *et al.* (2016) The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat Microbiol* 1(11):16163.

29. Galand PE, *et al.* (2009) Archaeal diversity and a gene for ammonia oxidation are coupled to oceanic circulation. *Environ Microbiol* 11(4):971-980.

30. Lovejoy C & Potvin M (2011) Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. *J Plankton Res* 33(3):431-444.

31. Ward BA, Dutkiewicz S, Moore CM, & Follows MJ (2013) Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. *Limnol Oceanogr* 58(6):2059-2075.

32. Michel C, *et al.* (2015) Arctic Ocean outflow shelves in the changing Arctic: A review and perspectives. *Prog Oceanogr* 139:66-88.

33. Moisander PH, Beinart RA, Voss M, & Zehr JP (2008) Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *ISME J* 2(6):954-967.

34. Zehr JP & McReynolds LA (1989) Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl Environ Microbiol* 55:2522-2526.

35. Moonsamy PV, *et al.* (2013) High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array System for simplified amplicon library preparation. *Tissue Antigens* 81(3):141-149.

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

37. Jayakumar A, *et al.* (2017) Biological nitrogen fixation in the oxygen-minimum region of the eastern tropical North Pacific ocean. *ISME J* 11(10):2356-2367.

38. Dabundo R, *et al.* (2014) The contamination of commercial $^{15}N_2$ gas stocks with 15 N–labeled nitrate and ammonium and consequences for nitrogen fixation measurements.

39. Montoya JP, Voss M, Kahler P, & Capone DG (1996) A simple, high-precision, high-sensitivity tracer assay for $N_2$ fixation. *Appl Environ Microbiol* 62:986-993.

40. Gradoville MR, *et al.* (2017) Diversity and activity of nitrogen-fixing communities across ocean basins. *Limnol Oceanogr* 62(5):1895-1909.

## Figures

**Fig. 1.** UCYN-A lineages are distributed throughout surface waters of the Western Arctic Ocean. Background colors represent sea surface temperature on September 10, 2016 (www.esrl.noaa.gov). UCYN-A was quantified with qPCR. UCYN-A2 (black circles) is present at more stations, but UCYN-A1 (white circles) had the highest maximum abundance. Red stars indicate nanoSIMS $^{15}N$ uptake rates measure locations.

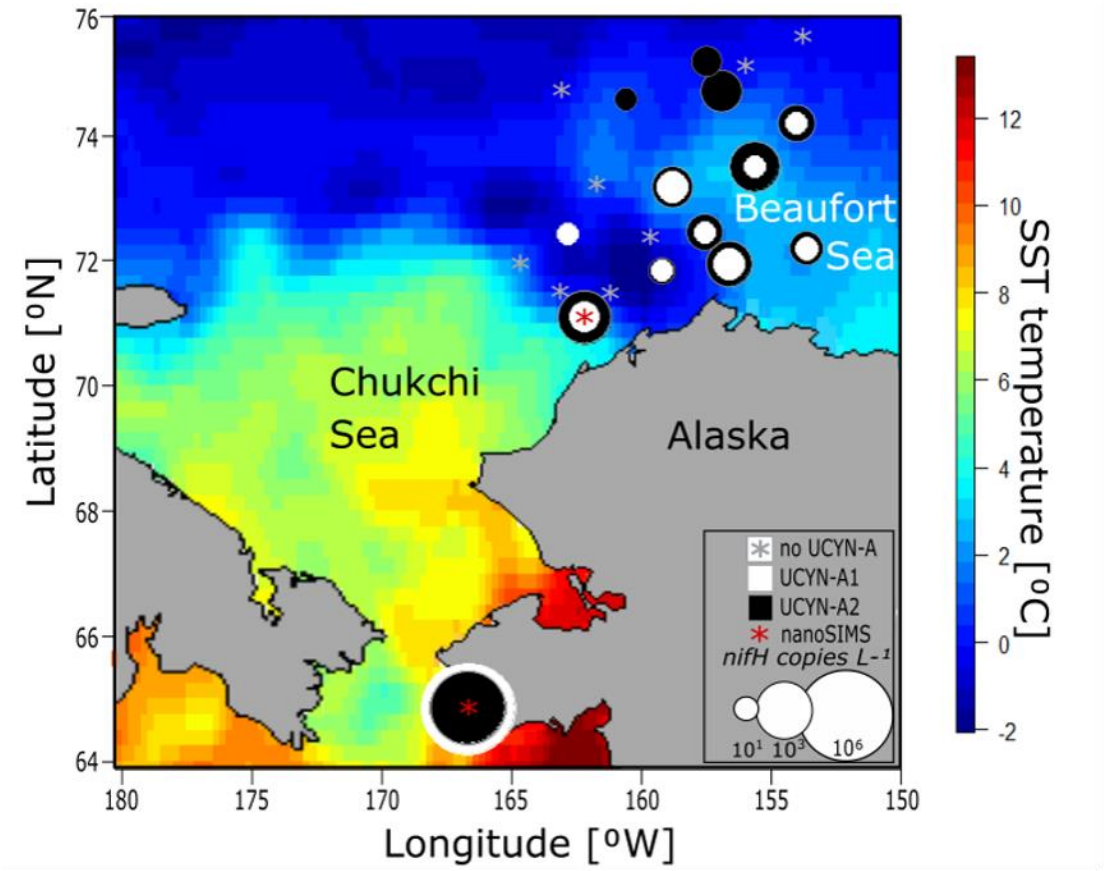


**Fig. 2**. Arctic UCYN-A *nifH* sequences are identical to broadly distributed and abundant sequence types. Maximum likelihood phylogenetic tree of UCYN-A microdiversity based on partial *nifH* nucleotide sequences from a recent global survey (12). UCYN-A sequences from the Bering and Arctic Seas (red dots) are identical to dominant sequence types found in all major ocean basins. Regions where each sequence type has been found, are specified by colored dots according to the legend, sequence counts from the global survey are also plotted. Nodes with bootstrap support >70 are identified with a diamond. Adapted from (12).

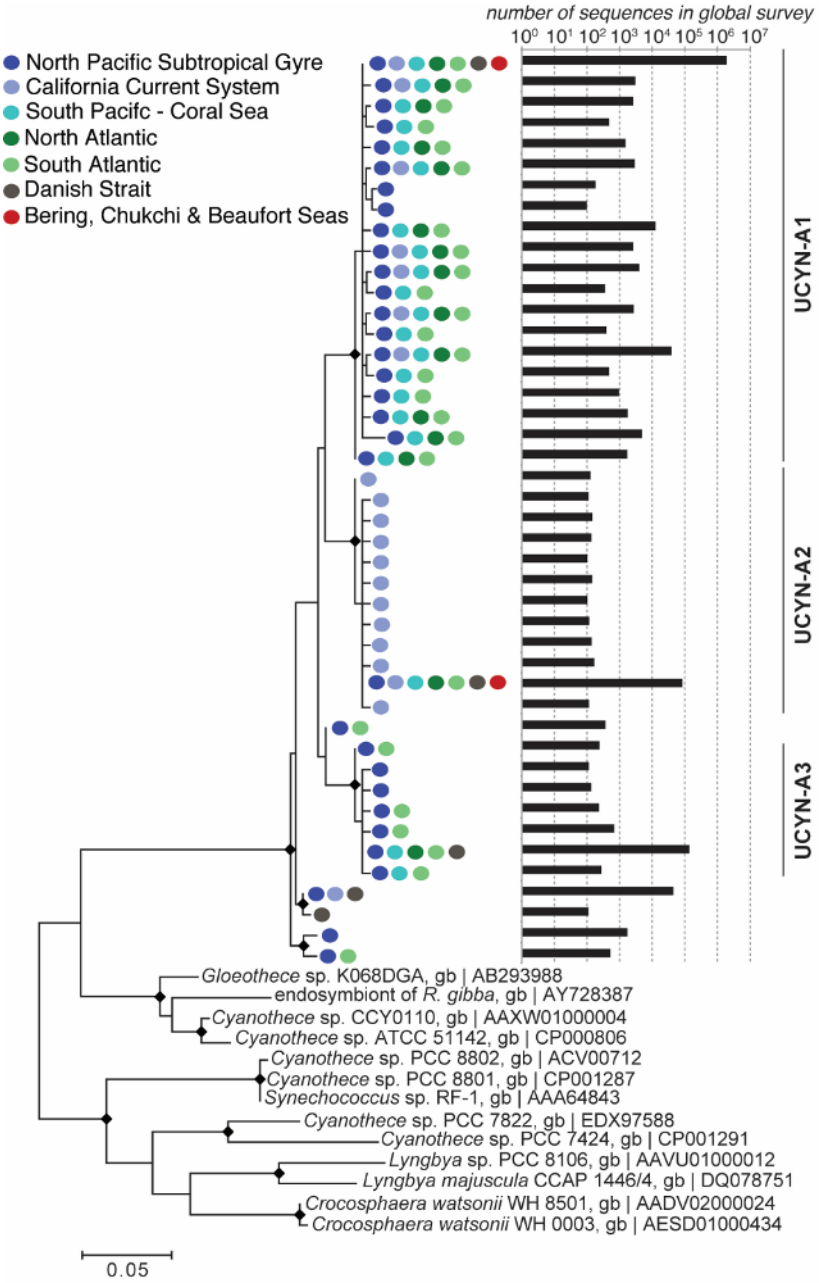


**Fig. 3**. Morphologies of Arctic UCYN-A symbioses are indistinguishable from subtropical strains. Double CARD-FISH comparison of UCYN-A lineages from the Bering Sea (A, C) and water collected at the Scripps Institute of Oceanography (SIO) Pier in La Jolla, CA (B, D) show similar sizes and morphologies. CARD-FISH images show the symbiosis is intact, with both the haptophyte host (green & blue) and cyanobacteria (red). Scale bar is 5 µm.

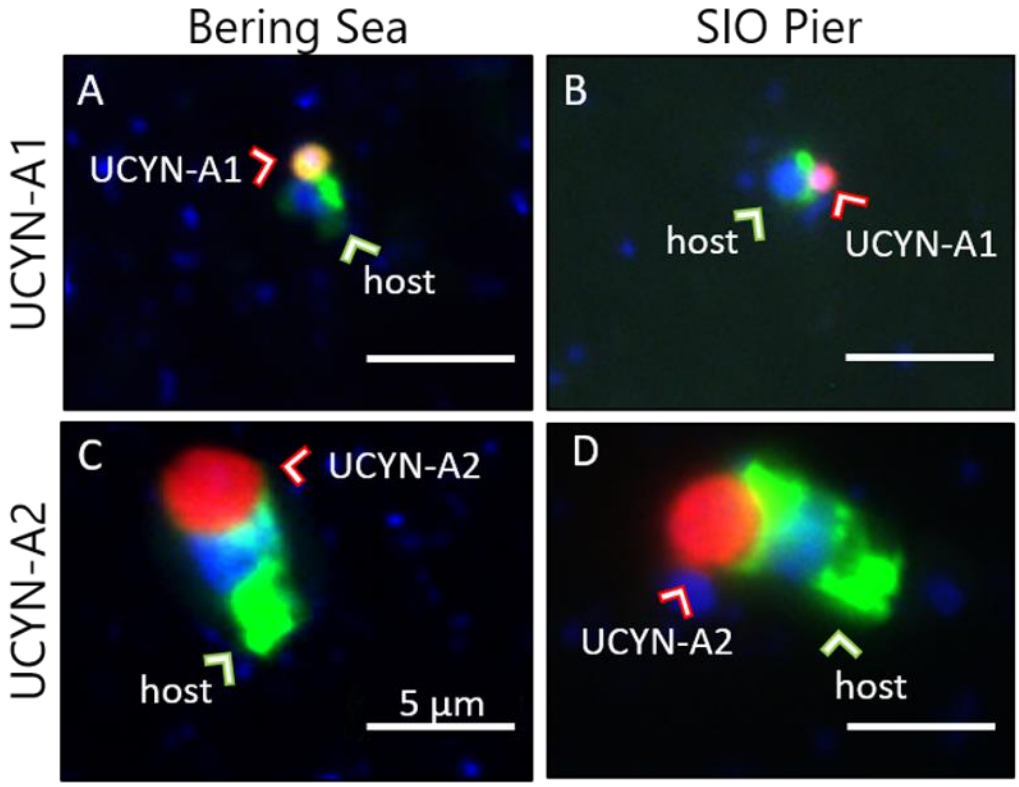
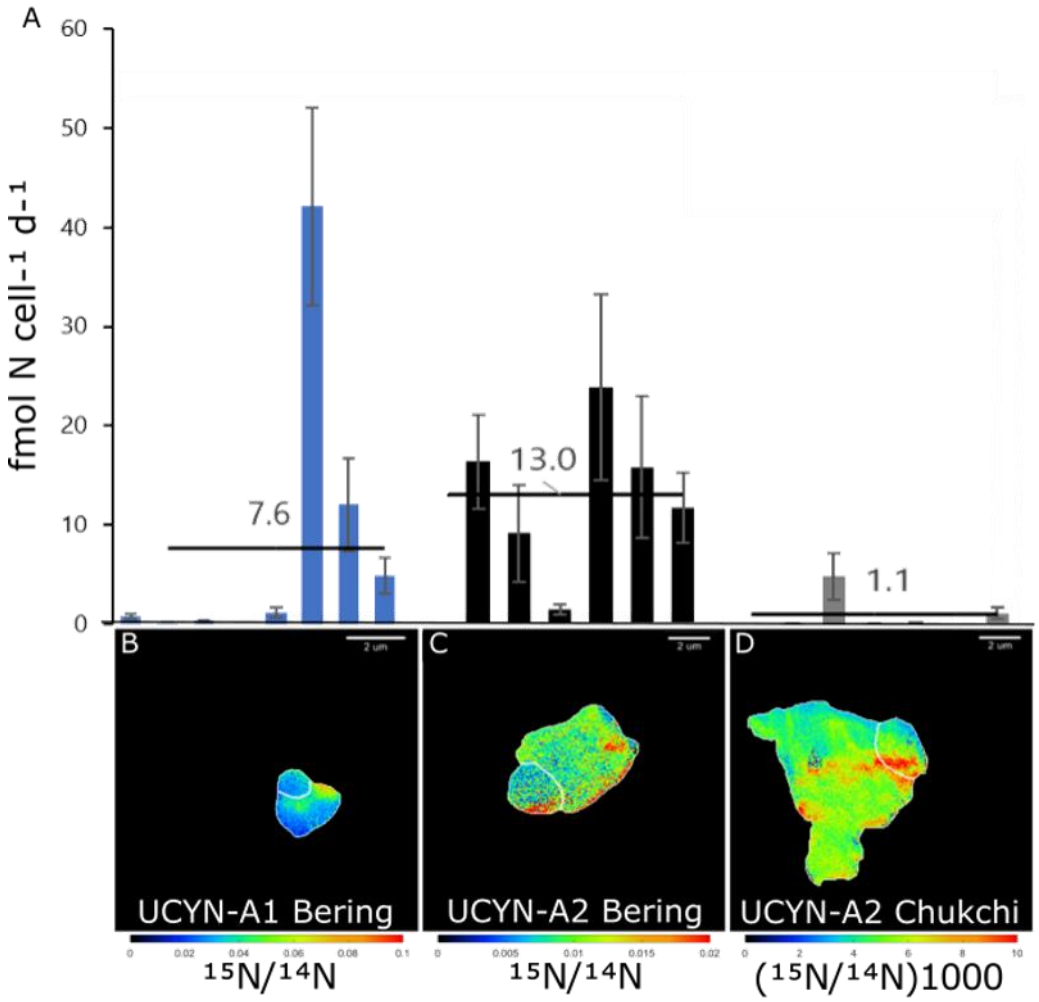


**Fig. 4.** UCYN-A symbioses fix $^{15}N_2$ in western Arctic waters. UCYN-A cell-specific $^{15}N_2$ fixation rates (A) and $^{15}N$ enrichment (B-D) from nanoSIMS measurements after incubating natural populations in seawater with $^{15}N_2$. Bars of the same color (A) represent rates measured in individual symbioses (UCYN-A with host alga) from a single station and lineage (noted in underlying cell image, scale bar is 2 µm), averages shown by the horizontal black line. Error bars are the standard deviation of the cell-specific rate between the host and UCYN-A. Note the colored axes differ in scale on nanoSIMS images (B-D).



# Chapter 3

## Direct cell-specific measurements show $N_2$ fixation by particle-attached non-cyanobacterial diazotrophs in the North Pacific Subtropical Gyre

Harding, K., Turk-Kubo, K.A., Mak, E.W.K., Weber, P.K., Mayali, X., Zehr, J.P. (2021) Submitted to: *Nature Communications*

**Abstract**

Biological nitrogen ($N_2$) fixation is an essential source of nitrogen (N) input for low nutrient surface marine waters. $N_2$-fixing (diazotrophic) cyanobacteria have been established as the primary contributors to this process, but the contribution of non-cyanobacterial diazotrophic organisms in oxygenated surface water has yet to be demonstrated. Previous commonly used approaches such as coupling community $N_2$ fixation rate measurements with functional marker gene sequencing cannot evaluate the contribution of the detected non-cyanobacterial diazotrophs to the measured community $N_2$ fixation rates. In this study, we used simultaneous $^{15}N_2$ and $^{13}C$-bicarbonate incubations combined with nanoSIMS analysis to screen tens of thousands of cells collected from the North Pacific Subtropical Gyre. These dual-isotope incubations allowed us to distinguish between non-cyanobacterial and cyanobacterial $N_2$-fixing microorganisms and to measure cell-specific $N_2$ fixation rates for those populations. This approach led to the first direct demonstration of $N_2$ fixation by non-cyanobacterial diazotrophs in the oxygenated surface ocean, which were associated with particles, and 3.2% of the analyzed particles (<210 μm) contained at least one active non-cyanobacterial diazotroph at four out of seven locations sampled. The non-cyanobacterial $N_2$ fixation rates presented here (avg: 1.09 ± 1.61 fmol N cell$^{-1}$ d$^{-1}$) demonstrate these organisms are capable of fixing $N_2$ in oxygenated surface water while attached to particles and contribute to oceanic $N_2$ fixation.

## Introduction

Primary productivity in the oceans is commonly limited by nutrient availability, and nitrogen (N) is the limiting nutrient in large regions of the surface oceans (Falkowski et al., 1998). Biological $N_2$ fixation is an energetically expensive process that converts atmospheric dinitrogen ($N_2$) to bioavailable forms of N (ammonia and amino acids), supporting primary production (Capone et al., 2005; Karl et al., 2002) and 26-47% of particulate N export (Bottjer et al., 2017). Quantifying biological $N_2$ fixation rates is critical for predicting carbon and nitrogen fluxes, yet there is high variability among biogeochemical model estimates of $N_2$ fixation, sometimes several orders of magnitude (Wang et al., 2019), suggesting these processes are not well constrained. This disconnect is likely due to many factors but two related yet distinct factors will be mentioned here. First, only a few $N_2$-fixing microorganisms are included in these models, and there are likely many other organisms yet to be discovered that fix $N_2$ in the surface ocean. Second, the physiology, genetic diversity, and *in situ* activity of the known diazotrophs are not well constrained, making them difficult to model at the ocean scale (Bombar et al., 2016; Horner-Devine and Martiny, 2008). Diverse diazotrophic Bacteria and Archaea have been shown to have the potential to fix $N_2$ as identified through amplification of their *nifH* gene (Zehr et al., 1998), which encodes a subunit of the nitrogenase enzyme used in $N_2$ fixation (Howard and Rees, 1996). Cyanobacteria, such as *Trichodesmium*, heterocyst forming symbionts of diatoms, and unicellular

cyanobacteria (*Crocosphaera* and UCYN-A), have been shown to be the dominant $N_2$ fixers in warm, low-nutrient, surface ocean waters through culture-based studies and single-cell analyses (Capone et al., 1997; Carpenter and Capone, 1992; Villareal, 1994; Zehr et al., 2001; Zehr and Bombar, 2015). However, amplification of *nifH* genes from ocean waters shows that there are abundant and diverse non-cyanobacterial diazotroph (NCD) *nifH* sequences (Farnelid et al., 2010; Riemann et al., 2010; Zehr et al., 1998) that often exceed the relative abundance of amplified cyanobacterial *nifH* genes (Farnelid et al., 2011; Riemann et al., 2010). Yet, the relative abundance of genes is not equivalent to volumetric cell abundances, and it has not yet been directly demonstrated how many of the NCDs fix $N_2$ or what their contribution is to measured community $N_2$ fixation rates.

NCDs, which could be heterotrophic or photoheterotrophic $N_2$-fixing Bacteria or Archaea, have been largely considered insignificant in biological $N_2$ fixation in marine surface waters (Bombar et al., 2016). The significance of NCD $N_2$ fixation has been questioned because of the low concentration of organic matter and relatively high concentrations of dissolved oxygen in surface waters, which inhibits this process (Bombar et al., 2016; Turk-Kubo et al., 2014). Heterotrophic $N_2$ fixation has been suggested to occur in or on particles (Bombar et al., 2013; Church et al., 2005a; Paerl and Prufert, 1987; Pedersen et al., 2018; Riemann et al., 2010) which could provide a rich source of carbon (Riemann et al., 2010) and in particles $\geq$ 1 mm in diameter microbial respiration-induced microaerobic zones (Klawonn et al., 2015a; Ploug, 2001; Riemann et al., 2010), which could support heterotrophic diazotrophy. In

agreement with this hypothesis, NCD *nifH* genes have been found on individual particles in the North Pacific Subtropical Gyre (NPSG) (Farnelid et al., 2019). However, the presence of the *nifH* gene does not ensure that the genes are transcribed or that there is active nitrogenase protein, and the diversity of amplified *nifH* genes does not necessarily reflect the quantitative abundance of an organism or its activity. Nitrogenase proteins from putative NCDs have been visualized associated with particles by immunolabeling in estuarine samples (Geisler et al., 2020), but $N_2$ fixation rates from these particle attached NCDs have yet to be measured.

NCDs could be a significant component of marine $N_2$ fixation, but their activity and quantitative significance have yet to be directly demonstrated. The potential $N_2$ fixation rates of NCDs are difficult to assess since most open ocean NCDs do not have cultured representatives. Marine NCD $N_2$ fixation rate measurements are limited to a few cultured representatives from estuarine environments with rates of 0.02 to 1.1 fmol N $cell^{-1}$ $d^{-1}$ (scaled to per day rates assuming 24 hours of $N_2$ fixation) (Bentzon-Tilia, Severin and Hansen, 2015; Paerl *et al.*, 2018). Unicellular, surface ocean cyanobacteria such as UCYN-A have much higher single cell $N_2$ fixation rates of 2 to 220 fmol N $cell^{-1}$ $d^{-1}$ (Gradoville et al., 2020; Martínez-Pérez et al., 2016). Additionally, a few indirect measurements of community $N_2$ fixation rates (0.7 to 8 nmol N $l^{-1}$ $d^{-1}$) have been inferred from locations where cyanobacteria were reportedly absent (reviewed in Moisander *et al.*, 2017). Nanoscale secondary ion mass spectrometry (nanoSIMS) analysis has been used extensively to identify single-cell activity, including $N_2$ fixation by uncultured

diazotrophs (Dekas et al., 2009; Woebken et al., 2012). Diazotrophs for which there are 16S rRNA gene sequences can be identified by catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH) or related methods (Behrens et al., 2008; Dekas et al., 2009; Thompson et al., 2012) and shown to fix $N_2$ by nanoSIMS analysis measuring cellular $^{15}N$ incorporation. However, most NCDs are only known by their *nifH* gene sequence, so visualization and identification using 16S rRNA-based CARD-FISH is not possible. Furthermore, FISH-nanoSIMS is generally low throughput, making surveys for potentially rare organisms, such as NCDs, impractical, due to the difficulty in identifying and mapping rare cells.

In this study, we use a dual-isotope nanoSIMS approach to *d*etermine if NCDs were fixing $N_2$ in surface waters of the NPSG. We incubated seawater samples in $^{15}N_2$ and $^{13}C$-bicarbonate to distinguish between cyanobacteria and NCDs, the former fixing both $CO_2$ and $N_2$ and the latter only $N_2$. This untargeted approach does not require identifying NCDs by CARD-FISH or other means. We also measured community (bulk water) $N_2$ fixation rates using $^{15}N_2$ and determined diazotroph diversity by *nifH* gene sequencing. We present the first direct measurements of NCD $N_2$-fixation rates in the surface ocean.

**Methods**

Seven locations (Fig. 1) spanning the NPSG between Guam and San Francisco (November 2019) were analyzed for the presence of $N_2$-fixing organisms by *nifH* sequencing, community $N_2$ fixation rate measurements, and cell-specific activities of NCD cells.

*Sample collection*

Samples were collected using Niskin bottles attached to a CTD profiler from surface seawater (15 m). All water samples were collected in acid-cleaned polycarbonate bottles rinsed 3 times with local seawater. Large grazers were removed while the bottles were filled using a 210 µm Nitex™ plankton netting (BioQuip, Rancho Dominguez, CA). Samples collected from each station include natural isotope abundance (triplicate samples of 2 L), DNA (replicates of 2 L), and isotope incubations (triplicates of 4.4 L). NanoSIMS samples (0.1 – 0.5 L) were subsampled from isotope incubations. Bottles for isotope incubations were immediately placed in a surface seawater flow-through incubator shaded to approximate 15 m light intensities until isotope addition (~1 hour after initial sampling). Natural isotope abundance and DNA water samples were filtered immediately.

*DNA extraction and nifH sequencing*

Diazotroph diversity was assessed by PCR amplification and sequencing of the *nifH* gene. Surface seawater (2 L, 15 m, <210 µm) was filtered onto 0.2 µm filters (PE 25mm; Supor-200; Pall Life Sciences, Port Washington, NY, USA) using peristaltic pumps. Filters were flash-frozen before storage at -80$^{o}$C. DNA was extracted with DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) with modifications for increased cell lysis (Moisander et al., 2008). The *nifH* gene was amplified by PCR using the universal *nifH* outer primers, YANNI/450, and common-sequencer linkers with inner primers, up/down, in a nested reaction (Zehr and Mcreynolds, 1989). Amplicons were sequenced with Illumina MiSeq sequencing (2 x 300 bp, with a

targeted sequencing depth of 20,000 per sample) at the University of Illinois Genome Research Core Facility. Raw sequences were processed as described in Cabello et al. (2020). Briefly, sequences were quality controlled and clustered at 97% nucleotide identity using Qiime (Caporaso et al., 2010), representative sequences were assigned phylogeny with BLASTX, and OTU tables were rarified according to the sample with the lowest sequence recovery (738 sequences).

*$^{15}N_2$ incubations*

To investigate the $N_2$ fixation activity of NCDs, triplicate 15 m depth seawater samples (<210 um, 4.4 L) were injected with a $^{15}N_2$ gas bubble (98%+, lot # - I-22779?AR0720534, Cambridge Isotopes, Tewksbury, MA, USA, ) following the bubble release method (Klawonn et al., 2015b). Bottles were rolled back and forth to equilibrate the $^{15}N_2$ bubble for 20 minutes (~25 rpm) before bubble release and $^{13}C$-bicarbonate addition (60 µM, 99%, Cambridge). Triplicate incubations were transferred back to the incubators shaded to the light intensity at 15 m depth (7 stations) or complete darkness (6 stations) for 24 hours.

*Community $N_2$ fixation rates*

Isotope incubations were used for particulate organic nitrogen measurements for community $N_2$ fixation rates. Membrane inlet mass spectrometry (MIMS) samples were taken at the end of 24-hour incubations from each bottle to measure the dissolved $^{15}N_2$-gas enrichment available (2.8 At% ± 0.4). Incubation bottles were opened, and MIMS samples were siphoned into glass vials (15 mL) which were then stoppered and crimped with aluminum caps. Samples were kept at 4$^{o}$C until

measured. After subsampling for nanoSIMS samples, biological triplicates were vacuum filtered (3.5-3.8 L) through pre-combusted (4 hours at 450ºC) 25 mm GF/F (Whatman®) and flash-frozen before storage at -80ºC. Additional water samples (2 L) were collected from each location in triplicate and immediately filtered with a separate vacuum filtration system for $^{15}N$ natural abundance measurements, except for station 14, which only had duplicate samples available. All filters were dried at 75ºC for 72 hours before being pelleted in tin foil disks (30 mm, Elemental Analysis, Okehampton, UK). Samples were not acidified to remove inorganic C before pelleting, as the $^{13}C$ isotope addition was only in the nanoSIMS analysis. The $^{15}N$ enrichment and natural abundance of particulate organic N were measured along with blanks using a Carlo-Erba EA NC2500 coupled with Thermo Finnigan DeltaPlus XP at the University of Hawaii Stable Isotope Facility. $N_2$ fixation rates, as well as the minimum quantifiable rate (MQR; lowest rate quantified with confidence provided the propagated errors for replicate samples) and limit of detection (LOD; lowest value considered detected, final At% - At% of natural abundance > 0.00146 At%), were calculated as in Gradoville et al., (2017) according to Montoya (1996); all values are listed in Table S1. Rate values that fell below the LOD but above the MQR are reported. Community $N_2$ fixation incubations and reporting follow recommendations from White et al. (2020): all samples were run in triplicate except as described above ($T_o$ Stn 14), and $^{15}N$ incubation filters contained >10 μg N per filter. Our experimental design differed from recommendations by White et al. (2020) in that

incubations were not initiated before dawn, our 24-hr incubations started at various times throughout the day.

