Physiological ecology of *Trichodesmium* and its microbiome in the oligotrophic ocean

Kyle R. Frischkorn

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

Columbia University

2018

Abstract

Physiological ecology of *Trichodesmium* and its microbiome in the oligotrophic ocean

Kyle R. Frischkorn

The colonial, N_2 fixing cyanobacterium *Trichodesmium* is a keystone species in oligotrophic ocean ecosystems. *Trichodesmium* is responsible for approximately 50% of the total biologically fixed N_2 in the ocean, and this "new" nitrogen fuels primary productivity and the amount of carbon sequestered by the ocean. *Trichodesmium* does not exist in isolation. Colonies occur ubiquitously with an assemblage of epibiotic microorganisms that are distinct from planktonic microbes and modulated across environments, yet the implications of this relationship have not been explored. In this thesis, the ecology, physiology, and potential geochemical impact of interactions within the *Trichodesmium* host-microbiome system were examined across three different oligotrophic ocean environments. First, to establish the metabolic diversity contributed by the microbiome to *Trichodesmium* consortia, a whole community metagenomic sequencing approach was used across a transect the western North Atlantic. This study demonstrated that the microbiome contributes a large amount of unique functional potential and is modulated across a geochemical gradient. In the following study, metatranscriptomics was used to show that such metabolic potential in *Trichodesmium* and the microbiome was expressed and modulated across the environment. Colonies were sampled in the western tropical South Pacific and gene expression dynamics indicated co-limitation by iron and phosphorus, and revealed a mechanism for phosphate reduction by *Trichodesmium* and subsequent utilization by the microbiome. These activities were verified with phosphate reduction rate measurements and indicated cryptic phosphorus cycling within colonies. Next, the suite of potential physiological interactions between host and microbiome was assessed with metatranscriptome sequencing on high frequency samples of *Trichodesmium* colonies from the North Pacific subtropical

gyre. Synchronized day-night gene expression periodicity between consortia members indicated tightly linked metabolisms. The functional annotations of these synchronous genes indicated intra-consortia cycling of nitrogen, phosphorus and iron, as well as a microbiome dependence on *Trichodesmium*-derived cobalamin—interactions that could alter the transfer of these resources to the surrounding water column. In the final study, the effect of the microbiome on Trichodesmium N_2 fixation was assessed. Using colonies obtained from the North Atlantic, activity in the microbiome was selectively modified using quorum sensing acyl homoserine lactone cell-cell signaling, a mechanism that *Trichodesmium* itself does not possess. These experiments indicated that the microbiome has the potential to increase or decrease *Trichodesmium* N_2 fixation to a degree that rivals the effects of alterations in nutrient concentration, but at a more rapid rate. In all, the research presented in this thesis demonstrates the integral importance of the microbiome to *Trichodesmium* physiology and ecology, highlighting the importance of an unexplored facet of marine microbial systems that likely influences the biogeochemistry of the planet.

Table of contents

List of figures **iv**

List of tables **vi**

1 Introduction **1**
1.1 Background and motivation . . . 2
1.2 Microbial ecology in the oligotrophic ocean . . . 3
1.3 The biogeochemical importance of the cyanobacterium *Trichodesmium* . . . 4
1.4 The *Trichodesmium* holobiont: a consortia of microbes . . . 6
1.5 Thesis overview . . . 9

2 Epibionts dominate metabolic functional potential of *Trichodesmium* colonies from the oligotrophic ocean **12**
2.1 Abstract . . . 13
2.2 Introduction . . . 13
2.3 Materials and methods . . . 14
2.3.1 Field sampling . . . 14
2.3.2 Chemical analyses . . . 15
2.3.3 DNA extraction and sequencing . . . 16
2.3.4 Sequence assembly and analysis . . . 16
2.4 Results and discussion . . . 18
2.4.1 Composition of the *Trichodesmium* holobiont . . . 18
2.4.2 The microbiome dominates holobiont functional potential . . . 24
2.4.3 Resource niche partitioning in the *Trichodesmium* holobiont . . . 29
2.5 Conclusions . . . 32

3 *Trichodesmium* physiological ecology and phosphate reduction in the western Tropical South Pacific **33**
3.1 Abstract . . . 34
3.2 Introduction . . . 34
3.3 Materials and methods . . . 37

3.3.1 Biogeochemical analyses . . . 37
3.3.2 *Trichodesmium* clade sampling and analysis . . . 38
3.3.3 *Trichodesmium* colony sampling . . . 39
3.3.4 Phosphate reduction in *Trichodesmium* colonies . . . 39
3.3.5 DNA extraction and metagenome sequencing . . . 40
3.3.6 Metagenomic assembly and analysis . . . 40
3.3.7 RNA extraction and metatranscriptome sequencing . . . 42
3.3.8 Metatranscriptomic sequence analysis . . . 43
3.4 Results . . . 44
3.4.1 Biogeochemistry . . . 44
3.4.2 Expression profiling of WTSP *Trichodesmium* consortia . . . 48
3.4.3 Evidence of phosphonate biosynthesis in *Trichodesmium* . . . 52
3.5 Discussion . . . 55
3.5.1 Trichodesmium distributions in the oligotrophic WTSP . . . 55
3.5.2 Expression of metabolic potential in *Trichodesmium* and its microbiome 57
3.5.3 Expression of resource-related signals in *Trichodesmium* . . . 58
3.5.4 Phosphonate biosynthesis by *Trichodesmium* in the WTSP . . . 60
3.6 Conclusions . . . 63

4 Coordinated gene expression between *Trichodesmium* and its microbiome over day-night cycles in the North Pacific Subtropical Gyre 65
4.1 Abstract . . . 66
4.2 Introduction . . . 66
4.3 Materials and methods . . . 68
4.3.1 Sample collection . . . 68
4.3.2 RNA extraction and sequencing . . . 69
4.3.3 Metatranscriptome analysis . . . 69
4.4 Results and discussion . . . 72
4.4.1 Coordinated expression in the *Trichodesmium* holobiont . . . 72
4.4.2 Coordinated expression in nitrogen and carbon pathways . . . 74
4.4.3 Transcriptional evidence of shared phosphorus and iron demand . . . 78
4.4.4 Microbiome dependence on *Trichodesmium* cobalamin . . . 80
4.5 Conclusions . . . 83

5 The *Trichodesmium* microbiome can modulate host N_2 fixation 84
5.1 Abstract . . . 85
5.2 Introduction . . . 85
5.3 Materials and methods . . . 87
5.4 Results and discussion . . . 90
5.4.1 Biological interactions are a driver of *Trichodesmium* N_2 fixation . . . 90

5.4.2 Microbiome transcriptional patterns varied with shifts in *Trichodesmium* N_2 fixation 91
5.4.3 Conclusions 96

6 Conclusion **97**
6.1 Dissertation summary 98
6.2 Future directions 100

Bibliography **104**

Appendices **122**

A Chapter 2 supplemental material **122**
A.1 Supplemental figures 122
A.2 Supplemental tables 124

B Chapter 3 supplemental material **126**
B.1 Supplemental figures 126
B.2 Supplemental tables 132

C Chapter 4 supplemental material **133**
C.1 Supplemental figures 133

D Chapter 5 supplemental material **139**
D.1 Supplemental figures 139
D.2 Supplemental tables 141

List of figures

2.1 Holobiont composition 20
2.2 Principal component analysis 21
2.3 Epibiont SEED subsystems 23
2.4 Distribution of OGs 25
2.5 Cell diagram 27

3.1 Transect map and geochemistry 45
3.2 *Trichodesmium* clade distribution 47
3.3 Holobiont OG expression 49
3.4 Resource responsive OGs in *Trichodesmium* 51
3.5 Phosphonate synthesis gene cassette 53

4.1 Principal component analysis 70
4.2 Resource-related co-expression 76
4.3 Carbon-related co-expression 77
4.4 Cobalamin-related co-expression 81

5.1 Sampling location and N_2 fixation data 88
5.2 Theoretical microbiome-induced N_2 fixation changes 92
5.3 Microbiome functional responses 95

A.1 Genome bin completeness 122
A.2 KEGG functional annotations 123

B.1 Assembly and annotation pipeline 127
B.2 Holobiont metabolic potential 128
B.3 Reads mapped to *ppm* cassette 129
B.4 Ppm protein alignment 130
B.5 Ppm protein phylogenic tree 131

C.1 Coordinated expression patterns 134
C.2 KEGG profile of periodic orthologous groups 135

C.3 Coordinated diel gene expression 136
C.4 Diel oscillation of nitrogen and phosphorus genes 137
C.5 Cobalamin-related gene expression 138

D.1 Responsive microbiome OG venn diagrams 140

List of tables

5.1 Microbiome response to +AHL amendment 93

A.1 Sequencing and environmental metadata 124
A.2 Frequency of key genes 124
A.3 Microbiome glycoside hydrolases detected 125

B.1 Geochemistry data 132

D.1 N_2 fixation and geochemistry data 141

Acknowledgments

My research and education at Columbia University and the Lamont-Doherty Earth Observatory have been supported by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP; DGE-16-44869). The work of the chapters included herein was also supported by grants from the NSF (OCE-1332912 to Sonya Dyhrman; OCE-1332898 and OCE-1536346 to Benjamin Van Mooy) and from the Simons Foundation's Simons Collaboration on Ocean Processes and Ecology (SCOPE; award ID 329108 to Sonya Dyhrman and Benjamin Van Mooy). The research presented in Chapter 3 is a contribution to the OUTPACE (Oligotrophy to UlTra-oligotrophy PACific Experiment) project which was funded by the French national research agency (ANR-14-CE01-0007-01), the LEFECyBER program (CNRS-8 INSU), the GOPS program (IRD) and the CNES (BC T23, ZBC4500048836). The research presented in Chapters 3 and 4 is a contribution of SCOPE.

First and foremost, I thank my advisor Sonya Dyhrman. I have said many times that she would take a bullet for one of her people, and that's barely hyperbole. Through the rough waves and the smooth sailing, vexing datasets or cocktail hour off the coast of Suriname, her support of my endeavors in the lab (and beyond) was a constant thread. Her door was always open, and I owe my successes in graduate school to her mentorship. The dedication, attention to detail, and scientific drive are all traits that I will aspire to for the rest of my career—whatever that career path may be. Sonya, I recognize that I was a handful... but I'm so grateful you put up with me.

I'm also thankful for the advice and support of my thesis committee, Hugh Ducklow, Kevin Griffin, Benjamin Van Mooy, Duncan Menge, and (sometimes) Ajit Subramaniam. Their research often takes them far from New York City, and I sincerely appreciate the time they took to make the sometimes equally arduous trek across the river to Lamont.

The best part about graduate school was having an amazing home base in the Dyhrman Lab. The expertise, ideas, and good humor of Harriet Alexander, Mónica Rouco, Matthew Harke, Gwenn Hennon, and Maria Hernandez have been so important to me. They made me excited to come into the lab each day. I also owe an immense, heartfelt thank you to

Sheean Haley: the gumby suit of the Dyhrman Lab. I am quite sure that I would not have survived these past five years without her. Sheean has enviable organization skills, the ability to pack a Zarges box like a Tetris master, a sixth sense about everything that's happening in the lab, and a knack for packing presents at sea to buoy spirits. These qualities were instrumental in facilitating the research I conducted. On the day-to-day, even more important to my graduate school experience was Sheean's sense of humor, her baking skills, and our daily debriefings. Really Sheean, I'd be out to sea without you. Thanks!

I also owe many thanks to the talented bioinformaticians who provided invaluable assistance wielding the "big data" that I generated as part of my thesis research. Without them, this data would just be random letters in a text file. It takes a village to translate results into something meaningful, and not reinvent the wheel while doing do. Most notably, I am thankful for Carrie Ganote of the National Center for Genome Analysis Support of Indiana University. She was my supercomputer guru, on call throughout the years of my PhD to load modules, troubleshoot scripts and talk down my frustrations—her bioinformatics wizardry made the alchemy of my data analyses possible. A special thanks to Melissa Lewis: we met for 2 hours at a café in Portland, Oregon when I was very lost, but the skills she taught me over one cup of coffee enabled me to code my way to this point.

Nearly all of the data in my thesis was collected during research expeditions at sea. I am indebted to the captains, crews and scientists with whom I sailed on the R/V *Atlantic Explorer*, the R/V *L'Atalante*, the R/V *Kilo Moana*, and the R/V *Ka'imikai-O-Kanaloa*. These research expeditions were the most rewarding and exciting experiences in my career and I am honored and thankful to have been given a top bunk on all of them. I owe a special thanks to Andi Krupke, my co-conspirator on an adventure across the South Pacific. Merci, Andi!

Balancing out the science were the friends I've made since moving to New York City. I met some wonderful people in the Department of Earth and Environmental Sciences. It has been a joy to learn about rocks, deep time, and the atmosphere with all of them, especially Laura, Logan, Kassandra, Sam, Ali, Lorelei, Frankie, Maayan, and last but not least, my kindred spirit Olivia. You all inspire me. Let's collaborate together, OK? A huge thanks

to my friends outside of science, my Brooklyn family (and the few stragglers who live in Manhattan and Queens). You all inspire me, too. Nosferatu!

Thanks also to my family: my dad Dave for the scientific drive, my mom Ann for the creativity and unyielding strength to tackle that science, and my siblings Eric, Kate and Mara for always having my back. The ability to do this work is due in large part to the foundations they helped me lay. It goes without saying that I am particularly grateful for Cale. For the past four years, he has known me only as a graduate student, and for some strange reason he still wanted to be my partner. I am a better scientist and an all around better person because of the balance he instilled in me... but also because he taught me how to make good coffee, cram 30 people into a Brooklyn apartment for a dinner party, and not take myself too seriously. Thank you for being there for me. I am so excited for whatever comes next!

And finally, hardly a day has gone by during my science career that I have not been thankful for meeting and working with Michael Carlson as a waffling undergraduate at the University of Washington. I dedicate this thesis to you! Thanks for lending me the Rip Curl Girl wetsuit, boarding the ferry, and starting this oceanography adventure.

For my friend, Michael Carlson

Chapter 1

Introduction

1.1 Background and motivation

In 1832, while sailing through the Caribbean Sea aboard the *Beagle*, Charles Darwin wrote that their ship had passed through bands of reddish-brown water measuring 10 m wide and stretching in length further than the eye could see. "The whole surface of the water, as it appeared under a weak lens, seemed as if covered by chopped bits of hay," he writes in *The Voyage of the Beagle*[1]. Similar occurrences had been noted during every voyage through the tropical and subtropical marine waters, from the logs of Captain James Cook aboard the H.M.S. *Endeavour*, back in time to the first historical accounts of those living on the coastal Red Sea—a body of water named for the very same enigmatic blooms of crimson-colored water. In the earliest records, these blooms were thought to have supernatural origins and were treated as omens from the gods. Later, the occurrence of red seawater was attributed to cosmological phenomenon, or even to occur as a result of migrating butterflies (Ehrenberg, 1830). With the invention of the microscope by Antonie van Leeuwenhoek in the 17th Century, the invisible, microscopic world came into focus. Tropical waters did not turn blood red on the whims of the gods, because of cosmic dust, or the flight of winged insects. Instead, the culprit was a microscopic plant, a phytoplankter.

Using a rudimentary microscope aboard the *Beagle*, Darwin viewed what he described as "minute cylindrical confervae," a term used to refer to filamentous phytoplankter, floating within a sample collected from the slick of reddish-brown water. He identified this organism as the newly described *Trichodesmium erythraeum* (Ehrenberg, 1830). With the *Beagle* adrift in an endless expanse of *Trichodesmium*, Darwin mused about the ecology of this enigmatic phytoplankton: "Their numbers must be infinite," he wrote[1]. Since Darwin's time, understanding of the ecology of this phytoplankton has been propelled to new heights. Though *Trichodesmium* might not be supernatural, it is now appreciated that this organism is a critical cog in global biogeochemical cycles that sustain the habitability of the planet. Humans have been peering over boats and puzzling over *Trichodesmium* for millennia. Even in light of the advancements of modern science, today is no different—there is still much about *Trichodesmium* and how it influences our planet yet to be discovered.

[1]Darwin, Charles, 1809-1882. *The Voyage of the Beagle.* New York; Modern Library, 2001. Print.

1.2 Microbial ecology in the oligotrophic ocean

The remote, subtropical gyres north and south of the equator are referred to as the oligotrophic oceans. In spite of the seemingly limitless expanses of water in every direction, these regions appear almost desert-like owing to their absence of any readily visible lifeforms. These oligotrophic oceans are hardly devoid of life, however. Covering 40% of the surface of Earth, they are considered the largest contiguous ecosystems on the planet (Karl, 1999, 2007). Here, life exists in great abundance at the microscopic scale: bacteria, archaea and unicellular eukaryotes inhabit seawater in concentrations of approximately 1 million cells per milliliter. Nearly half of the total primary production on Earth is fueled by the photosynthesizers among those marine microbes, the majority of which reside in the oligotrophic regions (Karl, 1999). The primary productivity performed by these phytoplankton captures CO_2 from the atmosphere and converts it into biomass; they are the very first link in the marine food web, and the arbiters of how much CO_2 is captured from the atmosphere and subsequently sequestered to the deep sea (Ducklow et al., 2001; Eppley and Peterson, 1979). Thus, understanding the factors that govern the processes of marine phytoplankton is a fundamental question of the field of biological oceanography.

A defining characteristic influencing microbial ecology and primary productivity in the oligotrophic oceans is the availability of nutrients (Arrigo, 2005). Nitrogen, phosphorus, iron, and a suite of other micronutrients are critical to primary producers, yet they variably occur at such low concentrations as to limit the growth and activities of the organisms that rely on them (Howarth and Marino, 1988). Which nutrient constrains productivity over geological timescales is a topic of much debate (Tyrrell, 1999), however in the modern ocean the concentration of nitrogen has a large and unequivocal impact on the distribution and activities of phytoplankton (Falkowski, 1997; Ryther and Dunstan, 1971; Zehr and Ward, 2002).

The elemental nitrogen cycle is complex, owing to the fact that this element occurs in more chemical species than other elements. The most abundant form in the atmosphere and the ocean is dinitrogen gas—a compound that is not readily available for biological use. The forms of nitrogen that can be taken up and used by microbes, termed “fixed” nitrogen and defined as all compounds other than dissolved N_2, are consumed by phytoplankton in the

sunlit surface ocean. With the death of these organisms, particulate matter sinks to depths below the photic zone, and there heterotrophic bacteria remineralize those compounds back to the inorganic, labile forms from which they were derived. The resulting dissolved inorganic nitrogen is subsequently returned to the surface ocean through physical processes, and the loop of the nitrogen cycle starts anew (Azam et al., 1983). In the ocean ecosystem however, this recycling of nitrogen is not enough to support the primary productivity that takes place, and must be balanced by an alternate process that delivers an alternate source of "new" nitrogen to the system (Falkowski et al., 1998).

In the low nutrient oligotrophic ocean, a critical source of this new nitrogen to the surface ocean is biological N_2 fixation, an enzymatically mediated process that converts dinitrogen gas to ammonium (Capone, 2001). This capability is restricted to bacteria and archaea, and in the surface ocean is carried out predominantly by cyanobacteria (Raymond et al., 2004). N_2 fixing, or diazotrophic, cyanobacteria encompass a large diversity of lifestyles, from free-living unicellular species, to symbionts that live exclusively within host diatom cells (Zehr, 2011). Among these cyanobacterial diazotrophs, the colonial, filamentous genus *Trichodesmium* is of paramount importance (Bergman et al., 2013; Capone et al., 1997).