*NanoSIMS analyses*

Subsamples of the isotope incubations were filtered (0.1-0.45 L) through 0.2 µm pore-size silver membrane filters (25mm, Cole-Parmer, Vernon Hills, IL, USA) then fixed (1.8% formaldehyde) at room temperature for 2 hours. Filters were flash-frozen before storage at -80$^{o}$C until nanoSIMS analysis (CAMECA NanoSIMS 50, Lawrence Livermore National Laboratory). NanoSIMS measurements were conducted with both untargeted and particle-targeted approaches. The untargeted approach sequentially scanned contiguous areas of the filter using an automated analysis routine ("chained analysis"). To target particles, their locations were mapped using the secondary electron imaging in the nanoSIMS and then analyzed using an automated analysis routine. These analyses only include cells at the particle surface. Several of the targeted particles were subsequently analyzed at multiple depths by serially eroding into the particle with a high $Cs^+$ current and collecting data. Measurements were made with a ~2 pA $Cs^+$ primary beam, and data were collected for the masses of $^{12}C_2^-$, $^{12}C^{13}C^-$, $^{12}C^{14}N^-$ and $^{12}C^{15}N^-$ where $^{12}C^{13}C^-/^{12}C_2^- = 2 \times {}^{13}C/{}^{12}C$; $^{12}C^{15}N^-/^{12}C^{14}N^- = {}^{15}N/{}^{14}N$ (Pett-Ridge and Weber, 2012). Before each surface analysis, the area was sputtered to a depth of ~60 nm to establish sputtering equilibrium. Data were then collected with a raster of 25-40 µm with 256 x 256 pixels and 30-60 frames for each image. Images were processed using L'Image software (developed by L. Nittler, Carnegie Institution of Washington, Washington, DC,

USA). Cells were identified as regions of interest (ROIs) based on counts of $^{12}C^{14}N$ (150). This threshold was set relatively low to allow for the duration of individual analyses and small cells. Cells were considered to have incorporated a statistically significant amount of an isotopically labeled substrate if the isotope enrichment was greater than that of unlabeled reference cells (*Pseudomonas*) by 3x the error associated with the minimum acceptable count for an ROI; these thresholds correspond to 1.23 At% $^{13}C$ and 0.47 At% $^{15}N$, which are relatively conservative. The enrichment of our reference cells was statistically indistinguishable from the unenriched cells in this study.

Enrichment values of $^{13}C$ and $^{15}N$ were used to calculate net assimilation percent (Xnet%, where X can be C or N), which we use to estimate newly synthesized biomass relative to total biomass, assuming no change in cell stoichiometry. For N, Nnet% = $[F_s/(F_s + F_i)] *100$, where $F_s$ is the N fraction derived from the isotope substrate and $F_i$ is the N fraction of the original biomass (Dekas et al., 2019). $F_s$ and $F_i$ are further defined by Popa et al. (Popa et al., 2007) in terms of the isotopic ratio of the initial pool ($R_i$), spiked pool ($R_s$), and final ratio in the cell of interest ($R_f$), where Xnet% = $Fx_{net}$%/($Fx_{net}$%+100)

$$Fx_{net} = \frac{R_f\left(1 - \frac{R_i}{R_i + 1}\right) - \frac{R_i}{R_i + 1}}{\frac{R_s}{R_s + 1} - R_f\left(1 - \frac{R_s}{R_s + 1}\right)} \cdot 100\%$$

Cell-specific $N_2$ fixation rates (fmol N $cell^{-1}$ $d^{-1}$) were calculated based on nanoSIMS-measured atom percent (At%) of $^{15}N$-enriched cells following Krupke et al. (2015). T-initial values were based on the average IRMS value of natural isotope

abundance samples from each location, while MIMS values were used to define the amount of $^{15}N$ available in the enriched substrate. The carbon content per cell was based on a spherical cell volume from the defined ROI using the C content per cell, according to Verity et al. (1992). Estimates for N content per cell were adjusted for heterotrophic cells by using an average C:N ratio measured in cultured heterotrophic bacterial cells of 5.2 (Vrede et al., 2002).

*Estimating NCD contribution from single-cell measurements*

We carried out calculations in two different ways (with slightly different assumptions) to scale the nanoSIMS data to a volumetric context (summarized in Table 1) and provide a lower and upper bound approximation of the contribution of NCD diazotrophy to total activity at the stations sampled. These calculations combined single-cell nanoSIMS data, the frequency of particles on which we found at least one diazotroph, and particle abundances from each station. In the first calculation (conservative), we assumed the portion of NCDs on the surface of particles are representative of the total NCD abundance and simply multiplied the number of particles by the average number of NCDs found per particle and the average $N_2$ fixation rate of NCD cells. This first method clearly underestimates the abundance of NCDs because our depth profile analyses of selected particles showed no difference in the relative abundance of NCDs at the surface of particles compared to the interior. Therefore, in the second set of calculations, we estimated the abundance of NCDs based on the abundance per unit volume of analysis, which we consider to be 1 micron in-depth for an average particle size (15 μm x 15 μm) times

an average particle depth of 4 µm. This estimate is more realistic but is still conservative because the actual depth of analysis is less than 1 micron, and using an average particle volume is assumptive. The mean NCD $N_2$ fixation cell-specific rates (fmol N $cell^{-1}$ $d^{-1}$) represent the means based on a gamma distribution (Stns. 5 and 10, $\alpha = 0.05$) or the mean of normal distributions (Stns. 20 and 22). Particles (> 5 µm) were quantified by staining filter pieces corresponding to nanoSIMS samples with DAPI (4′,6-diamidino-2-phenylindole,1µg $\mu l^{-1}$) and counting with microscopy (Zeiss Axioplan epifluorescence microscope). Unattached cyanobacterial-like diazotroph abundances were estimated based on the volume of water equivalent to the area analyzed with nanoSIMS.

**Results and Discussion**

Diazotroph diversity was investigated by sequencing *nifH* genes amplified from filtered seawater samples (Fig. 1). We detected both cyanobacterial and NCD *nifH* sequences at all locations sampled. Three locations had comparable relative abundances of cyanobacterial and NCD *nifH* sequences (Stns. 5, 10, and 20), while the western-middle stations were dominated by either cyanobacterial sequences (Stn. 14) or NCD sequences (Stn. 17). The two eastern most stations (Stns. 22 and 23) had similar compositions, with approximately ¾ of the total sequences identified as cyanobacterial and the remaining ¼ of sequences identified as NCD. The most common non-cyanobacterial *nifH* sequence was the γ-proteobacterium known as "Gamma A" (Langlois et al., 2008; Zehr et al., 1998), a putative heterotroph also identified as γ-247211A (Moisander et al., 2008). Relative abundances of Gamma A ranged from 34 to 99% of the total non-cyanobacterial *nifH* sequences at each

location (Fig. S1). The non-cyanobacterial sequence with the second-highest relative abundance was also a putative γ-proteobacterium (cluster 1G) and ranged from 1 to 48% of total NCD sequences per sample. Sequences from the 1O/1P cluster (namely b-proteobacteria) were also present but at much lower relative abundances (0 to 6% of NCD sequences). It is important to note that whole seawater *nifH* sequencing alone does not provide information about which NCDs may be particle-attached versus free-living. In fact, some particle-attached sequences may have been missed entirely as previous studies found that sequences from individual particles (>20 μm) were not well-represented in the whole water column diazotrophic community composition (Farnelid et al., 2019; Pedersen et al., 2018). Other studies have further indicated the possibility that some diazotrophic diversity may be missed or misrepresented due to PCR biases in *nifH* amplification and sequencing (Delmont et al., 2018; Turk-Kubo et al., 2014). Regardless, sequencing shows the potential for NCD $N_2$ fixation through the presence and relatively high proportions of NCD *nifH* genes.

To determine the activity of the diazotrophic populations at specific locations, we measured community (bulk water) $N_2$ fixation rates, which were relatively low (0.11 to 0.51 nmol N $l^{-1}$ $d^{1}$) (Fig. 1) compared to the well-studied Station ALOHA site in the NPSG (0.3 to 21 nmol N $l^{-1}$ $d^{-1}$) (Bottjer et al., 2017). The natural-light $N_2$ fixation rates were all above the MQR (3 stns > LOD), as were 4 out of 6 all-dark $N_2$ fixation rates (1 stn > LOD). Stn. 17 & 20 rates were below MQR and all-dark data for Stn. 23 are not available). Quantifiable all-dark $N_2$-fixation rates accounted for 24 to 69% of the natural-light $N_2$ fixation rates. For comparison, all-dark to natural-light

$N_2$ fixation in the NCD-dominant South Pacific Gyre ranges from 28% to over 100% (Halm et al., 2012) and 63% to over 100% in a temperate estuary (Bentzon-Tilia, Severin, and Hansen 2015). The comparison of all-dark to natural-light community $N_2$ fixation rates could indicate photosynthesis independent $N_2$ fixation. However, cyanobacteria that fix $N_2$ in the dark such as *Crocosphaera* and *Cyanothece* (Mohr et al., 2010), both present in this study (*Crocosphaera:* 0-39% *Cyanothece* <1%, Fig. Stn. 1), make all-dark $N_2$ fixation rates unreliable for estimating $N_2$ fixation by NCDs. All stations in this study had both NCD and cyanobacterial *nifH* sequences present, thus preventing conclusions of NCD $N_2$ fixation without making cell-specific measurements.

No unattached $N_2$-fixing NCDs were found with the nanoSIMS approach. Initially, our analyses were untargeted, collecting data for all unattached and particle-attached cells within a given filter area. During the untargeted analysis, we found several $^{15}$N-labeled NCD cells on particles but no unattached NCDs. Our untargeted analyses correspond to approximately 0.07 mL of seawater per station, or approximately 2.9 x $10^3$ cells screened. Based on the absence of unattached NCDs, we can estimate that total unattached NCDs were less than $10^3$ to $10^4$ cells $L^{-1}$. These unattached NCD abundances are within the range of *nifH* based qPCR abundances, from 3.0 x $10^2$ *nifH* copies $L^{-1}$ in the Eastern Tropical South Pacific (Turk-Kubo et al., 2014) to 3.0 x $10^4$ *nifH* copies $L^{-1}$ in the North Pacific Subtropical Gyre (Moisander et al., 2014) for the potentially free-living fraction (less than 3 µm or 10 µm size fraction). On the other hand, the unattached NCD abundances are orders of magnitude

lower than primer-free metagenomic based abundances of $1.3x10^6$ cells $L^{-1}$ (< 1.6 or 3 µm) from the Tara Ocean Database in the oligotrophic North Pacific (Delmont *et al.*, 2017). The large difference in nanoSIMS versus metagenomic free-living NCD abundance estimates could imply a large proportion of the NCDs detected with metagenomics may not be actively fixing $N_2$ and would therefore not be quantified using this nanoSIMS approach. Alternatively, there may have been a lower abundance of unattached NCDs during this study than when sampled using the metagenomic method. While particle-attached NCD $N_2$ fixation was detectable, we found no evidence for $N_2$ fixation by unattached NCDs in the surface ocean.

As we did not find unattached NCDs in the 7 sample locations using the untargeted nanoSIMS approach, we focused subsequent nanoSIMS analyses on particles, mapping and analyzing ~150 particles from each station. Of the total sample area analyzed, 75% of the nanoSIMS analysis was targeted for particles, and 25% was the untargeted approach, which included both unattached cells and a few particles. The number of areas analyzed with nanoSIMS plays a role in the probability of finding a NCD cell. Yet, the particle-targeted analysis increased the NCD encounter rate from 0.5 NCD per 100 analysis areas (untargeted approach, although the NCDs were found on particles) to 2.1 NCDs per 100 analysis areas (particle directed, high volume approach). This is further evidence that NCD $N_2$ fixation was more prevalent on particles than in unattached cells.

All NCD cells measured using nanoSIMS analyses were associated with particles. A total of 69 out of 101 $N_2$-fixing cells were identified as NCD as they were

enriched in $^{15}N$ but lacked $^{13}C$ enrichment, indicating no detectable $CO_2$ fixation occurred. The 69 NCDS were found associated with 20 particles over 4 out of 7 stations (Stns. 5, 10, 20, and 22). The remaining 32 cells (from the 101 total) were cyanobacterial-like diazotrophs as they were enriched in both $^{15}N$ and $^{13}C$ (Stns. 5, 14, 22, and 23) which were found during particle analysis. NCD and cyanobacterial-like cells were both found at stations 5 and 22, while stations 10 and 20 only had NCDs, and stations 14 and 23 only had cyanobacterial-like diazotrophs. The majority of NCDs (66 cells) were clearly particle-attached (Fig. 2a), while a small fraction of measured NCDs (3 cells) were particle-adjacent. Particle-adjacent NCDs were not directly visualized attached to particles but were found in the vicinity (<4 μm) of a particle during nanoSIMS analysis. Their proximity to a particle suggests the NCD may have dislodged from the particle or could have been connected to the particle by a polymeric substance that was lost during sample preparation. Particles available for analysis ranged from 5 to 210 μm in diameter, the latter being the pore size of the incubation pre-filter. We do not know if NCDs would have been found on larger particles that provide more microaerophilic zones conducive to $N_2$ fixation. Further, the size of the particle we analyzed could only be measured up to the raster size (20 to 40 μm) of the nanoSIMS analysis area, so some of these particles were likely larger. We found NCDs on a wide range of particle sizes, from 6 μm to larger than 40 μm. Particles with one or more associated NCD accounted for 3.6% of the total particles analyzed at Stn. 5, 4.1% at Stn. 10, 1.1% at Stn. 20, and 4.0% at Stn. 22. Approximately half of the particles with an associated NCD contained a single NCD,

although this is likely an underestimate as we did not acquire data through entire particles (nanoSIMS is a surface analysis technique). To examine this further, we randomly picked four particles with an associated NCD and analyzed these particles at multiple depths, finding additional NCDs in 3 of the 4 particles. As such, NCD abundances are underestimates due to missed $N_2$-fixing cells that were within or on the underside of the analyzed particles.

We estimated the percent of new biomass synthesized from fixed N and C relative to total biomass, Nnet%, and Cnet%, respectively (Dekas et al., 2019; Popa et al., 2007), for each $N_2$-fixing cell. The cell enrichment minimum thresholds (4x the standard deviation of unlabeled cells) for both $^{15}N$ and $^{13}C$ correspond to Nnet% = 3.1% and Cnet% = 9.4% (Fig. 2b). NCD cells exhibited a broad range of Nnet% from 3.8 to 98%, while Cnet% ranged from -8.7 to 5.7%. The Nnet% of cyanobacterial-like cells ranged from 3.8 to 28%, while Cnet% had a minimum of 9.9% and a maximum of 40%. For most cells, Nnet% and Cnet% were not equivalent, indicating the cells obtained either their N or C requirements from other sources. The cyanobacterial-like cells, for example, fixed both N and C but exhibited higher Cnet% values, indicating that they likely obtained their N from other sources in addition to diazotrophy. For the NCDs, 19% exhibited Nnet% values of 50% or higher, indicating that based on N, half or more of the cellular biomass had been synthesized during the isotope incubation. At this rate, NCDs could fulfill their full cellular N requirements through $N_2$ fixation in 24-hours. The highest observed Nnet%, 98% (measured in one cell), is equivalent to 4 cell divisions in 24-hours. Such high growth rates have been observed

in NCD cultures under nutrient replete conditions, but this is unusually fast compared to the average generation time of marine planktonic and particle attached bacteria (9-18 hours) (Bentzon-Tilia, Severin, and Hansen 2015; Ploug and Grossart 1999; L. Riemann, Steward, and Azam 2000) and may therefore reflect a smaller number of asymmetric divisions such as occurs in Planctomycetes (Delmont et al. 2018; Rivas-Marín, Canosa, and Devos 2016). A Nnet% lower than 50% could indicate a growth rate slower than the incubation period (24-hours) or that the cells use other sources of N in addition to $N_2$.

In addition to new biomass, we also calculated cell-specific $N_2$ fixation rates for each particle-associated NCD to quantify net $N_2$ fixed over the duration of the $^{15}N_2$ incubation by individual cells. NCD-specific $N_2$ fixation rate interquartile ranges at each station were 0.38-0.74, 0.08-0.30, 4.07 (mean of 2 cells), 1.06-2.73 fmol N cell$^{-1}$ d$^{-1}$, west to east (Fig. 2; Table 1). Cell-specific rates ranged from 0.03 to 8.61 fmol N cell$^{-1}$ d$^{-1}$ (Fig. 2c) with an overall average of 1.09 ± 1.61 fmol N cell$^{-1}$ d$^{-1}$. The cell-specific average was higher than those measured for NCD isolates from the Baltic Sea with a maximum of 0.06 fmol N cell$^{-1}$ d$^{-1}$ in minimal culture media and up to 0.2 fmol N cell$^{-1}$ d$^{-1}$ with replete labile C and $NH_4^+$ (Bentzon-Tilia, Severin, and Hansen 2015), which could be present at high concentrations on marine particles. However, the cell-specific average was similar to a microaerobic NCD isolate from oxygenated Baltic Sea water with a cell-specific rate of 1.1 fmol N cell$^{-1}$ d$^{-1}$ (Paerl et al., 2018). The NCD cell-specific rates show particle attachment allowed NCDs to fix $N_2$ in the oxygenated surface ocean contributing to the total fixed N available, a

process that was previously only hypothesized, but not demonstrated, to occur in the oligotrophic ocean.

We used the single-cell nanoSIMS data to estimate the contribution of NCDs to the total $N_2$ fixation in our samples (Summarized in Table 1). Our volumetric estimates of NCD abundance ranged from $7.9 \times 10^2$ to $7.7 \times 10^4$ cells $L^{-1}$, which is within the range of previously reported abundances of Gamma A, a commonly occurring NCD *nifH* sequence (Langlois et al., 2015; Moisander et al., 2014) and the most relatively abundant NCD *nifH* sequences recovered from our samples. Scaling the cell-specific rate averages from each station using the NCD abundance can provide a rough estimate of NCD contribution to community $N_2$ fixation rates, noting that our calculations are based on underestimations. The more conservative NCD $N_2$ fixation totals were up to 0.011 nmol N $L^{-1}$ $d^{-1}$, accounting for up to 8.2% of the total community $N_2$ fixation rates (natural-light incubations) and 16.4% of the all-dark community $N_2$ fixation rates. The less conservative particle-volume-based calculations resulted in an NCD $N_2$ fixation total of up to 0.067 nmol N $L^{-1}$ $d^{-1}$ accounting for up to 22.1% of the total community $N_2$ fixation rates of the natural-light incubations and up to 44.2% of the all-dark community $N_2$ fixation rates. We note that if the NCD activity represented the entire community $N_2$ fixation rate, in the hypothetical absence of cyanobacteria, the NCD contribution values would fall below the MQR of the bulk measurements. Yet, if particle-attached NCD $N_2$ fixation is light-independent, these low cell-specific rates and abundances may be significant

when depth-integrated, considering the large expanse of the dark ocean (Benavides et al., 2018b; Kuhn et al., 2018; Moisander et al., 2017; Rahav et al., 2013).

The calculated contributions of NCD activity to total fixed N discussed above are based on analyses of only several thousands of cells from seven stations on one cruise and a number of assumptions, and therefore are only rough estimates of the potential importance of this underexplored phenomenon in the world's oceans. Our underestimation mainly derives from NCD and cyanobacterial abundances when scaling active $N_2$-fixing cell numbers to volumetric abundances (cells $L^{-1}$). The areas scanned during nanoSIMS analysis are missing $N_2$-fixing cells that were in the interior or underside of particles, lowering the estimated cell abundances. Evidence of this is demonstrated by particle depth profiles (Fig. S2) in which 3 out of 4 particles harbored additional NCDs within the interior volume of the particle that were not measurable from the surface of the particle. To amend our estimates, we calculated NCD cells per unit volume using an average particle volume (15 µm by 15 µm by 4 µm) which increases the estimated NCD $N_2$ fixation contribution by 1.3 to 8.3x. However, the total volumes of the particles are still unknown. Our cell-specific $N_2$ fixation rates may further be underestimated due to dilution of the $^{15}N$ (also impacting the corresponding $^{13}C$ dilution for cyanobacteria) caused by the fixation process determined by Meyer et al. (2021) *of up* to 12% dilution (not included in our calculations). The comparison of bulk $N_2$ fixation rates to community contribution from single-cell nanoSIMS analysis also supports that our calculations were underestimates since the bulk rates were greater than the total volume integrated NCD

and cyanobacterial single-cell data (Fig. S3), which was only a maximum of 24% of the bulk. Regardless of these underestimations, our empirical measurements still show that NCD activity can account for a portion of the community $N_2$ fixation rates, contributing to the fixed N in the system.

This is the first demonstration of NCDs fixing $N_2$ in the surface ocean, and all encountered NCDs were associated with particles. It has been hypothesized that particles may provide favorable microenvironments for NCD, including low oxygen (Church et al., 2005b; Klawonn et al., 2015a; Paerl and Prufert, 1987), high C concentrations, and low fixed N concentrations compared to C (Rahav and Giannetto, 2016; Riemann et al., 2010). For heterotrophic NCDs, particle attachment would provide a readily available source of organic carbon compared to the water column. However, microaerophilic environments may not be a factor for the small particles surveyed in this study, as estimates for particle size ranges needed to harbor a microanoxic zone are ≥1 mm in diameter (Klawonn et al., 2015a; Ploug, 2001), and we found active NCDs on particles as small as 6 µm diameter. If microanoxic zones are required for $N_2$ fixation, it is possible small particles with associated NCD may have initially been large enough to harbor microaerobic zones but were fractured during filtration or analysis preparation. Particle-attached $N_2$ fixation affect the microscale biogeochemistry of the particles themselves, but on a larger scale, this process can transfer fixed N from the surface ocean to depth as the particles sink.

We found surprisingly low numbers of diazotrophic cyanobacteria-like cells compared to NCD cells (32:69). It is possible the particle-focused analysis

contributed to the lower number of cyanobacterial cells found as we would not necessarily expect to find free-living cyanobacteria associated with particles. However, a study comparing *nifH* sequences from the water column to particles at 150 m found that while the water column had a higher relative abundance of cyanobacteria, the sequences recovered from particles were 44.5% cyanobacterial (*Crocosphaera,* UCYN-A *Richelia,* and *Trichodesmium)* (Farnelid et al., 2019). Cyanobacterial *nifH* sequences were also found at 4000 m (*Crocosphaera* and *Richelia)* (Poff et al., 2021), yet the cyanobacterial diazotrophic activity on particles is unknown. As such, the lower cyanobacterial abundances may be due to the possibility of typically C-fixing cyanobacteria being able to use heterotrophic C acquisition while attached to particles and would therefore be categorized as a NCD in this study. Both *Trichodesmium* and *Cyanothece* have shown evidence of mixotrophy by incorporating dissolved organic C while maintaining their $N_2$ fixing capability (Benavides et al., 2017; Feng et al., 2010). Although the absolute abundance of $N_2$-fixing organisms found with nanoSIMS was underestimated in this study, the proportion of $N_2$-fixing cyanobacterial diazotrophs to $N_2$-fixing NCDs on particles should not be affected, resulting in a surprising lack of C-fixing diazotrophs. Future studies on particle-associated diazotrophs that can incorporate cell identity as well as C and $N_2$ fixation rates would lead to valuable insights into the C-fixing potential of particle-attached cyanobacteria.

Our confirmation of $N_2$ fixation by NCDs attached to particles in surface waters of the oligotrophic ocean sheds light on how these organisms may overcome

the difficulties of heterotrophically-fueled $N_2$ fixation in this low nutrient environment. Additionally, NCDs attached to particles could sink out of the euphotic zone extending the depth range of $N_2$ fixation. The particle-associated $N_2$ fixation would have local effects on particle dynamics, allowing for C remineralization even after initial bioactive N sources were depleted (Pedersen et al., 2018). Our data showing previously unidentified particle-associated heterotrophic $N_2$ fixation in the surface ocean demonstrates the need for a more careful evaluation of microscale interactions in these processes.

**Acknowledgments**

We gratefully acknowledge Ate Visser for providing access to the MIMS and for helpful discussion, Fangjuan Huang for shipboard assistance, Colin Carney for providing EA-IMRS consultation, Rosie Gradoville for $^{15}N$ incubation and calculation discussions, and Jonathan Magasin for statistical advice. We greatly appreciate the captain and crew of the R/V Sally Ride for field logistical support and Lasse Reimann (University of Copenhagen) for helpful discussions.

*Author contributions*

K.H. and X.M. conceived and designed the study. K.H., E.W.K.M, and K.A.T collected and processed samples and performed incubations. K.H., K.A.T, X.M, and P.K.W. analyzed data. K.H. and J.P.Z. wrote the initial drafts with revision from all authors.

*Data and materials availability*

All data is provided in manuscript and supplemental materials, with the exception of the raw *nifH* sequences, which were deposited in the Sequence Read

Archive at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/sra) under Bioproject ID PRJNA730862.

*Funding*

This research was funded by the Simons Foundation (SCOPE Award ID 724220, J.Z.) and NSF grants (OCE-1559165 to J. Z.), the Lawrence Livermore Graduate Research Scholars Programs, and the DOE-BER Genomics Sciences LLNL biofuels Science Focus Area grant #SCW1039. Ship time was granted by the UC Ship Funds Program supported by UC San Diego, Scripps Institution of Oceanography. Work at LLNL was performed under the auspices of the US Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.

## References


Behrens, S., Lösekann, T., Pett-Ridge, J., Weber, P. K., Ng, W. O., Stevenson, B. S., … Spormann, A. M. (2008). Linking microbial phylogeny to metabolic activity at the single-cell level by using enhanced element labeling-catalyzed reporter deposition fluorescence in situ hybridization (EL-FISH) and NanoSIMS. *Applied and Environmental Microbiology*, *74*(10), 3143–3150. https://doi.org/10.1128/AEM.00191-08

Benavides, M., Berthelot, H., Duhamel, S., Raimbault, P., & Bonnet, S. (2017). Dissolved organic matter uptake by *Trichodesmium* in the Southwest Pacific. *Nature Publishing Group*, *7*(41315), 3–8. https://doi.org/10.1038/srep41315

Benavides, M., Bonnet, S., Berman-frank, I., & Riemann, L. (2018). Deep Into Oceanic $N_2$ Fixation. *Frontiers in Marine Science*, *5*(April), 3–6. https://doi.org/10.3389/fmars.2018.00108

Bentzon-Tilia, M., Severin, I., & Hansen, L. H. (2015). Genomics and Ecophysiology of Heterotrophic Nitrogen-Fixing Bacteria Isolated from Estuarine Surface Water. *MBio*, *6*(4), 1–11. https://doi.org/10.1128/mBio.00929-15.Editor

Bombar, D., Paerl, R. W., & Riemann, L. (2016). Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. *Trends in Microbiology*, *24*(11), 16–927. https://doi.org/10.1016/j.tim.2016.07.002

Bombar, D., Turk-Kubo, K. A., Robidart, J., Carter, B. J., & Zehr, J. P. (2013). Non-cyanobacterial *nifH* phylotypes in the North Pacific Subtropical Gyre detected by flow-cytometry cell sorting. *Environmental Microbiology Reports*, *5*(5), 705–715. https://doi.org/10.1111/1758-2229.12070

Bottjer, D., Dore, J. E., Karl, D. M., Letelier, R. M., Mahaffey, C., Wilson, S. T., … Church, M. J. (2017). Temporal variability of nitrogen fixation and particulate nitrogen export at Station ALOHA. *Limnology and Oceanography*, *62*, 200–216. https://doi.org/10.1002/lno.10386

Cabello, A. M., Turk-Kubo, K. A., Hayashi, K., Jacobs, L., Kudela, R. M., & Zehr, J. P. (2020). Unexpected presence of the nitrogen-fixing symbiotic cyanobacterium UCYN-A in Monterey Bay, California. *Journal of Phycology*, *56*(6), 1521–1533. https://doi.org/10.1111/jpy.13045

Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., & Carpenter, E. J. (1997). *Trichodesmium*, a Globally Significant Marine Cyanobacterium. *Science*, *276*(5316), 1221–1229. https://doi.org/10.1126/science.276.5316.1221

Capone, Douglas G, Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., … Carpenter, E. J. (2005). Nitrogen fixation by *Trichodesmium* spp .: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, *19*.

https://doi.org/10.1029/2004GB002331

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. https://doi.org/10.1038/nmeth0510-335

Carpenter, E., & Capone, D. G. (1992). Nitrogen Fixation in *Trichodesmium* Blooms. In *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs* (pp. 1–12).

Church, M J, Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Vertical distributions of nitrogen-fixing phylotypes at Station ALOHA in the oligotrophic North Pacific Ocean. *Aquatic Microbial Ecology*, *38*, 3–14. https://doi.org/10.3354/ame038003

Church, Matthew J, Short, C. M., Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Applied and Environmental Microbiology*, *71*(9), 5362–5370. https://doi.org/10.1128/AEM.71.9.5362

Dekas, A. E., Parada, A. E., Mayali, X., Fuhrman, J. A., Wollard, J., & Pett-ridge, J. (2019). Characterizing Chemoautotrophy and Heterotrophy in Marine Archaea and Bacteria With Single-Cell Multi-isotope NanoSIP. *Frontiers in Microbiology*, *10*(December). https://doi.org/10.3389/fmicb.2019.02682

Dekas, A. E., Poretsky, R. S., & Orphan, V. J. (2009). Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science*, *326*(October), 422–426.

Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T. M., Rappé, M. S., … Eren, A. M. (2018). Nitrogen-fixing populations of *Planctomycetes* and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, *3*(July). https://doi.org/10.1038/s41564-018-0176-9

Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science*, *281*(5374), 200–206. https://doi.org/10.1126/science.281.5374.200

Farnelid, H., Andersson, A. F., Bertilsson, S., Al-Soud, W. A., Hansen, L. H., Sørensen, S., … Riemann, L. (2011). Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0019223

Farnelid, H., Helle, K. T., Justin, P., James, E. O., Mooy, B. A. S. Van, & Zehr, J. P. (2019). Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *The ISME Journal*, 170–182. https://doi.org/10.1038/s41396-018-0259-x

Farnelid, H., Tarangkoon, W., Hansen, G., Hansen, P. J., & Riemann, L. (2010). Putative $N_2$-fixing heterotrophic bacteria associated with dinoflagellate-cyanobacteria consortia in the low-nitrogen Indian Ocean. *Aquatic Microbial Ecology*, *61*(2), 105–117. https://doi.org/10.3354/ame01440

Feng, X., Bandyopadhyay, A., Berla, B., Page, L., Wu, B., Pakrasi, H. B., & Tang, Y. J. (2010). Mixotrophic and photoheterotrophic metabolism in *Cyanothece* sp. ATCC 51142 under continuous light. *Microbial Ecology*, *April* 2566–2574. https://doi.org/10.1099/mic.0.038232-0

Geisler, E., Bogler, A., Bar-zeev, E., Rahav, E., & Farnelid, H. (2020). Heterotrophic Nitrogen Fixation at the Hyper-Eutrophic Qishon River and Estuary System. *Frontiers in Microbiology*, *11*(June), 2012–2021. https://doi.org/10.3389/fmicb.2020.01370

Gradoville, M. R., Bombar, D., Crump, B. C., Letelier, R. M., Zehr, J. P., & White, A. E. (2017). Diversity and activity of nitrogen-fixing communities across ocean basins. *Limnology and Oceanography*, *62*, 1895–1909. https://doi.org/10.1002/lno.10542

Gradoville, M. R., Farnelid, H., White, A. E., Turk-Kubo, K. A., Stewart, B., Ribalet, F., … Zehr, J. P. (2020). Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnology and Oceanography*, *65*, 1858–1875. https://doi.org/10.1002/lno.11423

Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., … Kuypers, M. M. M. (2012). Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *The ISME Journal*, *6*(6), 1238–1249. https://doi.org/10.1038/ismej.2011.182

Horner-Devine, M., & Martiny, A. C. (2008). News About Nitrogen. *Science*, *320*, 757–758. https://doi.org/10.1126/science.1147012

Howard, J. B., & Rees, D. C. (1996). Structural Basis of Biological Nitrogen FIxation. *Chemical Reviews*, *96*(7), 2965–2982. https://doi.org/10.1021/cr9500545

Karl, D. M., Michaels, A., Bergman, B., & Capone, D. (2002). Dinitrogen fixation in the world's oceans. *Biochemistry*, *57*(58), 47–98. https://doi.org/10.1023/A:1015798105851

Karlusich, J. J. P., Pelletier, E., Carsique, M., Picheral, M., Pepperkok, R., Karsenti, E., … Foster, R. A. (2020). Global distribution patterns of marine nitrogen-fixers by imaging and molecular methods.

Klawonn, I., Bonaglia, S., Bruchert, V., & Ploug, H. (2015). Aerobic and anaerobic nitrogen transformation processes in $N_2$ -fixing cyanobacterial aggregates. *The*

*ISME Journal*, *9*(15), 1456–1466. https://doi.org/10.1038/ismej.2014.232

Klawonn, I., Lavik, G., Böning, P., & Marchant, H. K. (2015). Simple approach for the preparation of $^{15-15}N_2$-enriched water for nitrogen fixation assessments : evaluation, application, and recommendations. *Frontiers in Microbiology*, *6*(August), 1–11. https://doi.org/10.3389/fmicb.2015.00769

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

Kuhn, A. M., Fennel, K., & Berman-frank, I. (2018). Modelling the biogeochemical effects of heterotrophic and autotrophic $N_2$ fixation in the Gulf of Aqaba (Israel), Red Sea. *Biogeosciences*, *15*, 7379–7401.

Langlois, R., Großkopf, T., Mills, M., Takeda, S., & LaRoche, J. (2015). Widespread Distribution and Expression of Gamma A (UMB), an Uncultured, Diazotrophic, γ-Proteobacterial nifH Phylotype. *Plos One*, *10*(6), e0128912. https://doi.org/10.1371/journal.pone.0128912

Langlois, R. J., Hümmer, D., & LaRoche, J. (2008). Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean. *Applied and Environmental Microbiology*, *74*(6), 1922–1931. https://doi.org/10.1128/AEM.01720-07

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., … Kuypers, M. M. M. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, *1*(11). https://doi.org/10.1038/nmicrobiol.2016.163

Meyer, N. R., Fortney, J. L., & Dekas, A. E. (2021). NanoSIMS sample preparation decreases isotope enrichment: magnitude, variability, and implications for single-cell rates of microbial activity. *Environmental Microbiology*, *23*(1), 81–98. https://doi.org/10.1111/1462-2920.15264

Mohr, W., Großkopf, T., Wallace, D. W. R., & Laroche, J. (2010). Methodological Underestimation of Oceanic Nitrogen Fixation Rates. *PLoS ONE*, *5*(9). https://doi.org/10.1371/journal.pone.0012583

Moisander, P. H., Beinart, R. a, Voss, M., & Zehr, J. P. (2008). Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *The ISME Journal*, *2*, 954–967. https://doi.org/10.1038/ismej.2008.84

Moisander, P. H., Benavides, M., Bonnet, S., & Berman-frank, I. (2017). Chasing after Non-cyanobacterial Nitrogen Fixation in Marine Pelagic Environments. *Frontiers in Microbiology*, *8*(September).

https://doi.org/10.3389/fmicb.2017.01736

Moisander, P. H., Serros, T., Paerl, R. W., Beinart, R. a, & Zehr, J. P. (2014). Gammaproteobacterial diazotrophs and *nifH* gene expression in surface waters of the South Pacific Ocean. *The ISME Journal*, *8*(10), 1962–1973. https://doi.org/10.1038/ismej.2014.49

Montoya, J. P., Voss, M., Kahler, P., & Capone, D. G. (1996). A Simple, High-Precision, High-Sensitivity Tracer Assay for $N_2$ Fixation. *Applied and Environmental Microbiology*, *62*(3), 986–993. https://doi.org/10.1128/AEM.62.3.986.993

Paerl, H. W., & Prufert, L. E. (1987). Oxygen-Poor Microzones as Potential Sites of Microbial $N_2$ Fixation in Nitrogen-Depleted Aerobic Marine Waters. *Applied and Environmental Microbiology*, *53*(5), 1078–1087. https://doi.org/10.1128/AEM.53.5.1087.1987

Paerl, R. W., Hansen, T. N. G., Henriksen, N. N. S. E., & Olesen, A. K. (2018). $N_2$ fixation and related $O_2$ constraints on model marine diazotroph *Pseudomonas stutzeri* BAL361. *Aquatic Microbial Ecology*, *81*, 125–136. https://doi.org/10.3354/ame01867

Pedersen, J. N., Bombar, D., Paerl, R. W., Riemann, L., & Seymour, J. R. (2018). Diazotrophs and $N_2$-Fixation Associated With Particles in Coastal Estuarine Waters. *Frontiers in Microbiology*, *9*(November), 1–11. https://doi.org/10.3389/fmicb.2018.02759

Pett-Ridge, J., & Weber, P. K. (2012). NanoSIP: NanoSIMS applications for microbial biology. *Methods in Molecular Biology*, *881*, 375–408. https://doi.org/10.1007/978-1-61779-827-6_13

Ploug, H. (2001). Small-scale oxygen fluxes and remineralization in sinking aggregates. *Limnology and Oceanography*, *46*(7), 1624–1631. https://doi.org/10.4319/lo.2001.46.7.1624

Ploug, H., & Grossart, H. P. (1999). Bacterial production and respiration in suspended aggregates - a matter of the incubation method. *Aquatic Microbial Ecology*, *20*, 21–29. https://doi.org/10.3354/ame020021

Poff, K. E., Leu, A. O., Eppley, J. M., Karl, D. M., & Delong, E. F. (2021). Microbial dynamics of elevated carbon flux in the open ocean's abyss. *PNAS*, *118*(4), 1–11. https://doi.org/10.1073/pnas.2018269118

Popa, R., Weber, P. K., Pett-ridge, J., Finzi, J. A., Fallon, S. J., Hutcheon, I. D., … Capone, D. G. (2007). Carbon and nitrogen fixation and metabolite exchange in and between individual cells of Anabaena oscillarioides. *ISME Journal*, 354–360. https://doi.org/10.1038/ismej.2007.44

Rahav, E, & Giannetto, M. J. (2016). Contribution of mono and polysaccharides to heterotrophic $N_2$ fixation at the eastern Mediterranean coastline. *Nature Publishing Group*, (December 2015), 1–11. https://doi.org/10.1038/srep27858

Rahav, Eyal, Bar-Zeev, E., Ohayon, S., Elifantz, H., Belkin, N., Herut, B., … Berman-Frank, I. (2013). Dinitrogen fixation in aphotic oxygenated marine environments. *Frontiers in Microbiology*, *4*(August), 1–11. https://doi.org/10.3389/fmicb.2013.00227

Riemann, L., Steward, G. F., & Azam, F. (2000). Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Applied and Environmental Microbiology*, *66*(5), 578–587. https://doi.org/10.1128/AEM.66.5.2282-2282.2000

Riemann, Lasse, Farnelid, H., & Steward, G. F. (2010). Nitrogenase genes in non-cyanobacterial plankton : prevalence, diversity, and regulation in marine waters. *Aquatic Microbial Ecology*, (December). https://doi.org/10.3354/ame01431

Rivas-Marín, E., Canosa, I., & Devos, D. P. (2016). Evolutionary Cell Biology of Division Mode in the Bacterial *Planctomycetes - Verrucomicrobia - Chlamydiae* Superphylum. *Frontiers in Microbiology*, *7*(December), 1–11. https://doi.org/10.3389/fmicb.2016.01964

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., … Zehr, J. P. (2012). Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. *Science (New York, NY)*, *337*(September), 1546–1550. https://doi.org/10.1126/science.1222700

Turk-Kubo, K. A., Karamchandani, M., Capone, D. G., & Zehr, J. P. (2014). The paradox of marine heterotrophic nitrogen fixation: Abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. *Environmental Microbiology*, *16*(10), 3095–3114. https://doi.org/10.1111/1462-2920.12346

Verity, P. G., Robertson, C. Y., Tronzo, C. R., Andrews, M. G., Nelson, J. R., & Sieracki, M. E. (1992). Relationships between cell volume and the carbon and nitrogen content of marine photosynthetic nanoplankton. *Limnology and Oceanography*, *37*(7), 1434–1446. https://doi.org/10.4319/lo.1992.37.7.1434

Villareal, T. A. (1994). Widespread Occurrence of the Hemiaulus-Cyanobacterial Symbiosis in the Southwest North Atlantic Ocean. *Bulletin of Marine Science*, *54*(1), 1–7.

Vrede, K., Heldal, M., Norland, S., & Bratbak, G. (2002). Elemental Composition (C, N, P) and Cell Volume of Exponentially Growing and Nutrient-Limited Bacterioplankton, *68*(6), 2965–2971. https://doi.org/10.1128/AEM.68.6.2965

Wang, W., Moore, J. K., Martiny, A. C., & François, W. (2019). Convergent

estimates of marine nitrogen fixation. *Nature*, 205–213. https://doi.org/10.1038/s41586-019-0911-2

White, A. E., Granger, J., Selden, C., Gradoville, M. R., Potts, L., Bourbonnais, A., … Chang, B. X. (2020). A critical review of the $^{15}N_2$ tracer method to measure diazotrophic production in pelagic ecosystems. *Limnology and Oceanography*, 129–147. https://doi.org/10.1002/lom3.10353

Woebken, D., Burow, L. C., Prufert-bebout, L., Bebout, B. M., Hoehler, T. M., Pett-ridge, J., … Singer, S. W. (2012). Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *The ISME Journal*, *6*, 1427–1439. https://doi.org/10.1038/ismej.2011.200

Zehr, J. P., & Bombar, D. (2015). Cyanobacteria and Archaea Marine Nitrogen Fixation : Organisms, Significance, Enigmas, and Future Directions. In *Biological Nitrogen Fixation* (Vol. 2, pp. 857–871).

Zehr, J. P., & Mcreynolds, L. A. (1989). Use of Degenerate Oligonucleotides for Amplification of the *nifH* Gene from the Marine Cyanobacterium *Trichodesmium thiebautii*. *Applied and Environmental Microbiology*, 2522–2526. https://doi.org/10.1128/AEM.55.10.2522-2526.1989

Zehr, J. P., Mellon, M. T., Zani, S., & York, N. (1998). New Nitrogen-Fixing Microorganisms Detected in Oligotrophic Oceans by Amplification of Nitrogenase (*nifH*) Genes. *Applied and Environmental Microbiology*, *64*(9), 3444–3450. https://doi.org/10.1128/AEM.64.9.3444-3450.1998

Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., … Karl, D. M. (2001). Unicellular cyanobacteria fix $N_2$ in the subtropical North Pacific Ocean. *Nature*, *412*(6847), 635–638. https://doi.org/10.1038/35088063

## Figures

**Figure 1**: $N_2$ fixation rates and community composition by station. Map inset shows stations sampled across the North Pacific Subtropical Gyre. The bar chart shows community $N_2$ fixation rates for natural-light (yellow) and all-dark (dark grey) incubations. Error bars are the standard deviations of the averages of triplicate incubations. $N_2$ fixation rates marked with asterisks fall below the LOD, but above the MQR, all-dark values at S17 and S20 were below LOD and MQR (red), no data were available for S23 all-dark incubation. The lower pie charts show the relative proportion of cyanobacterial (green) and NCD (light grey) *nifH* sequences at each station.

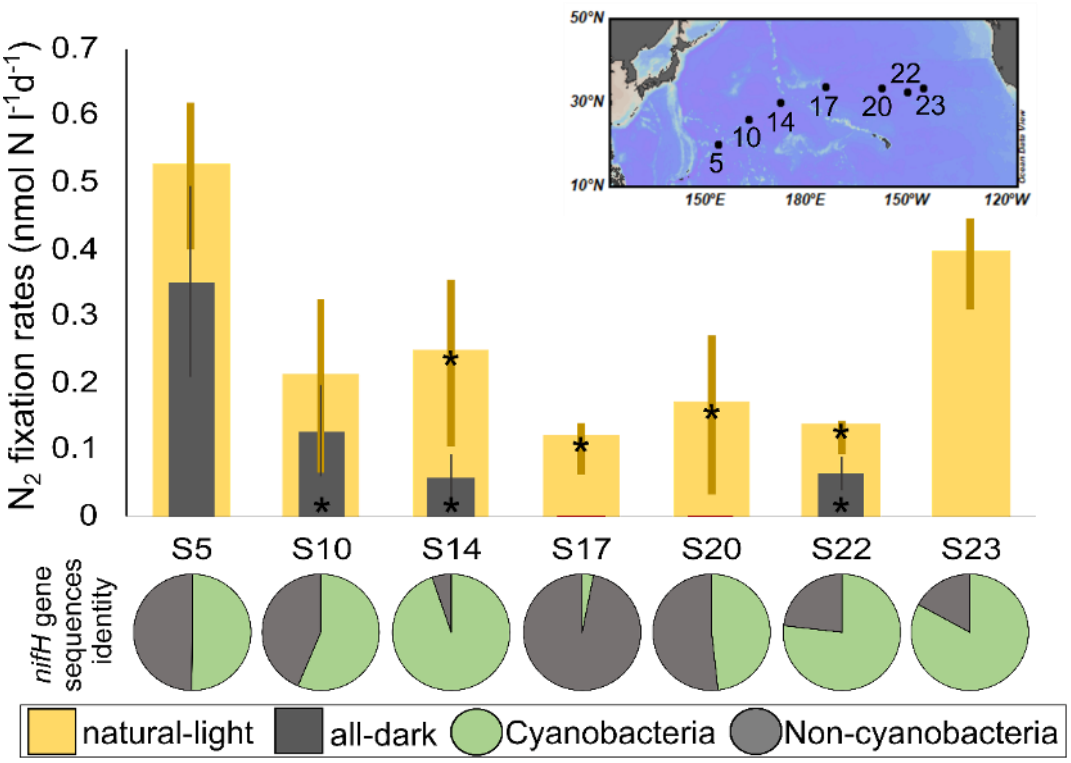


**Figure 2**: NanoSIMS data. A) Example images of 2 particles with associated NCDs (red arrows). Images show particles with attached cells enriched in 15N (A1 and A3) but not enriched in 13C (A2 and A4). Color scale bars correspond to At% enrichment for the given isotope. Green arrows show cells enriched in 13C but not 15N, similar to what we would expect for non-diazotrophic autotrophs. Scale bar is 5 µm. B) Cnet% versus Nnet% of cells enriched in 15N, lines show Cnet% and Nnet% minimum enrichment thresholds. The minimum enrichment thresholds section the data points into quadrants. Data points in the upper left quad are cells enriched in 15N and counted as NCD. Points in the upper right quadrant are cells enriched in 13C and 15N and classified as cyanobacterial-like cells. The lower two quadrants constitute heterotrophic cells (bottom left) and C-fixing (autotrophic) cells (bottom left) which make up >99% of the data analyzed (data not plotted). C) Cell-specific $N_2$ fixation rates, by station. Data points represent individual cells, and the horizontal line shows the mean from each station.

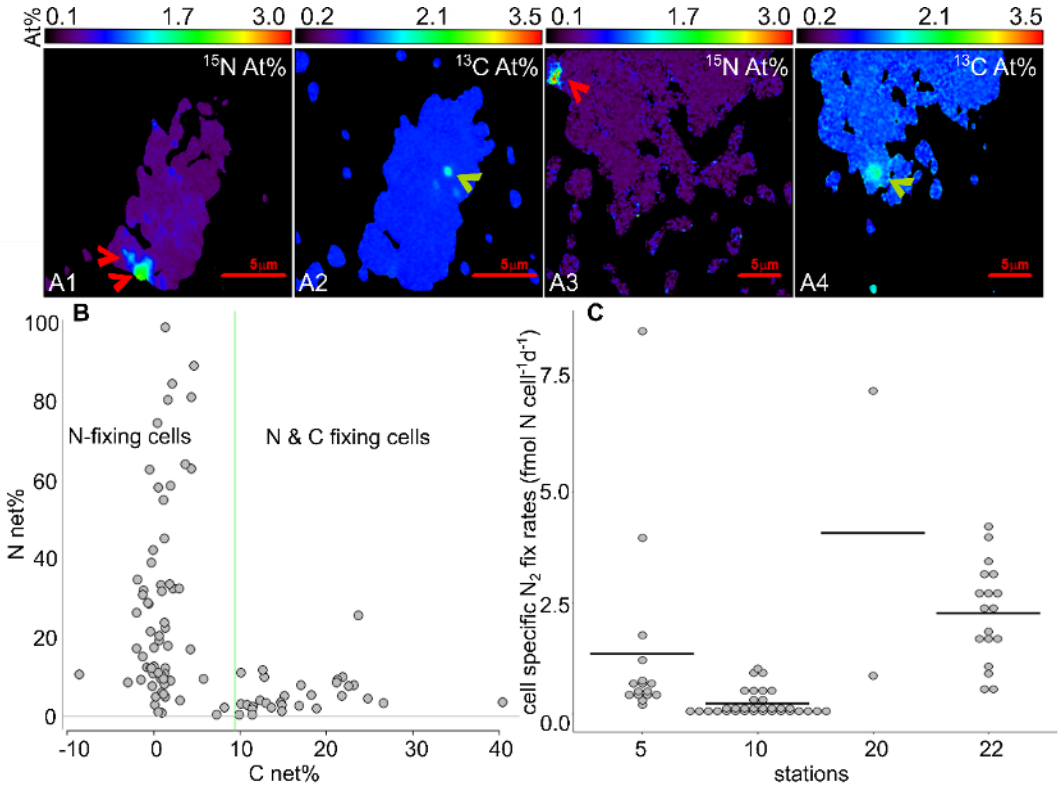


**Table 1**: Summary of community $N_2$ fixation rates and nanoSIMS calculations by station. Calculations represent values scaled-up from a limited analysis and should be considered as rough estimates included for context. Estimates were calculated 2 ways, one using the percent of analyzed particles with an associated NCD ('NCD abundance', 'NCD $N_2$ fixation rate contribution', '% NCD $N_2$ fix to community light/dark $N_2$ fix rates') and the other by using NCD per unit volume particle assuming an average particle volume (grey columns; 'particle volume NCD abundance', 'particle volume NCD $N_2$ fixation rate contribution', 'particle volume % NCD $N_2$ fix to community light/dark $N_2$ fix rates'). Values represent mean (± standard deviation) of the mean at each station, $N_2$ fixation rates marked with an asterisk are above the MQR but below the LOD.



| station | natural-light community $N_2$ fixation rate (nmol N $l^{-1}$ $d^{-1}$) | all-dark community $N_2$ fixation rate (nmol N $l^{-1}$ $d^{-1}$) | particles $l^{-1}$ x $10^4$ (5-150 µm) | total particles analyzed by nanoSIMS | % particles found with NCD | # NCD cells per station | NCD $N_2$ fixation rates (fmol N cell$^{-1}$ $d^{-1}$) | NCD abundance (cell $L^{-1}$) | particle volume estimated NCD abundance (cell $L^{-1}$) | NCD $N_2$ fixation rate contribution (nmol N $L^{-1}$ $d^{-1}$) | particle volume NCD $N_2$ fixation rate contribution (nmol N $L^{-1}$ $d^{-1}$) | % NCD $N_2$ fix to community light/dark $N_2$ fix rates | particle volume % NCD $N_2$ fix estimate to community light/dark $N_2$ fix rates |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.51 ± 0.11 | 0.35 ± 0.14 | 7.57 (± 1.07) | 196 | 4.08 | 16 | 1.35 (± 1.87) | 6.18E+03 | 4.97E+04 | 0.008 | 0.067 | 1.6 / 2.4 | 13.2 / 19.2 |
| 10 | 0.19 ± 0.13 | 0.12 ± 0.06* | 20.4 (± 8.98) | 146 | 4.11 | 34 | 0.23 (± 0.05) | 4.89E+04 | 7.66E+04 | 0.011 | 0.018 | 6.7 / 10.6 | 9.3 / 14.7 |
| 14 | 0.23 ± 0.12* | 0.06 ± 0.03* | 7.79 (± 4.30) | 123 | - | 0 | - | - | - | - | - | - | - |
| 17 | 0.10 ± 0.03* | < MQR | 3.21 (± 0.84) | 217 | - | 0 | - | - | - | - | - | - | - |
| 20 | 0.15 ± 0.11* | < MQR | 7.22 (± 1.22) | 182 | 0.55 | 2 | 4.07 (± 4.54) | 7.93E+02 | 3.19E+03 | 0.003 | 0.013 | 2.2 / na | 8.7 / na |
| 22 | 0.12 ± 0.02* | 0.06 ± 0.02* | 3.28 (± 0.31) | 126 | 3.97 | 17 | 2.22 (± 1.11) | 3.91E+03 | 1.19E+04 | 0.009 | 0.027 | 8.2/ 16.4 | 22.1 / 44.2 |
| 23 | 0.38 ± 0.07 | NA | 4.63 (± 1.63) | 115 | - | 0 | - | - | - | - | - | - | - |

**Chapter 4**

# Visualization of two uncultivated marine diazotrophs, the non-cyanobacterium 'Gamma-A' and the symbiotic cyanobacterium UCYN-A, using fluorescent in situ hybridization of the *nifH* gene (geneFISH).

**Abstract**

Marine nitrogen ($N_2$)-fixing microorganisms are an important source of N in marine ecosystems. Many marine $N_2$-fixing organisms (diazotrophs) are uncultivated, and little is known about their cell morphology, physiology, or lifestyle. Catalyzed reporter deposition-fluorescence *in-situ* hybridization (CARD-FISH) can be used to visualize and identify marine bacteria with a known 16S rRNA gene sequence. However, the 16S rRNA gene sequence is unknown for many uncultivated marine bacteria, including many non-cyanobacterial diazotrophs (NCD), which have only been identified by the sequence of a nitrogenase encoding gene, *nifH*. One such sequence type is Gamma A, a widespread, commonly occurring NCD that currently lacks a known 16S rRNA gene sequence and therefore cannot be visualized with traditional CARD-FISH approaches. Relatively little is known about Gamma A aside from the information gathered based on *nifH* gene presence (PCR) and transcription (reverse transcriptase PCR). To learn more about the lifestyle and morphology of Gamma A, we visually identified Gamma A cells by targeting the *nifH* gene for *in situ* hybridization. In this study, we employed geneFISH to link cell identity with the presence of the *nifH* gene in individual cells. Successful identification of Gamma A revealed rod-shaped cells (~1.5 x 0.6 μm), the majority of which were associated with particles. To validate the approach, we also targeted the symbiotic unicellular, cyanobacterial $N_2$-fixing group A (UCYN-A), which has been previously visualized by CARD-FISH. GeneFISH allowed us to clearly visualize the UCYN-A2 sublineage, which also showed for the first time that these cyanobacterial cells

contain multiple copies of the *nifH* gene. GeneFISH promises to be a useful tool for studying uncultivated $N_2$-fixing bacteria in the environment.

## Introduction

Biological nitrogen fixation (BNF) is an important source of bioavailable nitrogen (N) to ecosystems both on land and in the oceans (Galloway et al., 2004, 2008; Luo et al., 2012; Sohm et al., 2011). BNF is performed by diverse groups of Bacteria and Archaea known as diazotrophs. Diazotrophs reduce atmospheric dinitrogen ($N_2$) to ammonia, a bioavailable form of N, and this input of fixed N is an important source for ecosystem primary productivity in terrestrial and aquatic environments (Capone et al., 2005; Galloway et al., 2004, 2008; Karl et al., 2008; Zehr and Kudela, 2011). Although some diazotrophs have been cultivated, many are yet uncultivated, so nothing is known about their morphology, physiology, or lifestyle (such as symbiosis, attachment to particles, or free-living cells). The application of molecular techniques such as amplifying the *nifH* gene, a functional gene marker that encodes a subunit of the Fe protein of the nitrogenase enzyme, substantially expanded the known diversity of diazotrophs (Zehr et al., 1998). In the pelagic marine environment, $N_2$ fixation was historically ascribed to the filamentous nonheterocyst-forming cyanobacterium *Trichodesmium* and heterocyst-forming symbionts of diatoms (*Richelia*). The application of *nifH* gene amplification and sequencing methods over the last thirty years has shown a great diversity of other diazotrophs present in the oceans, including unicellular cyanobacterial and non-cyanobacterial diazotrophs (Zehr et al., 1998). Non-cyanobacterial diazotrophs have the genetic potential to fix $N_2$ based on the presence of the *nifH* gene and are putative heterotrophs or photoheterotrophs, meaning they do not fix carbon.

Cyanobacterial diazotrophs, including *Trichodesmium* and the more recently discovered UCYN-A symbiosis, are well documented to be significant contributors to BNF in the surface ocean (Church et al., 2009; Goebel et al., 2010; Martínez-Pérez et al., 2016; Montoya et al., 2004; Turk et al., 2011), while the importance of non-cyanobacterial diazotrophs (NCD) remains in question (Bombar et al., 2016; Moisander et al., 2017; Riemann et al., 2010). Diazotroph diversity determined by the *nifH* sequences shows NCDs are the most commonly recovered *nifH* sequences globally (Delmont et al., 2018; Farnelid et al., 2011), implying that these organisms could be ecologically meaningful. However, most commonly occurring marine NCDs are only known by their amplified *nifH* gene sequence, as only a handful of cultured representatives exist from coastal systems (Bentzon-Tilia et al., 2015; Farnelid et al., 2014; Martinez-Perez et al., 2017; Paerl et al., 2018). Little is known about marine NCD physiology or their significance in BNF, and what is known is primarily based on studies of *nifH* gene sequences, which show that NCDs are active (mRNA) and abundant (DNA) in the open ocean (Bird et al., 2005; Church et al., 2005b; Delmont et al., 2017; Farnelid et al., 2011; Halm et al., 2012; Riemann et al., 2010). While the few NCD cultured representatives show BNF activity in laboratory-based suboxic conditions (Bentzon-Tilia et al., 2015; Farnelid et al., 2014; Martinez-Perez et al., 2017; Paerl et al., 2018), only one Alphaproteobacterium has been measured directly for BNF activity *in situ* but did not show evidence of $N_2$ fixation (Martinez-Perez et al., 2017). Determining the NCD $N_2$ fixation contribution to total community $N_2$ fixation rates is difficult since $N_2$ fixation rates are generally measured in bulk water

samples containing mixed assemblages of species, and it is unknown if the NCDs fix $N_2$ or if the $N_2$ fixed is quantitatively significant relative to fixed N input by cyanobacterial diazotrophs.

The NCD sequence type γ-2474A11 also called Gamma A (Moisander et al., 2008), is one of the most well-studied NCDs due to its widespread distribution and stable abundances based on the *nifH* gene (Langlois et al., 2015; Moisander et al., 2014). The widespread abundance of Gamma A *nifH* genes and transcripts indicate that they may be important contributors to open ocean $N_2$ fixation. Gamma A is a putative Gammaproteobacterial diazotroph identified by the *nifH* gene sequence and is found throughout the oligotrophic tropical and subtropical ocean (Langlois et al., 2015; Zehr et al., 1998). The mean abundances ($10^2$ *nifH* copies $L^{-1}$ ) of Gamma A are comparable to those of the globally significant cyanobacterial diazotroph UCYN-A (Moisander et al., 2014; Shiozaki et al., 2017) over large areas. Gamma A maximum abundances rarely exceed $10^4$ *nifH* gene copies $L^{-1,}$ making them rare compared to other cyanobacterial diazotrophs, but Gamma A sequences are regularly recovered in low latitude sampling efforts (Langlois et al., 2015; Shiozaki et al., 2017). Gamma A abundances are positively correlated with oxygen ($O_2$), light, and warm, low-nutrient waters (Langlois et al., 2015; Moisander et al., 2014).