1.3 The biogeochemical importance of the cyanobacterium *Trichodesmium*

Trichodesmium has a cosmopolitan distribution throughout the tropical and subtropical oligotrophic oceans (Capone et al., 1997). Chains of *Trichodesmium* cells, or filaments, cluster together to form "puff" or "raft" shaped colonies that are large enough to be visible to the naked eye. Under the appropriate conditions—during summer months when winds slow and the surface ocean stabilizes—colonies of *Trichodesmium* can accumulate to form extensive blooms, the largest of which can extend over $2x10^6$ km^2, an area approximately 20% of the entire Arabian Sea (Capone et al., 1998). Although such massive proliferations of *Trichodesmium* have long been a source of wonder and scientific curiosity (Bergman et al., 2013; Capone et al., 1997; Ehrenberg, 1830), modern interest in this organism was propelled by the discovery that not only is *Trichodesmium* a diazotroph (Dugdale et al., 1961), but

that it is estimated to be responsible for approximately 50% of the total biologically available nitrogen in the ocean (Capone et al., 1997; Mahaffey et al., 2005; Westberry and Siegel, 2006).

The *Trichodesmium* genus encompasses multiple species, but surveys suggest that in the environment species belonging to phylogenetic Clade I (*T. thiebautii, T. tenue, T. hildebrandtii*, and *T. spiralis*) are more abundant than those of Clade III (*T. erythraeum* and *T. contortum*) (Hynes et al., 2012; Rouco et al., 2014, 2016b). Within the geochemically suitable environments where *Trichodesmium* species are detected, their occurrence is largely dependent on nitrogen concentration. In regions with higher bioavailable nitrogen concentrations, *Trichodesmium* fails to outcompete other phytoplankton. Conversely, where nitrate levels are depleted, *Trichodesmium* can serve as the dominant source of new, fixed N_2 (Subramaniam et al., 2008). Because *Trichodesmium* is freed from the constraints of nitrogen limitation, its ecology and physiology are controlled by other resources that vary depending on geographical location (Sohm et al., 2011).

The predominant geochemical controls on *Trichodesmium* distribution and N_2 fixation are iron and phosphorus (Sohm et al., 2011). *Trichodesmium* has high iron quotas due to the demands of enzymes involved in N_2 fixation and photosynthesis that use this element as a cofactor (Berman-Frank et al., 2001). In regions where iron is supplied sufficiently, for example in the North Atlantic where deposition of dust from the Sahara Desert carries ample iron, the supply of fixed N_2 through diazotrophy increases the nitrogen to phosphorus ratio, and subsequently phosphorus becomes limiting (Deutsch et al., 2007; Sañudo-Wilhelmy et al., 2001). Under very oligotrophic conditions and in certain regions of the ocean, iron and phosphorus can even co-limit N_2 fixation by *Trichodesmium* (Mills et al., 2004; Walworth et al., 2017b).

Over the course of its evolution in the marine environment, *Trichodesmium* has amassed an arsenal of physiological strategies to survive in the variable geochemical conditions of the oligotrophic oceans. For example, iron stress can be mitigated through *Trichodesmium's* ability to use iron complexed within desert dust (Polyviou et al., 2018), a process that can involve actively transporting comparatively large dust particles to the center of colonies where iron dissolution then occurs (Rubin et al., 2011). Other iron stress mitigation strate-

gies include increased production of proteins responsible for transport of ferrous and ferric inorganic iron, transport of ligand-bound organic iron, and the use of the electron carrier flavodoxin which can be switched for ferredoxin, its iron-requiring counterpart (Chappell and Webb, 2010; Snow et al., 2015b). Similar to iron, *Trichodesmium* employs a number of strategies to survive in low phosphorus conditions. Phosphorus stress yields the induction of molecular systems that enable use of a diverse suite of organic and reduced phosphorus compounds. Such forms are generally in higher abundance than inorganic phosphate in the oligotrophic ocean (Karl, 2014). These strategies include enzymes that hydrolyze organic ester-bound phosphate (Orchard et al., 2009) and phosphonates (Dyhrman et al., 2006), genes that enable the uptake and utilization of the inorganic phosphorus species phosphite (Polyviou et al., 2015), and the storage of phosphate in the form of polyphosphate (Orchard et al., 2010). Under low phosphorus conditions, *Trichodesmium* is also known to substitute phospholipids with sulfolipids (Van Mooy et al., 2009). Overall, these low iron and low phosphorus mechanisms are known to be implemented by *Trichodesmium* across the oligotrophic ocean and are induced to suit the prevailing environmental conditions (Chappell et al., 2012; Dyhrman et al., 2006; Hynes et al., 2009; Rouco et al., 2018). This interplay between *Trichodesmium* and the geochemical environment controls the physiological processes that supply critical fixed carbon and N_2 to the ecosystem, resources that go on to fuel further productivity by other phytoplankton. As such, in the low nutrient, desert-like oligotrophic ocean, *Trichodesmium* colonies are oases of biological activity. As such, this cyanobacterium does not exist in isolation, but rather with a consortium of tightly-associated microbes that are likely drawn to the resources *Trichodesmium* supplies. Unlike the previously described geochemical dynamics that are known to influence *Trichodesmium* physiology and ecology, the implications of potential biological interactions with these tightly-associated microbes have yet to be described.

1.4 The *Trichodesmium* holobiont: a consortia of microbes

The first scientific analysis of microbes living on *Trichodesmium* was made by Hans Paerl and colleagues in 1989 (Paerl et al., 1989). High magnification scanning electron mi-

crographs of colonies revealed a diverse assemblage of filamentous and rod-shaped bacteria in close association with chains of the much larger *Trichodesmium* cells (Paerl et al., 1989). Visualization was paired with microautoradiographic measurement of carbohydrate uptake that demonstrated how epibiotic bacteria were metabolically active, yet physiologically distinct from *Trichodesmium* that did not appreciably take up sugars (Paerl et al., 1989). It was hypothesized that the tight associations between *Trichodesmium* and associated bacteria could facilitate the exchange of fixed carbon and N_2, vitamins, or iron-chelating compounds—a currency that might underscore a beneficial symbiotic relationship (Paerl et al., 1989). Alternatively, but not necessarily exclusively, *Trichodesmium* could secrete antibiotic or allelopathic compounds in order to keep epibionts in check and to prevent skewing towards antagonistic interactions or invasion by parasitic organisms. Similarly, the epibionts could do the same to exclude colonization by potential competitors (Paerl et al., 1989).

These early microscopic observations of epibionts on *Trichodesmium* were only expounded upon two decades after this initial study. New molecular techniques facilitated determination of the taxonomic composition of these associated microbes. Results showed that the microorganisms on naturally occurring colonies were dominated by heterotrophic bacteria in the phylum Proteobacteria (Hewson et al., 2009a; Hmelo et al., 2012). These epibionts were distinct from free-living bacterioplankton, and also encompassed less taxonomic biodiversity overall (Hewson et al., 2009a; Hmelo et al., 2012). Subsequent taxonomic investigations showed that organisms associated with *Trichodesmium* were not limited to heterotrophic bacteria, however, as eukaryotic microorganisms like copepods, radiolarians, diatoms, dinoflagellates, foraminifera, other photosynthetic N_2-fixing cyanobacteria, and even viruses were shown to be associated with colonies across multiple ocean basins (Brown et al., 2013b; Hewson et al., 2009a; Momper et al., 2014). These findings underscore *Trichodesmium's* role as a hotspot of biological activity in the oligotrophic ocean, and suggest that colonies are ubiquitously associated with a distinct community of microbes.

Further research indicated that composition of epibiont communities was unique to different *Trichodesmium* morphologies and across ocean environments. In the Sargasso Sea, puff-type *Trichodesmium* colonies were shown to have taxonomically distinct epibiont com-

munity compositions than those associated with raft-type colonies, suggesting that physical or physiological differences between *Trichodesmium* morphologies might select for different epibiont assemblages (Hmelo et al., 2012). These results were supported by high throughput amplicon sequencing of epibiont 16S ribosomal DNA genes associated with *Trichodesmium* colonies from the North Atlantic, North Pacific, and South Pacific, indicating that to some extent epibiont communities were distinct across ocean basins and might be selected according to their benefit to the consortium as a whole (Rouco et al., 2016a).

Recently, the benefits conferred to *Trichodesmium* communities from the epibionts have also been delved into, building on the initial hypotheses of Paerl and colleagues (Paerl et al., 1989). One potential interaction explored was the exchange of iron between consortia members. *Trichodesmium*, like other marine cyanobacteria, has not been shown to synthesize siderophores, organic ligands that many bacteria use to chelate iron and facilitate uptake from deficient environments (Hopkinson and Morel, 2009). Unlike *Trichodesmium*, epibiont species do produce siderophores, though iron bound to the siderophores desferrioxamine and aerobactin were not as bioavailable to *Trichodesmium* as other forms of iron (Roe et al., 2012). The role of other siderophores or types of iron chelating agents in facilitating *Trichodesmium* success in low iron regions remains to be explored.

Evidence also suggests that activity of the associated epibionts might also mediate phosphorus cycling within consortia. For example, in one study, naturally occurring *Trichodesmium* colonies were amended with bacterial quorum sensing molecules called acyl homoserine lactones (AHLs), a form of cell-cell signaling that *Trichodesmium* itself does not possess (Van Mooy et al., 2012; Vannini et al., 2002). Amended colonies showed stimulated alkaline phosphatase activity, an enzyme that hydrolyzes organic ester-bound phosphorus compounds, indicating that interactions between epibionts could facilitate colony phosphorus uptake (Van Mooy et al., 2012). Adding to the idea that epibionts may be important facilitators of iron and phosphorus acquisition was the discovery that growth in low iron and low phosphorus conditions stimulates colony formation by *Trichodesmium* filaments (Tzubari et al., 2018). Formation of macroscopic colonies could provide more surface area on which epibionts could attach and form biofilms, and subsequent bacterial respiration within the colony microenvironment could create zones of anoxia beneficial for protecting

the oxygen-sensitive nitrogenase enzymes (Paerl and Bebout, 1988). Finally, while the process of recruitment of individual epibionts from the water column to colonies remains unknown, the formation of biofilms by epibionts is indeed stimulated by chemical signals produced by *Trichodesmium* (Rao et al., 2015), underscoring the idea of selective, beneficial recruitment of epibionts.

The ubiquitous co-occurrence of epibionts with *Trichodesmium*, the way these populations are curated across different environments, and the potential mechanisms of nutrient exchange and interactions are evidence of integral relationships within consortia. In other biological systems, the importance of interactions between a host and its associated microbes has rapidly gained appreciation (Blaser, 2014; Bordenstein and Theis, 2015; Engel et al., 2012). In humans, for example, such interactions are now appreciated to be critical and ubiquitous drivers of health and disease states (Blaser, 2014; Sampson and Mazmanian, 2015). In effect, no organism can be fully understood without taking into account its so-called microbiome[2]. *Trichodesmium*, with its consortia of epibionts, is likely no different. Given this organism's importance to primary productivity and global elemental cycles, *Trichodesmium's* physiology and ecology must be considered as a host-microbiome system.

1.5 Thesis overview

Together, decades of research has indicated the tantalizing potential for important interactions between *Trichodesmium* and its associated epibionts. These interactions have yet to be mechanistically examined, however. The broad objectives of the research presented in this thesis aim to metabolically characterize the *Trichodesmium* microbiome and to highlight potential interactions between host and microbiome that likely influence physiology, ecology, and biogeochemical impact.

Research on *Trichodesmium* is hindered by the limited availability and derived nature of epibionts associated with laboratory cultures, as well as the challenges inherent to probing the intricacies of microscopic biological systems. While *Trichodesmium* has been main-

[2]The word Microbiome is a newly coined term used to describe the assemblage of microscopic organisms—bacteria, archaea, viruses, and fungi—that live on and within a host organism.

tained for nearly three decades (Prufert-Bebout et al., 1993), axenic cultures do not exist. Because there is no record of how the epibiont populations have changed in culture since the early 1990s, the results of any laboratory experiments investigating the role of the microbiome may not be directly applicable to the environment. As such, studying biological interactions within *Trichodesmium* consortia could only be reasonably conducted in the field. The research in this thesis was enabled by expeditions to the western North Atlantic, the western tropical South Pacific, and the North Pacific Subtropical Gyre, and implemented multi-disciplinary approaches that paired *Trichodesmium in situ* samples and experimental incubations with geochemical parameter measurements. Metabolic functional potential and physiology in *Trichodesmium* and the microbiome was also assessed, using biochemical rate measurements and high throughput community genome and gene sequencing—metagenomics and metatranscriptomics—so-called "next generation" approaches.

The first challenge in understanding the dynamic relationship between *Trichodesmium* and their microbiome was the wholly unexplored metabolic functions of these epibionts—the physiological capabilities that govern biochemical transformation, biological interactions, and interfacing with the geochemical environment. In Chapter 2, I used metagenomic sequencing to reconstruct the metabolic functional potential encoded by the *Trichodesmium* host and microbiome across the western North Atlantic. I used these data to assess how the microbiome changes across a geochemical gradient and to predict how these tightly associated organisms might underscore consortia survival in the oligotrophic ocean. Metagenomic techniques provide an unprecedented wealth of whole community metabolic characterization, but cannot provide clues about how functions are wielded in the environment. In Chapter 3, I combined metagenomic, metatranscriptomic and nutrient uptake and reduction measurements to gauge how *Trichodesmium* and their microbiome tailor their physiologies to suit the oligotrophic western tropical South Pacific. This multi-disciplinary approach was optimized to uncover resources limiting consortia physiology and mechanisms of elemental cycling in colonies, both of which likely impact the geochemical environment. Next, metabolic pathways that govern interactions between host and microbiome were elucidated. In Chapter 4, I used high frequency Lagrangian sampling of *Trichodesmium* colonies in the North Pacific Subtropical Gyre to search for coordinated gene expression patterns that

would underscore the critical metabolic currencies within consortia. Finally, the last question addressed in this thesis is whether or not the activities of the microbiome can influence the host in ecologically or biogeochemically important ways. In Chapter 5, I answered this question by selectively modifying *Trichodesmium* colonies with AHL quorum sensing molecules and measuring changes in community gene expression and N_2 fixation across the western North Atlantic. Overall, the results of the studies presented in this thesis illustrate that the previously overlooked biological interactions between *Trichodesmium* and the microbiome are critical drivers of the ecology and physiology of this keystone marine microorganism.

Chapter 2

Epibionts dominate metabolic functional potential of *Trichodesmium* colonies from the oligotrophic ocean

This chapter was originally published as Frischkorn, K.R., Rouco, M., Van Mooy, B.A.S., and Dyhrman, S.T. (2017). Epibionts dominate metabolic functional potential of *Trichodesmium* colonies from the oligotrophic ocean. *ISME J.* **11**, 2090-2101.

2.1 Abstract

Trichodesmium is a genus of marine diazotrophic colonial cyanobacteria that exerts a profound influence on global biogeochemistry, by injecting 'new' nitrogen into the low nutrient systems where it occurs. Colonies of *Trichodesmium* ubiquitously contain a diverse assemblage of epibiotic microorganisms, constituting a microbiome on the *Trichodesmium* host. Metagenome sequences from *Trichodesmium* colonies were analyzed along a resource gradient in the western North Atlantic to examine microbiome community structure, functional diversity and metabolic contributions to the holobiont. Here we demonstrate the presence of a core *Trichodesmium* microbiome that is modulated to suit different ocean regions, and contributes over 10 times the metabolic potential of *Trichodesmium* to the holobiont. Given the ubiquitous nature of epibionts on colonies, the substantial functional diversity within the microbiome is likely an integral facet of *Trichodesmium* physiological ecology across the oligotrophic oceans where this biogeochemically significant diazotroph thrives.

2.2 Introduction

The colonial, diazotrophic cyanobacterium *Trichodesmium* has a cosmopolitan distribution throughout the tropical and subtropical oceans where it has a keystone role in oligotrophic ecosystems because of its ability to supply biologically available nitrogen through N_2 fixation and fixed carbon through photosynthesis (Capone et al., 1997). Models suggest that *Trichodesmium* N_2 fixation accounts for roughly half of the total 100–200 Tg of biologically fixed N_2 annually (Bergman et al., 2013), a supply that fuels the uptake of carbon by the broader community of photoautotrophs and ultimately the export of carbon to the deep sea (Arrigo, 2005). In the oligotrophic oceans where *Trichodesmium* occurs, there is intense competition for resources such as phosphorus and iron, which can limit *Trichodesmium* N_2 fixation (Sañudo-Wilhelmy et al., 2001) and growth (Krishnamurthy et al., 2007).

Trichodesmium cells grow as filaments, which aggregate to form colonies up to a millimeter in diameter, creating stable substrates that concentrate fixed carbon and nitrogen relative to surrounding seawater (Capone et al., 1997). A hallmark of these colonies is their

ubiquitous association with a diverse assemblage of microorganisms that are dominated by heterotrophic bacteria, as well as photosynthetic and even other N_2 fixing bacteria (Hmelo et al., 2012; Momper et al., 2014; Rouco et al., 2016a; Sheridan et al., 2002). Collectively, these tightly associated organisms are referred to as epibionts, and they form the *Trichodesmium* microbiome, a distinct community that is taxonomically different from planktonic microbes in surrounding seawater (Hmelo et al., 2012).

Despite the global biogeochemical significance of *Trichodesmium* and the ubiquitous presence of a community of tightly associated microorganisms, ecophysiological studies of *Trichodesmium* have rarely considered this consortium of co-occurring organisms as a holobiont. The microbiome's taxonomic diversity, functional diversity and the interplay between host and epibionts within the holobiont is only beginning to be explored (Hewson et al., 2009a; Rouco et al., 2016a; Van Mooy et al., 2012), but could help explain the fundamental unknowns that persist regarding *Trichodesmium* distribution and activities across different environments. Exploring the metabolic functional potential contained within the *Trichodesmium* microbiome is a key step toward gaining a mechanistic understanding of how this relationship influences the fitness of the holobiont and subsequently the fate of fixed carbon and nitrogen in the oligotrophic ocean. Here we use metagenomic sequencing of *Trichodesmium* colonies collected from stations along a gradient of phosphorus in the western North Atlantic to examine microbiome composition, functional diversity and metabolic contributions to the holobiont.

2.3 Materials and methods

2.3.1 Field sampling

Trichodesmium colonies were collected with surface water net tows along a cruise transect in the western North Atlantic aboard the R/V *Atlantic Explorer* (AE1409) during May 2014. Sampling occurred at the same time each day (approximately 0730–0830 hours) using nets with a mesh size of 130 μm. Nets were deployed and hauled through the surface water column six times before recovery, such that each sample represented thousands of liters of water. Individual *Trichodesmium* colonies were isolated and washed three times by succes-

sive transfer through fresh 0.2 μm sterile-filtered local surface seawater to remove all but tightly associated epibionts. A pooled sample of colonies was isolated and processed from each station. For each sample, an average of approximately 30 cleaned colonies were transferred onto 47 mm 5 μm pore size polycarbonate filters, gently vacuum filtered to remove excess liquid, flash frozen and stored in liquid nitrogen until extraction and sequencing. There were no discernable changes in average colony size from one station to another across the transect. In order to broadly assess the microbiome composition of the North Atlantic *Trichodesmium* populations, colony composition was sampled to reflect the distribution of *Trichodesmium* colony morphology found in net tows. At all stations, raft type colonies were much more abundant than puff or bowtie variants with approximately 30 rafts to 2 puff/bowtie colonies. As such, the data largely reflect the dominant raft morphology.