It has been hypothesized that Gamma A may be associated with particles (Benavides et al., 2016, Langlois et al., 2015). Gamma A *nifH* sequences are more abundant in the larger, 10 to 3 µm size fraction compared to the 3-0.2 µm size fraction where free-living bacteria are generally collected (Benavides et al., 2016;

Gradoville et al., 2020; Langlois et al., 2015). Based on metagenomic evidence, Gamma A is also found in size fractions up to 2000 µm (Cornejo-Castillo and Zehr, 2021). Additionally, Gamma A sequences have been found on individual and bulk particles (Farnelid et al., 2019) at 150 m depth. Depending on the origin, history, and size of the particles, they may provide ideal microenvironments for heterotrophically fueled $N_2$ fixation because of the availability of fixed carbon (C) and internal microaerobic microzones. Microaerobic microzones caused by microbial respiration within the particle would help protect the $O_2$ sensitive nitrogenase enzyme from irreversible inactivation (Paerl and Prufert, 1987; Ploug, 2001).

One of the major limitations in characterizing the uncultivated NCD Gamma A is the direct visualization and identification of cells. Visualization of uncultivated Gamma A cells could help to determine if they are free-living, symbiotic, or attached to particles, but currently, there is no way to visualize Gamma A since it is not known if they have unique pigments and the16S rRNA gene sequence of this group has not yet been obtained, which would otherwise facilitate visualization by rRNA based fluorescent *in situ* hybridization (FISH) approaches. To address this problem, we explored the use of the *in situ* hybridization approach known as geneFISH (Moraru et al., 2010). Gene targeted FISH methods have been successfully used to identify cells based on functional genes of interest, including genes involved in ethene oxidation (Richards and Mattes, 2021), antibiotic resistance (Gallego et al., 2020), biodegradation (Matturro and Rossetti, 2015), denitrification (Masuda et al., 2020) and nitrification (Moraru et al., 2010). GeneFISH allows the presence of a functional

gene (e.g., *nifH gene*) and the class level identity of a cell (e.g., Gammaproteobacterial 16S rRNA) to be linked in environmental samples (Moraru et al., 2010). We designed *nifH* geneFISH probes for the globally significant cyanobacterial diazotroph UCYN-A to validate the approach. We then developed a Gamma A *nifH* gene probe for Gamma A to visualize Gamma A cells in seawater samples from the North Pacific Subtropical Gyre (NPSG). The results of the successful hybridizations of UCYN-A and Gamma A provide new information on the lifestyle of these uncultivated diazotrophs and show the promise of geneFISH for marine $N_2$ fixation studies.

**Methods**

*DNA samples*

DNA samples for clone library preparations of UCYN-A were collected during a cruise on the R/V Ka'imikai-O-Kanaloa (KOK), 19 April 2016 to 3 May 2016 at 26.3$^{o}$N, 158$^{o}$W. Water samples (2 L) were collected with Niskin® bottles attached to a rosette with an attached conductivity, temperature, depth (CTD) sensor. A peristaltic pump then filtered samples through 25 mm, 0.2 μm pore-size Supor membranes (Pall Corporation, New York, USA); samples were flash-frozen in liquid nitrogen and stored at -80$^{o}$C until processing. DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) with modifications for increased cell lysis (Moisander et al., 2008). The *nifH* gene was amplified by PCR using degenerate universal *nifH* primers *nifH*3/*nifH4* and *nifH*1/*nifH*2 in a nested reaction (Zehr and Mcreynolds, 1989). Amplification was performed on a Bio-Rad MyCycler

thermocycler. The initial round of PCR reactions included: 1× Platinum Taq DNA polymerase PCR buffer (Thermo Fisher, Waltham, MA, USA), 250 µM dNTPs mixture, 4 mM MgCl2, 2 units Platinum ® Taq DNA polymerase (Thermo Fisher), 0.5 µM of each primer (*nifH*3 and *nifH*4), 2 µL of environmental DNA sample as a template and up to 25 µl 5 kD-filtered nuclease-free water (Amicon® Ultra-15 Centrifugal Filter, Millipore Sigma, Burlington, MA, USA). The thermocycler protocol was set to; initial denaturation of 95ºC for 3 minutes followed by 25 cycles of denaturation at 95ºC for 30 seconds, annealing at 55ºC for 30 seconds, elongation at 72ºC for 45 seconds, with a final single elongation at $72^0$C for 7 minutes. The second round of the nested PCR followed the same reagent recipe as round 1 except 2 µl of the round 1 PCR product was used as the template, and primers *nifH*1/*nifH*2 were used. The thermocycling conditions were as described, except the annealing temperature was 55ºC. Duplicate negative controls were run in parallel. PCR amplification products were verified by running them in 1.4% agarose gel and excising the band for gel extraction using QIAquick Gel Extraction Kit (Qiagen) following kit protocol including steps for increased yield. Purified fragments were cloned using a pGEM®-T Easy Vector system (Promega, Madison, WI, USA), following the manufacturer's instructions. Clones were grown in 96 well plates followed by plasmid purification using Quick Plasmid Miniprep Kit (Invitrogen-Life Technologies, Carlsbad, CA, USA). Sanger sequencing of partial *nifH* amplicons was carried out at the UC Berkeley DNA Sequencing Center (UC Berkeley, CA, USA).

DNA samples for clone library preparations of Gamma A were collected from environmental water samples (2 L) on cruise KOK1507 in the North Pacific Subtropical Gyre in August 2015 at 24 32.561ºN, 156 17.488ºW from 25m depth. DNA was extracted as described, and a fragment of the *nifH* gene was amplified with PCR using a y-24774A11-specific primer (Moisander et al., 2014). Amplification was done according to Invitrogen Platinum Taq PCR protocol in short (1x Platinum Taq DNA polymerase PCR Buffer (minus Mg), 0.2 mM dNTP mixture, 1.5 mM MgCl2, 0.2 µM of each primer, 1 unit Platinum® Taq DNA Polymerase, 1 µl DNA template, and up to 50 µl 5 kD-filtered nuclease-free water (Amicon®). The thermocycler protocol is described previously in Moisander et al. (2014) in summary; initial denaturation of 95ºC for 2 minutes followed by 30 cycles of denaturation at 94ºC for 30 seconds, annealing at $56^0$C for 30 seconds, elongation at 72ºC for 30 seconds, with a final single elongation at $72^0$C for 7 minutes. PCR products were purified, cloned, and sequenced as described above.

*Environmental geneFISH samples*

Environmental water samples for geneFISH application were collected from various cruises in the North Pacific Ocean. Samples collected in the California Coastal Current (31.5º N, 120.3ºW) on cruise SP1727 aboard the R/V Gordon Sproul in October 2017 were used to optimize the geneFISH protocol for Gamma A. Gamma A was further observed in samples from the western North Pacific Subtropical Gyre (NPSG: 25.7ºN, 162.7ºE) on cruise SR1917 aboard the R/V Sally Ride in November 2019. Environmental samples for UCYN-A geneFISH visualization were collected

from the Bering Strait (64.9$^{o}$N, 167.1$^{o}$W) aboard the R/V Sikuliaq in August 2017 and California Coastal Current (31.5$^{o}$ N, 120.3$^{o}$W). Environmental water samples were collected with Niskin bottles attached to a CTD rosette from 3 to 15 m depth, 95 ml of water was fixed (1.8% formaldehyde) at 4$^{o}$C for up to 72 hours and filtered onto 0.6um polycarbonate filters (Millipore Sigma) under gentle vacuum filtration. Samples were flash-frozen in liquid $N_2$ and stored at -80$^{o}$C until use.

*Quantitative-PCR*

UCYN-A and Gamma A volumetric abundances (*nifH* gene copies $L^{-1}$) were calculated using quantitative PCR (qPCR) with TaqMan chemistry and specific primer and probe sets. UCYN-A2, a coastal sublineage (2.5 µm diameter symbiont; 7 µm diameter symbiosis including host cell) of the UCYN-A group (Cornejo-Castillo et al., 2016; Turk-Kubo et al., 2017), was quantified with UCYN-A2 specific primers and probes (Thompson et al., 2014). Gamma A abundances were evaluated using Gamma A specific qPCR primers and probes (Moisander et al., 2010). An Applied Biosystems 7500 Real-Time PCR system thermocycler was used for amplification. Reaction mixtures for both UCYN-A2 and Gamma A qPCR assays follow: 12.5 µL 2x TaqMan Mastermix (Applied Biosystems, Carlsbad, CA, USA), 1 µL each primer (forward and reverse, 10 µM), 0.5 µL of the probe (10 µm) and 2 µL of the template DNA, and 5 kD-filtered nuclease-free water (Amicon® ) with a final volume of 25 µL. Samples were run in duplicate, and each run (96 well plates) contained replicate no template controls, inhibition controls, which included the $10^5$ standard added to the sample template, and replicate standards ($10^0$ to $10^7$ *nifH* gene copies per reaction) of

the target gene sequence. The limit of detection (LOD) for the qPCR assay based on the volume of water sampled and reactions containing 2 µL template is 2 x $10^1$ *nifH* copies $L^{-1}$ while the limit of quantification (LOQ) is 1 x $10^2$ *nifH* copies $L^{-1}$ (Goebel et al., 2010).

*CARD-FISH*

A double hybridization CARD-FISH was used to quantify UCYN-A2 in environmental samples. The UCYN-A CARD-FISH protocol has been previously described in detail (Cabello et al., 2015; Cornejo-Castillo et al., 2016). In summary, *B. bigelowii* host cells and the UCYN-A2 symbiont were identified using 5′-horseradish peroxidase (HRP)- labeled oligonucleotide probes (Biomers.net, Inc., Ulm, Germany) specific to each cell followed by Tyramide signal amplification (TSA). Unlabeled helper and competitor oligonucleotides were added to facilitate hybridization and discriminate against non-target similar sequence types, respectively. Hybridization and the TSA reaction for the host cells were first followed by hybridization and TSA reaction for UCYN-A. The *B. bigelowii* host rRNA was targeted first with an HRP-labeled probe, URADO-69, with associated helpers (Helper A PRYM, Helper B PRYM) and competitors (UBRADO-69 competitor) using a 40% FA hybridization buffer. UCYN-A was targeted using the HRP-labeled probe, UCYN-A2 732, in combination with associated helpers (Helper A-732, Helper B-732) and competitor against UCYN-A1 an unlabeled oligonucleotide, UCYN-A2 competitor, using a 50% FA hybridization buffer. Briefly, cells were embedded on filters with 0.1% Ultrapure agarose (Life Technologies, Carlsbad, CA), then

permeabilized in a lysozyme (Millipore Sigma) solution. Hybridization temperatures for the HRP-labeled probes were 46°C for the host (overnight) and 35°C for the symbiont (3 hours) in the respective hybridization buffers. Samples were washed to remove un-hybridized probes before TSA. The tyramides used for the host signal amplification were Alexa 488 fluorophore (Biomers.net), while UCYN-A2 was labeled with the Cy3 fluorophore (Biomers.net). After the initial host signal amplification, introduced probe HRPs were deactivated with 0.01M HCL preceding UCYN-A probe hybridization. After UCYN-A2 hybridization and amplification, filters were air-dried and stained with DAPI (4′, 6- diamidino-2-phenylindole, 1 µg $mL^{-1}$) for 10 minutes at room temperature, followed by 3x washing in MilliQ water and air drying. Once dry, filter pieces were mounted on a glass slide with ProLong™ Diamond Antifade Mountant (Molecular Probes, Eugene, OR, USA) and stored at -20$^{o}$C until viewed with epifluorescence microscopy.

*Gamma A geneFISH probe design*

The Gamma A geneFISH probe is designed based on a collection of Gamma A *nifH* sequences that represent the microdiversity within the sequence type. Gamma A partial *nifH* gene sequences were trimmed to remove vector and primer sequences and quality controlled using Sequencher (Gene Codes Co., Ann Arbor, MI, USA). Sequences were further quality checked and aligned in ARB (Ludwig et al., 2004), nucleotide sequences were translated into amino acid sequences, and any sequences with misplaced stop codons or unable to fit alignment were removed. Gamma A *nifH* sequences from the constructed clone library (81 clones) plus additional reference

sequences (42) cover the spectrum of diversity within the Gamma A operational taxonomic unit (OUT), defined as 97% sequence similarity. Sequences were used as a database to design Gamma A specific geneFISH probes using Polypro (Moraru et al., 2011). All Gamma A *nifH* fragment sequences available in *E. coli* clones were marked as potential probe templates, while environmental sequences and clones were marked as targets. Maximum percent mismatch of 5 % (%MM) and melting temperature (delta Tm) of 0.05$^{o}$C. The best matching available probe sequences is GAGCAAACAATGTAGATTTCCTGAGCCTTATTC TCACGAATAGGCATTGC AAATCCGCCACAAACCACATCACCTAGAACATCATAAAAAACGAAATCT AAGTCGTCTGAATATGCACCTTCCTCTTCAAGGAAGTTAATCGCGGTAAT AACACCACGACCAGCACAGCCAACTCCAGGCTCAGGACCACCTGACTCCA CGCACTTAATGTCGCCAAAGCCAACCTTAAGCACTTCTTCAAGCTCAAGA TCCTCTACCGTGCCAGCCTGTGCAGCCATCTCCATTATTGTGTTTTGCGCTT TTGAGTGCAGAATAAGACG.

*GeneFISH probe synthesis*

Synthesis of the double-stranded DNA probe used in geneFISH is dependent on the chosen sequence template and PCR. The UCYN-A *nifH* phylotype is highly conserved even among different sublineages of UCYN-A (95% nucleotide similarity between UCYN-A1 versus UCYN-A2). Based on the UCYN-A *nifH* sequences available in our clone library, we chose the highest quality sequence that matched the UCYN-A specific PCR primer set. The probe sequence selected was UCYN-A1, an open ocean sublineage that is 1 µm in diameter but is associated with a 3 µm diameter

haptophyte host (Thompson et al., 2014). Double-stranded UCYN-A DNA probes were synthesized according to Moraru et al. (2010) from *E. coli* clones containing the chosen template *nifH* gene sequences by incorporating digoxigenin (DIG)-11deoxyuridinetriphosphate (dUTP) by PCR. Replicate reactions were run using the Dig Probe Synthesis Kit (Cat. No. 11636090910, Roche, Basel, Switzerland) and UCYN-A specific primers (univ_UCYN-A_F/ univ_UCYN-A_R, Turk-Kubo et al., 2017) at an annealing temperature of 55ºC, following the manufacturer instructions. The reaction mixtures contained 2.5 µl 10x PCR Buffer with $MgCl_2$, 2.5 µl PCR DIG Probe Synthesis Mix, 2. 5 µl of each UCYN-A specific forward and reverse primers (10 µM), 0.375 µl Taq Enzyme mix, 1 µl DNA template, and up to 25 µl with 5 kD-filtered nuclease-free water (Amicon®). Thermocycler conditions started with an initial denaturation of 95ºC for 2 minutes followed by 30 cycles of denaturation 95ºC for 30 seconds, annealing 55ºC for 30 seconds, and elongation at 72ºC for 40 seconds, and a final elongation of 7 minutes at 72ºC. PCR reactions without the incorporation of Dig were run in parallel as well as negative template controls. PCR products were verified by agarose gel electrophoresis (1.4% agarose); Dig incorporated PCR products are larger than the Dig free counterparts, visibly evident as the PCR products migrate through the agarose gel matrix. Dig incorporated PCR products were purified with column purification (genElute PCR purification Kit, Sigma-Aldrich, St. Louis, MO, USA), the resulting DNA concentrations were measured spectrophotometrically using a NanoDrop instrument (Thermo Fischer),

and probes were stored at -20°C in water. Probes are stable for over 1 year and were allowed three freeze-thaw cycles.

Gamma A double-stranded DNA geneFISH probes were synthesized using the chosen Gamma A *nifH* containing plasmid. Gamma A probes were synthesized, purified, and stored following the same protocol described above for UYCN-A, substituting Gamma A-specific PCR primers (γ-24774A11-specific, Moisander et al., 2014) and an annealing temperature of 56°C.

*E. coli clone controls*

*Escherichia coli (E. coli)* clones containing Gamma A *nifH* plasmids were used as controls to test the Gamma A probe functionality. Several *E. coli* clones containing different Gamma A *nifH* sequences were grown in LB Buffer (Thermo Fisher) + Carbenicillin (100 μg ml $^{-1}$). The clones were not induced to produce multiple plasmids, and geneFISH samples were taken when clone growth reached the late exponential phase before entering the stationary phase. A volume of 0.2 ml of *E. coli* culture was fixed with formaldehyde (1.8%), vacuum filtered onto 0.6 pore size polycarbonate filters (GTTP, Millipore Sigma), and stored at -20°C until use. Gamma A *nifH E. coli* clones used to check for specificity included the clone containing the plasmid of the probe template (100% identical to the probe), a Gamma A clone containing a plasmid 98.7% identical to the probe, and the most dissimilar clone available to us having a 97.9% identity over the length of the probe. Clones were additionally used to test optimum formamide concentration (10, 15, 20, 35, 40, 50% FA). Based on the number of cells with positive signal and higher signal to noise

ratio, 50% FA was chosen for further hybridization buffer concentrations. *E. coli* clones containing UCYN-A *nifH* plasmids and *E. coli* without *nifH* plasmids were similarly prepared for use as negative controls for Gamma A.

*GeneFISH protocol summary*

The geneFISH protocol used in this study is described previously in Moraru et al. (2010), but an overview of the method is summarized here. Sample filters are cut into 1/6 sections, followed by cell permeabilization, then inactivation of endogenous peroxidases with HCl before gene hybridization. Hybridization consists of a pre-hybridization step, followed by denaturation of double-stranded probes at 85$^{o}$C, then a final overnight hybridization at 46$^{o}$C. After gene hybridization, excess probes are washed away before antibody binding in which HRP conjugated anti-DIG antibodies bind to the DIG-dUTPs synthesized into the *nifH* gene probe. Lastly, tyramide signals are amplified by CARD using the HRPs attached to the target *nifH* gene. The introduced gene-bound HRPs are then inactivated (HCl), and a second typical CARD-FISH protocol follows, targeting the rRNA of a broader group of target cells. This double hybridization procedure allows a gene of interest (*nifH*) to be co-localized within a target group (Gammaproteobacterial) identified through the 16S rRNA gene.

*Modifications of the geneFISH protocol*

Several adjustments were made to optimize the protocol for our study system. All buffers and procedures follow Moraru (2010) unless described below. Before hybridization, the cells underwent more permeabilization with 10 mg ml$^{-1}$ lysozyme for 1 hour at 37$^{o}$C followed by 60 U ml$^{-1}$ achromopeptidase for 30 min at 37$^{o}$C. Initial

inactivation of the endogenous peroxidases was done with 0.05 M HCL for 10 min at room temperature, and gene hybridization and amplification were done before the rRNA CARD-FISH protocol.

To prepare the cells (pre-hybridization) for gene hybridization, filter sections were incubated in hybridization buffer (50% Formamide (FA)) without probes for 1 to 3 hours at 46$^{o}$C. Filter pieces were transferred to the hybridization solution (50% FA hybridization buffer + gene probe) containing 5 pg $\mu l^{-1}$ of the polynucleotide gene probe and incubated at 85$^{o}$C for 50 min to denature the double-stranded polynucleotide probes before incubation at 46$^{o}$C for 18 to 20 hours. Hybridization washing buffers (I and II) are made as described in Moraru et al. (2010). Washing was performed with Washing buffer I (2x 10 min) followed by washing buffer II (2x 1 min), followed by washing buffer II for 1.5 hours with gentle shaking, with all washing steps done at 46$^{0}$C.

The amplification of tyramides corresponding to the gene hybridization was adjusted for increased signal allowing for better visualization of rare cells. Samples were incubated in geneFISH specific amplification buffer (Moraru et al., 2010) containing 0.0015% $H_2O_2$ and 10 µl $ml^{-1}$ $Alexa_{594}$ (Invitrogen) at 37$^{o}$C for 1 hour.

The degradation of rRNA step of the protocol was skipped as potential probe attachment to rRNA would only increase the signal intensity, and DNA-specific binding was not necessary for this application. Although, as part of the proof of concept using UCYN-A2, the rRNA was degraded to verify that the signal originated from DNA and not rRNA. UCYN-A rRNA was degraded with RNAase. The RNAse

solution contained 4 µl (800 µg $ml^{-1}$) RNAse (Qiagen) into 500 µl (1 M Tris-HCL)) and water up to 5 ml, sampled were incubated in the solution overnight at 37$^{o}$C or 30 minutes at 65$^{o}$C.

Introduced HRPs from gene hybridization were inactivated before the second hybridization (rRNA hybridization) by incubating samples in 0.2 M HCL for 10 minutes at room temperature.

The geneFISH protocol for UCYN-A was the same as described for Gamma A, except samples were only permeabilized with lysozyme as described. The amplification of the gene signal was 10 minutes at room temperature, according to Moraru et al. (2010).

*geneFISH rRNA hybridization*

After gene hybridization and amplification, the rRNA of the cells are targeted for hybridization to connect cell identity to the presence of the gene. The *E coli* controls were hybridized with the general CARD-FISH bacterial probe EUB I-III (Amann et al., 1990) using a 35% FA hybridization buffer. In environmental samples with UCYN-A2, the attached *B. bigelowii* host rRNA was targeted with URADO-69 probe with associated helpers (Helper A PRYM, Helper B PRYM) and competitors (UBRADO-69 competitor) using 40% FA hybridization buffer following description by Cornejo-Castillo et al. (2016). The environmental Gamma A samples were hybridized with a general Gammaproteobacterial probe Gam42a with a 35% FA hybridization buffer (Manz et al., 1992). The rRNA hybridizations were incubated at 46$^{o}$C for 2 hours before washing, followed by CARD amplification. For CARD

amplification, samples were incubated in amplification buffer containing 1 µl $ml^{-1}$ $Alexa_{488}$ (Invitrogen) and 0.0015% $H_2O_2$ for 45 minutes at 37°C. Hybridization, washing, and amplification buffers were as described in Pernthaler et al. (2004). After the amplification procedure, samples were stained with DAPI and mounted on a glass slide described above for CARD-FISH. Slides were stored at -20°C until viewed with epifluorescence microscopy.

*Negative controls*

Various negative controls were included to verify the steps of the protocol and the specificity of the hybridizations. Initial endogenous peroxidase inactivation and secondary introduced peroxidase inactivation controls were included, samples were inactivated and then amplified to ensure all non-target peroxidases were inactivated. *E. coli* cells containing *nifH* plasmids identified as UCYN-A2 were used as specific controls, and *E. coli* cells without plasmids were used to control false-positive signals. Additionally, *E. coli* clones containing Gamma A *nifH* plasmids that had been digested with DNAase (0.2 U $µl^{-1}$, Qiagen) overnight at 37°C followed by DNAse inactivation (25min @ 75°C) and 2x washing in 1xPBS (as described in Moraru 2010) were used as a negative control. Environmental samples with no or undetectable abundances of Gamma A as measured by qPCR were used as negative environmental controls.

*Imaging*

Samples and controls were imaged under epifluorescence with a Zeiss Axioplan or Zeiss Axio Imager.Z2 microscope equipped with digital imaging at the

University of California, Santa Cruz. Counts of hybridized cells used to calculate volumetric cell abundance (UCYN-A) were performed in triplicate transects (~8.0 × 0.1 $mm^2$ each) across the length of the filter pieces (Cabello et al., 2015). The number of Gamma A geneFISH hybridized cells were counted over the whole filter section used for hybridization.

*NanoSIMS*

To investigate Gamma A $N_2$ fixation activity in the Western Pacific, triplicate 15 m depth seawater samples (<210 um, 4.4 l) were injected with a $^{15}N_2$ gas bubble (98%+, lot # - I-22779?AR0720534, Cambridge Isotopes, Tewksbury, MA, USA, ) using the bubble release method (Klawonn et al., 2015b). The $^{15}N_2$ bubble was equilibrated with the seawater by rolling the bottle back and forth for 20 minutes (~25 rpm) before the bubble was released and $^{13}$C-bicarbonate was added (60 µM, 99%, Cambridge). Triplicate incubations were transferred to flow-through deck-board incubators shaded to the approximate in situ light intensity at 15 m depth. After 24 hours, membrane inlet mass spectrometry (MIMS) samples were taken from each of the triplicate bottles to measure the dissolved $^{15}N_2$-gas enrichment available (2.8 At% ± 0.4). MIMS samples were siphoned into glass vials (15 ml) after 3x volume overflow. The vials were stoppered and crimped with aluminum caps and kept at 4$^o$C until measured with a MIMS. Samples for geneFISH were collected by fixing 95 ml of each $N_2$ fixation incubation in 1.8% formaldehyde (final) for up to 72 hours at 4$^o$C, filtering onto 0.6 µm pore size polycarbonate filters (Millipore Sigma) and flash freezing in liquid N before storage at -80$^o$C until processing. Gamma A was

visualized in the $N_2$ fixation incubation samples using the geneFISH procedure as described above. Cells were transferred by freezing an inverted filter piece on top of a silicon (Si) chip with alphanumeric labeled grids (Ted Pella, Redding, CA, USA) for 3 minutes at -80$^{o}$C before removing the filter piece. Visual maps of positive cells on the Si chip were created using a Zeiss Axioplan epifluorescence microscope. The Si chip was gold coated and analyzed with a nanoscale secondary ion mass spectrometer (nanoSIMS, Cameca50, Lawrence Livermore National Laboratory) and a $Cs^+$ ion beam. Areas of interest (30 µm raster) were sputtered to a depth of 60 nm before data collection of $^{12}C_2^-$, $^{13}C^{12}C^-$, $^{14}N^{12}C^-$, $^{15}C^{12}C^-$, and secondary electrons for 80 cycles. Cells identified as Gamma A by microscopy were analyzed for $N_2$ fixation as evident by enrichment of $^{15}N$ ($^{15}N^{12}C/(^{15}N^{12}C+^{14}N^{12}C)$) (Pett-Ridge and Weber, 2012) compared to surrounding non-Gamma A cells analyzed within the same image frame. Cells outlined as regions of interest (ROI) were defined automatically based on $^{14}N^{12}C^-$ patterns. NanoSIMS data analysis and ROI outlining were done using L'image (developed by L. Nittler, Carnegie Institution of Washington, Washington, DC, USA)). The LOD for $^{15}N$ enrichment ($N_2$ fixation activity) of Gamma A cells was defined as $^{15}N$ enrichment values distinctly above all other non-Gamma A cells provided the ROI had acceptable instrumentation error (<30%) and sufficient $^{15}N$ counts (>50).

**Results**

Since it is challenging to validate hybridization approaches for uncultivated microorganisms in environmental samples, geneFISH was first used to target the

well-studied $N_2$-fixing cyanobacterium UCYN-A before targeting Gamma A. UCYN-A is an uncultivated symbiont of a haptophyte host, *Braarudosphaera bigelowii*. We used a UCYN-A *nifH* gene probe to visualize the UCYN-A2 symbiosis (~7µm diameter, host and symbiont together) in samples from the Bering Strait and California Coastal Current (Figure 1) that had abundances of ~$10^5$ UCYN-A2 *nifH* gene copies $l^{-1}$ as determined by qPCR. Following successful hybridization to the geneFISH probe, the presumed cyanobacterial cell (~2.5 um) in the symbiosis had multiple distinct hybridization signals (Figure 1b-d). There were 0-9 signals with an average of 3.7 (±1.7) in the Bering Strait and 4.9 (±1.7) signals per symbioses in the California Coastal Current (Figure 1). The average number of geneFISH signals from each location was not statistically different. It was not visually evident whether the multiple signals were due to multiple cells, multiple gene copies or chromosomes, or a technical artifact. GeneFISH probes are double-stranded DNA synthesized by PCR and could potentially bind to *nifH* mRNA, although previous studies found no evidence of mRNA bound to probes (Moraru 2010). To confirm that the signal originated from DNA in our samples, controls were treated with RNase to digest mRNA prior to the geneFISH probe hybridization. RNAse digested samples had, on average, the same number of positive gene signals within UCYN-A as the non-RNAse digested samples (3.7 ± 1.7) in the Bering Strait. This implies that hybridization to mRNA was not the cause of multiple signals and that multiple copies of the *nifH* gene must be present. A comparison of UCYN-A2 abundances determined by qPCR and CARD-FISH indicates there is approximately 5 times

higher number of the *nifH* gene (qPCR) than occurrences of the symbioses as a whole (CARD-FISH) (Supplemental Figure 1). Loss of DNA during DNA extraction likely results in a lower number of UCYN-A *nifH* gene copies $L^{-1}$, so the discrepancy between qPCR and CARD-FISH abundances is most likely higher than the 5 times observed here. The multiple gene signals may be due to multiple cyanobacterial cells or a single cell with multiple genome copies (polyploidy) (Sargent et al., 2016). Regardless, the results of the UCYN-A geneFISH show that geneFISH can be successfully used to visualize specific uncultivated diazotrophs.

To visualize the uncultivated Gamma A cells in environmental samples, we applied a Gamma A geneFISH approach. To do this, we designed a Gamma A geneFISH probe to target Gamma A, which is a group defined by closely related *nifH* sequences (>97% ID similarity). However, polynucleotide geneFISH probes cannot be as specific as the more common oligonucleotide probes (Matturro and Rossetti, 2015). Therefore, the Gamma A geneFISH probe would likely bind to other γ-proteobacterial NCDs outside the Gamma A microdiversity. Reference Gamma A sequences from previous studies are summarized in Table 1 and have at least 98% similarity to the designed probe. The most dissimilar Gamma A sequence to the probe was 97.5% identical. A collection of other Gammaproteobacterial *nifH* gene sequences had a maximum sequence identity to the Gamma A geneFISH probe of 74-97.5% over the probe length.