2.3.2 Chemical analyses

Total dissolved phosphorus was determined on 0.2 μm filtrates of surface water (approximately 5 m depth) samples collected by a Niskin rosette outfitted with conductivity, temperature, and depth sensors into acid-clean polycarbonate bottles. Samples were processed at the SOEST Laboratory for Analytical Biogeochemistry at the University of Hawaii, Honolulu, HI, USA, according to facility protocols. Alkaline phosphatase activity samples were obtained by placing 2–5 cleaned *Trichodesmium* colonies on 5 μm PC filters, gently vacuum filtering away excess liquid, then storing in 47 mm plastic Petri dishes at –20 °C until analysis. Samples were processed as previously described (Dyhrman and Ruttenberg, 2006) using 6,8-difluoro-4-methylumbeliferyl phosphate (DiMufP) on a Synergy H1 Hybrid plate reader using the Gen5 software package (BioTek, Winooski, VT, USA) (Dyhrman and Ruttenberg, 2006). N_2 fixation was measured using the acetylene reduction technique as previously described (Capone, 1993; Paerl, 1994). Briefly, approximately 20 *Trichodesmium* colonies were placed in a 60 mL polycarbonate bottle containing 60 mL of filtered seawater. A 1 mL aliquot of acetylene was injected into the bottle through a septum cap, the bottle was gently inverted and allowed to incubate in an on-deck incubator at ambient temperature and light. The headspace of the bottle was analyzed for ethylene approximately every 30 minutes and the rate of ethylene production through acetylene reductionwas determined

by linear regression. All incubations were conducted in triplicate and harvested between approximately local noon and 1400 hours.

2.3.3 DNA extraction and sequencing

Total genomic DNA was extracted from samples using the MoBio Power Plant Pro DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA, USA) following the manufacturer instructions. An average concentration of 30 ng μL^{-1} of genomic DNA for each sample was sequenced at the Argonne National Lab (Lemont, IL, USA). Genomic DNA was quantified using the Invitrogen (Carlsbad, CA, USA) Qubit and sheared using the Covaris Sonicator (Woburn, MA, USA) to the desired size range. Libraries were then generated using WaferGen's Apollo324 automated library system and Illumina (San Diego, CA, USA) compatible PrepX ILMN DNA library kits following the manufacturer's instructions. Resulting libraries were then size-selected using the Sage BluePippin (Beverly, MA, USA) and sequenced on one 2x100 bp lane of the Illumina HiSeq2000. During library preparation, an average insert size of approximately 750 base pairs (bp) was targeted. Metagenomic reads from the six samples are available on the NCBI Sequence Read Archive under BioProject number PRJNA330990.

2.3.4 Sequence assembly and analysis

Metagenomic reads were first trimmed using Sickle with default settings (github.com/najoshi/sickle). Trimmed forward and reverse reads were then converted to fasta with the fq2fa command in IDBA-UD (Peng et al., 2012). Reads from the six samples were assembled into scaffolds to create a merged assembly, using IDBA-UD under default parameters in order to yield robust assembly of the western North Atlantic *Trichodesmium* holobiont, modeling our approach after similar environmental metagenomic investigations (Dombrowski et al., 2016; Handley et al., 2012). This merged assembly of pooled colony metagenomic reads from across six stations was used for the bulk of the analyses presented herein.

Scaffolds produced by the merged assembly were clustered into genome bins by tetranucleotide frequency and read coverage of individual samples using MaxBin 2.0 set with default parameters (Wu et al., 2015). Genome completeness was estimated at >65% using MaxBin,

resulting in robust gene set comparisons for the majority of genome bins. Relative abundance estimates were calculated by multiplying the length of contigs in each bin by the number of reads recruited (coverage), then summing across genome bins. This method has shown good correlation with 16S-based results in other metagenomic data sets (Aylward et al., 2014). Binned scaffolds were translated into predicted proteins using Prodigal on the metagenomic setting (Hyatt et al., 2010). The resulting protein sequences were annotated using the blastp program of DIAMOND against the NCBI nr database (Buchfink et al., 2015; Suzek et al., 2007). The organismal identity of each bin was determined by assessing the nr database taxonomic affiliation of the best hits, with identity determined as the majority (>70%) taxonomic affiliation of predicted proteins in a bin. Proteins from each epibiont bin were also annotated against the SEED subsystems using the RAST online annotation program to assess differences in functional categories (Aziz et al., 2008; Overbeek et al., 2014). Although the majority of our analyses were based on the merged, six station metagenome assembly, we also examined regional difference in epibiont community structure by grouping northern (n=2) and southern (n=4) stations as replicates. Variance in this north–south epibiont community structure (species relative abundance) was visualized using principal component analysis in R (www.r-project.org), using the rda function in the vegan package (Oksanen et al., 2015). A permutational multivariate analyses of variance test ($P<0.1$) was performed to examine differences in community structure between the two northern stations versus the four southern stations, using a Bray–Curtis dissimilarity distance matrix (Anderson, 2001). Unpaired t-tests were calculated using GraphPad (La Jolla, CA, USA).

To prepare for orthologous group (OG) clustering, the six individual station assemblies were translated into predicted proteins as described above and merged together. These proteins were then filtered to remove sequences <70 amino acids and clustered into OGs by performing a reciprocal blastp with DIAMOND and then using MCL (Markov cluster algorithm), set to an inflation parameter of 1.4 as previously described (Bertrand et al., 2015). Taxonomic composition of the individual station assembly OG clustered proteins was determined by DIAMOND blastp using the previously assembled genome bins as a reference. OGs were classified as 'epibiont only' if no genes making up the group had best blast hits

to the four *Trichodesmium* genome bins or to proteins that are taxonomically affiliated with *Trichodesmium*. OGs defined as 'both' were composed of epibiont and *Trichodesmium* identified proteins.

Functional annotation of the predicted proteins clustered into OGs followed a tiered protocol. First, the size filtered and clustered proteins were annotated using blastp search with DIAMOND against the UniRef90 database (Buchfink et al., 2015) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) with the online Automatic Annotation Server using the single-directional best-hit method targeted to prokaryotes and with the metagenomic option selected. Single functional annotations for entire OGs were determined by taking the majority annotation for all proteins clustered into that group. To refine the annotations of select proteins, curated databases and protein models were used. The alkaline phosphatase-identified OG annotations were refined by performing DIAMOND searches against representative proteins from COGs 3211 (PhoX), 1785 (PhoA) and 3540 (PhoD) (Luo et al., 2009), as well as the protein sequences of three previously identified putative alkaline phosphatases in the IMS101 genome (PhoA: YP723031, PhoX: YP723360, and PhoX2: YP723924) (Orchard et al., 2009). Putative alkaline phosphatase metal cofactors were determined based on previous investigations (Luo et al., 2009; Rodriguez et al., 2014; Yong et al., 2014). PepM OGs were obtained by DIAMOND blast against representative proteins from PFAM 13714 (PEP_mutase). Blast results were accepted if the e-value was $<1x10^{-5}$ with a bit score >50. For comparisons against the putative PhoA protein from IMS101, OGs within the metagenomes were considered homologous to this protein if they passed the blast requirements above, and contained UniRef blast homologs to the YP723031 gene from *T. erythraeum* IMS101 and other putative alkaline phosphatase genes identified through KEGG or UniRef annotation.

2.4 Results and discussion

2.4.1 Composition of the *Trichodesmium* holobiont

High-throughput paired-end sequencing (Supplemental Table A.1) was performed on total genomic DNA extracted from *Trichodesmium* colonies collected at six stations in the

western North Atlantic (Figure 2.1a). A merged genomic DNA assembly of sequences from all stations was used to identify 12 unique taxonomic genome bins conserved across all 6 North Atlantic stations, 9 of which were estimated to be over 65% complete based on single copy marker gene presence (Supplemental Figure A.1), results that are similar to recoveries from other environmental metagenomic data sets (Dombrowski et al., 2016; Handley et al., 2012). All bins were identified down to the lowest definitive taxonomic level possible. Of the 12 taxonomic bins, there were 4 identified as *Trichodesmium*, which suggests that multiple *Trichodesmium* species could be present in these samples. Consistent with this observation, analyses of surface water and colonies a similar transect in the western North Atlantic detected species from at least three co-occurring clades of *Trichodesmium* in this region (Rouco et al., 2014, 2016a). The remaining eight genomic bins were identified as heterotrophic epibionts including two in the Bacteroidetes genus *Microscilla*, one Gammaproteobacterium, one Alphaproteobacterium in the order Rhodobacterales and four Alphaproteobacteria in the order Rhodospirillales (Figure 2.1b). These results contribute new information about the *Trichodesmium* microbiome, building upon a previous 16S clone library survey of epibiont diversity from the North Atlantic (Hmelo et al., 2012), recent high-throughput assessment of epibiont taxonomic diversity across three ocean basins (Rouco et al., 2016a), and metatranscriptomic profiles of Trichodesmium communities from the South Pacific (Hewson et al., 2009a). Similarities in the taxonomic groups dominating samples in all of these aforementioned studies indicate that the epibiont metagenomes detected here are likely from core members of the *Trichodesmium* holobiont. In addition, this microbiome community does not merely represent general particle-associated microbes, as the community recovered here was distinct from those previously found on sinking particles, which have been shown to be enriched with Deltaproteobacteria, Planctomyces and Bacteroidetes in genera other than *Microscilla* (Fontanez et al., 2015).

To estimate changes in relative abundance of holobiont members across the cruise transect, reads from each station were mapped to the taxonomic bins (Aylward et al., 2014). All microbiome members were detected at each station, confirming that these epibionts represent core components of the microbiome, although the relative abundance was variable from station to station (Figure 2.1c). Using the relative abundance of different epibiont

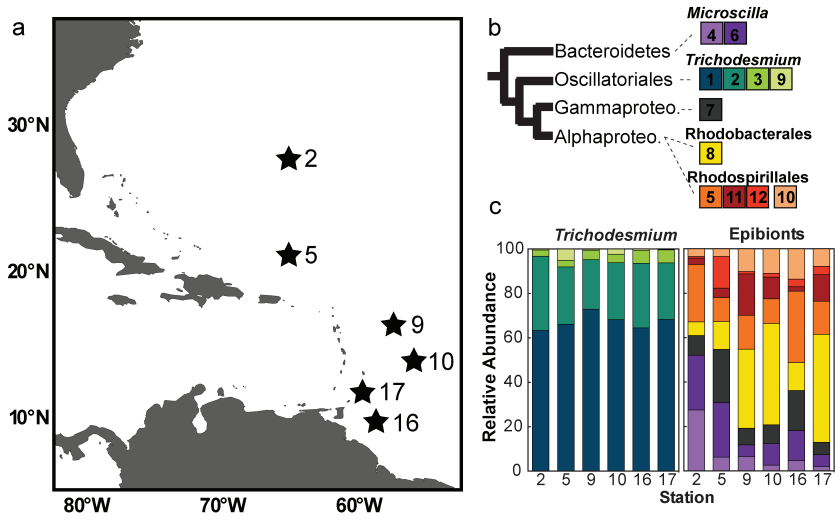

Figure 2.1: Sampling locations, genome bin identity and relative community composition of *Trichodesmium* holobiont members in the western North Atlantic. (a) *Trichodesmium* colonies were collected from near surface water using hand held nets at six stations during spring 2014 in the western North Atlantic. (b) Taxonomic affiliations of the 12 genome bins generated from a merged metagenome assembly and represented on a simplified phylogenetic tree to the class level. Gammaproteo., Gammaproteobacteria. Alphaproteo., Alphaproteobacteria. (c) The relative community composition of the holobiont members along the transect, noting that values were determined using reads from single samples of pooled colony metagenomic libraries. Values were calculated by multiplying contig lengths in each bin by read mapping coverage (Aylward et al., 2014).

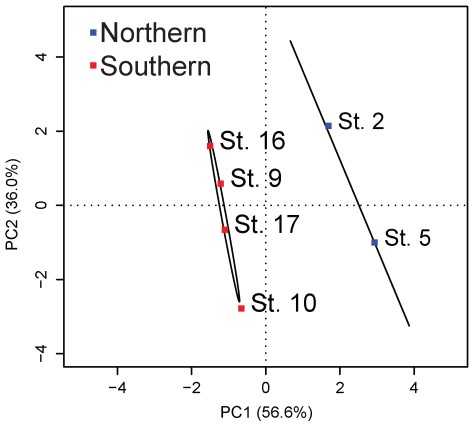

Figure 2.2: Principal component analysis of the relative abundance of *Trichodesmium* microbiome members. The 95% confidence intervals between northern and southern stations are indicated by black ellipses. The relative abundance of epibionts in the microbiome were significantly (P<0.1) different between the two northern and four southern stations (permutational multivariate analyses of variance P=0.067).

members, microbiome communities clustered according to their sampling site, with the two northernmost stations separated from the southernmost stations along the PC1 axis of a principal component analysis (Figure 2.2). A permutational multivariate analyses of variance analysis (P=0.067), confirmed a differential community structure between the two northern and four southern stations. This may be driven in part by the abundance of *Microscilla* (bins 4 and 6), which was significantly different (t-test, P=0.010) between the two northern stations and the four southern stations. There were no significant differences between the northern and southern stations in N_2 fixation rate, colony PO_4 turnover or alkaline phosphatase activity, however, the two northernmost stations had significantly lower total dissolved phosphorus compared with the four southern stations (0.08 and 0.2 μM total dissolved phosphorus, respectively, t-test, P=0.0157; Supplemental Table A.1). As such, *Microscilla* relative abundance was higher at the stations with increased oligotrophy. Although further work measuring more parameters over a greater range of conditions is warranted to address the consistency of these relationships, the differential distribution of members of the core microbiome between regions suggests that variations in epibiont community structure could be modulated by the geochemical environment.

In order to examine the diversity of metabolic pathways in the Trichodesmium microbiome, the genome bins were functionally annotated against the SEED subsystems to assess differences in broad functional categories between epibionts (Overbeek et al., 2014). This functional analysis showed that while key capacities like ammonia assimilation, phosphate metabolism, transport of organic compounds and aerobic metabolic pathways like the tricarboxylic acid cycle were largely uniformly present in all epibionts, there were certain SEED subsystems that were enriched or uniquely present in discrete groups (Figure 2.3). The *Microscilla* genome bins were enriched in functional categories related to nitrogen metabolism, such as nitrate, and nitrogen stress functions, as well as the synthesis and utilization of reduced phosphorus compounds (Figure 2.3), which could help explain their relative abundance at the most oligotrophic stations (2 and 5) (Supplemental Table A.1). The Gammaproteobacterium and Rhodobacterales (bins 7 and 8) were enriched with functions related to the uptake and exchange of genetic information (gene transfer agents and bacterial secretion systems), whereas the Rhodospirillales bins were enriched in motility-related functions, the uptake of tungstate and the utilization of plant-derived sugars like fructose (Figure 2.3). In marine bacteria, secretion systems and motility functions have been implicated in the transfer of toxins between adjacent cells and pathogenicity (Salomon et al., 2015) and the modulation of these activities could influence relationships between epibionts or the nature of the host–microbiome relationship. Finally, homologs to genes encoding the light-mediated proton pump proteorhodopsin were found in *Microscilla*, as well as the Rhodospirillales (Figure 2.3). In sum, these data show that there are differences in functional metabolic capacity between epibiont groups present in the core microbiome.

Iron and phosphorus-related SEED subsystems within the microbiome (Figure 2.3), like siderophore and heme-related functions, or reduced phosphorus utilization pathways, may be particularly critical to *Trichodesmium* physiological ecology given that iron and phosphorus are known drivers of *Trichodesmium* activities in the study region (Chappell et al., 2012; Sañudo-Wilhelmy et al., 2001). The relative proportion of microbiome genome bins in which key marker genes were found for phosphonate metabolism, heme and siderophore utilization, were compared with free-living microbial communities in the Sargasso Sea region of the western North Atlantic. The *Trichodesmium* microbiome was enriched nearly

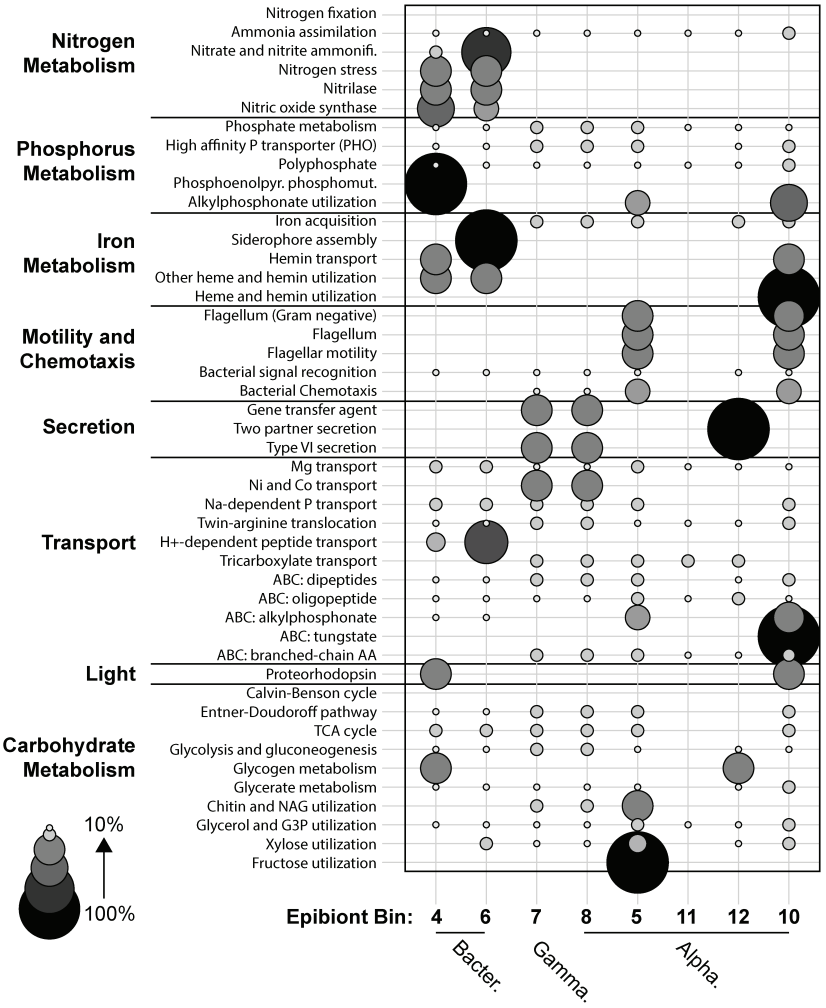

Figure 2.3: Enrichment of functional pathways recovered from epibiont genome bins. The distribution is based on RAST annotation against the SEED subsystems (Aziz et al., 2008; Overbeek et al., 2014). The contribution of each epibiont to a given SEED subsystem is scaled relative to the percentage of genes within each subcategory found in each genome bin. Ammonifi., Ammonification; ABC, ABC transporter; AA, amino acid. Alpha., Alphaproteobacteria; Bacter., Bacteroidetes; Gamma., Gammaproteobacterium.

two fold in the phosphonate utilization marker *phnJ* relative to genome equivalents in the Global Ocean Survey data set from the Sargasso Sea (Karl et al., 2008) (Supplemental Table A.2). Phosphonates are known to be an important source of bioavailable phosphorus for *Trichodesmium* in the North Atlantic (Dyhrman et al., 2006) and the enrichment of *phnJ* in the microbiome indicates the importance of this bond class of phosphorus to the holobiont relative to free-living microbes in the upper water column. Heme transporters and siderophore/vitamins transporters were present in all core microbiome genome bins, relative to communities of free-living Sargasso Sea microbes in which these functions were present in roughly 2% and 18% of genome equivalents, respectively (Tang et al., 2012) (Supplemental Table A.2). This enrichment is again suggestive of the importance of iron to the holobiont relative to free-living microbes in the upper water column. The apparent enrichment of phosphorus and iron functions in the microbiome may provide a competitive advantage to the holobiont relative to co-occurring free-living microbes and underlie possible syntrophic interactions between *Trichodesmium* and the microbiome, but this enrichment could also enhance internal competition within the holobiont. Taken together, the taxonomic and functional diversity within the *Trichodesmium* microbiome, and its variation along the transect, raises the possibility that the fitness of holobiont members or regional biogeochemistry could influence the distribution and activities of the holobiont.