The Gamma A *nifH* gene probe was tested for functionality and specificity using *E. coli* clones. The large majority (97.2 ± 2.5%) of *E. coli* clones containing

Gamma A *nifH* plasmids had at least one positive signal (Figure 2A). However, the presence of multiple *nifH* plasmid copies per *E.coli* cell likely resulted in a higher hybridization efficiency (percentage of cells showing a gene signal) than cells containing a single chromosomal copy of the *nifH* gene. To check the Gamma A probe specificity, an *E. coli* clone containing UCYN-A1 *nifH* plasmids (78% sequence identity to the Gamma A probe) was hybridized with the Gamma A probe (Figure 2b). There were on average 4.8% (± 2.1%) positive signals observed in the UCYN-A *nifH* clones. Negative controls included *E. coli* clones with a Gamma A *nifH* plasmid digested with DNAse and *E. coli* cells without a *nifH* plasmid. Both controls were negative and allowed us to calculate the false-positive rates of 4.7% (± 1.8%) and 6.3 (± 3.0%), respectively. Since the UCYN-A *nifH* clone had a positive signal rate within the range of false-positive found in negative controls, UCYN-A *nifH* clones were considered a successful negative control. They showed the Gamma A probe specificity is greater than 78% identical sequences. Low background signal noise was observed (Figure 2B) but did not interfere with the true signal as the background signal did not overlap with the secondary hybridization of the general bacterial CARD-FISH probe (EUBI-III) and should not be confused with a positive signal. The Gamma A probe was determined to be functional, and further optimization of the protocol was performed on environmental samples with natural populations of Gamma A.

Gamma A geneFISH was employed to visualize Gamma A cells in seawater samples to gain insight into the lifestyle of these cells. Samples used for Gamma A

visualization originated from the western NPSG. Environmental samples that did not contain Gamma A (determined by qPCR and *nifH* gene sequencing) were used as negative controls and did not have any observable false positive signals. Positive Gamma A hybridizations were observed in rod-shaped cells 1.5 (± 0.22) µm in length and 0.6 (± 0.12) µm in width (Figures 3 and 4). Gamma A was observed as single free-living cells, pairs of free-living cells attached tip to tip, and cells attached to particles. The majority of Gamma A cells observed were associated with particles (61% of observed cells, n=23). Cells were attached to particles of a range of diameters (4 µm to 50 µm in diameter). Gamma A was found on loosely aggregated particles (Figure D), dense particles (Figure 3A, B), and in one observation on a larger eukaryotic cell (Figure 3C). Gamma A cells attached to particles were usually individual cells. However, in 2 particles, there were multiple Gamma A cells, and in another, there were pairs of connected Gamma A cells. The majority of unattached cells (not associated with particles) were found in pairs (90% of non-particle attached cells) attached end to end (Figure 4). A few solitary, detached Gamma A cells were also observed.

To determine if Gamma A cells were fixing $N_2$, individual cells were also analyzed by nanoSIMS after incubation with $^{15}N_2$. Eight cells from NPSG samples were identified as Gamma A, hybridizing with the general Gammaproteobacterial CARD-FISH and Gamma A *nifH* probes. $N_2$ fixation would be demonstrated by $^{15}N$ isotope enrichment compared to the natural isotopic abundance present in non-diazotrophic cells. The Gamma A cells were not enriched in $^{15}N$ above non-Gamma

A cells within the same analysis image. Although the $^{15}N$ enrichment was not statistically significant as defined by the threshold, the Gamma A cells were in the highest 2% of $^{15}N$ enrichment of all qualified ROIs in 4 out of the 8 analyzed images. This suggests the possibility that Gamma A cells were fixing $N_2$ at low rates but below the detection limits of our method; however, our data remain inconclusive.

**Discussion**

*geneFISH method development*

Hybridization tests of the *nifH* geneFISH method indicated that it was successful. Hybridization to the well-characterized uncultivated symbiotic UCYN-A cyanobacteria showed that hybridization to a gene in single cells of known identity could be detected (Figure 1). Hybridization to cloned genetically diverse Gamma A *nifH* sequences showed that probes could effectively be designed to hybridize with uncultivated species (Figure 2). Applications of probes to seawater samples containing Gamma A showed that Gamma A cells could be identified *in situ* (Figure 3 and 4).

We tested if $N_2$-fixing organisms could be identified microscopically in environmental samples by first designing geneFISH probes for the UCYN-A *nifH* gene. UCYN-A has previously been visualized with CARD-FISH (Figure 1a), making it a suitable target for geneFISH validation since the morphology is known and identity can be confirmed (Cabello 2015, Cornejo-Castillo 2016). Less than 100% of the UCYN-A2 cells observed by host targeted CARD-FISH had positive geneFISH signals (80 to 86%), confirming that geneFISH cannot be used

quantitatively (Moraru 2010). Nevertheless, geneFISH was successfully used to visualize the UCYN-A *nifH* gene in environmental samples.

Gamma A geneFISH probe design included optimizing the length of the probe. The successful geneFISH probe for Gamma A was 321 base pairs (bp) in length which produced a bright signal over most of the cells after amplification. Initially, shorter probes (102 bp and 113 bp) that sequentially covered the length of the *nifH* gene were tested, as their smaller size would enter the cell more easily, and shorter probes can have proportionally higher probe specificity to the target region (Gallego 2020, Moraru 2010). However, the shorter probes did not result in a sufficient signal. Longer polynucleotide probes contain more (Dig)-labelled-dUTP allowing more anti-Dig conjugated HRPs to bind to the probe, producing a higher rate of tyramide deposition during amplification and consequently a brighter and larger signal (Matturro 2015). Previous studies also found better success with longer probe lengths (Gallego, 2020). The longer Gamma A probes provided a balance of probe specificity and signal intensity.

The chosen Gamma A probe successfully targeted the spectrum of Gamma A *nifH* sequence diversity as shown by hybridization to *nifH* cloned into *E. coli* plasmids. Additionally, the Gamma A probe did not hybridize with the UCYN-A *nifH* clones (negative controls) or UCYN-A symbioses in environmental samples (78% sequence identity to the probe). Determining the specificity threshold is challenging without sufficient diverse representative cloned sequences. Previous studies found that similar length polynucleotide probes had specificity values that ranged from 82 –

89% (Kawakami 2012, Moraru 2011). Although it was not possible to test the Gamma A probe specificity threshold, the Gamma A geneFISH probe would likely hybridize with similar non-Gamma A Gammaproteobacterial *nifH* sequences, possibly with lower efficiency (Kawakami, 2012). To avoid visualization of other Gammaproteobacteria, we chose an environmental sample where Gamma A was the dominant Gammaproteobacterial *nifH* sequence based on relative abundance by at least a factor of 60 compared to the next most abundant Gammaproteobacterial sequence (see chapter 3, Fig. S2, station 10). As Gamma A abundances were $10^3$ *nifH copies* $L^{-1}$ at that location, an organism with 60 times lower relative abundance has a very low probability of even being present on the filter section or detected by geneFISH. Based on the abundance differences, the stringency of the protocol (50% FA, rapid temperature changes), and conserved cellular morphologies observed, we are confident the rod-shaped cells identified were Gamma A. Future Gamma A specific studies in areas with high abundances of other similar Gammaproteobacterial diazotrophs should check probe specificity for the *nifH* sequences in the study area.

Multiple negative controls were used in the Gamma A geneFISH protocol to confirm that the hybridization signals observed were true positive signals. GeneFISH generally has a higher rate of false positives than CARD-FISH or similar methods due to the small dot-like signal and high background (Bernhard et al., 2012; Moraru et al., 2010). The false-positive rate observed in our negative *E. coli* controls (4.7 to 6.3%, Figure 2B) is within the range of false-positive rates for the method in different systems (4-7%) (Bernhard et al., 2012; Moraru et al., 2010). We found no evidence of

false positives in negative environmental controls that would confuse a true positive signal, i.e., the cell having both rRNA and geneFISH signal. Although nonspecific binding is higher in geneFISH compared to CARD-FISH, the use of dual hybridizations greatly reduces the probability of nonspecific binding being mistaken for a positive signal. For multiple reasons, including creating a larger separation between noise and signal, the amplification protocol was increased in environmental samples to obtain a brighter signal, which is likely why we observed no evidence of false positives in the negative environmental controls. The geneFISH signal was considered positive when both rRNA and gene hybridizations were apparent, and the gene hybridization signal was distributed over the majority of the cellular membrane.

We did not attempt to calculate Gamma A volumetric abundances (cells $L^{-1}$) with geneFISH as the hybridization is not quantitative. Moraru et al. (2010) previously reported a hybridization efficiency (percentage of cells showing a gene hybridization signal) of 46%, while Bernhard et al. (2012) calculated a hybridization efficiency up to 65%, while a similar approach known that increases the amplification of the signal known as geneCARD-FISH had a detection efficiency of up to 99% (Matturro 2015). GeneFISH on Gamma A *nifH* clones in this study had a hybridization efficiency of 98% (Figure 2A). However, this efficiency may not apply to environmental samples in which probes are hybridized to genomic DNA, and there is likely a single copy of the *nifH* gene per chromosome, which would decrease the hybridization efficiency. We made a rough comparison of Gamma A cells identified using geneFISH to Gamma A *nifH* gene abundances enumerated using qPCR. Based

on qPCR, there should be 23-30 Gamma A cells per filter piece; by scanning the entire filter, we found an average of 11 Gamma A cells. If qPCR represents the true abundance of Gamma A cells, this will equate to a hybridization efficiency of 36-48%. However, qPCR-based abundances are likely underestimated due to incomplete DNA extraction efficiency (Matturro and Rossetti, 2015; Sargent et al., 2016)), which would result in a lower geneFISH efficiency than estimated here. The qualitative comparison of Gamma A geneFISH abundances compared to qPCR abundances further validated the probe specificity as we do not find more positively identified geneFISH cells than expected.

*UCYN-A nifH geneFISH*

Application of *nifH* geneFISH on the symbiotic cyanobacterium UCYN-A2 revealed multiple copies of the *nifH* gene in the UCYN-A2 symbiosis. The geneFISH hybridization signal was observed in the cyanobacterial partner of the symbiosis, where there were several visually distinct dot-like signals within the cyanobacterium (Figure 1). Multiple signals could suggest multiple UCYN-A2 cells per *B. bigelowii* host cell, a single UCYN-A2 cell with multiple copies of the genome (polyploid), or that multiple *nifH* copies are present in a single UCYN-A2 genome. While the closed genome of the closely related UCYN-A1 only has a single copy of the *nifH* gene present (Tripp et al., 2010), the genome of UCYN-A2 is incomplete, and therefore the number of *nifH* gene copies per genome is unknown (Bombar et al., 2014). To observe the potential presence of multiple UCYN-A2 cells, samples were stained with a cell membrane stain, but interior membranes corresponding to multiple

cyanobacterial cells were not observed (data not shown). Although the cause of multiple hybridization signals could not be determined in this study, polyploidy is a possibility as the cyanobacterial diazotroph *Trichodesmium* is polyploid with more than 100 genome copies per cell (Sargent et al., 2016). Additionally, a strain of the non-diazotrophic cyanobacteria *Synechococcus* has an average of 4 genome copies per cell (Perez-Sepulveda et al., 2018). GeneFISH was successful in identifying UCYN-A and revealed new information on the biology of the symbiosis.

The UCYN-A2 cells observed with geneFISH had an overall average of 4.3 *nifH* gene signals per symbiosis. Although geneFISH is not considered quantitative, a study that compared host abundance (18S rRNA gene) to the cyanobacterial abundance (*nifH* gene) via qPCR found a similar ratio average of 3.3:1 UCYN-A2 to *B. bigelowii* cells (Thompson et al., 2014). The presence of multiple copies of the 18S rRNA gene is unknown in *B. bigelowii*; however, if there is gene multiplicity in the host, it would result in more UCYN-A *nifH* gene copies per individual host cell. Similarly, a study that identified UCYN-A2 symbioses with CARD-FISH found 3 to 10 UCYN-A2 cells per *B. bigelowii* host. The multiple cyanobacterial cells were observed as DAPI staining within the CARD-FISH labeled cyanobacteria (Cornejo-Castillo et al., 2016). Similarly, our CARD-FISH-based UCYN-A2 symbiosis counts are ~5 times lower than the corresponding qPCR-based UCYN-A2 abundances (Supplementary Figure 1). Although comparing these two methods is problematic due to the high likelihood of incomplete DNA extraction, a higher extraction efficiency would increase the number of *nifH* genes (qPCR) per CARD-FISH identified

symbiosis, making our UCYN-A *nifH* copies to *B. bigelowii* cell ratio an underestimation. Our results and the results of previous studies could not distinguish between multiple cyanobacterial cells, a single polyploid cell, or the presence of multiple copies of the *nifH* gene in the genome. However, our results suggest that attempts to use *nifH* qPCR abundances as a proxy for UCYN-A2/*B. bigelowii* symbiosis abundances should be viewed with caution.

*Gamma A nifH geneFISH*

The goal of the geneFISH identification of Gamma A was to gain insight into the lifestyle of this organism through visualization of cell morphology. Based on previous studies that investigated the presence of Gamma A *nifH* genes in size-fractionated samples, it has been hypothesized that Gamma A may have a particle attached lifestyle (Benavides et al., 2016; Cornejo-Castillo and Zehr, 2021; Langlois et al., 2015; Moisander et al., 2014). Through geneFISH identification, we observed that the majority of Gamma A were associated with particles (61%, Figure 3) but were also found non-particle attached, primarily as pairs of cells (35%, Figure 4) and a few cases of individual cells (4%). While an individual Gamma A cell is now known from this study to be 1.5 um long and 0.7 um wide and would be expected to be found in the free-living fraction (< 3 μm), in the few studies that targeted this specific size fraction, Gamma A *nifH* genes were not detected (Benavides et al., 2016; Cornejo-Castillo and Zehr, 2021; Gradoville et al., 2020). qPCR data for the < 3 μm range and our geneFISH observations both suggest that occurrences of free-living individual cells are rare. Most of the non-particle attached Gamma A cells appeared

as pairs of cells attached pole to pole (≥ 3 µm in length) and would likely end up in the >3 µm size fraction. This observation is supported in size fractionation data, where maximal Gamma A abundances have been reported in the >3 µm size fraction (Benavides et al., 2016; Cornejo-Castillo and Zehr, 2021; Gradoville et al., 2020). However, studies with higher resolution size fractionation found the highest Gamma A abundances in even larger > 10 µm (Benavides et al., 2016) and 5 to 20 µm size fractions (Cornejo-Castillo and Zehr, 2021) which should be larger than even pairs of Gamma A cells. Additionally, Gamma A *nfiH* sequences were present in all size fractions from >3 µm up to 2000 µm based on metagenomic analysis (Cornejo-Castillo and Zehr, 2021), which is in line with our geneFISH results showing that the majority of Gamma A cells were attached to particles.

GeneFISH identification of Gamma A allowed us to directly visualize cells attached to particles and compare our observations to previous findings. Recently another visual-based investigation of NCDs reported many NCDs attached to a single particle in estuarine samples through immunolabeling of the nitrogenase enzyme (Geisler et al., 2019, 2020). Although taxonomic identity and $N_2$ fixation activity of NCDs were not determined, these findings suggest that particle attachment in surface water may be a common lifestyle for NCDs. In an alternative lifestyle, a NCD isolated from the surface waters of the Baltic Sea (*Pseudomonas stutzeri BAL361)* was found to produce extracellular organic polymers, which facilitates aggregation of cells and limits $O_2$ diffusion (Bentzon-Tilia et al., 2015a; Paerl et al., 2018). A similar strategy had been hypothesized for Gamma A (Benavides et al., 2016), but it unlikely

that the particles Gamma A were attached were from organic polymers produced by Gamma A itself, as all but two particles had a single observable Gamma A cell. Additionally, although organic polymer production for aggregation is found in both terrestrial and marine NCDs (Bentzon-Tilia et al., 2015a; Paerl et al., 2018; Ueda and Saneoka, 2015), this would be energetically challenging in the oligotrophic ocean (Inomura et al., 2017). We found no evidence of a specific eukaryotic partner, although eukaryotic nuclei were visible in three particles. Surprisingly for a particle-attached cell, maximum abundances of Gamma A were at 15 m depth (Langlois et al., 2015; Moisander et al., 2014), which may indicate that not all Gamma A cells stay attached to sinking particles due to either cell motility or cell mortality. In the oligotrophic North Pacific, Gamma A sequences were among the top 10 sequences recovered from individual and bulk particles at 150 m depth (Farnelid et al., 2019), but were not found in sediment trap samples from 4000 m depth (Poff et al., 2021). Additionally, maximum abundance at 15 m depth may imply a light-dependent lifestyle such as photoheterotrophy, which could help fuel energetically expensive $N_2$ fixation (Inomura et al., 2017; Moisander et al., 2017), but visualization by geneFISH could not provide any evidence of this strategy.

Particles have been hypothesized to provide an ideal microenvironment for NCD in the C-poor, oxygenated ocean as they would provide abundant C and low $O_2$ (Bombar et al., 2016; Church et al., 2005b; Moisander et al., 2014; Ploug, 2001; Riemann et al., 2010). C is commonly a limiting nutrient for heterotrophic bacteria in the oligotrophic ocean (Kirchman et al., 2000), and $N_2$ fixation is an energetically

costly process (Howard and Rees, 1996) requiring sufficient C to fuel $N_2$ fixation. One reason cyanobacteria have long been considered the only relevant $N_2$-fixing organisms in the surface ocean is their unlimited supply of C through photosynthesis (Karl et al., 2002). Particle attachment for an NCD would supply a reliable source of C (Bombar et al., 2016; Inomura et al., 2017; Klawonn et al., 2015a), and larger particles could also theoretically provide an interior microaerobic zone (particles >1 mm), caused by high rates of microbial respiration (Klawonn et al., 2015a; Paerl and Prufert, 1987; Ploug, 2001). Low $O_2$ zones would protect the $O_2$-sensitive nitrogenase enzyme (Paerl and Prufert, 1987; Ploug, 2001). However, more studies are needed to determine the physiology and activity of Gamma A on particles, as cells were observed on particles smaller than 1 mm, and the presence of Gamma A does not reflect $N_2$ fixation activity.

Unfortunately, detection of $N_2$ fixation by Gamma A was inconclusive in this study. Out of the 8 Gamma A cells measured using nanoSIMS, 4 had higher $^{15}N$ enrichment values than the other non-Gamma A cells in the image, yet differences were not statistically significant. These results may be due to very low or no $N_2$ fixation or may be due to isotope dilution caused by the geneFISH protocol. Previous studies demonstrate that CARD-FISH protocols can dilute the heavy isotope enrichment of target cells (Meyer et al., 2021; Musat et al., 2014; Woebken et al., 2012). A single CARD-FISH procedure can dilute the $^{15}N$ enrichment of cells up to 60% (Meyer et al., 2021). The geneFISH protocol likely creates a much higher dilution. In addition to a full CARD-FISH protocol, it includes geneFISH probe

hybridizations, HRP-conjugated anti-Dig binding, and increased amplification. There would also be further dilution caused by the increased cell permeabilization protocol needed to allow the large geneFISH probes to enter the cell. The $^{15}N$ isotope is diluted with the addition of any non-labeled substrates (such as HRP-conjugated anti-Dig or tyramides) or the removal of enriched cellular material (cell permeabilization). For future studies wanting to identify cells by geneFISH for single-cell analysis with nanoSIMS, we recommend starting with high $^{15}N$ enrichment in the $N_2$ fixation incubations to compensate for $^{15}N$ lost due to dilution through geneFISH.

**Conclusion**

Direct visualization of cells is a powerful tool to gain insight into an organism's lifestyle. Our results demonstrate geneFISH is a viable method to identify and study uncultured diazotrophs with unknown 16S rRNA genes in environmental samples. GeneFISH allowed for the first visual identification of Gamma A cells and revealed multiple copies of the *nifH* gene within UCYN-A2 cells. Visualization of Gamma A confirmed the hypothesis that Gamma A is associated with particles in the surface ocean. Although $N_2$ fixation activity could not be confirmed for Gamma A, its particle attached lifestyle highlights the potential role that particles may play as loci for $N_2$ fixation in the surface ocean.

## References


Amann, R., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., & Stahl, D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology*, *56*(6), 1919–1925. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid =184531&tool=pmcentrez&rendertype=abstract

Benavides, M., Moisander, P. H., Daley, M. C., Bode, A., & Arístegui, J. (2016). Longitudinal variability of diazotroph abundances in the subtropical North Atlantic Ocean. *Journal of Plankton Research*, *0*, fbv121. https://doi.org/10.1093/plankt/fbv121

Bentzon-Tilia, M., Severin, I., & Hansen, L. H. (2015). Genomics and Ecophysiology of Heterotrophic Nitrogen-Fixing Bacteria Isolated from Estuarine Surface Water. *MBio*, *6*(4), 1–11. https://doi.org/10.1128/mBio.00929-15.Editor

Bernhard, J. M., Edgcomb, V. P., Casciotti, K. L., Mcilvin, M. R., & Beaudoin, D. J. (2012). Denitrification likely catalyzed by endobionts in an allogromiid foraminifer. *ISME Journal*, *6*, 951–960. https://doi.org/10.1038/ismej.2011.171

Bird, C., Martinez, M., Donnell, A. G. O., & Wyman, M. (2005). Spatial distribution and transcriptional activity of an uncultured clade of planktonic diazotrophic-proteobacteria in the Arabian Sea. *Applied and Environmental Microbiology*, *71*(4), 2079–2085. https://doi.org/10.1128/AEM.71.4.2079

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

Bombar, D., Paerl, R. W., & Riemann, L. (2016). Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection. *Trends in Microbiology*, *24*(11), 16–927. https://doi.org/10.1016/j.tim.2016.07.002

Bonnet, S., Dekaezemacker, J., Turk-Kubo, K. A., Moutin, T., Hamersley, R. M., Grosso, O., … Capone, D. G. (2013). Aphotic $N_2$ Fixation in the Eastern Tropical South Pacific Ocean. *PLoS ONE*, *8*(12), e81265. https://doi.org/10.1371/journal.pone.0081265

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

Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., … Carpenter, E. J. (2005). Nitrogen fixation by *Trichodesmium*

spp .: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, *19*. https://doi.org/10.1029/2004GB002331

Church, M J, Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Vertical distributions of nitrogen-fixing phylotypes at Station ALOHA in the oligotrophic North Pacific Ocean. *Aquatic Microbial Ecology*, *38*, 3–14. https://doi.org/10.3354/ame038003

Church, Matthew J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., & Karl, D. M. (2009). Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochemical Cycles*, *23*(2), n/a-n/a. https://doi.org/10.1029/2008GB003418

Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sa, P., Lima-mendez, G., Hingamp, P., … Zehr, J. P. (2016). Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nature Communications*, (March). https://doi.org/10.1038/ncomms11071

Cornejo-Castillo, F. M., & Zehr, J. P. (2021). Intriguing size distribution of the uncultured and globally widespread marine non-cyanobacterial diazotroph Gamma-A. *The ISME Journal*, 124–128. https://doi.org/10.1038/s41396-020-00765-1

Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T. M., Rappé, M. S., … Eren, A. M. (2018). Nitrogen-fixing populations of *Planctomycetes* and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, *3*(July). https://doi.org/10.1038/s41564-018-0176-9

Farnelid, H., Andersson, A. F., Bertilsson, S., Al-Soud, W. A., Hansen, L. H., Sørensen, S., … Riemann, L. (2011). Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0019223

Farnelid, H., Harder, J., Bentzon-Tilia, M., & Riemann, L. (2014). Isolation of heterotrophic diazotrophic bacteria from estuarine surface waters. *Environmental Microbiology*, *16*(10), 3072–3082. https://doi.org/10.1111/1462-2920.12335

Farnelid, H., Helle, K. T., Justin, P., James, E. O., Mooy, B. A. S. Van, & Zehr, J. P. (2019). Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *The ISME Journal*, 170–182. https://doi.org/10.1038/s41396-018-0259-x

Gallego, S., Barkay, T., & Fahrenfeld, N. L. (2020). Tagging the *vanA* gene in wastewater microbial communities for cell sorting and taxonomy of *vanA* carrying cells. *Science of the Total Environment*, *732*, 138865. https://doi.org/10.1016/j.scitotenv.2020.138865

Galloway, J N, Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., … Vo, C. J. (2004). *Nitrogen cycles : past, present, and future*.

Galloway, James N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., … Sutton, M. A. (2008). Transformation of the Nitrogen Cycle : *Science*, *320*(May), 889–892. https://doi.org/10.1126/science.1136674

Geisler, E., Bogler, A., Bar-zeev, E., Rahav, E., & Farnelid, H. (2020). Heterotrophic Nitrogen Fixation at the Hyper-Eutrophic Qishon River and Estuary System. *Frontiers in Microbiology*, *11*(June), 2012–2021. https://doi.org/10.3389/fmicb.2020.01370

Geisler, E., Bogler, A., Rahav, E., & Bar-Zeev, E. (2019). Direct detection of heterotrophic diazotrophs associated with planktonic aggregates. *Scientific Reports*, 1–9. https://doi.org/10.1038/s41598-019-45505-4

Goebel, N. L., Turk, K. a., Achilles, K. M., Paerl, R. W., Hewson, I., Morrison, A. E., … Zehr, J. P. (2010). Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to $N_2$ fixation in the tropical Atlantic Ocean. *Environmental Microbiology*, *12*(12), 3272–3289. https://doi.org/10.1111/j.1462-2920.2010.02303.x

Gradoville, M. R., Farnelid, H., White, A. E., Turk-Kubo, K. A., Stewart, B., Ribalet, F., … Zehr, J. P. (2020). Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnology and Oceanography*, *65*, 1858–1875. https://doi.org/10.1002/lno.11423

Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., … Kuypers, M. M. M. (2012). Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *The ISME Journal*, *6*(6), 1238–1249. https://doi.org/10.1038/ismej.2011.182

Howard, J. B., & Rees, D. C. (1996). Structural Basis of Biological Nitrogen Fixation. *Chemical Reviews*, *96*(7), 2965–2982. https://doi.org/10.1021/cr9500545

Inomura, K., Bragg, J., & Follows, M. J. (2017). A quantitative analysis of the direct and indirect costs of nitrogen fixation : a model based on *Azotobacter vinelandii*. *ISME Journal*, 166–175. https://doi.org/10.1038/ismej.2016.97

Karl, D M, Michaels, A., Bergman, B., & Capone, D. (2002). Dinitrogen fixation in the world's oceans. *Biochemistry*, *57*(58), 47–98. https://doi.org/10.1023/A:1015798105851

Karl, David M, Bidigare, R. R., Church, M. J., Dore, J. E., Letelier, R. M., Mahaffey, C., & Zehr, J. P. (2008). *The Nitrogen Cycle in the North Pacific Trades Biome : An Evolving Paradigm*. https://doi.org/10.1016/B978-0-12-372522-6.00016-5

Kirchman, D. L., Meon, B., Cottrell, M. T., Hutchins, D. A., Weeks, D., & Bruland, K. W. (2000). Carbon versus iron limitation of bacterial growth in the California upwelling regime. *Limnology and Oceanography*, *45*(8), 1681–1688. https://doi.org/10.4319/lo.2000.45.8.1681

Klawonn, I., Bonaglia, S., Bruchert, V., & Ploug, H. (2015). Aerobic and anaerobic nitrogen transformation processes in $N_2$-fixing cyanobacterial aggregates. *The ISME Journal*, *9*(15), 1456–1466. https://doi.org/10.1038/ismej.2014.232

Klawonn, I., Lavik, G., Böning, P., & Marchant, H. K. (2015). Simple approach for the preparation of $^{15-15}N_2$-enriched water for nitrogen fixation assessments : evaluation, application, and recommendations. *Frontiers in Microbiology*, *6*(August), 1–11. https://doi.org/10.3389/fmicb.2015.00769

Langlois, R., Großkopf, T., Mills, M., Takeda, S., & LaRoche, J. (2015). Widespread Distribution and Expression of Gamma A (UMB), an Uncultured, Diazotrophic, γ-Proteobacterial *nifH* Phylotype. *Plos One*, *10*(6), e0128912. https://doi.org/10.1371/journal.pone.0128912

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, … Schleifer, K.-H. (2004). ARB: a software environment for sequence data. *Nucleic Acids Research*, *32*(4), 1363–1371. https://doi.org/10.1093/nar/gkh293

Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., … Zehr, J. P. (2012). Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth System Science Data*, *4*(1), 47–73. https://doi.org/10.5194/essd-4-47-2012

Manz, W., Amann, R., Ludwig, W., Wagner, M., & Schleifer, K.-H. (1992). Phylogenetic Oligodeoxynucleotide Probes for the Major Subclasses of Proteobacteria: Problems and Solutions. *Systematic and Applied Microbiology*, *15*(4), 593–600. https://doi.org/https://doi.org/10.1016/S0723-2020(11)80121-9

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., … Kuypers, M. M. M. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, *1*(11). https://doi.org/10.1038/nmicrobiol.2016.163

Martinez-Perez, C., Mohr, W., Schwedt, A., Durschlag, J., Callbeck, C. M., Schunck, H., … Kuypers, M. M. M. (2017). Metabolic versatility of novel $N_2$-fixing Alphaproteobacterium isolated from oxygen minimum zone. *Environmental Microbiology*, *49*(1), 1–41. https://doi.org/10.1111/1462-2920.

Masuda, S., Nagaosa, K., & Kato, K. (2020). Direct Detection of Denitrifying Bacteria in Groundwater by GeneFISH. In *Behavior of Radionuclides in the Environment* (pp. 99–113). Springer, Singapore. https://doi.org/doi.org/10.1007/978-981-15-0679-6_4

Matturro, B., & Rossetti, S. (2015). GeneCARD-FISH: Detection of *tceA* and *vcrA* reductive dehalogenase genes in Dehalococcoides mccartyi by fluorescence in situ hybridization. *Journal of Microbiological Methods*, *110*, 27–32. https://doi.org/10.1016/j.mimet.2015.01.005

Meyer, N. R., Fortney, J. L., & Dekas, A. E. (2021). NanoSIMS sample preparation decreases isotope enrichment: magnitude, variability and implications for single-cell rates of microbial activity. *Environmental Microbiology*, *23*(1), 81–98. https://doi.org/10.1111/1462-2920.15264

Moisander, P. H., Beinart, R. a, Hewson, I., White, A. E., Johnson, K. S., Carlson, C. a, … Zehr, J. P. (2010). Unicellular cyanobacterial distributions broaden the oceanic $N_2$ fixation domain. *Science (New York, N.Y.)*, *327*(5972), 1512–1514. https://doi.org/10.1126/science.1185468

Moisander, P. H., Beinart, R. a, Voss, M., & Zehr, J. P. (2008). Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *The ISME Journal*, *2*, 954–967. https://doi.org/10.1038/ismej.2008.84

Moisander, P. H., Benavides, M., Bonnet, S., & Berman-frank, I. (2017). Chasing after Non-cyanobacterial Nitrogen Fixation in Marine Pelagic Environments. *Frontiers in Microbiology*, *8*(September). https://doi.org/10.3389/fmicb.2017.01736

Moisander, P. H., Serros, T., Paerl, R. W., Beinart, R. a, & Zehr, J. P. (2014). Gammaproteobacterial diazotrophs and *nifH* gene expression in surface waters of the South Pacific Ocean. *The ISME Journal*, *8*(10), 1962–1973. https://doi.org/10.1038/ismej.2014.49

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

Montoya, J. P., Holl, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., & Capone, D. G. (2004). High rates of $N_2$ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature*, *430*(August), 1027–1031. https://doi.org/10.1038/nature02744.1.

Moraru, C., Lam, P., Fuchs, B. M., Kuypers, M. M. M., & Amann, R. (2010). GeneFISH- an in situ technique for linking gene presence and cell identity in environmental microorganisms. *Environmental Microbiology*, *12*(11), 3057–3073. https://doi.org/10.1111/j.1462-2920.2010.02281.x

Moraru, C., Moraru, G., Fuchs, B. M., & Amann, R. (2011). Concepts and software for a rational design of polynucleotide probes. *Environmental Microbiology Reports*, *3*(1), 69–78. https://doi.org/10.1111/j.1758-2229.2010.00189.x

Musat, N., Stryhanyuk, H., Bombach, P., Adrian, L., Audinot, J. N., & Richnow, H. H. (2014). The effect of FISH and CARD-FISH on the isotopic composition of $^{13}C$- and $^{15}N$-labeled *Pseudomonas putida* cells measured by nanoSIMS. *Systematic and Applied Microbiology*, *37*(4), 267–276. https://doi.org/10.1016/j.syapm.2014.02.002

Paerl, H. W., & Prufert, L. E. (1987). Oxygen-Poor Microzones as Potential Sites of Microbial $N_2$ Fixation in Nitrogen-Depleted Aerobic Marine Waters. *Applied and Environmental Microbiology*, *53*(5), 1078–1087. https://doi.org/10.1128/AEM.53.5.1087.1987

Paerl, R. W., Hansen, T. N. G., Henriksen, N. N. S. E., & Olesen, A. K. (2018). $N_2$ fixation and related $O_2$ constraints on model marine diazotroph *Pseudomonas stutzeri* BAL361. *Aquatic Microbial Ecology*, *81*, 125–136. https://doi.org/10.3354/ame01867

Perez-Sepulveda, B., Pitt, F., N'Guyen, A. N., Ratin, M., Garczarek, L., Millard, A., & Scanlan, D. J. (2018). Relative stability of ploidy in a marine *Synechococcus* across various growth conditions. *Environmental Microbiology Reports*, *10*(4), 428–432. https://doi.org/10.1111/1758-2229.12614

Pernthaler, A., Pernthaler, J., & Amann, R. (2004). Sensitive multi-color fluorescence in situ hybridization for the identification of environmental microorganisms. *Molecular Microbial Ecology Manual*, *3*(11), 711–726.

Pett-Ridge, J., & Weber, P. K. (2012). NanoSIP: NanoSIMS applications for microbial biology. *Methods in Molecular Biology*, *881*, 375–408. https://doi.org/10.1007/978-1-61779-827-6_13

Ploug, H. (2001). Small-scale oxygen fluxes and remineralization in sinking aggregates. *Limnology and Oceanography*, *46*(7), 1624–1631. https://doi.org/10.4319/lo.2001.46.7.1624

Poff, K. E., Leu, A. O., Eppley, J. M., Karl, D. M., & Delong, E. F. (2021). Microbial dynamics of elevated carbon flux in the open ocean's abyss. *PNAS*, *118*(4), 1–11. https://doi.org/10.1073/pnas.2018269118

Richards, P. M., & Mattes, T. E. (2021). Detection of an alkene monooxygenase in vinyl chloride-oxidizing bacteria with GeneFISH. *Journal of Microbiological Methods*, *181*. https://doi.org/10.1016/j.mimet.2021.106147

Riemann, L., Farnelid, H., & Steward, G. F. (2010). Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. *Aquatic Microbial Ecology*, (December). https://doi.org/10.3354/ame01431

Sargent, E. C., Hitchcock, A., Johansson, S. A., Langlois, R., Moore, C. M., LaRoche, J., … Bibby, T. S. (2016). Evidence for polyploidy in the globally important diazotroph Trichodesmium. *FEMS Microbiology Letters*, *363*(21), 1–

7. https://doi.org/10.1093/femsle/fnw244

Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., … Furuya, K. (2017). Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. *AGU Publications*, 996–1009. https://doi.org/10.1002/2017GB005681

Sohm, J. A., Webb, E. A., & Capone, D. G. (2011). Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, *9*(7), 499–508. https://doi.org/10.1038/nrmicro2594

Thompson, A. W., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F., & Zehr, J. P. (2014). Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environmental Microbiology*, *16*, doi:10.1111/1462-2920.12490. https://doi.org/10.1111/1462-2920.12490

Turk-Kubo, K. A., Farnelid, H., Shilova, I. N., Henke, B., & Zehr, J. P. (2017). Distinct ecological niches of marine symbiotic $N_2$-fixing cyanobacterium Candidatus Atelocyanobacterium thalassa sublineages. *Journal of Phycology*, *53*(2), 451–461. https://doi.org/10.1111/jpy.12505

Turk, K. a., Rees, A. P., Zehr, J. P., Pereira, N., Swift, P., Shelley, R., … Gilbert, J. (2011). Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic. *The ISME Journal*, *5*(7), 1201–1212. https://doi.org/10.1038/ismej.2010.205

Ueda, A., & Saneoka, H. (2015). Characterization of the Ability to Form Biofilms by Plant-Associated Pseudomonas Species. *Current Microbiology*, *70*(4), 506–513. https://doi.org/10.1007/s00284-014-0749-7

Woebken, D., Burow, L. C., Prufert-bebout, L., Bebout, B. M., Hoehler, T. M., Pett-ridge, J., … Singer, S. W. (2012). Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *The ISME Journal*, *6*, 1427–1439. https://doi.org/10.1038/ismej.2011.200

Zehr, J. P., & Kudela, R. M. (2011). Nitrogen Cycle of the Open Ocean : From Genes to Ecosystems. *Annual Review of Marine Science*. https://doi.org/10.1146/annurev-marine-120709-142819

Zehr, J. P., & Mcreynolds, L. A. (1989). Use of Degenerate Oligonucleotides for Amplification of the *nifH* Gene from the Marine Cyanobacterium *Trichodesmium thiebautii*. *Applied and Environmental Microbiology*, 2522–2526. https://doi.org/10.1128/AEM.55.10.2522-2526.1989

Zehr, J. P., Mellon, M. T., Zani, S., & York, N. (1998). New Nitrogen-Fixing Microorganisms Detected in Oligotrophic Oceans by Amplification of Nitrogenase (*nifH*) Genes. *Applied and Environmental Microbiology*, *64*(9), 3444–3450. https://doi.org/10.1128/AEM.64. 9.3444-3450.1998

**Figures**

**Table 1:** Gamma A sequences in the green rows are within the range of probe similarity verified with positive hybridization to the probe, denoted in the Hybridization column as yes. Grey rows are unlikely to hybridize based on dissimilarity value tested in the form of UCYN-A (78% similarity) controls (Hybridization = no). The decreasing similarity of white rows represents a spectrum non-Gamma A cell with possible hybridizations (Hybridization = unknown) if the organisms co-occur in samples.

| OTU group | Accession # | % Similarity to probe | Reference | Hybridization |
|---|---|---|---|---|
| Gamma A | AY896371.1 | 99.7 | Langlois et al., 2015 | yes |
| Gamma A | AF059623.1 | 99.1 | Zehr et al., 1998 | yes |
| Gamma A | AY706889.1 | 99.4 | Church et al., 2005 | yes |
| Gamma A | EU052413.1 | 98.1 | Moisander et al., 2008 | yes |
| Gamma B | HQ611810.1 | 97.5 | Turk-Kubo et al., 2011 | unknown |
| Gamma C | KF151762.1 | 91.9 | Turk-Kubo et al., 2014 | unknown |
| Gamma D | HQ456037.1 | 87.5 | Kong et al., 2011 | unknown |
| Gamma E | AY896456.1 | 84.1 | Langlois et al., 2005 | unknown |
| Gamma 3 | HM210397.1 | 83.3 | Halm et al., 2012 | unknown |
| Gamma ETSP2 | KF151819.1 | 81.9 | Turk-Kubo et al., 2014 | unknown |
| Gamma ETSP1 | KF151567.1 | 78.8 | Turk-Kubo et al., 2014 | no |
| Gamma ETSP3 | KF151661.1 | 77.1 | Turk-Kubo et al., 2014 | no |
| Gamma 4 | HM210363.1 | 76.8 | Halm et al., 2012 | no |
| Gamma 1 | HM210377.1 | 74.8 | Halm et al., 2012 | no |

**Figure 1**: UCYN-A2 hybridized with CARD-FISH (A) and geneFISH (B-D). A) CARD-FISH image showing UCYN-A2 attached to haptophyte host B. Bigelowii, red is the UCYN-A rRNA probe, green is the rRNA probe for the host, and blue is DNA staining with DAPI. B-D) GeneFISH images of UCYN-A2 symbiosis, red represents positive UCYN-A *nifH* gene signal within the cyanobacterial portion of the symbiosis. Green is the CARD-FISH signal targeting the host cell. The orange/red coloring is the host chloroplast. The scale bar is 10 µm

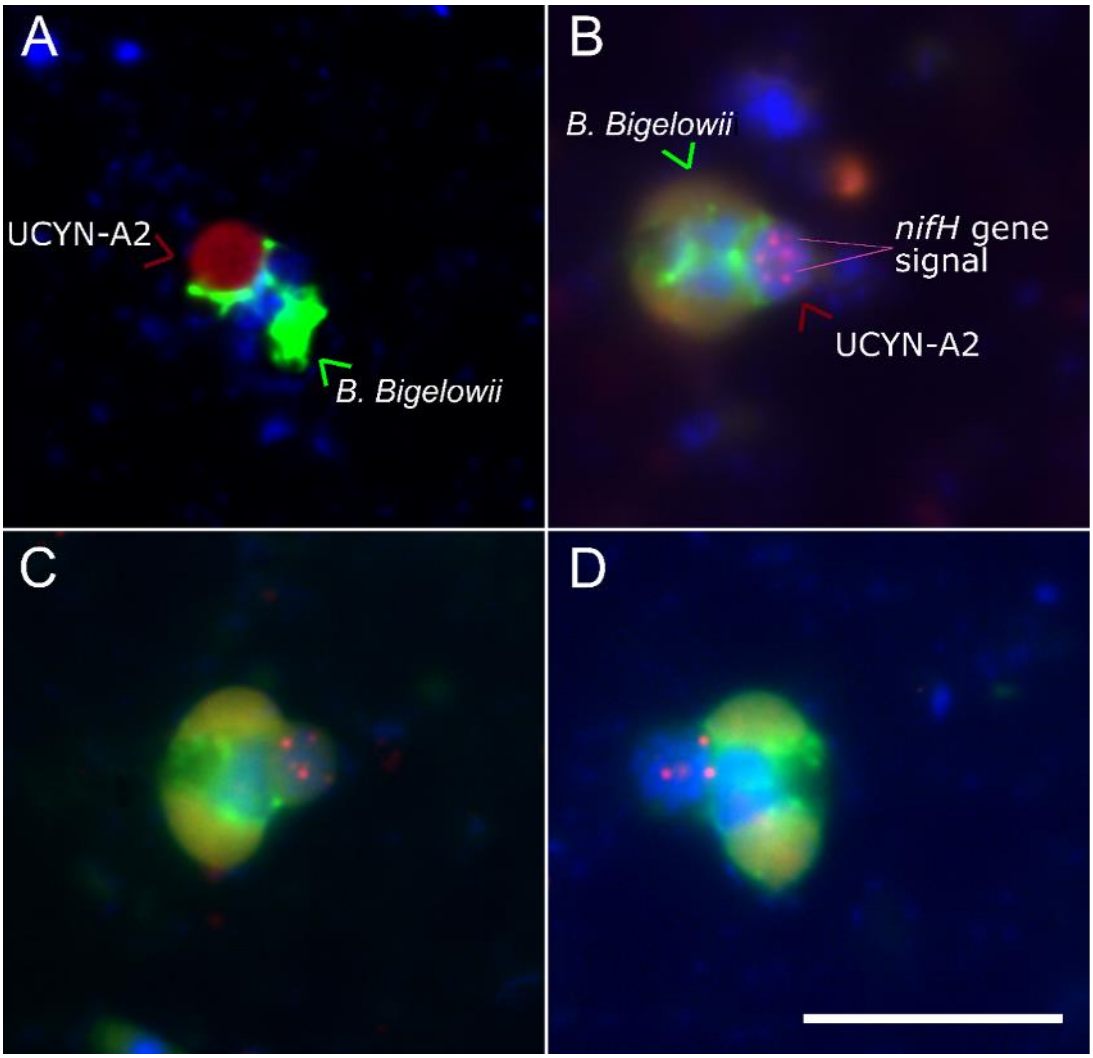


**Figure 2:** Gamma A geneFISH hybridization with E. coli clone controls. A) Gamma A nifH clones B) UCYN-A nifH clones. The red color shows hybridization to the nifH geneFISH probe, while the green is hybridization to the 16S rRNA general bacterial probe (EUBI-III). A) There is a positive gene signal in nearly all (97%) of GammaA nifH clones. B) UCYN-A nifH clones (negative controls) showed minimal hybridization (4.8%). The red arrow indicates a false positive signal, while the grey arrows indicate nonspecific background signals that do not overlap with cells. The scale bar is 10 µm.

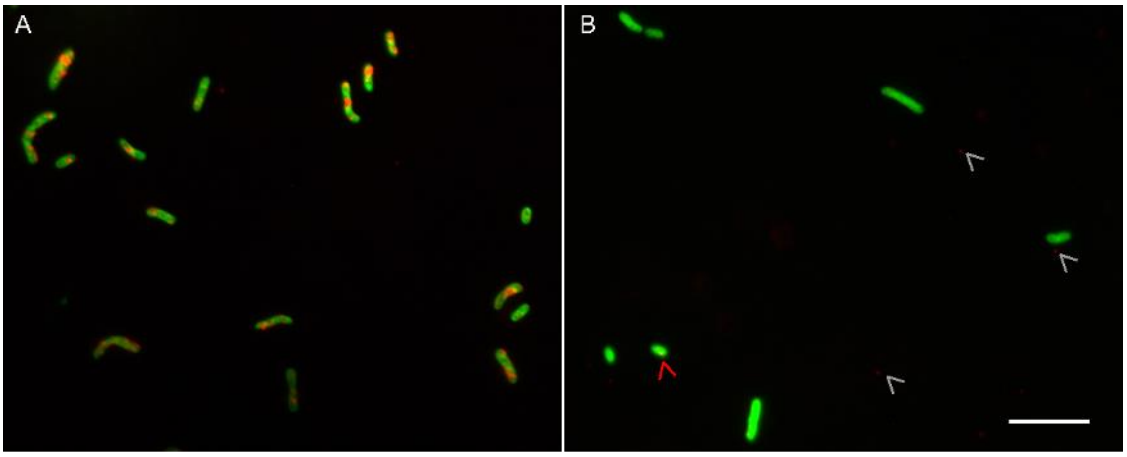


**Figure 3:** GeneFISH identified Gamma A cells attached to particles. Gamma A cells were found associated with particles in 61% of observations. Particles with an attached Gamma A had various densities and sizes. Red shows the *nifH* geneFISH hybridization signal, bright green shows 16S rRNA Gammaproteobacterial hybridization signal, while light green is unspecific fluorophore sticking to particle components. The scale bar is 10 µm.

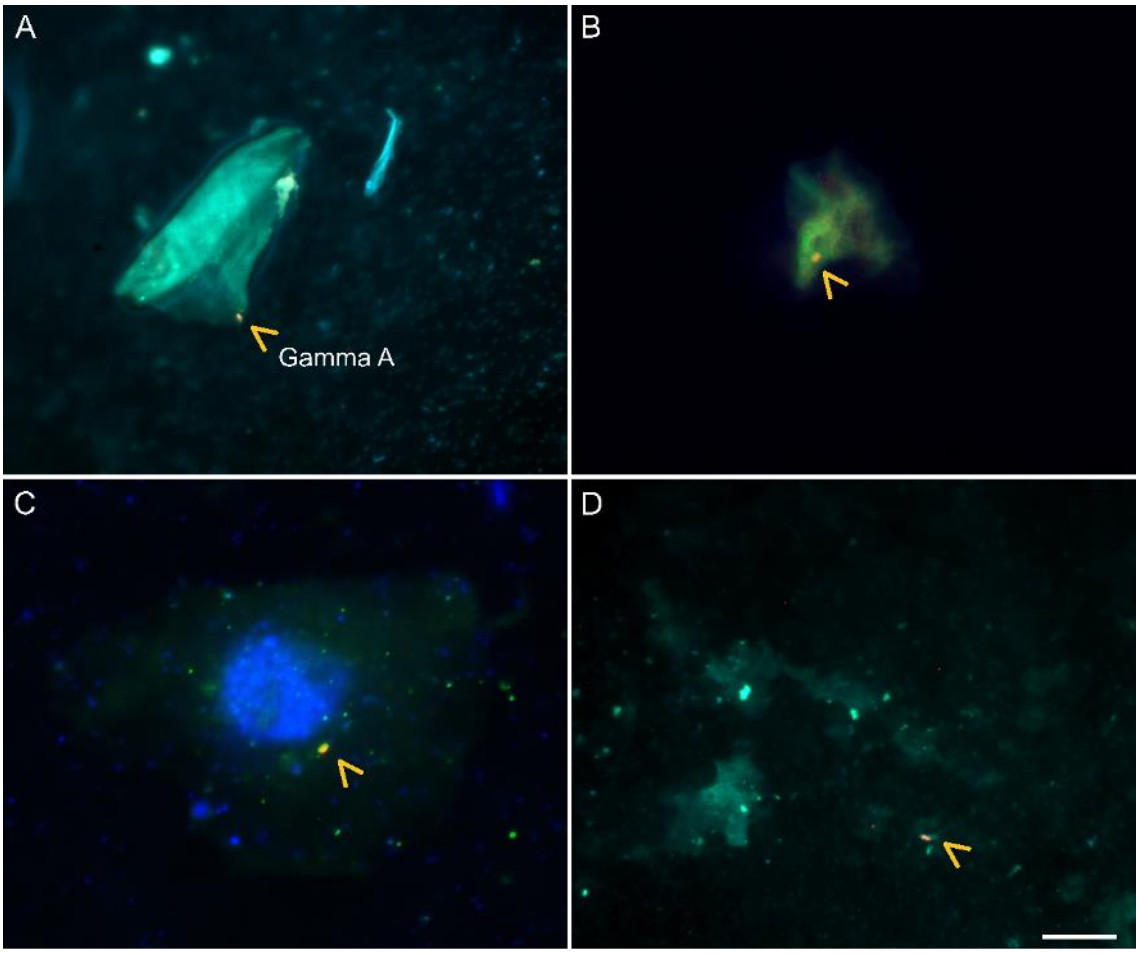
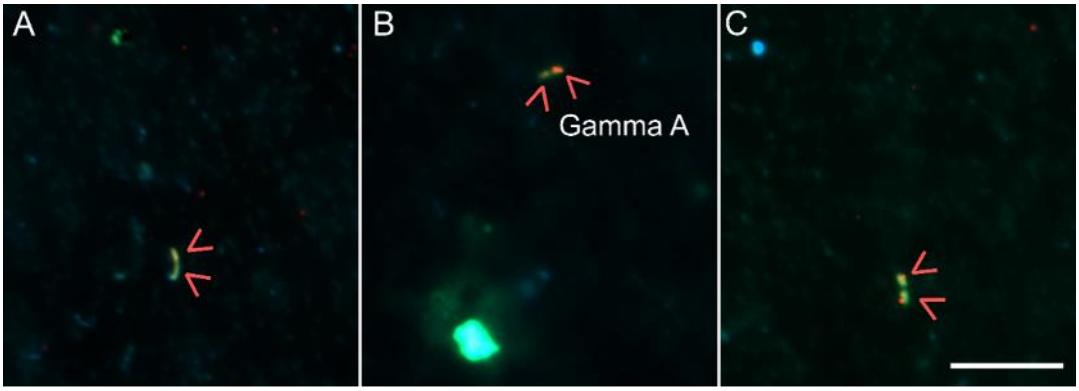


**Figure 4:** Examples of pairs of non-particle attached Gamma A cells. Most free Gamma A cells were found as pairs of cells attached end to end (4A-C). Red shows hybridization to the *nifH* geneFISH probe, while green shows hybridization to the 16S rRNA Gammaproteobacterial probe (Gam42a). The orange color is due to the overlap of red and green signals. The scale bar is 10 µm.



## Chapter 5: Conclusions and future directions

BNF fixation is an important source of N input in the oligotrophic waters, fueling primary productivity and enabling C sequestration to the deep sea. A better understanding of the organisms involved in BNF and their distribution and lifestyle are fundamental components for estimating primary productivity and parameterizing predictive future global ocean models. This dissertation summarizes investigations of uncultured unicellular diazotrophs at the single-cell level, thus improving our understanding of their roles and potential as sources of fixed N globally. The results in this dissertation expand the known temperature bounds of which UCYN-A is capable of fixing $N_2$, present the first direct confirmation that NCD can fix $N_2$ in the fully oxygenated surface ocean, and reveal the morphology and particle-attached lifestyle of the well-studied NCD, Gamma A. These results add to the growing body of research that is changing the idea that marine $N_2$ fixation is limited to cyanobacteria in the warm oligotrophic ocean.

The results of chapter 2 show that the commonly occurring subtropical cyanobacterial diazotroph, UCYN-A are present in the Bering, Chukchi, and Beaufort Seas and are capable of $N_2$ fixation in these low temperatures environments. Quantification and sequencing of the *nifH* gene show that two sublineages, UCYN-A1 and UCYN-A2, are present in polar waters. The polar UCYN-A1 and UCYN-A2 sequence types are the same sequence types that are found in temperate waters. These results are evidence that UCYN-A has a much broader temperature range than most other marine cyanobacterial diazotrophs. This study also employed nanoSIMS

allowing direct measurements of cell-specific $N_2$ fixation rates in UCYN-A, which showed both sublineages fix $N_2$ at temperatures down to 10$^o$C and UCYN-A2 down to 4$^o$C. Identification of N2-fixing UCYN-A in the Arctic provides insight into the unique physiology of UCYN-A and sheds light on an unexpected N source in the nutrient-depleted Arctic.

$N_2$ fixation in cold marine waters is a relatively new idea, and the extent and constraints are not well defined. Future studies that focus on the seasonality of $N_2$ fixation and factors that control diazotroph distribution would be valuable in understanding BNF in polar regions. Additionally, investigations into whether UCYN-A is transported in the Arctic or is present year-round could provide insights into the physiology of the cyanobacteria and haptophyte host. UCYN-A is a widespread and significant contributor to marine BNF; therefore, determining the physiology and controlling factors of UCYN-A would provide valuable information in estimating and constraining global BNF.

The data presented in chapter 3 demonstrate the first direct evidence of $N_2$ fixation by non-cyanobacterial diazotrophs in the open ocean. Dual-labeled isotope incubations allowed nanoSIMS analysis to distinguish between cyanobacterial diazotrophs enriched in $^{15}$N and $^{13}$C and non-cyanobacterial diazotrophs only enriched in $^{15}$N. The lack of $^{13}$C enrichment shows the cells are capable of heterotrophically (or photoheterotrophically) fueled $N_2$ fixation in the oxygenated surface ocean. Using an un-targeted nanoSIMS approach allowed detection of all possible NCD regardless of phylogenetic affiliation, all of which were found attached

to particles. The results of this study show that particle-attached NCDs are capable of fixing N in the surface ocean, contributing to the BNF in that area.

Chapter 3 results show the importance of particles as loci for $N_2$ fixation, yet the microscale conditions on the particle that facilitate NCD $N_2$ fixation is unknown. As many studies speculate $O_2$ acts as a controlling factor in NCD BNF, future studies that address $O_2$ concentrations within particles would be particularly informative. Additionally, phylogenetic identification of the NCD responsible for particle attached BNF would provide a means to study their distribution and BNF potential over larger areas without relying on the time and cost-intensive nanoSIMS method.

The work in chapter 4 shows geneFISH is a viable method to study uncultured marine diazotrophs when only the *nifH* gene is known. The goal of geneFISH was to identify Gamma A, which was done successfully in samples from the Western North Pacific Subtropical Gyre. Gamma A was found to be a rod-shaped cell primarily attached to particles. Direct visualization of Gamma A allowed us to address several hypotheses about its lifestyle and confirm that the majority of Gamma A cells observed were attached to particles of various sizes and densities. Although cell-specific $N_2$ fixation rates were inconclusive, the presence of Gamma A on particles highlights the importance of particles in NCD $N_2$ fixation in the surface ocean.

The design, optimization, and successful visualization of geneFISH on Gamma A will allow future studies to study Gamma A specifically. For example, geneFISH identification of Gamma A for cell-specific $N_2$ fixation rates would be valuable in determining their contribution to BNF. However, future studies should

use high enrichment of $^{15}N$ in initial incubations to account for isotope dilution through the geneFISH protocol. Additionally, geneFISH identified cells could possibly be detached from particles and sorted for sequencing of the 16S rRNA gene. The 16S rRNA gene would allow for better phylogenetic identification and the potential to design CARD-FISH probes for a simplified visualization protocol and more reliable cell-specific $N_2$ fixation rate measurements.

# Appendix 1

**Supplementary Information for Chapter 2:** Symbiotic Unicellular Cyanobacteria Fix Nitrogen in the Arctic Ocean

**Supplementary Methods**

*Sample collection for DNA extractions*

Samples were taken in the Bering, Chukchi, and Beaufort Seas in September 2016 (Fig. S1). Water samples were collected from Niskin bottles attached to a CTD into clean polycarbonate bottles (washed 3x 10% HCl and 3x milliQ water) after 3x seawater sample rinses. Samples (2-4L) were filtered by a peristaltic pump onto sequential 3 and 0.2 µm polyphenylene ether (PPE) filters (0.2 µm 25mm Supor-200, Pall Life Sciences, Port Washington, NY, USA) in Swinnex filter holders. Filters were transferred into sterile bead beater polypropylene microcentrifuge tubes containing 0.5 and 1 mm glass beads (BioSpec) and immediately frozen in liquid nitrogen. Surface and the depth corresponding to chlorophyll maximum (DCM) were sampled at most stations, and a near bottom depth was taken if the station depth was more than ~20m deeper than the chlorophyll maximum. Only surface samples were collected at Station 1 (off the coast of Nome, AK). Replicate samples were collected for DNA at each depth and station. Samples for molecular analysis were stored at -80$^{o}$C until processing.

*DNA extraction and nifH amplification*

DNA was extracted using a modified DNeasy Plant Mini Kit (Qiagen, Germantown, MD) protocol, described in detail in (1) that includes steps for disruption of cell walls and automation of wash steps. The *nifH* gene, a functional marker for nitrogenase enzyme in diazotrophs, was PCR-amplified in duplicate with

one positive and at least two negative (water) PCR controls, using a degenerate universal *nifH* primers *nifH1*/*nifH2* (Yanni/450) and *nifH3/nifH4* (up/down) in a nested reaction (2). Briefly, the first round of *nifH* amplification consisted of 1x buffer, 4 mM $MgCl_2$, 200 µM dNTP mix (Invitrogen, Carlsbad, CA), 0.5 µM *nifH1*/*nifH2* primers (Sigma-Aldrich Oligos, St. Louis, MO), 1 µl Platinum Taq polymerase (Invitrogen), 1 µl DNA template and Ambion RT-PCR grade water (Invitrogen) for a total reaction volume of 15 µl. Thermocycler conditions were as follows: initial denaturation at 95ºC for 3 min., followed by 25 cycles of 95ºC for 30 s 55ºC for 30 s 72ºC for 45 s and final elongation at 72ºC for 7 min. The second round of *nifH* amplification used the *nifH3/nifH4* primers, each modified with common sequence (CS) linkers (3) using the same reaction conditions but with a final volume of 25 µl, and an additional 5x amplification cycles. Amplified products were checked with agarose gel electrophoresis, and replicates from samples with positive amplifications were pooled at equal volumes. Library preparation was carried out by the DNA Sequencing Core Facility at the University of Illinois at Chicago (http://rrc.uic.edu/cores/genome-research/sequencing-core/) and included an additional ten rounds of amplification to add multiplexing oligonucleotides and sequencing adaptors according to protocols detailed in (4). Amplicons were sequenced using Illumina MiSeq, to a sequencing depth of 40,000 sequences per sample.