2.4.2 The microbiome dominates holobiont functional potential

We evaluated the distribution of and diversity of physiological capabilities in the *Trichodesmium* host (genome bins 1–3 and 9) compared with members of their microbiome by functionally analyzing OGs of proteins. The taxonomic composition of predicted proteins in OGs was assessed to investigate the extent to which the *Trichodesmium* microbiome augments the global metabolic potential of the holobiont. Of the total 55,738 unique OGs, 4,546 OGs were composed solely of predicted proteins identified as *Trichodesmium* and 2,252 OGs were composed of predicted proteins from *Trichodesmium* and epibiont genome bins (Figure 2.4). The *Trichodesmium* spp. present in the field samples contained over two times the number of gene families (6,798) present in *T. erythraeum* IMS101 genome, which is predicted to encode 5,076 proteins (Walworth et al., 2015) and yielded 2,982 OGs

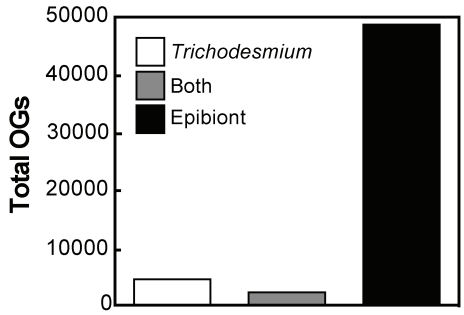

Figure 2.4: Distribution of OGs in the *Trichodesmium* holobiont. A total of 264,073 predicted proteins (>70 amino acids) were clustered into 55,738 OGs. OGs were considered '*Trichodesmium* only' or 'epibiont only' if they were composed of predicted proteins solely from those organisms. The 'both' category refers to OGs composed of predicted proteins from *Trichodesmium* and epibiont genome bins.

following the clustering protocol described herein. *Trichodesmium* does not exhibit genome streamlining like other oligotrophic cyanobacteria (Walworth et al., 2015), and the disparity in the number of OGs between type strain and field samples could be driven by the presence of multiple species with varying gene contents, or the fact that *T. erythraeum* is not common in *Trichodesmium* populations in the North Atlantic (Rouco et al., 2014).

Given the taxonomic diversity of the microbiome, we would expect the number of epibiont OGs to exceed that of *Trichodesmium*. Nearly 90% of the total 55,738 OGs were composed only of epibiont-identified proteins (Figure 2.4). This number exceeds what would be expected from the eight core microbiome members together and likely comes from less abundant epibionts that were not sequenced deeply enough to yield genome bins. These epibiont-only OGs represent functions without homologs in *Trichodesmium*, suggesting recruitment and selection in the microbiome could expand and alter the metabolic repertoire of the holobiont. To more conservatively examine functional content, OGs were assigned functional annotations using KEGG, and the epibiont-only OGs were found to have twice as many unique functions relative to those shared between epibionts and *Trichodesmium*, and nearly 10 times as many unique functions as those found in *Trichodesmium* alone (Supplemental Figure A.2). These differences are most notable within the genetic information

processing module, consistent with the broad taxonomic diversity within the microbiome and to some extent in the carbohydrate and lipid metabolism module (Supplemental Figure A.2), which is represented by an abundance of carbohydrate active enzymes not present in *Trichodesmium* (see below).

Specific OGs enriched or unique to the microbiome included proteins with functions related to cell–cell signaling and the processing of organic matter (Figure 2.5). Homologs to the proteins responsible for sensing and responding to quorum-sensing molecules like acyl homoserine lactone (AHL), the LuxR family, were detected in all epibionts (Figure 2.5). Although putative LuxR homologs identified as belonging to *Trichodesmium* genome bins shared sequence identity along the DNA-binding domain, it has been previously determined that these *Trichodesmium* genes do not share the characterized AHL-binding residues of these proteins (Patankar and Gonzalez, 2009; Van Mooy et al., 2012; Vannini et al., 2002). This indicates that *Trichodesmium* is either not involved in quorum sensing, or that the LuxR variant present in *Trichodesmium* is responding to a different quorum-sensing molecule. Homologs to proteins in the LuxQ family, which is part of a sensor kinase complex that detects the autoinducer-2 (AI-2) signaling molecule (Miller and Bassler, 2001) were detected in the Rhodospirillales (bins 5 and 11) (Figure 2.5). Homologs of LuxI, a gene responsible for synthesis of AHL quorum-sensing molecules, and a homolog to acyl-sn-glycerol-3-phosphate acyltransferase, a putative AHL synthase designated HdtS and characterized in *Pseudomonas fluorescens* (Case et al., 2008), were found in the Rhodobacterales and Rhodospirillales epibionts (Figure 2.5).

Quorum-sensing circuits, and subsequently the activities they modulate, can be broken through the secretion of molecules that break down AHLs. This process, termed quorum quenching, can be driven by quorum quenching enzymes like metallo-beta-lactamases, which degrade antibiotics and AHLs (Fetzner, 2014; Hong et al., 2012). Homologs of this enzyme were detected in *Microscilla*, Rhodospirillales and Rhodobacterales (Figure 2.5). Taken together, these data indicate that a suite of different quorum-sensing pathways are present within members of the microbiome. In fact, quorum-sensing molecules and quorum-quenching activity have been detected in *Trichodesmium* colonies from environmental samples, and quorum-sensing molecules added to sinking particles from *Trichodesmium* rich

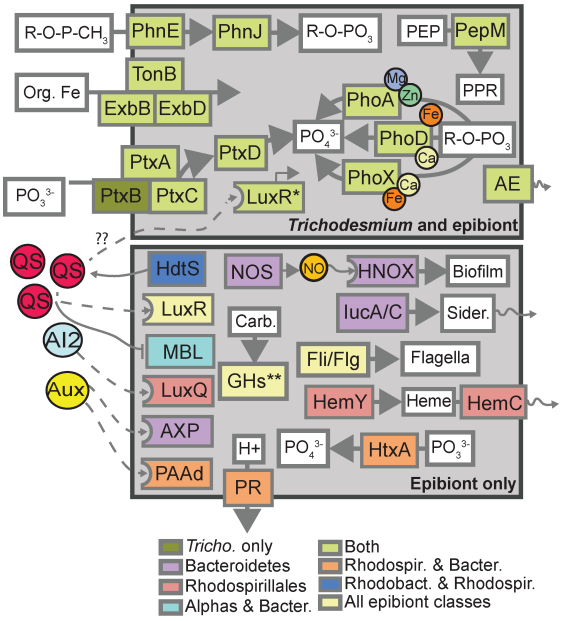

Figure 2.5: Cell diagram depicting OGs with key functions shared across the *Trichodesmium* holobiont, or unique to the microbiome. *Trichodesmium* LuxR is homologous only in DNA-binding region (*) and the GH protein represents many enzymes with specific targets unique to epibionts (**) (Supplemental Table A.3). The 'all epibiont' category refers to homologs that were found in all epibiont genome bins. AE, auxin efflux protein; AI2, auto- inducer 2; Aux, auxin; AXP, auxin-regulated protein; Fli/Flg, flagella biosynthesis and motility; GH, glycoside hydrolase; Carb., Carbohydrate; HemY, HemC, heme group synthesis and export; HNOX, heme-nitric oxide/oxygen binding protein; HtxA, alter- native phosphite dehydrogenase; HdtS, AHL synthase; IucA/C, siderophore biosynthesis; Sider., Siderophore; LuxQ, AI2 tran- scriptional activator; LuxR, QS transcriptional activator; MBL, Metallo-beta lactamase; NO, nitric oxide; NOS, nitric oxide synthase; PhnE, phosphonate transporter inner membrane sub- unit; PhnJ, C-P lyase; PepM, phosphoenolpyruvate mutase; PEP, phosphoenolpyruvate; PPR, 3-phosphonopyruvate; PR, proteorho- dopsin; PtxABC, phosphite transport system; PtxD, phosphite dehydrogenase; PhoA, PhoD, PhoX, alkaline phosphatases with putative metal cofactors (Mg, magnesium, Zn, zinc, Ca, calcium, Fe, iron); PAAd, phenylacetic acid degrading protein; QS, quorum sensing molecule; TonB-ExbB-ExbD, putative organic iron uptake system. *Tricho.*, *Trichodesmium*; Alphas, Alphaproteobacteria; *Micros.*, *Microscilla*; Rhodospir., Rhodospirillales; Rhodobact., Rhodobacterales.

environments stimulate the hydrolysis of organic matter (Hmelo et al., 2011; Krupke et al., 2016; Van Mooy et al., 2012). The concordance between metagenomic and *in situ* chemical evidence of these pathways constitutes strong evidence that quorum sensing is active in *Trichodesmium* colonies, likely regulates a range of microbiome functions, and mediates interactions between host and microbiome.

OGs with homologs of nitric oxide synthase (NOS) and the sensing protein (H-NOX) were found in both *Microscilla*-identified genome bins (Figure 2.5). The presence of both pathways have only previously been detected in an Alphaproteobacterial lineage (Rao et al., 2015), and their presence here in *Microscilla*, a genus in the Bacteroidetes class, suggests that the ability to both produce and sense NO is more widespread in marine bacteria than previously thought. The NO signaling pathway can be triggered by *Trichodesmium* to induce biofilm formation (Rao et al., 2015), and concomitant quorum sensing is used to modulate a range of responses including N_2 fixation, siderophore and enzyme biosynthesis, as well as motility, aggregation and biofilm formation in other systems (Bassler, 2012).

Finally, homologs were detected for the efflux and sensing of auxin, a family of plant hormones implicated in stimulating a suite of plant processes, including growth and division, which can be synthesized by both plants and bacteria (Spaepen and Vanderleyden, 2011)(Figure 2.5). In the microbiome, the *Microscilla* epibionts contained an auxin responsive gene, and both *Microscilla* and the Rhodospirillales contained genes for the degradation of phenylacetic acid (Figure 2.5), an auxin-like compound produced by certain plant-associated bacteria with known antimicrobial activity (Somers et al., 2005). For heterotrophic bacteria, auxin compounds have been shown to induce resistance to stress agents and biofilm formation, upregulation of the tricarboxylic acid cycle and amino-acid biosynthesis, and increased enzyme activity (Spaepen and Vanderleyden, 2011)[and references therein]. The known roles of auxin, NO and AHLs in cell–cell signaling suggests that these gene targets could be used to query potential interactions, such as mutualism or parasitism within the holobiont. Overall, the presence and known activation of a number of different signaling pathways in the *Trichodesmium* microbiome suggests that physiological activities may be modulated and coordinated within the holobiont, and could vary depend-

ing on the relative abundance of epibionts and with fluctuations in taxonomic composition of the microbiome or the fitness of the *Trichodesmium* host.

Another set of OGs enriched in the microbiome were identified as members of the glycoside hydrolase (GHs) family of carbohydrate active enzymes (CAZymes). The majority of GH OGs were unique to epibionts with over 80% (21 out of 26) of the unique GH-identified OGs composed solely of epibiont proteins (Supplemental Table A.3). These GHs are predicted to be active against a diverse suite of compounds including xylans, lichenin, chitin, xylose and arabinose (Supplemental Table A.3). GH enzymes have been implicated in the bacterial processing of algal-derived polysaccharides (Teeling et al., 2012) and the processing of *Trichodesmium* exudates by epibionts has been suggested to affect the rate of carbon and nitrogen transfer to the deep ocean (Herbst and Overbeck, 1978; Hmelo et al., 2012; Nausch, 1996). As such, microbiome metabolism and its potential modulation through cell–cell signaling, may influence organic matter processing within the holobiont, competition for resources, and the concomitant impact on the fate of carbon and nitrogen. The abundance of these GHs and other epibiont-only OGs representing key functions highlights how the microbiome of *Trichodesmium* expands holobiont functional diversity. The extent to which this metabolic potential increases or decreases host fitness may influence the cycling of both nitrogen and carbon.

2.4.3 Resource niche partitioning in the *Trichodesmium* holobiont

Iron and phosphorus bioavailability have been shown to influence the distribution and activities of diazotrophic communities (Deutsch et al., 2007; Krishnamurthy et al., 2007), and both are major drivers of *Trichodesmium* abundance and N_2 fixation rate in the Atlantic (Moore et al., 2008, 2009; Sañudo-Wilhelmy et al., 2001; Sohm and Capone, 2006; Webb et al., 2007; Wu et al., 2000). To evaluate the contribution of the microbiome to potential iron and phosphorus cycling in the holobiont, OGs were additionally screened for the presence of key processes including the metabolism of organic phosphorus, processing of reduced phosphorus, and metabolism and transport of organic iron. The alkaline phosphatase enzyme hydrolyzes phosphate from phosphoester bond dissolved organic phosphorus, and is central to microbial phosphorus bioavailability in the western North Atlantic

(Mahaffey et al., 2014), where dissolved organic phosphorus concentration is higher than the inorganic phosphate concentration (Lomas et al., 2010). Homologs for the alkaline phosphatases PhoA, PhoD and PhoX were found in both *Trichodesmium* and the microbiome (Figure 2.5). PhoX was found in *Trichodesmium*, as well as all epibionts with the exception of one Rhodospirillales (bin 12). Within the core microbiome, PhoX was more common than PhoA, which was only found in bins 5, 6, 7 and 10, and PhoD, which was only detected in the *Microscilla* genome bins (bins 4 and 6). This observation agrees with previous molecular surveys that found PhoX to be more prevalent than PhoA among oligotrophic planktonic marine bacteria (Sebastian and Ammerman, 2009).

OGs with proteins for the uptake and metabolism of reduced phosphorus were also present in both *Trichodesmium* and the microbiome (Figure 2.5). The phosphite dehydrogenase PtxD was detected in *Trichodesmium* and the microbiome, and a homolog of an alternate phosphite dehydrogenase enzyme was detected in the Rhodospirillales and in *Microscilla* in the form of HtxA, a gene characterized in *Pseudomonas stutzeri* that has a similar function to PtxD (Metcalf and Wolfe, 1998) (Figure 2.5). When paired to the ABC transporter PtxABC, PtxD allows *Trichodesmium* to use phosphite as a sole phosphorus source (Polyviou et al., 2015). In addition, *Trichodesmium* consortia at station 9 took up radiolabeled phosphite (Van Mooy et al., 2015) and the data herein suggest that phosphite is a substrate for both *Trichodesmium* and multiple members of the microbiome, which could contribute to resource competition between holobiont members.

Other pathways to metabolize reduced phosphorus, like those involved in phosphonate transport (Phn E) and hydrolysis were detected in both *Trichodesmium* and the microbiome. PhnJ, a component of the broad specificity C-P lyase, was found in *Trichodesmium*, consistent with its presence in the *T. erythraeum* IMS101 genome (Dyhrman et al., 2006), two Rhodospirillales epibionts and the Gammaproteobacterium (Figure 2.5). Finally, phosphoenolpyruvate mutase (PepM), the enzyme that catalyzes formation of a carbon–phosphorus bond through the conversion of phosphoenolpyruvate to phosphonopyruvate and thought to be a major source of biological phosphonate production (Metcalf and van der Donk, 2009), was detected in Microscilla bin 4, an unbinned epibiont species, as well as one of the *Trichodesmium* genome bins (Figure 2.5). Previous research showed

elevated rates of phosphonate compound biosynthesis at station 5 (Van Mooy et al., 2015), and the metagenome findings suggest that phosphonate biosynthesis may be driven by both *Trichodesmium* and members of the microbiome, in particular *Miroscilla* in this system. Overall, the abundance and diversity of phosphorus-related functions shared between *Trichodesmium* and the microbiome is consistent with the importance of active phosphorus cycling within the holobiont, as may be expected given the known role of phosphorus in limiting *Trichodesmium* populations in this region (Dyhrman et al., 2002; Sañudo-Wilhelmy et al., 2001).

Similar to phosphorus-related OGs, iron-related OGs were also enriched in the holobiont (Supplemental Table A.2), however, the majority of these were present only in epibionts, and with subtle differences in specific genes contributing to each OG between different members of the microbiome (Figure 2.5). Homologs to a siderophore biosynthesis protein were found exclusively in one *Microscilla* (bin 6) (Figure 2.5). In addition, homologs to heme group biosynthesis and exporter proteins were made up exclusively from those found within the *Microscilla* and Rhodospirillales (Figure 2.5). Homologs of the TonB-dependent transporter system predicted to transport heme- and siderophore-bound iron, were detected in *Trichodesmium* and all epibiont genome bins (Figure 2.5; Supplemental Table A.2), suggesting that while not all holobiont members produce these molecules, they could be taken up and potentially utilized by the core microbiome and *Trichodesmium*. *Trichodesmium* is rarely maintained in axenic culture, so direct comparisons of iron uptake between epibionts and the host are challenging. Previous work found that siderophore-bound iron was not as accessible to the *Trichodesmium* holobiont as inorganic forms, at least relative to select epibiont isolates, which could readily utilize most iron sources (Roe et al., 2012). However, uptake of different iron forms has been shown to vary between colonies (Achilles et al., 2003), suggestive of the fact that alterations in epibiont composition could influence the accessibility of different iron forms to the holobiont, and could increase competition for this nutrient in low iron, oligotrophic conditions. Overall, the production and uptake of different organic iron complexes by epibionts could shape the dynamics of iron-dependent processes like N_2 fixation, particularly in the many low iron regions of the ocean. In contrast to the redundant phosphorus-related functions, the epibiont-only iron functions highlight

how the microbiome contributes diverse and unique functions that have the potential to uniquely modulate the geochemical microenvironment within colonies. Iron is relatively more abundant in the study region than in other areas like the North Pacific (Sohm et al., 2011), and additional surveys of the functional composition of holobiont metagenomes from diverse ocean regimes would identify the consistency of the iron and phosphorus-related OG distributions within the *Trichodesmium* holobiont.

2.5 Conclusions

The importance of microbiomes to marine metazoans is well established (Hentschel et al., 2012), but the role of the *Trichodesmium* microbiome in shaping the distribution, activities and concomitant biogeochemical impact of these consortia is still in its infancy. Here we show that the substantial majority of the metabolic potential in the *Trichodesmium* holobiont is contained within the microbiome. This finding suggests that within the microbiome, there is a palette of functional diversity that could modulate host fitness and subsequent biogeochemical impact across different environments. This study also provides an annotated holobiont metagenome that could serve as a template for future metatranscriptomic and metaproteomic investigations focused on tracking physiological activities or interactions within the holobiont. The microbiome may underpin *Trichodesmium's* success in oligotrophic systems and could be an important facet determining its resilience in a future ocean that is likely to bring increased oligotrophic conditions (Riebesell et al., 2009). Overall, these results suggest that *Trichodesmium* should not be considered in isolation, but rather studied as a dynamic microbial holobiont.

Chapter 3

Trichodesmium physiological ecology and phosphate reduction in the western Tropical South Pacific

A modified version of this chapter has been submitted as Frischkorn, K.R., Krupke, A., Guieu, C., Louis, J., Rouco, M., Salazar Estrada, A.E., Van Mooy, B.A.S., and Dyhrman, S.T. (2018). *Trichodesmium* physiological ecology and phosphate reduction in the western tropical South Pacific. *Biogeosciences*, In review).

3.1 Abstract

N_2 fixation by the genus *Trichodesmium* is predicted to support a large proportion of the primary productivity across the oligotrophic oceans, regions that are considered among the largest biomes on Earth. Many of these environments remain poorly sampled, limiting our understanding of *Trichodesmium* physiological ecology in these critical oligotrophic regions. *Trichodesmium* colonies, communities that consist of the *Trichodesmium* host and their associated microbiome, were collected across the oligotrophic western tropical South Pacific (WTSP). These samples were used to assess host clade distribution, host and microbiome metabolic potential, and functional gene expression, with a focus on identifying Trichodesmium physiological ecology in this region. Gene sets related to phosphorus, iron, and phosphorus-iron co-limitation were dynamically expressed across the WTSP transect, suggestive of the importance of these resources in driving *Trichodsmium* physiological ecology in this region. A gene cassette for phosphonate biosynthesis was detected in *Trichodesmium*, the expression of which co-varied with the abundance of *Trichodesmium* Clade III, which was unusually abundant relative to Clade I in this environment. Coincident with the expression of the gene cassette, phosphate reduction to phosphite and low molecular weight phosphonate compounds was measured in *Trichodesmium* colonies. The expression of genes that enable use of such reduced phosphorus compounds were also measured in both *Trichodesmium* and the microbiome. Overall, these results highlight physiological strategies employed by consortia in an undersampled region of the oligotrophic WTSP, and reveal the molecular mechanisms underlying previously observed high rates of phosphorus-reduction in *Trichodesmium* colonies.

3.2 Introduction

The oligotrophic oceans extend over approximately 70% of the Earth and are characterized by chronically low nutrient concentrations that limit primary productivity (Moore et al., 2013). Within oligotrophic marine environments, N_2 fixing microorganisms can serve as a source of "new" nitrogen that is bioavailable to other organisms. Among these marine diazotrophs, the colonial, filamentous cyanobacterium *Trichodesmium* plays a dispropor-

tionately large role in the cycling of carbon, phosphorus and nitrogen: it supplies fixed carbon through photosynthesis, was recently found to be a hotspot of phosphate reduction (Van Mooy et al., 2015), and has been estimated to be responsible for approximately half of the biologically fixed N_2 in the ocean (Bergman et al., 2013; Capone et al., 1997). As such, the efficiency of the biological pump in sequestering carbon in the deep ocean is dependent in part on the distribution and activities of diazotrophic organisms like *Trichodesmium*, and an understanding of how this organism's physiology and ecology varies across diverse environments is a critical aspect of understanding present, and future, global biogeochemical cycles.

Diazotrophy frees *Trichodesmium* from nutrient limitation by nitrogen. As such, the distribution and activities of this cyanobacterium are predominantly influenced by the availability of phosphorus and iron in the surface ocean, which vary depending on the ocean basin and its proximity to supply of these resources (Moore et al., 2013; Sohm et al., 2011). Evidence of the intense competition for phosphorus and iron is evident in the suite of physiological strategies that this organism is known to employ. These strategies include the production of transporters and enzymes that take up and hydrolyze diverse organic and reduced phosphorus compounds (Dyhrman et al., 2006; Orchard et al., 2009; Polyviou et al., 2015), or enable the uptake and storage of organic and inorganic iron (Polyviou et al., 2018; Snow et al., 2015b). The genes encoding these functions are expressed in situ across diverse environments, indicating that competition for these resources is a critical aspect of *Trichodesmium* physiology (Chappell et al., 2012; Dyhrman et al., 2006; Rouco et al., 2018). Recent evidence from culture studies also suggests that *Trichodesmium* employs a unique set of physiological strategies to cope with co-limitation of phosphorus and iron that differs from that of either resource alone (Walworth et al., 2017b).

Survival in oligotrophic environments might also be enabled by biological interactions within *Trichodesmium* colonies. *Trichodesmium* has long been known to occur with tightly associated bacteria that are unique from those free-living in the water column (Hmelo et al., 2012; Paerl et al., 1989). Recent evidence suggests that these interactions are ubiquitous and taxonomically conserved across ocean basins (Lee et al., 2017; Rouco et al., 2016a) and that this epibiotic bacterial community, referred to as the *Trichodesmium* microbiome,

contains a large amount of metabolic potential that exceeds and complements that of the *Trichodesmium* host in populations from the western North Atlantic (Frischkorn et al., 2017). Coordinated gene expression patterns within the holobiont (*Trichodesmium* and its microbiome) suggest an interdependence of the microbiome on host-derived fixed carbon, N_2 and vitamins, and suggests microbiome respiration could create conditions that favour continued diazotrophy and photosynthesis (Frischkorn et al., 2018; Paerl and Bebout, 1988). The stability of these relationships in the future ocean is unknown, but they are likely to change. For example, incubation of cultured *Trichodesmium* colonies with an elevated carbon dioxide concentration resulted in significant changes in microbiome nutritional physiology (Lee et al., 2018). These microbiome changes have the potential to alter the amount of fixed N_2 and carbon that transfer from the colony to the environment at large. Overall, the continued appreciation of the importance of the microbiome in *Trichodesmium* ecology underscores that investigations must consider these microbial communities as a holobiont in order to fully understand and predict their role in the future environment.

Geochemical drivers of *Trichodesmium* distribution and N_2 fixation are increasingly well characterized in regions of the ocean where either phosphorus or iron are limiting such as the North Atlantic and North Pacific Subtropical Gyre (Rouco et al., 2018; Sañudo-Wilhelmy et al., 2001; Sohm et al., 2011). The western tropical South Pacific (WTSP) represents and understudied region of the world's oceans (Bonnet et al., 2018) with conditions characterized by chronically low concentrations of both iron and phosphate (Moore et al., 2013; Sohm et al., 2011). Despite low resources, this region can support high levels of N_2 fixation, with rates exceeding 700 μmol $m^{-}2d^{-1}$ where this resource transfers across diverse ecological groups and ultimately supplies up 90% of to the photic zone with new nitrogen (Bonnet et al., 2017; Caffin et al., 2018a,b). In this study, metagenomic and metatranscriptomic sequencing was leveraged along with taxonomic distribution, physiological activities and geochemical measurements to better understand *Trichodesmium* physiological ecology in an under-sampled but important region of the oligotrophic ocean.

3.3 Materials and methods

3.3.1 Biogeochemical analyses

Samples were collected across a transect of the western Tropical South Pacific (WTSP) during the OUTPACE cruise (Oligotrophy to UlTra-oligotrophy PACific Experiment, DOI: http://dx.doi.org/10.17600/15000900) (Moutin et al., 2017) aboard the R/V *L'Atalante* during austral summer (February – April) of 2015 (Figure 3.1). Samples for nutrient analyses were collected using a Titanium Rosette mounted with 24 Teflon-coated 12L GoFlos and operated along a Kevlar cable. Samples were filtered directly from the GoFlos through 0.2 μm cartridges (Sartorius Sartrobran-P-capsule with a 0.45 μm prefilter and a 0.2 μm final filter) inside a clean van and analysed for dissolved inorganic phosphorus (DIP) and total dissolved iron concentrations (DFe). DIP was analysed directly on board using a 2 m length Liquid Waveguide Capillary Cells (LWCC) made of quartz capillary following a method described previously (Pulido-Villena et al., 2010). Briefly, LWCC was connected to a spectrophotometer and the measurements were performed in the visible spectrum at 710 nm. The 2 m length LWCC allowed for a detection limit of 1 nM and a relative standard deviation of less than 10%. DFe concentrations were measured by flow injection with online preconcentration and chemiluminescence detection using the exact protocol, instrument, and analytical parameters as previously described (Blain et al., 2008).

Water column phosphate uptake rate was determined as previously described (Van Mooy et al., 2015), and is briefly outlined here. First, 50 mL aliquots of whole seawater collected in Niskin bottles were decanted into acid-washed polycarbonate vials. Next, 1 μCi of carrier-free ^{33}P-phosphoric acid was added to the bottles, which represented an amendment of approximately 10 pmol L^{-1} of phosphoric acid. Then, bottles were incubated in an on-deck incubator for 2-4 hours. Finally, the seawater in the bottles was filtered through a 25 mm diameter, 0.2 μm pore size polycarbonate membrane, and the radioactivity of the membrane was determined by liquid scintillation counting. The phosphate turnover rate in each incubation was calculated as the quotient of the ^{33}P-radioactivity of the membrane and the total ^{33}P-radioactivity of the amendment, divided by the time duration of the incubation. The phosphate uptake rate was calculated as the product of the phosphate

turnover and the DIP concentration. Data were visualized and contoured using Ocean Data View 4.6.2 with the DIVA grid method (R. Schlitzer; http://odv.awi.de).

3.3.2 *Trichodesmium* clade sampling and analysis

Samples for *Trichodesmium* clade distribution analysis were obtained at select short duration stations across the transect from water depths ranging from 5 m to 150 m using 12 L of water for each depth obtained from a Rosette sampling device filtered through 47 mm 10 μm pore size polycarbonate filters. Filters were flash frozen and stored in liquid nitrogen until processing. Quantification of absolute cell numbers in these samples was performed following a previously described protocol (Rouco et al., 2014). Briefly, DNA was extracted from filters and the abundance of *Trichodesmium* clade I (which encompasses *T. thiebautii, T. tenue, T. hildebrandtii,* and *T. spiralis*) and III (which encompasses *T. erythraeum* and *T. contortum*), was determined with quantitative polymerase chain reactions (qPCR) targeting the *rnpB* gene using clade specific primer sets (Chappell and Webb, 2010). Amplification of standards, no template controls (RNase-free water), and field samples were run in triplicate on a Bio-Rad CFX96 Real-Time System C1000 Touch Thermal Cycler using Bio-Rad SYBR Green SuperMix (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Standard curves were generated from DNA extracts performed on filters with known concentrations of *T. erythraeum* IMS101 and *T. thiebautii* VI-I. Concentrations were previously determined by 10 replicates of cell counting using a Sedgwick Rafter slide (Rouco et al., 2014). Reactions were run in final volumes of 25 μL, encompassing 12.5 μL SuperMix, 2 μL template, 9.5 μL sterile water, and 200 nmol L^{-1} forward and reverse primers. Reaction conditions were as follows: 2 minutes at 50 °C, 10 minutes at 95 °C; 40 cycles of 15 seconds at 95 °C, and 1 minute at 55 °C with a fluorescence measurement. Resulting CTs were averaged across the triplicates and compared against the standard curve to calculate the abundance of the *rnpB* gene, which is interpreted as absolute cell number. Clade distribution data was displayed and contoured using Ocean Data View 4.6.2 with the DIVA grid method (R. Schlitzer; http://odv.awi.de).

3.3.3 *Trichodesmium* colony sampling

Trichodesmium colonies were sampled across the transect at approximately the same time (between 8 AM and 10:50 AM local time). *Trichodesmium* samples were obtained with six manual hauls of a 130 μm mesh size net hand towed through surface sea water. The total time for six hauls of the net tow was approximately 15 minutes and likely filtered thousands of liters of seawater. Colonies were skimmed from the concentrated net tow sample with a Pasteur pipette from the surface layer of net towed samples and then washed two times with 0.2 μm sterile-filtered surface seawater on 5 μm, 47 mm polycarbonate filters with gentle vacuuming (<170 mbar) to remove non-tightly associated microorganisms. All colony samples were cleaned and processed within 15 minutes of collection. Filters with colonies were flash frozen and stored in liquid nitrogen for DNA or RNA extraction.

3.3.4 Phosphate reduction in *Trichodesmium* colonies

Phosphate uptake and synthesis rates of low-molecular-weight (LMW) reduced phosphate (P(+3)) compounds in *Trichodesmium* colonies were determined as described previously (Van Mooy et al., 2015). Briefly, phosphate uptake by *Trichodesmium* colonies was determined by filling acid-washed polycarbonate 50 mL bottles with filter-sterilized surface seawater and approximately 20 *Trichodesmium* colonies. Incubation in on-deck incubators and measurement by liquid scintillation counting proceeded as previously described above for the whole water analyses. In parallel, to measure the synthesis rates of LMW P(+3) compounds, at stations SD2, LDA, SD9, SD11, and LDB *Trichodesmium* incubations were not immediately measured by liquid scintillation counting. Instead, colonies were placed in a cryovial containing 1 mL of pure water and flash frozen in liquid nitrogen. These samples were then transported to the lab ashore and subjected to numerous freeze-thaw cycles to extract intracellular LMW P(+3) compounds. The LMW P(+3) compounds in the extracts were then isolated by preparative anion chromatography. Two fractions were collected in retention time windows consistent with retention times of pure standards of 1) methylphosphonic acid, 2-hydroxy ethylphosphonic acid, and 2-amino ethylphosphonic acid; and 2) phosphorous acid. The ^{33}P radioactivity in these two operationally-defined fractions is ascribed to LMW phosphonates and phosphite, respectively.

3.3.5 DNA extraction and metagenome sequencing

Genomic DNA was extracted from *Trichodesmium* colony samples (approximately 40 colonies per sample) obtained from one day each during LDA and LDB, as well as at station SD5 (Figure 3.1) using the MoBio Power Plant Pro DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA, USA) following the manufacturer instructions. Genomic DNA extracts were sequenced at the Argonne National Lab (Lemont, IL, USA) following a *Trichodesmium* consortium protocol previously described (Frischkorn et al., 2017). Briefly, DNA was sheared with a Covaris Sonicator (Woburn, MA, USA), transformed into libraries with WaferGen Apollo324 automated library system (Clonetech Laboratories, Mountain View, CA, USA) and Illumina compatible PrepX ILMN DNA kits (San Diego, CA, USA) following manufacturer instructions. An average insert size of approximately 750 base pairs was targeted. Sage BluePippin (Beverly, MA, USA) was used to size select libraries prior to sequencing all three samples split across one 2 x 100 bp lane of the Illumina HiSeq2000. Metagenomic reads from these three samples are available on the NCBI Sequence Read Archive under BioProject number PRJNA435427.

3.3.6 Metagenomic assembly and analysis

Raw reads were trimmed assembled into scaffolds and subsequently analysed following a protocol previously reported (Frischkorn et al., 2017) and summarized here in an assembly and analysis methods pipeline (Supplemental Figure B.1). Briefly, reads were trimmed with Sickle (https://github.com/najoshi/ sickle), converted into fasta format, merged together and co-assembled with IDBA-UD (Peng et al., 2012) so as to create a South Pacific *Trichodesmium* holobiont genomic template to which future metatranscriptomic reads could be mapped. Assembled scaffolds were partitioned between *Trichodesmium* and heterotrophic bacteria (hereafter referred to as the microbiome) and after clustering into genome bins using MaxBin 2.0 with default parameters (Wu et al., 2015). The taxonomic partitioning of binned scaffolds was carried out after translation of each scaffold into protein coding genes with the metagenomic setting of Prodigal (Hyatt et al., 2010), annotation of resultant proteins with the blastp program of DIAMOND (Buchfink et al., 2015) against the NCBI nr database, and classification with MEGAN6 (Huson et al., 2013) based on the phylogenetic

classification of the majority of proteins within a genome bin. Scaffolds from bins identified as phototrophic bacteria or eukaryotes were excluded from further analyses. Functional annotations for translated proteins in the *Trichodesmium* and microbiome identified genome bins were obtained by DIAMOND against the UniRef90 database (Suzek et al., 2007) with an e-value cut-off of 1 x 10^{-3}. Functional annotation was also carried out using the Kyoto Encyclopedia of Genes and Genomes (KEGG) with the online Automatic Annotation Server using the bi-directional best-hit method, the GHOSTX search program, and the prokaryote representative gene set options. KEGG definitions were obtained from the modules within the "Pathway module" and "Structural complex" categories and the submodules therein.

Proteins from the merged assembly were also clustered into gene families of similar function or orthologous groups (OGs) following a previously reported pipeline (Frischkorn et al., 2017) and summarized in Supplemental Figure B.1. Briefly, reciprocal blast of translated proteins greater than 70 amino acids were performed with the DIAMOND blastp program. Blast results were processed using the program MCL (Markov cluster algorithm) set to an inflation parameter or 1.4. UniRef was used for the consensus functional annotation of each OG. The final UniRef annotation and KEGG annotations for each OG represent the individual annotation that the majority of proteins within that OG were assigned to. Homologs to the phosphonate biosynthesis gene phosphoenolpyruvate phosphomutase (*ppm*) were found by screening against manually annotated and reviewed Ppm proteins from the Swiss-Prot database. These verified proteins were aligned with Muscle version 3.8.425 with default parameters (Edgar, 2004), converted into HMM profiles with hmmbuild and hmmpress and used as a the database for hmmsearch, all using HMMER version 3.1 (Eddy, 1995). This HMMER approach was used to screen proteins generated from this study, as well as a previously published *Trichodesmium* consortia metagenome assembly from the western tropical North Atlantic Ocean (Frischkorn et al., 2017) and protein sequences derived from the genome sequence of *Trichodesmium erythraeum* IMS101 available through the Joint Genome Institute. Sequence alignments were visualized using Geneious version 11.0.3 (Kearse et al., 2012) and important residues were obtained from previous crystal structure analyses (Chen et al., 2006). A reassembly of *Trichodesmium*-identified bins was performed to lengthen scaffolds in an attempt to provide genomic context to the ppm-containing scaffold. The

subset of all metagenomic reads mapping to *Trichodesmium* bins was selected using Samtools (Li et al., 2009) and the subseq program in Seqtk (https://github.com/lh3/seqtk) and then reassembled using IDBA-UD as described previously. Maximum likelihood phylogenetic analysis of Ppm and related proteins was performed using the FastTree plugin in Geneious with default settings (Price et al., 2010), following a protocol previously employed for the annotation of environmental Ppm proteins (Yu et al., 2013). The sequences used to generate the tree were composed of proteins with homology to the identified *Trichodesmium* Ppm proteins, as well as similar sequences pulled from the NCBI nr database after online blastp analysis. The Interactive Tree of Life program was used to edit phylogenetic trees (Letunic and Bork, 2016). Proteins were also screened for MpnS, a protein that produces methylphosphonate, using the same protocol described above.

3.3.7 RNA extraction and metatranscriptome sequencing

Prokaryotic RNA was extracted and sequenced from *Trichodesmium* colony samples obtained from SD2, SD6 and SD9 as well as colonies collected on three separate days at LDB (approximately 40 colonies per sample) (Figure 3.1) following a protocol described previously (Frischkorn et al., 2018). Briefly, the Qiagen RNeasy Mini Kit (Qiagen, Hildern, Germany) was used to extract total RNA following manufacturer instructions, with the addition of 5 minutes of bead beating with approximately 500 μL 0.5 mm zirconia/silica beads after addition of Buffer RLT. On column DNase digestion (RNase-free DNase Kit, Qiagen) was performed. A MICROBEnrich Kit (ThermoFisher Scientific, Waltham, MA, USA) was used to enrich the prokaryotic RNA fraction and ribosomal RNA was removed with the Ribo-Zero Magnetic kit for bacteria (Illumina), both following manufacturer instructions. Concentration and integrity of mRNA was assessed using a BioAnalyzer and the RNA 600 Nano Kit (Agilent Technologies, Santa Clara, CA, USA). Library preparation and sequencing was performed at the JP Sulzberger Genome Center at Columbia University. Libraries were generated with the Illumina TruSeq RNA sample preparation kit. Samples were chemically fragmented using the Fragment, Prime, Finish Mix reagent (Illumina) which generates fragments of 140-220 bp. An Illumina HiSeq 2500 was used to sequence 60 million paired end 100 bp reads for each sample. Metatranscriptomic reads from these

six samples are available on the NCBI Sequence Read Archive under BioProject number PRJNA435427.