*nifH amplicon sequence analysis and UCYN-A oligotyping*

Raw paired-end *nifH* amplicon reads (2 x 250 bp) were merged using Paired-End read (PEAR) merging software (5) with a Phred score of 20, a minimum length

allowance of 300 bp, the maximum length of 400 bp. Merged files were imported into QIIME (6), where chimeric sequences were removed using a *denovo* approach using UCHIME (7), and the remaining sequences were clustered into OTUs with 97% similarity using usearch6.1 (8). OTUs with less than ten sequences were removed. Representative sequences were imported into a curated *nifH* database (9) in Arb (10), where they were translated, non-*nifH* and sequences with stop codons were removed, then remaining sequences were aligned using Pfam-curated multiple alignments of the *NifH*/frxC family (Fer4_*NifH*) in HMMer (11). The nearest closest cultivated relative (based on nucleotide sequences) and *nifH* cluster were determined using blastx against a curated list of genome-derived *nifH* sequences.

UCYN-A OTUs were identified in the blastx analysis described above, and unique UCYN-A sequences were recovered after clustering all UCYN-A sequences at 100% nucleotide similarity using usearch6.1 (8). UCYN-A sequences were aligned against a reference UCYN-A database using pynast (12) and trimmed in galaxy to a length consistent with a global database of UCYN-A sequences (13). UCYN-A oligotypes – highly refined taxonomic units defined using nucleotide positions with high "entropy"– were defined using the alignment positions identified by Turk-Kubo et al. (13) and the oligotyping pipeline developed by Eren et al. (14). The following arguments were used in the oligotyping analysis: (i) a given oligotype was allowed to be present in only one sample (-s 1); (ii) a given oligotype was required to be present at a relative abundance of at least 0.1% in one sample (-a 0.1); and (iii) the most

abundant unique sequence defining an oligotype was required to have a sequence count > 100 across the whole data set (-M 100).

*Estimating UCYN-A abundances using qPCR*

UCYN-A1 and UCYN-A2 abundances were estimated using TaqMan® quantitative PCR (qPCR) chemistry and primers and probes specific for UCYN-A1 (15) and UCYN-A2 (16) and their respective haptophyte partners, UCYN-A1 host (UCYN-A1-host forward - 5’ - AGGTTTGCCGGTCTGCCGAT- 3'; UCYN-A1-host reverse - 5’ - ATCCGTCTCCGACACCCGCTC - 3'; UCYN-A1-host probe - 5’ - [6FAM]CTGGTAGAACTGTCCTTCC[TAMRA] - 3') and UCYN-A2 host (16), in samples positive for *nifH* amplification. All aspects of qPCR reaction conditions, thermocycling parameters, inhibition tests, and standard generation and calculations to determine abundances are described in detail in (17), with the exception of a 64$^{o}$C annealing temperature for the UCYN-A2 assay. The limit of detection (LOD) and limit of quantification (LOQ) were 0.5 *nifH* gene copies/µL DNA extraction and 4 *nifH* gene copies/µL DNA extraction, respectively. DNQs (detected, not quantified) were abundances that fell in between this range. When samples were below the LOQ (DNQs) but were above the LOD, they were given the minimum LOQ value, and values at the LOD were given the minimum LOD value for visualization purposes. Abundances were quantified separately for the 3.0 and 0.3 µm size fractions, and then averages were added to estimate total UCYN-A abundance in the sample.

*CARD-FISH sample preparation and hybridizations*

After 24-hour $^{15}N_2$ incubations, 100 ml of labeled sample was fixed with a final concentration of 1.8% formaldehyde (v/v) up to 48 hours at 4$^{o}$C to preserve cells

for CARD-FISH and subsequent nanoSIMS. Fixed samples were gently filtered onto 25mm, 0.6 µm pore polycarbonate filters (Millipore Isopore, EMD Millipore, Billerica, MA, USA) with 25mm, 0.8 µm pore polycarbonate cellulose acetate support filter (Sterlitech Corp. Kent, WA, USA) using low vacuum (<100 mm Hg). Filters were stored in 2 ml cryovials (Nalgene, Millipore, Sigma, Burlington, MA), flash-frozen in liquid nitrogen, and kept at -80$^{o}$C until analysis. To visualize both strains and their respective hosts (UCYN-A1/UCYN-A1 host and UCYN-A2/UCYN-A2 host), a double CARD-FISH protocol was used according to the protocols detailed in (18, 19). The full suite of HRP probes, competitor oligonucleotides, and helper probes are given in Table S1.

*Measuring cell-specific $N_2$ fixation rates using nanoSIMS*

Pieces of CARD-FISH were washed to remove mounting media according to (20). Briefly, filters containing targets hybridized using CARD-FISH as described above were transferred to silicon gridded wafers (Pelcotec Silicon SFG12, Ted Pella Inc., Redding, CA) by dampening filter pieces with 10 µl water, placing them face down on silicon wafer and freezing at -80$^{o}$C for 5 min, then gently removing the filter while still frozen. In order to locate targets for nanoSIMS analysis, gridded wafers were mapped using bright field and epifluorescence on an AxioVersion Epifluorescence Microscope. NanoSIMS analyses were performed at Stanford Nano Shared Facilities (SNSF; https://snsf.stanford.edu) on a Cameca NanoSIMS 50L at Stanford University, CA, USA**.** Once a target cell was located, the cell was exposed to 2-3 min of large diameter, $Cs^+$ beam to saturate cell ions with $Cs^+$ before analysis. Cells were rastered with a 16 keV $Cs^+$ beam and a current between 2 and 4 pA over a

10 - 15 μm$^2$ area with an image size of 256 x 256 pixels and a dwell time of ~1 ms per pixel. Negatively charged secondary ions of carbon ($^{12}C^-$, $^{13}C^-$), nitrogen (as $^{12}C^{14}N^-$, $^{12}C^{15}N^-$), and a secondary electron microscope image (1AU) were collected simultaneously for 30-45 frames for each individual cell. Data were imaged and analyzed using the look@nanoSIMS image analysis software (21).

Cell-Specific nitrogen fixation rates were determined by calculating the carbon content per cell based on a spherical cell volume (V) from the measured cell diameter determined by the ROI. Carbon content [c] was estimated using two equations for the different types of cells, UcynA_log[c]= -0.363+(0.863 x log(V)), haptophyte_log[c]= -0.422+(0.758 x log(V)), following the example of (*27*). The C:N ratio of 6.3 was measured in UCYN-A from the tropical North Atlantic (*28*) and was used in our calculation to estimate the N content of the cell. NanoSIMS measurements of $^{15}N$:$^{14}N$ correspond to the atom% enrichment within the cell. To determine true ion counts versus ratio noise, the measured image was split into 100 equally sized squares and the ion counts per pixel evaluated. Squares within the ROI had significantly higher counts per pixel compared to those outside the ROI. Imaged cell only includes ratio values inside the ROI as any ratio values occurring outside the cell were determined to be Poisson error due to low counts. After subtracting the natural abundance of $^{15}N$ occurring at each station and factoring in the atom percent $^{15}N_2$ labeling in the enrichment (MIMS data), cell-specific nitrogen fixation rates were calculated (fmol N cell$^{-1}$ day$^{-1}$). The limit of detection was determined to be three times the standard deviation of unenriched samples (0.02 At%), similar to LOD

determination in Jayakumar *et al.* (22). In other words, if the atom percent excess (At% of a cell minus the naturally occurring $^{15}N$ from bulk particulate nitrogen) was greater than three times the standard deviation of an average of unenriched samples (n=32), then samples were considered enriched above detection, and cell-specific rates were calculated. Due to limited Arctic samples, unenriched UCYN-A/haptophyte symbioses samples from subtropical latitudes were used to determine the limit of detection and were measured on the same nanoSIMS instrument. Cell-specific rates are based on the two volumetric spheres of the haptophyte and cyanobacteria together. Using qPCR abundance data, the cell-specific $N_2$-fixation rates were extrapolated to estimate the total volumetric $N_2$-fixation by UCYN-A (nmol N $L^{-1}$ $day^{-1}$) at a given station.

**Supplementary Text**

*Chukchi and Beaufort Sea UCYN-A oligotypes*

Oligotyping enables fine resolution of closely related phylotypes based on nucleic acid positions with high entropy (14). This approach has been applied to UCYN-A diversity in several recent studies (13, 23), and there are now over 100 UCYN-A oligotypes defined across multiple studies. A total of 84,156 UCYN-A partial *nifH* sequences were recovered from 29 Arctic samples, and the analysis identified only two oligotypes, which represented 96.71% of the sequences submitted for analysis, with a perfect purity score of 1.00. The UCYN-A oligotypes found in the Arctic were oligo3 (ATTCTATTTTCTT), which is the dominant oligotype from the UCYN-A2 strain, and oligo1 (ATCTCGCTTCTTT), which is the dominant oligotype from the UCYN-A1 strain (Fig. 1, Fig. S2; Turk-Kubo et al, 2016). Oligo1 is widely

distributed throughout the subtropical oceans, while oligo3 is found at high relative abundances in high latitude samples from the Danish Strait but does not appear to be a dominant strain in the subtropical oceans (13). In high latitude waters, oligo1 was found at high relative abundances in the Bering Sea (Off Nome, Station 1) and, interestingly, in the proximity of ice algae (Station 14). It was detected, however, in 10 of the 29 samples. In general, oligo3 (UCYN-A2) was more widely distributed, being detected in 23 of the 29 samples.

*UCYN-A abundances and distributions*

The UCYN-A1 symbiont was quantifiable at a single station (Station 1; Fig. S3, Table S2) in surface waters with an abundance of $1.6x10^6$ *nifH* copies $L^{-1}$, and although it was detected in samples in the Chukchi and Beaufort Seas, it was not above the limit of quantification. The UCYN-A2 symbiont was detected at 10 surface stations with an average of $2.4x10^3$ *nifH* copies $L^{-1}$ (n=41) (Fig. S3, Table S2), and was below quantification in the DCM and deep samples. Interestingly, UCYN-A was detected throughout the water column as a combination of UCYN-A1 and UCYN-A2 at a single station (Station 14) which was in proximity to ice-containing, ice algae. Higher abundances of UCYN-A1 were found in the 0.2 µm size fraction (average of $3.9x10^4$ *nifH* copies $L^{-1}$, n=21), although it was also detected in the 3 µm size fraction (average of $2.0x10^4$ *nifH* copies $L^{-1}$, n=21). The majority of UCYN-A2 occurred in the larger size fraction (> 3 µm) (average of $4.2x10^3$ *nifH* copies $L^{-1}$, n=21), although it could also be quantified in the 0.2 µm size fraction (average of $5.1x10^2$ *nifH* copies $L^{-1}$, n=21). In general, the distributions and abundances of UCYN-A1 and UCYN-A2 are consistent with recent findings of Shiozaki et al., which reported similar

abundances of both UCYN-A1 and UCYN-A2 in the Bering Sea in August 2014 (24), and the predominance of low abundances of UCYN-A2 throughout the Chukchi and Beaufort Seas in September-October 2015 (25).

The haptophytes associated with both UCYN-A strains were also quantified using qPCR. The UCYN-A1 host was not detected at any station or depth, but there is some evidence that the qPCR assay used in this study may not target a well-conserved region of the 18S rRNA gene (26). Further research is required to verify the 18S rRNA gene sequence of the UCYN-A1 host, but we can be confident that it is associated with a haptophyte that is very closely related to the host identified based on the results from CARD-FISH analyses. The UCYN-A2 host was detected in 5 surface samples with an average of $1.1x10^4$ *nifH* copies $L^{-1}$ (n=21). Consistent with reports from the western tropical South Pacific (26), the ratio of symbionts:host based on *nifH*:18S rRNA gene counts were greater than 1; the symbionts were detected in more stations than their respective hosts.

*UCYN-A found primarily in late-season meltwater*

Temperatures ranged from -1.6$^o$ C to 10.2$^o$C, salinity ranged from 26.5 to 33.5 for the whole water column, and nitrate was below detection in all surface samples but one. Ice floes were present at many of the sampling locations, mostly on the Chukchi Shelf, which created a low-density freshwater layer on the surface of the water (<1 m). The depth of the chlorophyll maximum ranged from 17 to 38 m with an average of 28.4 m, but at certain stations disappeared entirely due to full mixing of the water column. UCYN-A was found, either in *nifH* libraries or via qPCR (or both), in samples that span the entire range of temperature and salinity. However, UCYN-A

was primarily detected in surface water samples. No linear correlations were found between UCYN-A abundances and environmental factors presented. However, it is clear that UCYN-A was often found in late-season meltwater (27), based on temperature and salinity plots (Fig. S4). It must be noted, however, that this may be a result of sampling bias, given that the majority of surface samples occurred in this water mass. The highest abundances of UCYN-A were found in the only sample in the Alaskan Coastal Current, which has been proposed to be the current by which subtropical UCYN-A populations are advected from the North Pacific to the Bering and Chukchi Seas (25).

*Verification of active $N_2$ fixation in UCYN-A using CARD-FISH/nanoSIMS*

NanoSIMS analysis allows the direct measurement of $^{15}N$ uptake by individual UCYN-A cells. This unique method of measuring $N_2$ fixation rates enables making a direct link between the amount of $^{15}N$ uptake and the identity of the cell responsible for that uptake. Sequencing or qPCR data can be used to find the most probable diazotroph(s) responsible for observed rates based on abundances or transcripts; however, the presence or transcription of *nifH* alone does not verify that a given cell is fixing $N_2$.

We used nanometer scale secondary ion mass spectrometry (nanoSIMS) to measure $^{15}N$ uptake into individual cells to quantify nitrogen fixation rates. Isotopic enrichment of $^{15}N$ in the host or UCYN-A cells is attributed to $N_2$ fixation by UCYN-A, as the host lacks the ability to fix $N_2$, but the transfer of UCYN-A derived fixed N to the host cell, and host-derived fixed C to the UCYN-A cell has been well

established (28-30). Eight individual UCYN-A1/haptophyte symbioses from the Bering Sea (station 1) were measured, and symbioses had an average cell-specific nitrogen fixation rate of 7.6 ± 14.5 fmol N $cell^{-1}d^{-1}$. A total of six UCYN-A2/haptophyte symbioses from the Bering Sea were analyzed with an average cell-specific rate of 13.0 ± 7.7 fmol N $cell^{-1}d^{-1}$. A total of six UCYN-A2/haptophyte symbioses from the Chukchi Sea (station 2) were analyzed, two were above the limit of detection and resulted in an average cell-specific nitrogen fixation rate of 1.1 ± 1.95 fmol N $cell^{-1}d^{-1}$ (summarized in Table S3).

The UCYN-A1 cell-specific rates found here are similar to cell-specific rates found in the tropical north Atlantic of 0.45-12 fmol N $cell^{-1}$ $d^{-1}$ (29, 30). To our knowledge, these are the first reported cell-specific rates for the UCYN-A2/haptophyte symbioses. Rates for UCYN-A/haptophyte symbioses that were larger than the UCYN-A1/haptophyte symbioses were reported (31), but these are likely to be a third strain, UCYN-A3, based on their size class (intermediate between UCYN-A1 and UCYN-A2 symbioses) and the emerging biogeography of the UCYN-A2 symbioses (13).

Based on UCYN-A abundances, cell-specific $N_2$ rates were extrapolated to volumetric $N_2$ fixation rates (Table S3). UCYN-A1 in the Bering Sea (station 1) could be responsible for fixing up to 9.3 ± 17.6 nmol N $l^{-1}$ $d^{-1}$ in the surface water. UCYN-A2 abundances at the same station in the Bering Sea (station 1) could result in 1.2 ± 0.7 nmol N $L^{-1}$ $d^{-1}$. The bulk nitrogen fixation rate in the surface waters of the Bering Sea was measured at 6.8 ± 3.8 nmol N $L^{-1}$ $d^{-1}$. Extrapolated UCYN-A rates

from nanoSIMS data can account for up to 100% of particulate bulk nitrogen fixation rates in the Bering Sea. UCYN-A2 abundances in the Chukchi Sea result in volumetric $N_2$-fixation rates of up to 0.004 $\pm$ 0.007 nmol N $L^{-1}$ $d^{-1}$. Comparing these values to the bulk particulate $N_2$ fixation at this station (0.2 $\pm$ 0.2 nmol N $L^{-1}$ $d^{-1}$, DNQ), other $N_2$-fixing organisms could be responsible for the average bulk $N_2$-fixation measured at this location. Interestingly, UCYN-A2 *nifH* sequences at this station (station 2) accounted for almost half (46%, Table S2) of the total *nifH* sequences recovered, and from sequence data alone, UCYN-A could have been presumed as (one of) the primary contributors to $N_2$ fixation, showing the importance of cell-specific analysis.

**Supplemental Figures S1 - S4**

**Figure S1.** Station Map. Stations from the cruise in 2016 stations are organized from the Bering Sea to the Chukchi Sea and the Beaufort Sea, from nearshore to offshore.

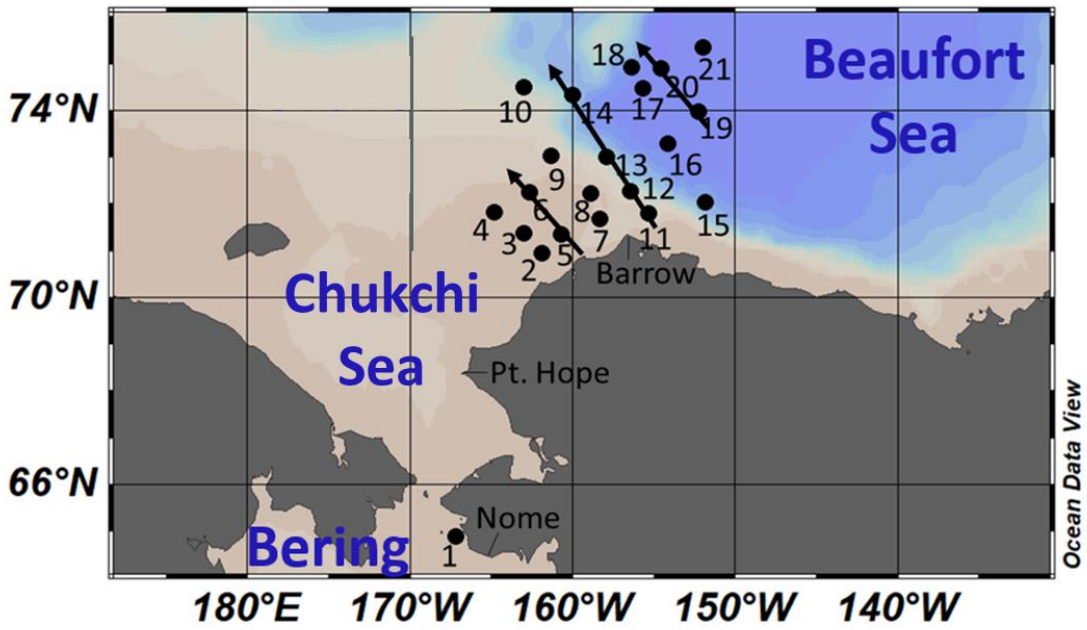


**Figure S2.** UCYN-A *nifH* oligotype distributions. Stations are numbered according to the above station map. The dominant UCYN-A1 (oligo1) is shown in blue, and the dominant UCYN-A2 (oligo3) is shown in black.

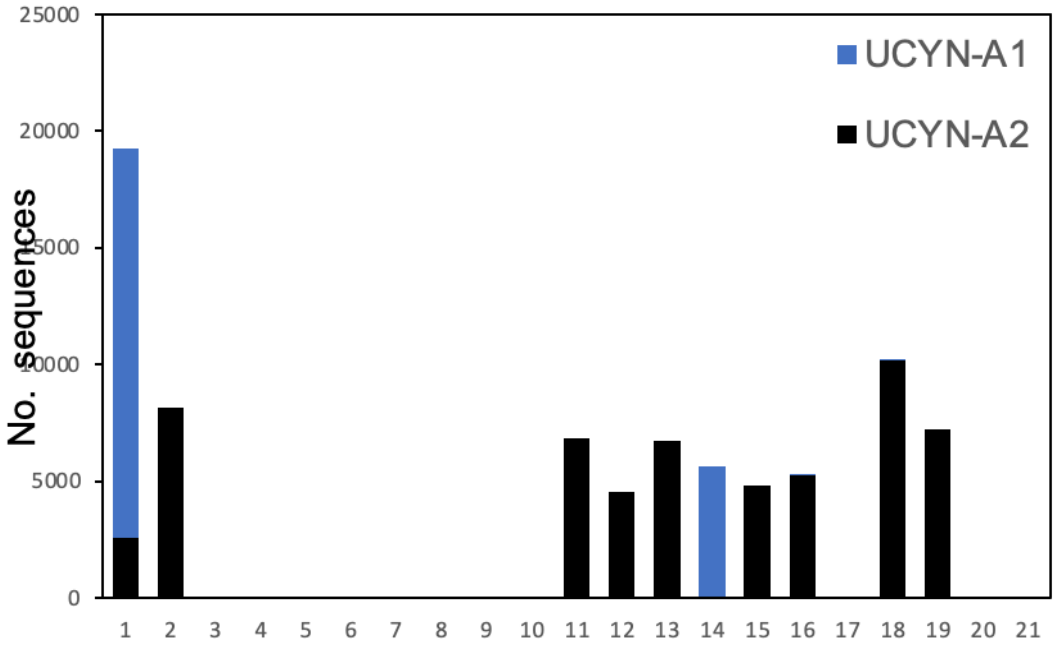


**Figure S3.** UCYN-A qPCR data. UCYN-A strain abundances and distribution in surface samples by qPCR. UCYN-A1 (blue) and UCYN-A2 (black) are present in the Bering Sea, Chukchi Sea, on the Chukchi Shelf, and the Beaufort Sea. Error bars represent the standard deviation of technical replicates. Bars below the dashed line are samples below the limit of quantification, given the minimum value for visualization purposes.

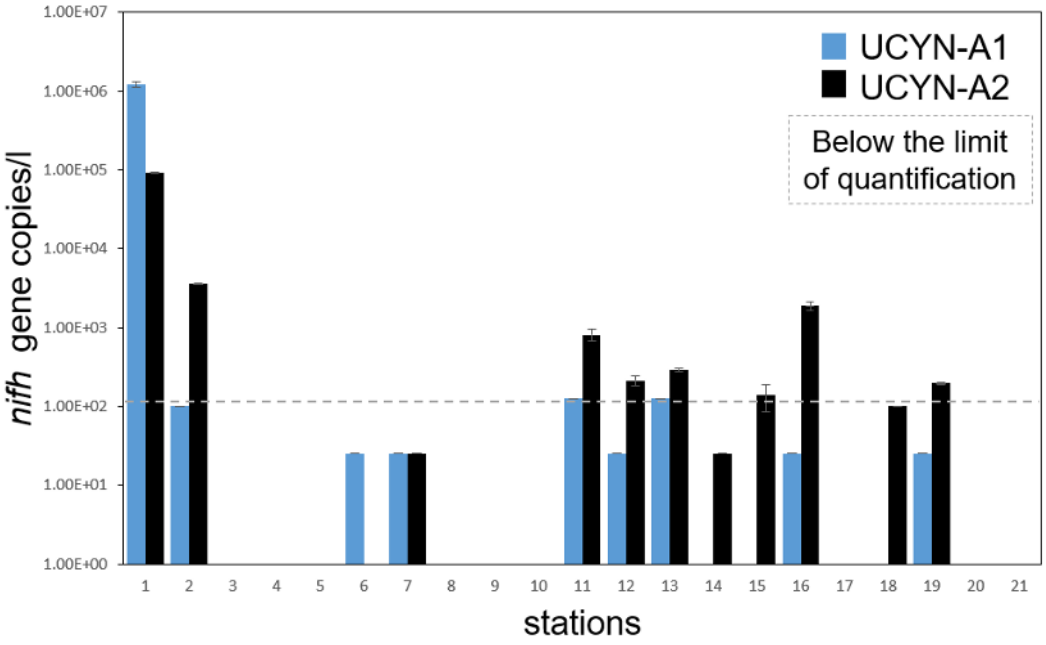
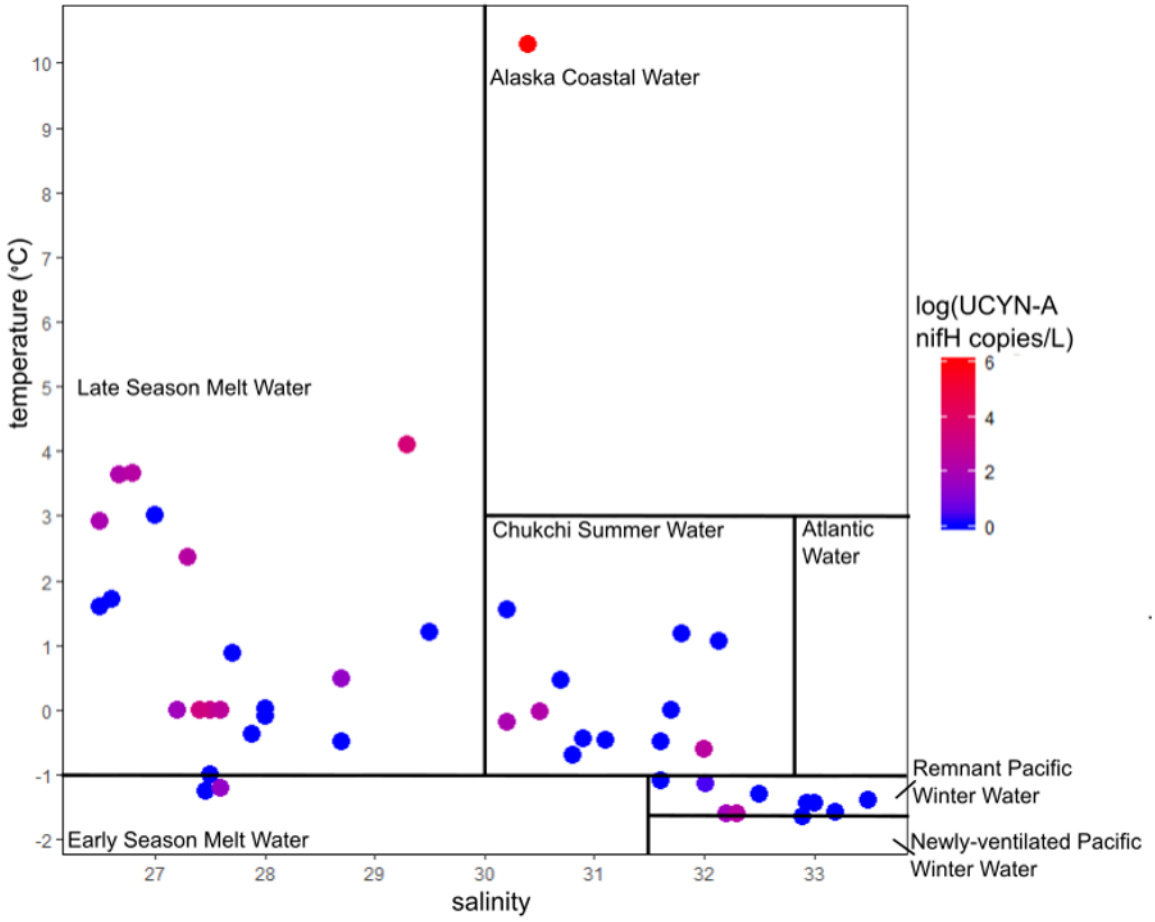


**Figure S4.** UCYN-A relationship to T/S. Water masses are defined as in (27). Water samples from all depths are categorized by temperature and salinity. The color gradient of points indicates total UCYN-A (UCYN-A1 + UCYN-A2) abundances determined by qPCR.