3.3.8 Metatranscriptomic sequence analysis

Metatranscriptomic reads were trimmed, normalized and mapped as previously described (Frischkorn et al., 2018) and are summarized in the analysis and assembly pipeline (Supplemental Figure B.1). Briefly, raw reads were pre-processed following the Eel Pond Protocol for mRNAseq (Brown et al., 2013a). Cleaned reads were mapped with RSEM (paired-end and bowtie2 options selected) (Li and Dewey, 2011) to the protein coding regions of the metagenomic scaffolds previously partitioned across *Trichodesmium* and the microbiome with an average of 10 million reads mapping per sample. Read counts were summed separately for *Trichodesmium* and microbiome fractions for all genes in an OG. Counts were normalized in each sample by calculating the transcript reads mapped per million (TPMs) separately for the *Trichodesmium* and microbiome fractions. Comparisons of relative enrichment across sets of nutrient responsive genes were made using Kolmogorov-Smirnov tests to examine the null hypothesis that the expression of gene sets at a given station did not deviate significantly from the average expression of that set across the transect. Prior to testing, TPM expression values for each OG were normalized to the average abundance of that OG across the six samples. This normalization equalized the relative contribution of individual OGs to the gene set as a whole, thereby avoiding bias caused by highly expressed individual OGs. P values less than 0.05 were considered significant. Pairwise correlation coefficients between clade abundance, DIP and OG expression (TPMs) were calculated using the cor function in R, and p values less than 0.05 were considered significant. In the case of the OG expression, LDB samples from separate days were averaged. The data did not span enough coincident samples to do pairwise correlations between OG expression and iron. *Trichodesmium* OG TPMs were hierarchically clustered with the Broad Institute Morpheus program (https://software.broadinstitute.org/morpheus/) using the "one minus Pearson correlation" metric and the "Average" linkage method. Only OGs with an average expression greater than 2 TPM across all samples were included in this analysis.

3.4 Results

3.4.1 Biogeochemistry

Across the study transect the DIP concentration in surface water (10 m depth, with the exception of the sample from LDA which was collected at 30 m) ranged between 2.3 nmol L^{-1} and 230 nmol L^{-1} (Figure 3.1, Supplemental Table B.1) and averaged 36 nmol L^{-1}. The phosphate turnover time in the water column microbial community was variable across the transect ranging between approximately 2 hours and 800 hours (Figure 3.1, Supplemental Table B.1), averaging approximately 220 hours across all stations sampled. The water column phosphate uptake rate was similarly variable, ranging from 0.006 nmol $L^{-1}hr^{-1}$ to 0.68 nmol $L^{-1}hr^{-1}$ with lower uptake measured at stations where turnover time was high and vice versa (Figure 3.1). Iron concentrations in the surface water (10 m depth, with the exception of the sample from SD12 which was collected at 30 m) ranged between 0.21 nmol L^{-1} and 1.16 nmol L^{-1} (Figure 3.1, Supplemental Table B.1) and averaged 0.6 nmol L^{-1}.

Trichodesmium (combined cell counts of Clade I and Clade III) was detected at every station and at all depths sampled across the transect with a maximum estimated concentration of over 172,000 cells L^{-1} at 11 m at SD6 (Figure 3.2a). Overall, abundance was markedly greater at the stations in the western half of the transect (stations west/left of 170 °W and LDB (Figure 3.2a). In this western region of the transect, *Trichodesmium* was concentrated in the surface ocean between approximately 5 - 20 m depth, where the average concentration was approximately 53,000 cells L^{-1}. To the east/right of station LDB and the transition into the ultra-oligotrophic region, the concentration of *Trichodesmium* dropped and the average concentration at the surface between approximately 5 - 20 m was 75 cells L^{-1} (Figure 3.2a). In addition to overall abundance of *Trichodesmium*, the contribution of Clade I (*T. thiebautii, T. tenue, T. hildebrandtii*, and *T. spiralis*) and Clade III (*T. erythraeum* and *T. contortum*) to the total *Trichodesmium* abundance at each station was determined (Figure 3.2b). Across the transect, the *Trichodesmium* communities were dominated by Clade I which made up approximately 80% of the cells measured on average at each station, while Clade III made up approximately 20% on average (Figure 3.2b). Al-

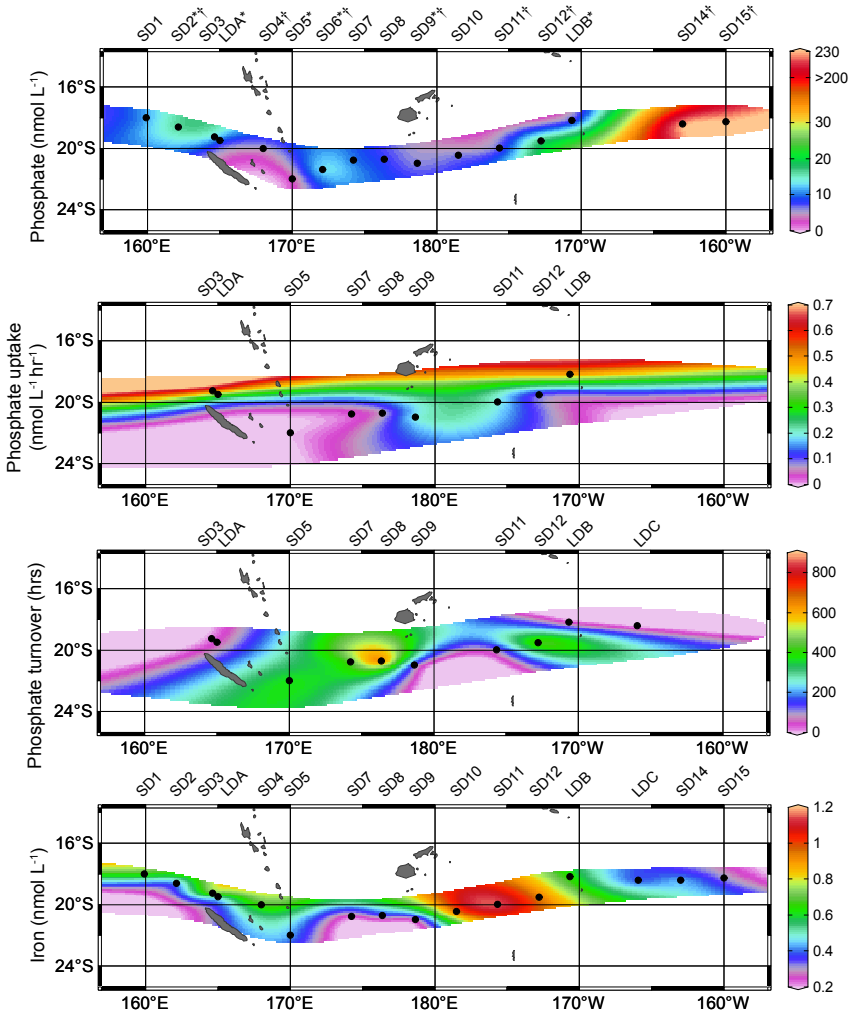

Figure 3.1: Surface water column dissolved inorganic phosphorus (DIP) concentration, community phosphate uptake rate, community phosphate turnover, and iron concentration measured at stations across the OUTPACE (Oligotrophy to UlTra-oligotrophy PACific Experiment) transect during austral summer (February – April) of 2015. All samples were obtained from 10 m depth, with the exception of the DIP measurement from LDA and the iron measurement from SD12 (30 m). Numbers above and below indicate the short duration (denoted SD) or long duration (denoted LDA, LDB or LDC) stations where samples were obtained across the transect. In the top panel, a star indicates stations where metagenomic or metatranscriptomic samples were obtained and a cross indicates stations where *Trichodesmium* clade distribution samples were obtained.

though Clade I was dominant overall, the percentage of Clade III rose to nearly 50% of the measured *Trichodesmium* community in some samples, and was highest between 4 and 40 m at SD4, SD6 and SD11 (Figure 3.2b). Clade III was not detected at any depth in the ultra-oligotrophic subtropical gyre stations (SD14 and SD15) (Figure 3.2b).

A merged metagenomic assembly of *Trichodesmium* consortia reads from selected stations across the OUTPACE transect (Figure 3.1) yielded 801,858 scaffolds in total. Taxonomic binning partitioned scaffolds into 48 genome bins with similar read coverage and tetranucleotide frequency. After phylogenetic analysis with MEGAN6, 18 of the bins were classified as *Trichodesmium*, 23 as heterotrophic bacteria, while the remaining bins had the majority of their proteins phylogenetically classified as Eukaryotes or other photosynthetic cyanobacteria. Subsequently, all *Trichodesmium*-identified scaffolds were merged and considered as the *Trichodesmium* fraction of the WTSP consortia sampled. Similarly, all heterotrophic bacteria-identified bins were merged together and considered as the microbiome fraction. Together, these taxonomically verified scaffolds were translated into 198,156 proteins, which clustered into 75,530 gene families of putatively similar function, or orthologous groups (OGs). Within the WTSP consortia, *Trichodesmium* and their microbiome possessed 9,790 and 68,538 OGs respectively. The majority of these OGs were unique to the microbiome, with 2,798 (3.7%) of the total OGs composed of proteins found in both the *Trichodesmium* and microbiome genome bins.

Functional annotation of OGs showed that the microbiome contained nearly 10 times more unique KEGG IDs (Supplemental Figure B.2). The greatest differences in functional gene capacity between *Trichodesmium* and the microbiome were found in the Environmental Information Processing and Carbohydrate and Lipid Metabolism modules (Supplemental Figure B.2). Unique microbiome KEGG submodules included glycan, fatty acid, and carbohydrate metabolism functions. In the Environmental Information Processing category the microbiome possessed unique functions pertaining to peptide, nickel, phosphate, amino acid, and ABC transporters as well as an enrichment in proteins related to bacterial secretion systems (Supplemental Figure B.2).

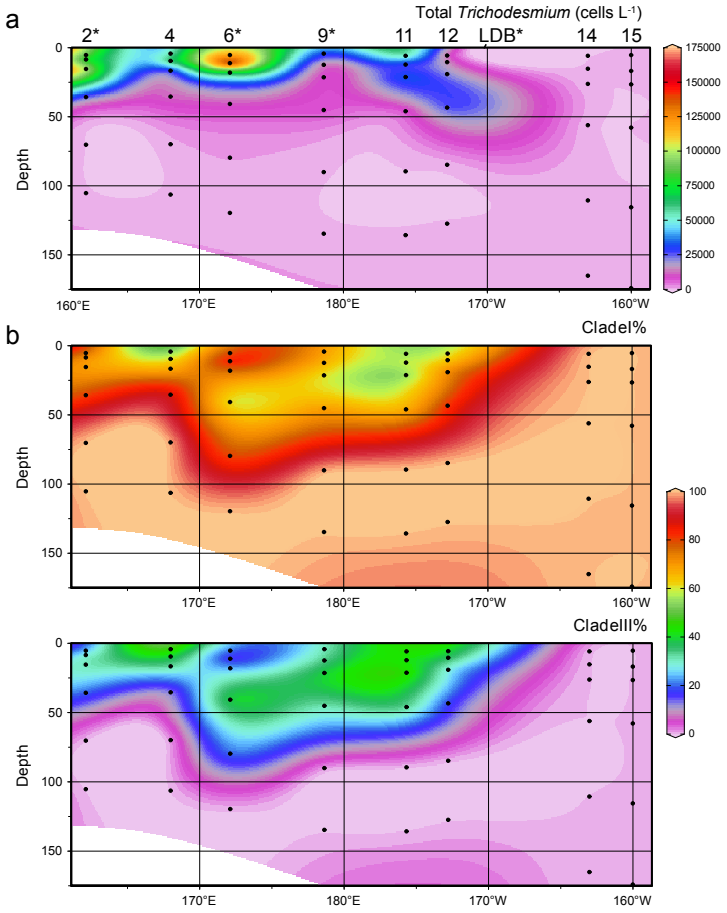

Figure 3.2: Abundance and clade distribution of *Trichodesmium*. Black dots denote the depths at which samples were taken, while station number is indicated above the panels. (a) Concentration of total *Trichodesmium* cells estimated from qPCR of the *rnpB* gene (cells L^{-1}). (b) Relative proportion of *Trichodesmium* Clade I (top panel) and Clade III (bottom panel) across the transect. Clade I includes *T. thiebautii, T. tenue, T. hildebrandtii,* and *T. spiralis*. Clade III includes *T. erythraeum* and *T. contortum*. Stars indicate stations where metatranscriptomic sequences were sampled. Clade samples were not obtained at LD stations.

3.4.2 Expression profiling of WTSP *Trichodesmium* consortia

A total of 7,251 *Trichodesmium* OGs and 21,529 microbiome OGs recruited metatranscriptome reads from at least one sample. Hierarchical clustering of expression patterns in KEGG annotated OGs identified variable patterns in *Trichodesmium* and the microbiome gene expression from station to station, and over three days of sampling at LDB (Figure 3.3). Submodules in the carbohydrate and lipid metabolism, genetic information processing, and nucleotide and amino acid metabolism modules had elevated expression in the SD stations relative to LDB. Similar patterns were generally observed in the expression of microbiome modules, most strikingly in the nucleotide and amino acid metabolism and environmental information processing modules. In *Trichodesmium*, exceptions to the trend included OGs in the ATP synthesis, nitrogen metabolism, phosphate and amino acid transport, two-component regulatory systems, and certain amino acid metabolism submodules, all of which had elevated relative expression during the three days of samples from LDB (Figure 3.3).

Hierarchical clustering of expression patterns in *Trichodesmium* OGs identified shifts in *Trichodesmium* gene expression from station to station, and over the three days of sampling at LDB (Figure 3.4a). The nitrogenase enzyme subunit *nifH* OG peaked at SD2 (Figure 3.4a), while a RuBisCO OG peaked at LDB 1 (Figure 3.4a). The expression of *Trichodesmium* OGs known to be responsive to low nutrient conditions also showed variation in expression pattern across the transect (Figure 3.4a). The low P responsive set was composed of OGs previously shown to have increased relative expression under conditions of phosphate stress and included the alkaline phosphatases *phoA* (Tery_3467) and *phoX* (Tery_3845; proteins identified as PhoX clustered into two separate OGs, as has been previously observed in natural populations of *Trichodesmium* (Rouco et al., 2018)), the high affinity phosphate binding protein *sphX* (which is homologous to and clustered into one OG with *pstS*) (Tery_3534), phosphite dehydrogenase *ptxD* (Tery_0368) and the carbon-phosphorus lyase gene marker *phnJ* (Tery_5000) (Dyhrman et al., 2006; Orchard et al., 2009; Polyviou et al., 2015). The iron responsive set included OGs previously shown to have increased transcript or protein expression in experimental low iron cultures and *in situ* in low iron environments and included the flavodoxins *fld1* and *fld2* which clustered into the same OG (Tery_1666,

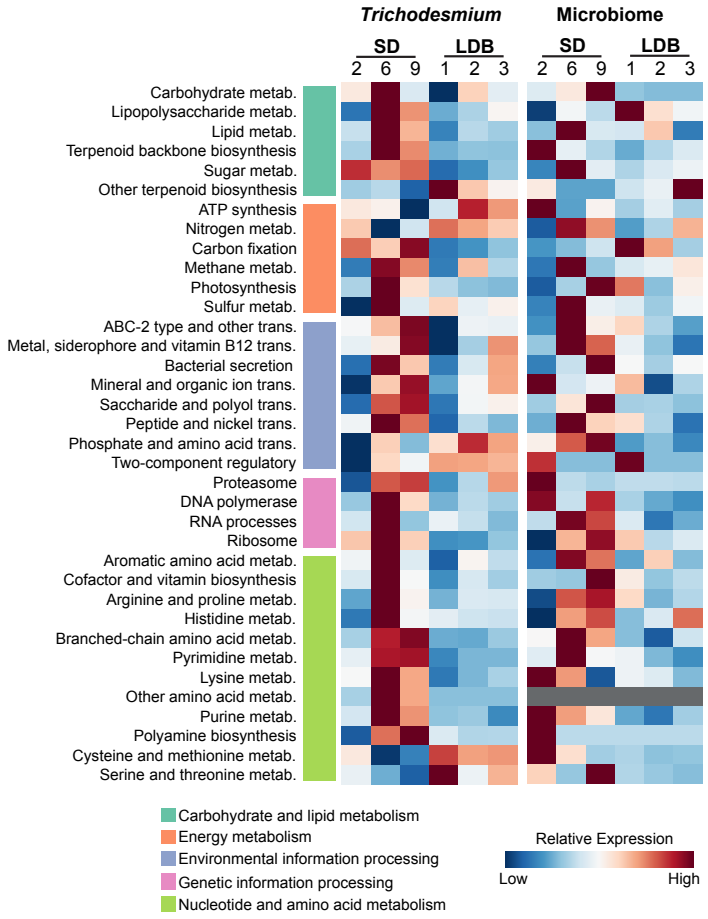

Figure 3.3: Heatmap of the summed relative expression of *Trichodesmium* and microbiome orthologous groups (OGs) belonging to KEGG modules (large categories) and submodules at the short duration (SD) and over three days at the long duration (LDB) stations. Relative expression values are normalized to row averages for either *Trichodesmium* or the microbiome for that particular OG. The carbohydrate and lipid metabolism module was simplified as follows: carbohydrate metabolism is the sum of carbohydrate metabolism, central carbohydrate metabolism, glycan metabolism and other carbohydrate metabolism submodules; lipid metabolism is the sum of fatty acid and lipid metabolism submodules, The genetic information processing module was simplified as follows: RNA processes is the sum of the RNA processing and RNA polymerase submodules. No microbiome OGs were detected in the other amino acid metabolism submodule.

Tery_2559), fructose bisphosphate aldolase class II *fbaA* (Tery_1687), the iron-stress induced protein *isiA* (Tery_1667), and the Fe-stress induced gene *idiA* (Tery_3377) (Chappell and Webb, 2010; Chappell et al., 2012; Snow et al., 2015b; Webb et al., 2001). Also assayed were a suite of OGs recently shown to be significantly enriched in cultures of *T. erythraeum* IMS101 following prolonged maintenance in co-limiting concentrations of phosphorus and iron (herein called the co-limitation responsive set) (Walworth et al., 2017b). These OGs included the flavin-containing monoxygenase *FMO* (Tery_3826) that hydrolyzes organic nitrogen, 5-methyltetrahydropteroyltriglutamate—homocysteine methyltransferase *metE* (Tery_0847), the 3-dehydroquinate synthase *aroB* (Tery_2977), and beta-ketoacyl synthase (*OXSM*, Tery_3819 and Tery_3821 which clustered into one OG). All OGs included in the three nutrient responsive *Trichodesmium* sets were expressed at each station In general, low P and low Fe responsive gene set expression patterns tracked together from sample to sample. Two exceptions to this trend were the alkaline phosphatase *phoA* and the phosphonate lyase *phnJ* (Figure 3.4a). There were no significant correlations between the low P responsive OGs and the DIP concentration. The majority of the OGs (*aroB*, *metE*, *FMO*) in the co-limitation responsive set had similar expression patterns to the majority of the low P responsive set, with the exception of *OXSM* (Figure 3.4a). Expression of the RuBisCO OG was also modulated with this pattern (Figure 3.4a). The expression of the *nifH* OG tracked with the Fe responsive set as well as *phoA*, and *OXSM* (Figure 3.4a).

The relative enrichment of the low P responsive gene set, low Fe responsive gene set, and co-limitation gene set were examined from station to station, and over the three days of sampling at LDB using Kolmogorov-Smirnov tests (Figure 3.4b). This approach estimated enrichment based on normalized expression levels of the resource responsive genes at individual stations relative to the average expression of the set across all stations sampled. In testing enrichment of the low P responsive set, only the four OGs whose expression clustered together across the transect (Figure 3.4a) were used. The signal of the low P responsive set was significantly enriched during the first day sampled at LDB, while the signal in the low Fe responsive set was significantly enriched at SD2 (Figure 3.4b). The co-limitation set was not found to significantly deviate from the average expression levels at any station sampled,

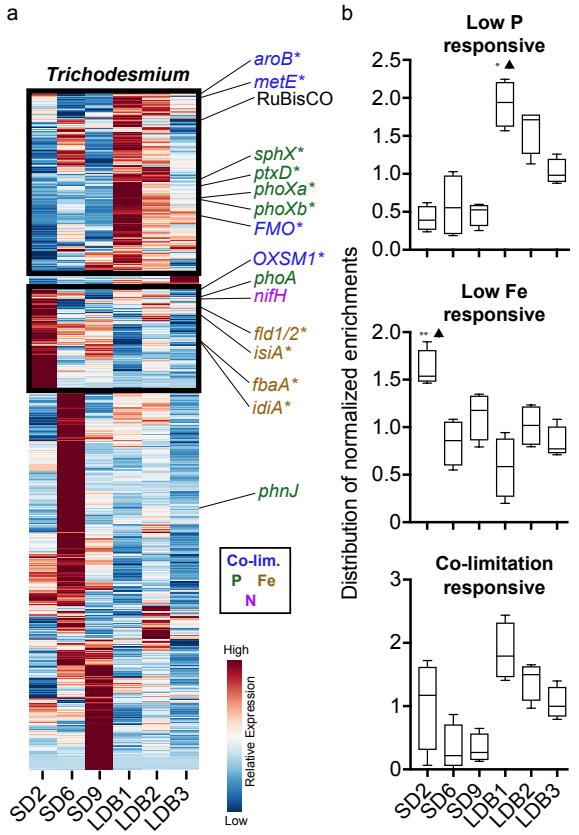

Figure 3.4: *Trichodesmium* orthologous group (OG) expression patterns. (a) Hierarchically clustered expression patterns of *Trichodesmium* OGs with key genes called out. Stars indicates OGs used in subsequent Kolmogorov-Smirnov tests. Colored text indicates phosphorus (green), iron (brown), co-limitation (blue), or nitrogen (purple) related OGs. (b) Distribution of expression patterns in OG sets known to be significantly responsive in Trichodesmium to low phosphorus (P), low iron (Fe), and P/Fe co-limiting conditions. Distributions for each set at each station were compared to the average distribution across all six stations using Kolmogorov-Smirnov tests to examine the null hypothesis that the expression of gene sets at a given station did not deviate significantly from the average expression of that set across the transect. Whiskers show the normalized enrichment level for the least and most enriched OG in that set. Boxes denote the upper and lower 25th percentiles, while the line indicates the median enrichment expression value. Asterisks indicate significance: * $p < 0.05$, ** $p < 0.005$. Black triangles denote whether significant stations were increased or decreased relative to the mean.

although the overall pattern in expression was similar to that of the low P responsive set (Figure 3.4b).

3.4.3 Evidence of phosphonate biosynthesis in *Trichodesmium*

A scaffold of approximately 7,400 bp containing a cassette of genes encoding a phosphonate biosynthesis pathway was found in the *Trichodesmium* partitioned genome bins (Figure 3.5a). This scaffold contained 6 protein coding regions. The first three genes were annotated as phosphoenolpyruvate phosphomutase (*ppm*), phosphonopyruvate decarboxylase (*ppd*), and 2-aminoethylphosphonate-pyruvate transaminase (*2-AEP-TA*). These first two genes are most similar to homologs in the UniRef databaste that belong to the non-heterocystous diazotroph *Planktothrix agardhii*, a member of the order Oscillatoriales along with *Trichodesmium*. The *2-AEP-TA* is homologous to a gene in a Gammaproteobacterial *Beggiatoa* sp. The fourth, fifth and sixth genes in this scaffold were annotated as a methyltransferase, a cytidylyltransferase (both homologs to genes in *P. agardhii*), and a group 1 glycosyl transferase (homologous to a gene from a *Tolypothrix* sp., a freshwater cyanobacterium in the order Nostocales).

In order to verify genome binning of the ppm-containing scaffold into the *Trichodesmium* fraction, reassembly was attempted to lengthen this scaffold. This effort extended the length of the scaffold upstream of the ppm cassette and resulted in the addition of 44 bp to the 5' end with 95% homology to a non-coding region in the *T. erythraeum* IMS101 genome, as well as a protein coding gene fragment with no known annotation (Figure 3.5a). Metagenomic reads mapped to this scaffold showed extensive coverage of this IMS101 region as well as mate pairs connecting it to genes downstream in the *ppm*-containing cassette (Supplemental Figure B.3).

Comparison of the putative *Trichodesmium ppm* gene's amino acid sequence against experimentally verified phosphonate bond forming enzymes in other organisms showed high sequence identity (Supplemental Figure 3.4). Across the length of the peptide sequence, the *Trichodesmium* Ppm was 72% and 66% similar to sequences from the freshwater cyanobacteria *Planktothrix* sp. and *Moorea producens*, 65% similar to that of the freshwater ciliate *Tetrahymena*, 63% similar to the blue mussel *Mytilus edulis*, and 35% similar to that of

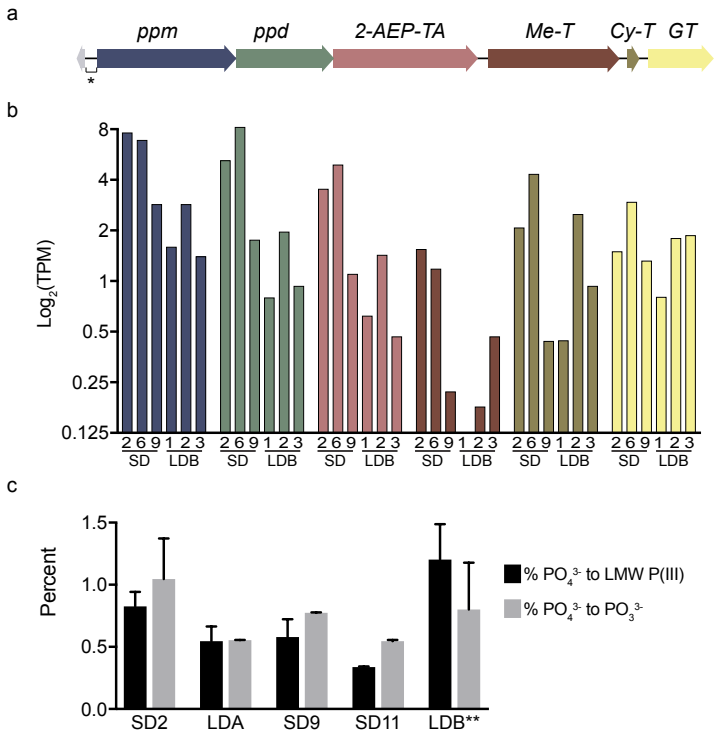

Figure 3.5: Annotations, topology and expression levels of a scaffold containing a phosphonate biosynthesis cassette recovered from a *Trichodesmium* identified metagenome bin. (a) Gene organization and annotations across the scaffold where arrows represent direction of transcription. The star denotes a region with 95% homology to a non-coding region of the *T. erythraeum* IMS101 genome. (b) Normalized expression (transcripts per million, TPM) of each OG on this ppm-containing scaffold at each short duration (SD) or long duration (LDB) station sampled. Column colours match those of the top panel (a). (c) Percentage of total radiolabeled phosphate taken up by *Trichodesmium* colonies and reduced into low molecular weight (LMW) phosphonate compounds (including methylphosphonate, phosphonoacetylaldehyde, and 2-aminoethylphosphonate) and phosphite. Error bars denote standard deviation. Abbreviations: *ppm*, phosphoenolpyruvate phosphomutase; *ppd*, phosphonopyruvate decarboxylase; *2-AEP-TA*, 2-aminoethylphosphonate-pyruvate transaminase; *Me-T*, SAM-dependent methyltransferase; *Cy-T*, cytidylyltransferase; *GT*, group 1 glycosyltransferase. **LDB values represent an average of three measurements obtained during sampling at this long duration station.

a bacterium in the *Streptomyces* genus. Furthermore, closer inspection of the sequences of all organisms showed 100% identity across residues involved in cofactor and substrate interactions, as well as strong conservation across other key residues involved in the enzyme's tertiary structure as determined by crystal structure analysis and comparison to similar enzymes (Supplemental Figure B.4) (Chen et al., 2006). Phylogenetic analysis of *Trichodesmium* Ppm placed it in a phylogenetic branch along with sequences from other N_2 fixing freshwater cyanobacteria (Supplemental Figure 3.5).

A metagenome derived from *Trichodesmium* consortia in the western tropical North Atlantic (Frischkorn et al., 2017) was also re-screened for proteins with homology to the *Trichodesmium* Ppm sequence from this dataset. Five homologous sequences were recovered from this North Atlantic metagenome, with two proteins falling in the branch of verified phosphonate producing Ppm proteins. One protein was located on a branch adjacent to the *Trichodesmium* Ppm sequences recovered from this South Pacific dataset (Supplemental Figure B.5). With the exception of the closely related North Atlantic sequence adjacent to the Ppm sequence recovered from this dataset, the other similar proteins from the South Pacific and North Atlantic consortia did not exhibit amino acid identity at the key conserved residues determined from crystal structure analysis (Chen et al., 2006).

The *T. erythraeum* IMS101 genome was also screened for homologs to the Ppm protein recovered here, but no sequences with conserved amino acid identity across important conserved residues were detected. Furthermore, another enzyme responsible for biosynthesis of methylphosphonate, MpnS, was not detected in proteins from the *Trichodesmium* or microbiome fractions.

All 6 genes in this scaffold recruited reads after metatranscriptome mapping of each sample, indicating that all genes in the phosphonate biosynthesis cassette were expressed at each station, with the exception of the methyltransferase, which recruited no reads from LDB1 (Figure 3.5b). Relative expression of *ppm* at the three SD stations and the average expression across the three LDB samples was significantly positively correlated with surface water column average abundance (5-25 m) of Clade III *Trichodesmium* ($R = 0.99$, $p = 0.006$, Pearson correlation). Furthermore, the concentration of DIP (10 m) was positively correlated with both Clade III abundance ($R = 0.99$, $p = 0.009$, Pearson correlation) and

expression of *ppm* (R = 0.97, $p = 0.03$, Pearson correlation). There was no significant correlation between *ppm* relative expression and Clade I abundance ($p = 0.1$), or between DIP concentration and Clade I abundance ($p = 0.13$). Coincident with expression of phosphonate biosynthesis genes at all stations sampled, phosphate reduction by *Trichodesmium* colonies was also detected in each sample analysed (Figure 3.5). At the five stations tested, approximately 2% of the radiolabeled phosphate taken up by colonies was reduced to either a LMW phosphonate compound (methylphosphonate, phosphonoacetylaldehyde, or 2-aminoethylphosphonate) or phosphite ($PO_3{}^{3-}$) (Figure 3.5c). Coincident with phosphonate production, both *Trichodesmium* and the microbiome possessed and expressed markers for reduced phosphorus metabolism, including the phosphonate (C-P) lyase *phnJ* (Tery_5000 in the *T. erythraeum* IMS101 genome) and *ptxD* (Tery_0368 in the *T. erythraeum* IMS101 genome), a gene identified as phosphite dehydrogenase that is implicated in oxidation of phosphite to phosphate (Polyviou et al., 2015).

3.5 Discussion

3.5.1 Trichodesmium distributions in the oligotrophic WTSP

The WTSP is considered to be among the most oligotrophic environments in the global ocean due to low concentrations of critical resources like nitrogen, phosphorus and iron, coupled with intense stratification that prevents upwelling of remineralized nutrients (Moutin et al., 2008). In spite of chronically depleted resources, a diverse assemblage of free-living and symbiotic diazotrophs thrive in this region (Stenegren et al., 2017). In the WTSP, *Trichodesmium* is typically abundant, and the *Trichodesmium* distribution determined here by clade-specific qPCR agreed with previous analyses along this transect using genus specific qPCR that showed high abundance in the west, decreasing sharply at the transition into the gyre (Stenegren et al., 2017). The absolute value of *Trichodesmium* cells estimated here must be interpreted with some caution, as cultures and field samples of *Trichodesmium* are known to exhibit polypoloidy (Sargent et al., 2016). However, the counts presented here are tabulated using a standard curve generated from cell counts performed on cultures of Clade I and Clade III (Rouco et al., 2014, 2016b), which would yield C_T values that take into

account polyploidy, unlike gene standard approaches. Regardless, the consistent trends in relative abundance observed here and with other methods (Stenegren et al., 2017) suggest that the observed patterns are robust and that the level of polyploidy does not likely vary drastically in *Trichodesmium* populations across the WTSP.

In general, low *Trichodesmium* relative abundance at the easternmost stations (SD14 and 15) sampled along this transect were found despite high DIP concentrations detected in this region. This could be due to the low and homogeneous iron concentrations (0.1-0.3 nM DFe) throughout the entire 0 to 500 m profile at these eastern stations (Guieu et al., 2018), limiting growth. This is corroborated by the fact that rates of water column N_2 fixation were lowest in the easternmost stations along this transect (Bock et al., 2018; Bonnet et al., 2018). Further, *Trichodesmium* distribution and overall rates of N_2 fixation in the water column were positively correlated with iron (Bonnet et al., 2018), the source of was from shallow hydrothermal vents west of the Tonga arc (at approximately 175 °W, near SD11) (Guieu et al., 2018). *Trichodesmium* biomass was too low to evaluate gene expression patterns for the easternmost stations, but regardless, these data are consistent with the importance of iron concentration as a driver of *Trichodesmium* distribution and activities in the WTSP. In addition to iron, phosphorus also likely exerts a strong influence over *Trichodesmium* in the WTSP. The surface water concentration of DIP measured along this transect was low in the context of the global ocean (Sohm et al., 2011) and measured phosphate turnover times in the water column at some stations on the order of hours indicated there was intense competition for phosphate (Van Mooy et al., 2009). The WTSP is poorly sampled relative to other oligotrophic ocean basins (Bonnet et al., 2018; Luo et al., 2012), and little is definitively known about the canonical resource controls that characterize this environment over prolonged time periods (Sohm et al., 2011). Herein, the *Trichodesmium* distributions were consistent with the potential roles of both iron and phosphorus in driving the physiological ecology of this genus in the WTSP.

3.5.2 Expression of metabolic potential in *Trichodesmium* and its microbiome

Trichodesmium does not exist in isolation, as filaments and colonies are associated with an assemblage of epibiotic microorganisms that co-occur ubiquitously in the environment (Lee et al., 2017; Rouco et al., 2016a) and contribute a large amount of metabolic potential that could underpin success in oligotrophic, low nutrient environments (Frischkorn et al., 2017). At the broadest functional level, the microbiome contained approximately 10 times the unique KEGG functions found within the *Trichodesmium* fraction of the WSTP, and these functions were largely consistent with those observed previously in the western North Atlantic (Frischkorn et al., 2017). The presence of unique microbiome transporter functions, especially those related to the transport of phosphate and metals including iron, reflect the importance of these resources within the colony microenvironment that is likely depleted in these key resources. The enrichment of microbiome functions related to the transport and subsequent metabolism of sugars, carbohydrates and lipids could reflect the transfer of fixed carbon from host to microbiome, as the genes encoding these functions are known to oscillate over day night cycles in lockstep with *Trichodesmium* photosynthesis and carbon fixation genes (Frischkorn et al., 2018). These oscillations may support respiration processes that help maintain an environment favorable for N_2 fixation, and the functional enrichment observed here could underpin interactions within the holobiont that help maintain N_2 fixation in the WTSP, despite the scarcity of resources.

In addition to distinct functions in the metagenome, the expression of broad functional categories varied for *Trichodesmium* and the microbiome. Expression of OGs that belonged to carbohydrate metabolism and nucleotide and amino acid metabolism KEGG modules were elevated at the three SD stations sampled, relative to the three days of samples obtained from LDB. Arrival at this long duration station coincided with the decline of a phytoplankton bloom that had been at this location for approximately two months (De Verneil et al., 2017). Coincident with this decline in sea surface chlorophyll *a*, heterotrophic bacterial populations at LDB differed from other stations both taxonomically and in their response during experimental incubation with increased dissolved organic compounds from copepods (Valdés et al., 2018). In *Trichodesmium*, the decreased relative expression at LDB samples

in functions related to carbon fixation, DNA and RNA replication, and a suite of amino acid metabolic functions that require nitrogen relative to the SD stations suggest shifts away from these energy intensive and cell-division processes with bloom demise. Similarly, in the microbiome, decreased relative expression of functional categories over the course of sampling at LDB could reflect shifts in physiology away from reliance on *Trichodesmium* or changes in community structure away from common colonizers of *Trichodesmium* to opportunistic or saprophytic species. For example, in the coral reef system, pulses of organic carbon similar to what could be released during a declining phytoplankton bloom, led to activity shifts in associated bacterioplankton including the increased expression of virulence factors (Cárdenas et al., 2018). Taken together, these data continue to reinforce that the microbiome both possesses and expresses unique metabolic potential relative to Trichodesmium alone, and as such, could play an important role in the physiological ecology of this important diazotroph.

3.5.3 Expression of resource-related signals in *Trichodesmium*

The expression of genes that lead to changes in activities like nitrogen fixation (*nifH*), or resource acquisition (e.g. *phoX*) can be used to assess the physiology of *Trichodesmium in situ*. Genes responsive to low phosphorus and iron conditions are particularly well-studied in *Trichodesmium* (Chappell and Webb, 2010; Chappell et al., 2012; Dyhrman et al., 2006; Orchard et al., 2009; Snow et al., 2015b) and a recent culture study assessed the physiological response of *Trichodesmium* to coupled low phosphate and low iron conditions, yielding a set of genes with significantly elevated expression under co-limitation conditions (Walworth et al., 2017b). Expression of many of these marker genes is heavily repressed in cultures grown under replete conditions (Chappell and Webb, 2010; Orchard et al., 2009) or in field samples with relatively high concentrations of resources like iron (Chappell et al., 2012). Expression of these resource-responsive OG sets was detected in *Trichodesmium* across all samples, indicating that there was intense scavenging of phosphorus and iron, consistent with the low levels of these resources at the stations analyzed for gene expression.