**Supplemental Tables S1 – S3**

**Table S1.** Double CARD-FISH probes and oligonucleotide competitors and helpers were used to visualize morphologies and target for nanoSIMS analysis. The last column indicates the presence (Y) or absence (N) of Horseradish-peroxidases (HRP) conjugated to the CARD-FISH probes.

| **Probe** | **Target** | **Sequence (5' to 3')** | **Reference** | **HRP?** |
|---|---|---|---|---|
| UCYN–A1 732 | Unicellular cyanobacteria UCYN-A1 | GTTACGGTCCAGTAGCAC | (29) | Y |
| UCYN–A2 732 | Unicellular cyanobacteria UCYN-A2, used as competitor | GTT**G**CGGTCCAGTAGCAC | (18) | Y |
| UPRYM-69 | UCYN-A1 host | CACATAGGAACATCCTCC | (18) | Y |
| UBRADO-69 | UCYN-A2 host | CACATTGGAACATCCTCC | (18) | Y |
| UCYN-A1 competitor | UCYN-A2 probe used as A2 competitor | GTT**G**CGGTCCAGTAGCAC | (18) | N |
| UCYN-A2 competitor | UCYN-A1 probe used as A2 competitor | GTTACGGTCCAGTAGCAC | (18) | N |
| Helper A–732 | Unicellular cyanobacteria UCYN-A | GCCTTCGCCACCGATGTTCTT | (29) | N |
| Helper B–732 | Unicellular cyanobacteria UCYN-A | AGCTTTCGTCCCTGAGTGTCA | (29) | N |
| UPRYM-69 competitor | UCYN-A2 host, used as competitor | CACAT**T**GGAACATCCTCC | (18) | N |
| UBRADO-69 competitor | UCYN-A1 host, used as competitor | CACAT**A**GGAACATCCTCC | (18) | N |
| Helper A-PRYM | *Haptophyta* | GAAAGGTGCTGAAGGAGT | (18) | N |
| Helper B-PRYM | *Haptophyta* | AATCCCTAGTCGGCATGG | (18) | N |

**Table S2.** Table summarizing stations sampled. Station numbers according to station map (Fig. S1) repeating station numbers are different depths sampled. Depths between 17 to 38 m correspond to the location of the chlorophyll maximum at that station. The location corresponds to the sea or hydrographical location of the sample, shelf corresponds to stations above the Chukchi Shelf but not in the Chukchi Sea proper. UCYN-A abundances were determined by qPCR. Relative % UCYN-A *nifH* sequence is the total UCYN-A (UCYN-A1 + UCYN-A2) *nifH* sequences as a percent of the total number of *nifH* sequences. Bulk N-fixation rates, the standard deviation of the rate, limit of detection (LOD), and minimum quantifiable rate (MQR) were calculated as described in methods as the sum of values 3 µm and 0.3 µm size fractions. In the $^{15}N_2$ fixation column, 'x' indicates no bulk measurements made at the station, ND (not detected) shows rate was below the limit of detection. If the rate was above detection, but below minimum quantifiable rate, the rate is included but marked with DNQ (detected, not quantified). Enrichment of $^{15}N_2$ in incubations, as determined by MIMS, is shown as the average of triplicate incubations for each station.

| Station # | Latitude | Longitude | sample depth (m) | location | salinity | temp (°C) | UCYN-A1 nifH copies/L | UCYN-A1 standard dev. | rel. % UCYN-A *nifH* gene seqs. | 15N2 fixation (nmol N/l day) | 15N2 fix St.Dev. | MQR | LOD | 15N2 At% enrichment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 64.9 | -167.2 | 2 | Bering | 30.4 | 10.3 | 1.2E+06 | 8.4E+04 | 34 | 6.85 | 3.84 | 4.36 | 0.32 | 3.14 |
| 2 | 72.0 | -164.6 | 2 | Chukchi | 29.3 | 4.1 | 1.0E+02 | 0.0E+00 | 45 | 0.17 (DNQ) | 0.19 | 0.23 | 0.09 | 7.23 |
| 3 | 73.7 | -164.0 | 2 | Chukchi | 28.5 | -0.1 | 0.0E+00 | 0.0E+00 | 0 | 0.14 | 0.08 | 0.07 | 0.13 | 7.37 |
| 4 | 71.8 | -164.8 | 2 | shelf | 27.7 | 0.9 | 0.0E+00 | 0.0E+00 | 0 | 0.23 (DNQ) | 0.21 | 0.24 | 0.13 | 4.73 |
| 5 | 73.2 | -161.0 | 2 | Chukchi | 27.5 | -1.3 | 0.0E+00 | 0.0E+00 | 0 | ND | - | 0.09 | 0.18 | 3.83 |
| 6 | 72.2 | -162.7 | 2 | shelf | 28.7 | 0.5 | 2.5E+01 | 0.0E+00 | 0 | 0.53 | 0.34 | 0.46 | 0.13 | 4.47 |
| 7 | 71.7 | -158.4 | 2 | shelf | 27.2 | 0.0 | 2.5E+01 | 0.0E+00 | 0 | 0.25 (DNQ) | 0.35 | 0.49 | 0.11 | 4.17 |
| 8 | 72.2 | -158.9 | 2 | shelf | 27.5 | -1.0 | 0.0E+00 | 0.0E+00 | 0 | 0.14 (DNQ) | 0.14 | 0.22 | 0.06 | 6.77 |
| 9 | 73.0 | -161.4 | 2 | shelf | 28.0 | 0.0 | 0.0E+00 | 0.0E+00 | 0 | 0.37 (DNQ) | 0.24 | 0.59 | 0.16 | 4.13 |
| 10 | 74.5 | -163.0 | 2 | shelf | 27.9 | -0.4 | 0.0E+00 | 0.0E+00 | 0 | ND | - | 0.11 | 0.05 | 10.93 |
| 11 | 71.8 | -155.4 | 2 | shelf | 27.5 | 0.0 | 1.3E+02 | 0.0E+00 | 0 | 0.42 | 0.14 | 0.19 | 0.07 | 7.47 |
| 12 | 72.3 | -156.4 | 2 | shelf | 27.3 | 2.4 | 2.5E+01 | 0.0E+00 | 31 | 0.27 | 0.08 | 0.12 | 0.05 | 6.70 |
| 13 | 73.0 | -157.9 | 2 | Beaufort | 27.6 | 0.0 | 1.3E+02 | 0.0E+00 | 0 | 1.19 | 1.38 | 1.15 | 0.09 | 5.93 |
| 14 | 74.3 | -160.0 | 2 | Beaufort | 27.6 | -1.2 | 0.0E+00 | 0.0E+00 | 0 | 0.70 (DNQ) | 0.62 | 0.98 | 0.17 | 3.40 |
| 15 | 72.0 | -151.8 | 2 | Beaufort | 26.7 | 3.6 | 0.0E+00 | 0.0E+00 | 0 | 0.14 (DNQ) | 0.11 | 0.16 | 0.05 | 6.20 |
| 16 | 73.3 | -154.1 | 2 | Beaufort | 27.4 | 0.0 | 2.5E+01 | 0.0E+00 | 0 | 0.37 | 0.19 | 0.17 | 0.08 | 8.13 |
| 17 | 74.5 | -155.7 | 2 | Beaufort | 26.7 | 3.0 | 0.0E+00 | 0.0E+00 | 0 | 0.53 | 0.22 | 0.28 | 0.07 | 4.97 |
| 18 | 74.9 | -156.4 | 2 | Beaufort | 26.5 | 2.9 | 0.0E+00 | 0.0E+00 | 0 | 0.39 | 0.23 | 0.22 | 0.04 | 5.27 |
| 19 | 74.0 | -152.3 | 2 | Beaufort | 26.8 | 3.7 | 2.5E+01 | 0.0E+00 | 0 | 0.40 | 0.17 | 0.14 | 0.06 | 5.20 |
| 20 | 74.9 | -154.6 | 2 | Beaufort | 26.5 | 1.6 | 0.0E+00 | 0.0E+00 | 0 | 0.28 | 0.15 | 0.19 | 0.04 | 4.00 |
| 21 | 75.4 | -152.0 | 2 | Beaufort | 26.6 | 1.7 | 0.0E+00 | 0.0E+00 | 0 | 0.20 | 0.09 | 0.17 | 0.03 | 5.13 |
| 2 | 72.0 | -164.6 | 27 | Chukchi | 31.80 | 1.2 | 0.0E+00 | 0.0E+00 | 0 | 0.42 (DNQ) | 0.42 | 0.45 | 0.09 | 5.70 |
| 3 | 73.7 | -164.0 | 25 | Chukchi | 32.00 | -1.3 | 0.0E+00 | 0.0E+00 | 0 | 0.21 | 0.16 | 0.20 | 0.20 | 7.47 |
| 4 | 71.8 | -164.8 | 24 | shelf | 32.10 | 1.1 | 0.0E+00 | 0.0E+00 | 0 | 0.49 | 0.28 | 0.36 | 0.14 | 5.87 |
| 5 | 71.4 | -160.7 | 20 | Chukchi | 32.00 | -1.1 | 0.0E+00 | 0.0E+00 | 0 | x | x | x | x | x |
| 6 | 72.2 | -162.7 | 22 | shelf | 28.70 | -0.5 | 0.0E+00 | 0.0E+00 | 0 | 0.57 | 0.35 | 0.46 | 0.21 | 4.20 |
| 7 | 71.7 | -158.4 | 27 | shelf | 30.80 | -0.7 | 0.0E+00 | 0.0E+00 | 0 | 0.24 (DNQ) | 0.41 | 0.73 | 0.13 | 4.70 |
| 8 | 72.2 | -158.9 | 26 | shelf | 31.70 | 0.0 | 0.0E+00 | 0.0E+00 | 0 | 0.14 (DNQ) | 0.18 | 0.18 | 0.07 | 6.83 |
| 9 | 73.0 | -161.4 | 29 | shelf | 31.60 | -0.5 | 0.0E+00 | 0.0E+00 | 0 | ND | - | 0.41 | 0.12 | 5.77 |
| 10 | 74.5 | -163.0 | 17 | shelf | 29.50 | 1.2 | 0.0E+00 | 0.0E+00 | 0 | 0.23 | 0.13 | 0.14 | 0.05 | 10.97 |
| 11 | 71.8 | -155.4 | 23 | shelf | 30.50 | 0.0 | 0.0E+00 | 1.5E+02 | 44 | 0.41 | 0.19 | 0.22 | 0.04 | 6.20 |
| 12 | 72.3 | -156.4 | 38 | shelf | 31.60 | -1.1 | 0.0E+00 | 0.0E+00 | 0 | 0.17 | 0.06 | 0.07 | 0.04 | 6.13 |
| 13 | 73.0 | -157.9 | 34 | Beaufort | 32.20 | -1.6 | 0.0E+00 | 5.0E+01 | 0 | 0.39 | 0.42 | 0.39 | 0.05 | 4.97 |
| 14 | 74.3 | -160.0 | 33 | Beaufort | 32.30 | -1.6 | 2.0E+02 | 0.0E+00 | 17 | 1.39 | 0.64 | 0.57 | 0.05 | 2.37 |
| 15 | 72.0 | -151.8 | 35 | Beaufort | 30.90 | -0.5 | 0.0E+00 | 0.0E+00 | 0 | 0.13 | 0.08 | 0.12 | 0.03 | 6.80 |
| 16 | 73.3 | -154.1 | 25 | Beaufort | 30.20 | -0.2 | 0.0E+00 | 1.0E+02 | 3 | 0.25 | 0.13 | 0.12 | 0.20 | 7.17 |
| 17 | 74.5 | -155.7 | 28 | Beaufort | 30.20 | 0.5 | 0.0E+00 | 0.0E+00 | 0 | 0.87 | 0.31 | 0.34 | 0.14 | 5.20 |
| 18 | 74.9 | -156.4 | 35 | Beaufort | 30.20 | 1.5 | 0.0E+00 | 0.0E+00 | 0 | 0.33 (DNQ) | 0.17 | 0.70 | 0.05 | 5.60 |
| 19 | 74.0 | -152.3 | 35 | Beaufort | 30.70 | 0.5 | 0.0E+00 | 0.0E+00 | 0 | 0.33 | 0.28 | 0.29 | 0.05 | 6.30 |

Station 1 $^{15}N_2$-fixation rate represents replicate samples, not triplicate. Only the 0.3 µm size fraction was measured for $^{15}N_2$-fixation rates at Station 3 at both depths listed.

**Table S3.** Volumetric calculations of UCYN-A derived $N_2$ fixation rates**.** Summary table comparing nanoSIM measured cell-specific $N_2$ fixation rates, calculated volumetric rates (from cell-specific $N_2$ fix rates and qPCR abundances), and bulk particulate $N_2$ fixation rates.

| Sample | UCYN-A specific $N_2$ fix (fmol N cell$^{-1}$day$^{-1}$) | *nifH* gene copies l$^{-1}$ | UCYN-A total $N_2$ fix rate (nmol N l$^{-1}$ d$^{-1}$) | measured bulk $N_2$ fix rate (nmol N l$^{-1}$ d$^{-1}$) |
|---|---|---|---|---|
| UCYN-A1 Bering | 7.6 | $1.2x10^6$ | 9.3 ± 17.7 | 6.9 ± 3.8 |
| UCYN-A2 Bering | 13.0 | $9.2x10^4$ | 1.2 ± 0.7 | |
| | | | | |
| UCYN-A2 Chukchi | 1.1 | $3.6x10^3$ | 0.004 ± 0.006 | 0.17 ± 0.19 |

**Supplemental References**


1. Moisander PH, Beinart RA, Voss M, & Zehr JP (2008) Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *ISME J* 2(6):954-967.

2. Zehr JP & McReynolds LA (1989) Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl Environ Microbiol* 55:2522-2526.

3. Moonsamy PV, *et al.* (2013) High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array System for simplified amplicon library preparation. *Tissue Antigens* 81(3):141-149.

4. Green SJ, Venkatramanan R, & Naqib A (2015) Deconstructing the Polymerase Chain Reaction: Understanding and Correcting Bias Associated with Primer Degeneracies and Primer-Template Mismatches. *PLoS ONE* 10(5):e0128122.

5. Zhang J, Kobert K, Flouri T, & Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End Read mergeR. *Bioinformatics* 30(5):614-620.

6. Caporaso JG, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5):1548-7091.

7. Edgar RC, Haas BJ, Clemente JC, Quince C, & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194-2200.

8. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):1367-4803.

9. Heller P, Tripp HJ, Turk-Kubo K, & Zehr JP (2014) ARBitrator: a software pipeline for on-demand retrieval of auto-curated *nifH* sequences from GenBank. *Bioinformatics*.

10. Ludwig W, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32(4):1363-1371.

11. Finn RD, *et al.* (2010) The Pfam protein families database. *Nucl. Acids Res.* 38(suppl_1):D211-222.

12. Caporaso JG, *et al.* (2010) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26(2):266-267.

13. Turk-Kubo KA, Farnelid HM, Shilova IN, Henke B, & Zehr JP (2017) Distinct Ecological Niches of Marine Symbiotic N-2-Fixing Cyanobacterium *Candidatus Atelocyanobacterium Thalassa* Sublineages. *J Phycol* 53(2):451-461.

14. Eren AM, *et al.* (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* 4(12):1111-1119.

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

16. Thompson A, *et al.* (2014) Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ Microbiol* 16(10):3238-3249.

17. Goebel NL, *et al.* (2010) Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to $N_2$ fixation in the tropical Atlantic Ocean. *Environ. Microbiol.* 12:3272-3289.

18. Cornejo-Castillo FM, *et al.* (2016) Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat Commun* 7:11071.

19. Cabello AM, *et al.* (2016) Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. *ISME J* 10(3):693-706.

20. Dekas AE & Orphan VJ (2011) Identification of diazotrophic microorganisms in marine sediment via fluorescence in situ hybridization coupled to nanoscale secondary ion mass spectrometry (FISH-NanoSIMS). *Methods Enzymol*, (Elsevier), Vol 486, pp 281-305.

21. Polerecky L, *et al.* (2012) Look@ NanoSIMS–a tool for the analysis of nanoSIMS data in environmental microbiology. *Environ. Microbiol.* 14(4):1009-1023.

22. Jayakumar A, *et al.* (2017) Biological nitrogen fixation in the oxygen-minimum region of the eastern tropical North Pacific ocean. *ISME J* 11(10):2356-2367.

23. Henke BA, Turk-Kubo KA, Bonnet S, & Zehr JP (2018) Distributions and Abundances of Sublineages of the $N_2$-Fixing Cyanobacterium *Candidatus Atelocyanobacterium thalassa* (UCYN-A) in the New Caledonian Coral Lagoon. *Frontiers in Microbiology* 9:554.

24. Shiozaki 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. Cy.* 31(6):996-1009.

25. Shiozaki T, *et al.* (2018) Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnol Oceanogr*.

26. Stenegren M, Caputo A, Berg C, Bonnet S, & Foster RA (2018) Distribution and drivers of symbiotic and free-living diazotrophic cyanobacteria in the western tropical South Pacific. *Biogeosciences* 15(5):1559.

27. Pickart RS, *et al.* (2016) Circulation of winter water on the Chukchi shelf in early Summer. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* 130:56-75.

28. Thompson AW, *et al.* (2012) Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. *Science* 337(6101):1546-1550.

29. Krupke A, *et al.* (2013) In situ identification and $N_2$ and C fixation rates of uncultivated cyanobacteria populations. *Systematic and Applied Microbiology* 36(4):259-271.

30. Krupke A, *et al.* (2015) The effect of nutrients on carbon and nitrogen fixation by the UCYN-A-haptophyte symbiosis. *ISME J* 9(7):1635-1647.

31. Martinez-Perez C, *et al.* (2016) The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat Microbiol* 1(11):16163.

## Appendix 2

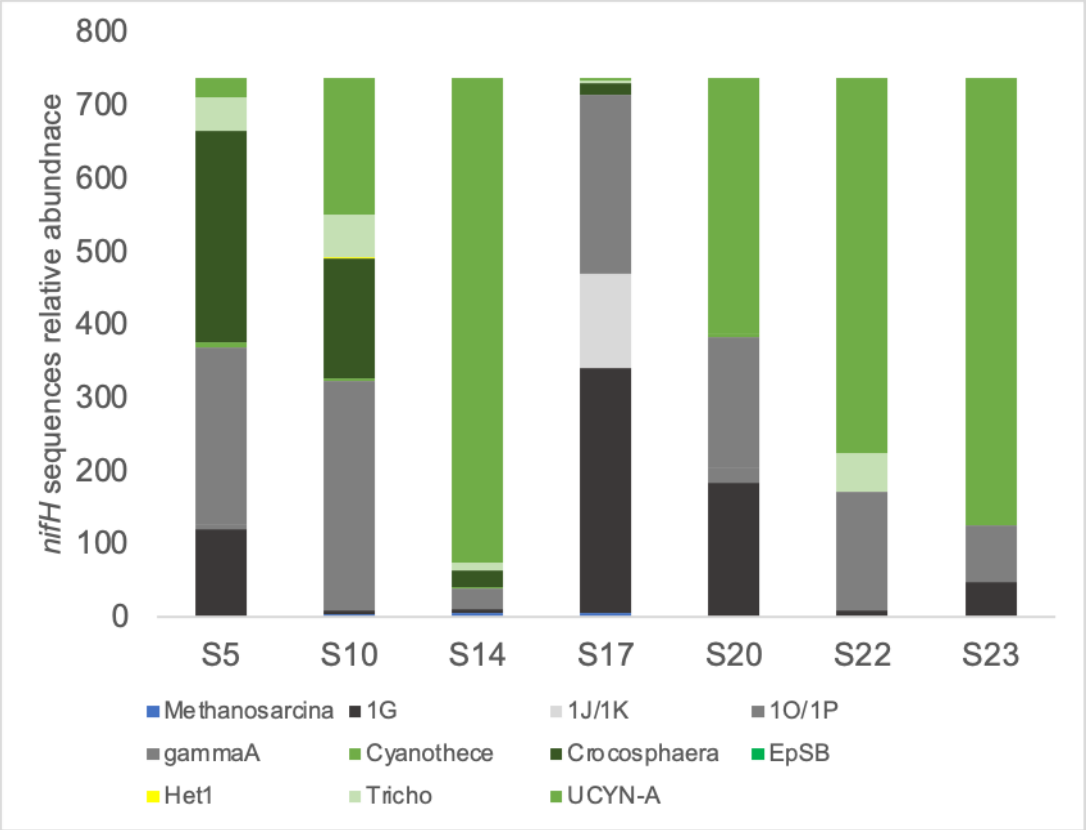
**Supplementary information for Chapter 3:** Direct cell-specific measurements show $N_2$ fixation by particle-attached non-cyanobacterial diazotrophs in the North Pacific Subtropical Gyre

**Supplementary Table 1:** Summary of natural-light and all-dark $N_2$ fixation rates (NFR). Average NFR is the average rate (nmol N $l^{-1}$ $d^{-1}$) from triplicate incubations and the associated standard deviation. Limit of detection (LOD) and minimum quantifiable rate (MQR) were calculated for each location and light treatment. Average NFR that were below the LOD (a conservative estimate) but above the MQR are denoted with asterisks.

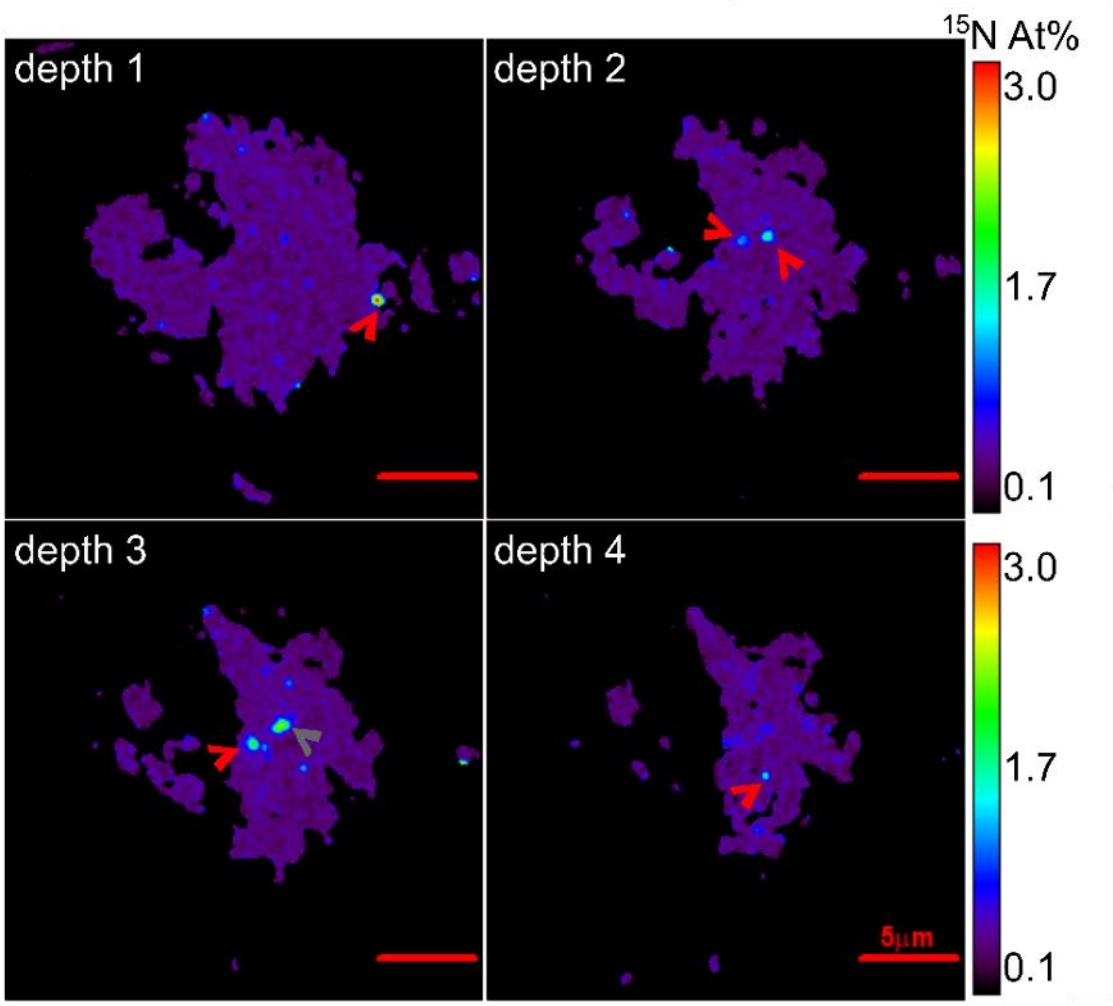
| Station - treatment | Average NFR (nmol N $l^{-1}$ $d^{-1}$) | Stdev NFR (nmol N $l^{-1}$ $d^{-1}$) | LOD (nmol N $l^{-1}$ $d^{-1}$) | MQR (nmol N $l^{-1}$ $d^{-1}$) | Detected? | Quantified? |
|---|---|---|---|---|---|---|
| S5 - light | 0.51 | 0.11 | 0.19 | 0.13 | yes | yes |
| S5 - dark | 0.35 | 0.14 | 0.16 | 0.15 | yes | yes |
| S10 - light | 0.19 | 0.13 | 0.12 | 0.13 | yes | yes |
| S10 - dark | 0.13* | 0.07 | 0.32 | 0.10 | No | yes |
| S14 - light | 0.23* | 0.12 | 0.29 | 0.13 | No | yes |
| S14 - dark | 0.06* | 0.04 | 0.08 | 0.05 | No | yes |
| S17 - light | 0.10* | 0.04 | 0.33 | 0.09 | No | yes |
| S17 -dark | 0.05 | 0.03 | 0.29 | 0.08 | No | No |
| S20 - light | 0.15* | 0.12 | 0.24 | 0.10 | No | yes |
| S20 - dark | 0.00 | 0.02 | 0.15 | 0.05 | No | No |
| S22 - light | 0.12* | 0.02 | 0.17 | 0.07 | No | yes |
| S22 - dark | 0.06* | 0.02 | 0.18 | 0.03 | No | yes |
| S23 - light | 0.38* | 0.07 | 0.16 | 0.08 | yes | yes |

**Supplemental Figures S1 – S3**

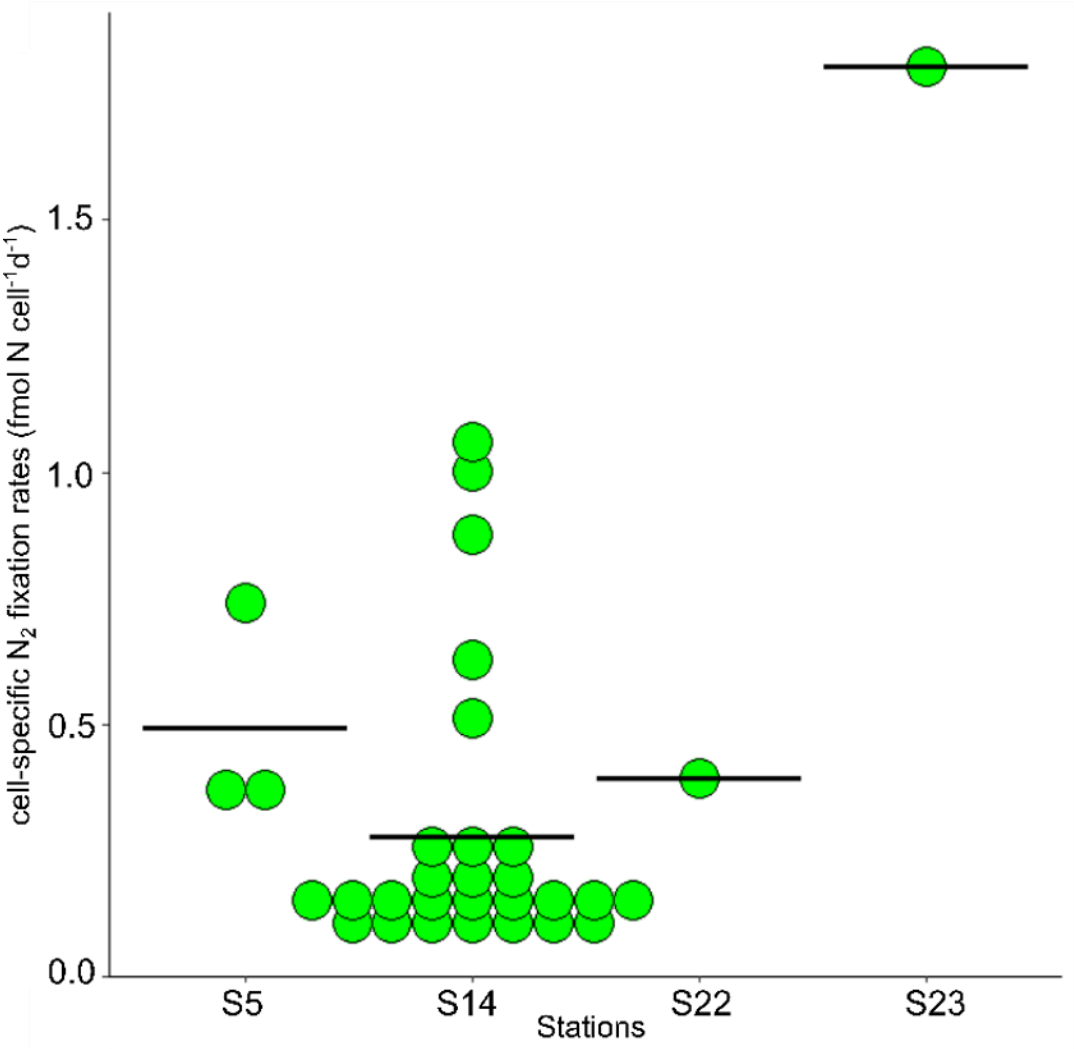
**Supplementary Figure 1:** Relative abundance of *nifH* gene sequences rarified to minimum sequence recovery (738 sequences). Green shading corresponds to cyanobacterial *nifH* sequences, while grey shades correspond to bacterial non-cyanobacterial diazotrophs, and blue (S17) is an archaeal non-cyanobacterial diazotroph—EpSB – *Epithemia pelagica* spherical body (Schvarcz et al., in press).



**Supplementary Figure 2:** Particle depth profile showing $^{15}N$ At%. A particle analyzed at multiple depths showing several additional NCD cells (red arrows) that were not identifiable from the particle surface. The grey arrow (depth 3) is likely the same cell that was identified in depth 2. Depth profiles demonstrate the likelihood of missing NCD cells on the interior or underside of particles.



**Supplementary Figure 3:** Cell-specific $N_2$ fixation rates (fmol N $l^{-1}$ $d^{-1}$) of cyanobacterial-like diazotrophs organized by location. Black horizontal lines represent the average cell-specific rate for that location. Cells were classified as cyanobacterial-like as they were enriched in $^{13}C$ as well as $^{15}N$.



## Appendix 3

**Supplementary information for Chapter 4:** Visualization of two uncultivated marine diazotrophs, the non-cyanobacterium 'Gamma-A' and the symbiotic cyanobacterium UCYN-A, using fluorescent in situ hybridization of the *nifH* gene (geneFISH).

**Supplementary Figure 1:** Comparison of UCYN-A2 CARD-FISH abundances (brown) to qPCR abundances (grey) in *nifH* copies $L^{-1}$ from corresponding samples. A comparison of the abundance differences is indicated with black brackets. Differences suggest multiple copies of the *nifH* gene (qPCR) for every 1 symbiosis (CARD-FISH). CARD-FISH values are the average of triplicate transects across the filter piece. QPCR abundances are the average of biological replicates. Error bars represent standard deviation