The expression of P-responsive OGs related to phosphate uptake (*sphX*), phosphoester hydrolysis (*phoX*) and the metabolism of phosphite (*ptxD*) tracked together, with RuBisCO,

and were significantly enriched at LDB 1. This pattern may indicate increased P stress at LDB 1 relative to other stations on the transect, and may be an indicator of shifts in *Trichodesmium* physiology associated with modulating carbon fixation and or the declining bloom in this region (De Verneil et al., 2017). For example, increased expression of *ptxD* may indicate increased metabolism of phosphite. Notably, the expression of this low P responsive set, which included OGs related to phosphate uptake (*sphX*), was significantly enriched at LDB 1 where the phosphate turnover time was among the lowest observed. Although *phnJ* has been shown to be regulated by phosphate concentration in culture studies (Dyhrman et al., 2006), the *phnJ* OG here deviated from the expression pattern of the other P-responsive genes like *phoX*. As a result, there may be some variability in *Trichodesmium* processing of phosphoesters and phosphonates over these stations. Regardless, the expression of *phoX*, *ptxD*, and *phnJ* OGs underscores the importance of organic phosphorus compounds, and phosphite in supporting *Trichodesmium* growth across the WTSP.

The OGs in the low Fe responsive set were also detected at all stations consistent with the sub-nanomolar concentrations of Fe observed across the transect. This set was significantly enriched at SD2, where iron concentration was roughly half that of LDB. Although a larger dataset would be needed to resolve patterns of iron stress, these data are suggestive of an increase in *Trichdoesmium* iron stress at SD2 compared to the other stations. Strikingly, expression of the alkaline phosphatase *phoA*, and the nitrogenase subunit *nifH* had similar expression patterns to this low Fe responsive set. Nitrogenase requires iron and its expression in *Trichodesmium* is tightly synchronized with iron processes (Frischkorn et al., 2018). Conversely, the enzyme PhoA does not require iron, instead using a zinc-magnesium cofactor, as opposed to *PhoX* which has an iron-calcium cofactor (Luo et al., 2009; Yong et al., 2014). In low iron environments, *Trichodesmium phoA* is known to show enriched expression relative to that of *phoX* (Rouco et al., 2018), a strategy that could free up iron for use in photosynthetic or nitrogen fixation enzymes. Collectively, these results suggests there is intense scavenging of iron by *Trichodesmium* in the WSTP, and in this and other environments where multiple resources can be low or co-limiting, co-factors like iron could play a role in the phosphorus acquisition strategies employed by *Trichodesmium*.

The OGs in the co-limitation responsive set were detected in all samples. The expression patterns of these co-limitation responsive OGs clustered among both the low P and low Fe sets, and there were no significant patterns of enrichment between stations. The lack of significant enrichment in the co-limitation set is consistent with the fact that the low P and low Fe responsive sets were not simultaneously enriched at the same station. Broadly however, the expression patterns between low P and co-limitation sets were more similar to each other than that shown by the low Fe responsive set, suggesting that across these samples phosphorus was a driver of expression of co-limitation OGs. More field observations over a greater range in iron and phosphate might further resolve these putative co-limitation signals, which have not been previously tracked in field populations. Collectively, the OGs expressed across this WTSP transect are suggestive of the importance of both iron and phosphorus in driving *Trichodsmium* physiological ecology in this region.

3.5.4 Phosphonate biosynthesis by *Trichodesmium* in the WTSP

Phosphate exists in vanishingly low concentrations in the oligotrophic surface ocean and the activity of diazotrophs increases the demand for phosphorus by relieving nitrogen stress—a process that is enhanced by periodic increases in iron availability (Moutin et al., 2005). In low phosphate environments, marine microbes, like *Trichodesmium*, can hydrolyse phosphate from organically bound compounds like phosphoesters and phosphonates, the concentration of which far surpasses phosphate in the oligotrophic ocean (Dyhrman et al., 2007). The production and hydrolysis of reduced compounds like phosphonates are of particular interest because the hydrolysis of methylphosphonate has the potential to release methane, a potent greenhouse gas (Karl et al., 2008; Repeta et al., 2016). Previous studies showing biosynthesis of phosphonates by certain *Trichodesmium* isolates (Dyhrman et al., 2009) as well as rapid phosphate reduction to phosphonate and phosphite and release by *Trichodesmium* colonies in the environment (Van Mooy et al., 2015) implicate this diazotroph as an important player in phosphonate biogeochemistry, yet the molecular mechanisms underlying phosphonate biosynthesis are poorly understood for this genus.

A *Trichodesmium* scaffold containing the full set of genes necessary to synthesize phosphonate compounds was recovered from metagenomes assembled from this WTSP transect.

The *Trichodesmium* origin of this scaffold is supported by tetranucleotide frequency and metagenomic read mapping coverage, as well as the presence of a stretch of non-coding DNA with homology to the *T. erythraeum* IMS101 genome. Furthermore, the protein in this scaffold identified as phosphoenolpyruvate phosphomutase (Ppm), the enzyme that carries out the formation of a carbon-phosphorus bond using phosphoenolpyruvate as a substrate (McGrath et al., 2013), was phylogenetically most similar to Ppm sequences from cyanobacteria like *Planktothrix* that are closely related to *Trichodesmium*. The phylogenetic distance between the *Trichodesmium* Ppm and those of heterotrophic bacteria further support that this scaffold was recovered from a *Trichodesmium* genome and not from a member of the microbiome.

The molecular machinery necessary to synthesize phosphonates is evolutionarily conserved and the biosynthesis of phosphonoacetaldehyde is the starting point from which a diverse suite of organic phosphonate compounds can be produced (McGrath et al., 2013). Based on the genes in this *Trichodesmium* scaffold, synthesis begins with the formation of the carbon-phosphorus bond after molecular rearrangement of phosphoenolpyruvate to phosphonopyruvate, catalysed by Ppm. Next, phosphonopyruvate decarboxylase (Ppd), the protein encoded by the following gene in the cassette, likely performs the irreversible conversion of phosphonopyruvate to phosphonoacetaldehyde which prevents reversion to the ester bond structure. Finally, the presence of the gene for 2-aminoethylphosphonic acid pyruvate-transaminase (*2-AEP-TA*) suggests that phosphonoacetaldehyde is further converted to 2-aminoethylphosphonate (2-AEP), the organophosphonate that occurs most commonly in the environment (McGrath et al., 2013).The *mpnS* gene mediates the production of methylphosphonate down stream of *ppm* in the marine microbes where it has been detected (Metcalf et al., 2012). There was no evidence of *mpnS* in *Trichodesmium* or the microbiome, but *Trichodesmium*-derived phosphonates could potentially be further modified to methylphosphonate by organisms not associated with colonies.

The ppm gene can be found in approximately 7% of microbial genome equivalents recovered from the Global Ocean Survey, and of these *ppm*-containing genomes, 20.6% are estimated to be cyanobacterial in origin (Yu et al., 2013). A protein with homology to Ppm was previously detected and attributed to *Trichodesmium* in metagenomic samples

from the western North Atlantic (Frischkorn et al., 2017), though this gene was not found to be part of a 2-AEP synthesis cassette. OGs derived from genes on the *ppm*-containing *Trichodesmium* scaffold identified here recruited metatranscriptomic reads from each sample sequenced, suggesting active use of these enzymes across the transect. Furthermore, reduced phosphonate compounds (which would include 2-AEP) were produced from radiolabeled phosphate taken up by *Trichodesmium* colonies at each station analysed. Together, these results clearly illustrate a pathway by which *Trichodesmium* synthesizes phosphonates, and that this pathway is active in the WTSP and likely other environments like the western North Atlantic, where high rates of phosphate reduction in *Trichodesmium* colonies has also been measured (Van Mooy et al., 2015).

The production of a recalcitrant form of phosphorus and its potential release into the oligotrophic environment could have important consequences for microbial communities. Phosphonates are a critical source of phosphorus in the oligotrophic ocean, and the ability to utilize this resource could influence microbial ecology in low nutrient environments. Across this transect, both *Trichodesmium* and their microbiome contained and expressed OGs related to phosphonate catabolism, including the marker of the C-P lyase enzyme complex, *phnJ*. *Trichodesmium* and microbiome *phnJ* genes have also been detected and expressed in *Trichodesmium* communities from the chronically low phosphate western North Atlantic ocean as well as the North Pacific subtropical gyre (Dyhrman et al., 2006; Frischkorn et al., 2017, 2018). In addition to the production and hydrolysis of phosphonate compounds, we also detected evidence of the use of other forms of reduced phosphorus. The expression of the *ptxD* gene which is responsible for the oxidation of phosphite, another reduced phosphorus compound (Polyviou et al., 2015), was also expressed by *Trichodesmium* and the microbiome at all stations sampled. This finding suggests active transformation and exchange of reduced phosphorus compounds between consortia members. In the low phosphorus western North Atlantic, up to 16% of the phosphate taken up by *Trichodesmium* colonies has been shown to be reduced and subsequently released from cells, an amount of phosphorus cycling that rivals the amount input to marine systems from allochthonous or atmospheric sources (Van Mooy et al., 2015). The evidence of utilization of these traits across additional geochemical environments, like in the WTSP, and the large quantities

of phosphorus they recycle suggests that phosphonate cycling composes an integral facet of the *Trichodesmium* holobiont's physiology, yet the reasons behind this cycling remain enigmatic. Not all microbes can metabolize phosphonates (Villarreal-Chiu et al., 2012)), therefore it could also be plausible that formation of such compounds creates a cryptic phosphorus pool that would in part restrict access to this critical nutrient by other microbes. Uptake and reduction to a more recalcitrant form as a mechanism of luxury storage exclusive to members of the *Trichodesmium* holobiont is supported by the significant positive correlation between expression of *ppm* and DIP concentration. In short, exchange of these compounds within the *Trichodesmium* holobiont, especially if through a cryptic pool, could help support N_2 fixation in *Trichodesmium* by modulating access to bioavailable phosphorus in the oligotrophic WTSP (Van Mooy et al., 2015).

Marine N_2 fixation is expected to increase in future oceans that are predicted to have higher temperatures and CO_2 concentration (Hutchins et al., 2007), and *Trichodesmium* cultures incubated in high CO_2 conditions exhibited irreversibly increased rates of N_2 fixation (Hutchins et al., 2015). In such conditions, cycling of phosphonate compounds that are not accessible to the full microbial community could support enhanced N_2 fixation if enough iron is available, the release of new nitrogen into the water column and subsequently fuel primary production. In future studies it will be important to assay how future ocean conditions will alter the clade distribution of *Trichodesmium* in the environment, as this could play a role in determining the potential flux of phosphonate compounds from colonies to the water column.

3.6 Conclusions

Marine microbes interact and alter the environment through abiotic transformations as well as through biotic interactions with one another and across trophic levels, and these processes work in tandem to influence global biogeochemical cycles. Understanding these processes *in situ* is of paramount importance to forecasting the ocean's role in the future climate, yet challenges persist with sampling remote locations and filling knowledge gaps surrounding the ecology and physiology of key species. The OUTPACE research expedition

afforded a unique opportunity to investigate communities of the keystone N_2 fixer *Trichodesmium* and their microbiome in the under-sampled South Pacific. Metagenomic and metatranscriptomic data showed a majority of unique physiological functions within the microbiome, many of which were expressed *in situ*, and these functions may be important to *Trichodesmium* physiological ecology in this environment. Patterns of OG expression in low Fe responsive, low P responsive, and co-limitation sets suggested that iron and phosphorus are highly scavenged and that *Trichodesmium* variably experienced changes in these resources, which could modulate growth and N_2 fixation *in situ*. A *Trichodesmium* gene cassette for the biosynthesis of the phosphonates, its expression, and corresponding phosphate reduction rate measurements suggested that *Trichodesmium* is producing reduced phosphate in the WTSP. This finding expands the environments where phosphate reduction has been detected, and confirms the role of *Trichodesmium* in this poorly understood aspect of phosphorus biogeochemistry. Collectively, these data underscore the importance of iron and phosphorus, and the microbiome, in jointly driving the physiological ecology of this key diazotroph in the WTSP.

Chapter 4

Coordinated gene expression between *Trichodesmium* and its microbiome over day-night cycles in the North Pacific Subtropical Gyre

This chapter was originally published as Frischkorn, K.R., Haley, S.T., and Dyhrman, S.T. (2018). Coordinated gene expression between *Trichodesmium* and its microbiome over day-night cycles in the North Pacific Subtropical Gyre. *ISME J.* **12**, 997-1007.

4.1 Abstract

Trichodesmium is a widespread, N_2 fixing marine cyanobacterium that drives inputs of newly fixed nitrogen and carbon into the oligotrophic ecosystems where it occurs. Colonies of *Trichodesmium* ubiquitously occur with heterotrophic bacteria that make up a diverse microbiome, and interactions within this *Trichodesmium* holobiont could influence the fate of fixed carbon and nitrogen. Metatranscriptome sequencing was performed on *Trichodesmium* colonies collected during high frequency Lagrangian sampling in the North Pacific Subtropical Gyre (NPSG) to identify possible interactions between the *Trichodesmium* host and microbiome over day-night cycles. Here we show significantly coordinated patterns of gene expression between host and microbiome, many of which had significant day-night periodicity. The functions of the co-expressed genes suggested a suite of interactions within the holobiont linked to key resources including nitrogen, carbon, and iron. Evidence of microbiome reliance on *Trichodesmium*-derived vitamin B12 was also detected in co-expression patterns, highlighting a dependency that could shape holobiont community structure. Collectively, these patterns of expression suggest that biotic interactions could influence colony cycling of resources like nitrogen and vitamin B12, and decouple activities, like N_2 fixation, from typical abiotic drivers of *Trichodesmium* physiological ecology.

4.2 Introduction

The cyanobacterium *Trichodesmium* is biogeochemically important in nutrient-poor oligotrophic ocean ecosystems because of its ability to supply fixed nitrogen to the water column (Bergman et al., 2013). Through biological N_2 fixation, *Trichodesmium* is estimated to supply approximately half of the total fixed nitrogen in the ocean (Bergman et al., 2013; Sohm et al., 2011). As such, *Trichodesmium* plays a keystone role in fueling marine primary production that is otherwise limited by the availability of nitrogen across much of the oligotrophic ocean gyres (Falkowski et al., 1998; Zehr, 2011). The dissolved organic matter produced through this primary production is one of the largest pools of reduced carbon on earth, and its fate in the ocean is hypothesized to be greatly influenced by microbe-microbe interactions (Moran et al., 2016). These interactions can span entire microbial commu-

nities and are subject to environmental and evolutionary pressures that work in tandem to influence ecology and subsequent biogeochemical impact (Brussaard et al., 2016). In marine microbial communities, photosynthetic and heterotrophic bacteria have been shown to exhibit synchronous periodic gene expression patterns over day-night cycles that are attributed to cascades in metabolic activity sparked by primary productivity (Aylward et al., 2015; Ottesen et al., 2014; Wilson et al., 2017). The taxonomic patterns of these responses in free-living communities were similar between coastal and open ocean systems, suggesting conserved ecological interactions and synergistic activities that could be used to inform or predict biogeochemical transformations (Aylward et al., 2015).

The presence of bacteria on *Trichodesmium* filaments was first noted in the 1980s (Paerl et al., 1989), and it is now appreciated that *Trichodesmium* colonies ubiquitously occur with these heterotrophic bacterial epibionts (Frischkorn et al., 2017; Gradoville et al., 2017; Lee et al., 2017; Rouco et al., 2016a). This microbiome is distinct from the bacteria found free-living in the water column (Hmelo et al., 2012), and varies across ocean basins (Rouco et al., 2016a). Such tight associations between photosynthetic diazotroph and heterotrophic bacteria, in conjunction with detection of microenvironments of depleted oxygen within colonies, led to the hypothesis that these interactions could facilitate exchange of organic carbon, fixed N_2, iron, and vitamins (Paerl et al., 1989). Metagenomic analyses of *Trichodesmium* consortia have begun to elucidate the functional underpinnings of the relationship between host and microbiome, with the abundance and diversity of microbiome genes related to phosphorus acquisition, iron cycling, carbon metabolism and vitamin B12 transport further suggesting that the exchange of critical resources between host and microbiome could occur (Frischkorn et al., 2017; Lee et al., 2017).

The functional potential recovered from *Trichodesmium* consortia indicates a suite of mechanisms that might underlie previously observed dynamics between host and microbiome. For example, selective modulation of the microbiome community with quorum sensing molecules, which *Trichodesmium* does not have receptors for, resulted in increased activity of the organic phosphate hydrolyzing enzyme alkaline phosphatase within colonies (Van Mooy et al., 2012). This finding suggests that the interactions and activities of the microbiome could modulate phosphorus cycling in colonies, and underscores gene expres-

sion analyses that indicate a phosphorus-limited microenvironment within colonies may be decoupled from the surrounding water column (Hewson et al., 2009a). The dynamics of how the other metabolic functions observed in metagenomes are partitioned and expressed between *Trichodesmium* and their microbiome may yield insight into the nature of this relationship, helping to determine whether epibionts are important members of these consortia or simply stochastic colonizers.

New evidence of conserved relationships between *Trichodesmium* and specific epibionts across ocean basins (Lee et al., 2017; Rouco et al., 2016a) suggest that these host-microbiome interactions are widely conserved and could be used to inform or predict biogeochemical transformations, as has been demonstrated for free-living communities (Aylward et al., 2015). For example, synchronization of carbon and nitrogen metabolic processes between *Trichodesmium* and the microbiome could alter the flux of limiting resources like nitrogen from the colony to the water column, and ultimately influence the primary production and carbon cycling these resources support. Here we used metatranscriptomic sequencing of *Trichodesmium* colonies collected during high frequency Lagrangian sampling over day-night transitions in the NPSG to identify the coordinated patterns in host and microbiome gene expression that underpin interactions within the holobiont.

4.3 Materials and methods

4.3.1 Sample collection

Sampling was carried out between 27 July and 30 July, 2015 in the NPSG (Figure 4.1a). *Trichodesmium* samples were obtained with six hauls of a 130 μm mesh size net tow through surface sea water every four hours starting at approximately 1:30 pm on 27 July. *Trichodesmium* colonies were isolated by hand and washed of non-tightly associated microorganisms by serial transfer through 0.2 μm sterile-filtered surface seawater as previously described (Frischkorn et al., 2017). Studies using a similar approach to colony collection have shown that the dominant heterotrophic bacterial species tightly associated with *Trichodesmium* colonies are conserved within, and to some extent across, ocean basins (Lee et al., 2017; Rouco et al., 2016a). Therefore, this approach should yield the sta-

ble, tightly associated microbiome, limiting the presence of opportunistic colonizers, which would contribute to colony function in a transient way. An average of 14 *Trichodesmium* colonies were isolated and preserved per time point. Sampling was performed under red light during the time points at night. Colonies were cleaned within roughly 15 minutes of collection and immediately filtered by gentle vacuum onto 5 μm, 47 mm polycarbonate filters and stored in liquid nitrogen until RNA extraction.

4.3.2 RNA extraction and sequencing

Prokaryotic RNA was extracted with the Qiagen RNeasy Mini Kit (Qiagen, Hildern, Germany) with a minor modification to the lysis step. Briefly, approximately 500 μL zirconia/silica beads (0.5 mm) were added to each sample tube after the addition of Buffer RLT, and the samples were vortexed for 5 minutes. The resulting lysate was processed as per the remainder of the manufacturer instructions, including on-column DNase digestion (RNase-free DNase Kit, Qiagen). Prokaryotic RNA was enriched in the eluted total RNA with a MICROBEnrich kit (ThermoFisher Scientific, Waltham, MA, USA) following manufacturer instructions. Ribosomal RNA was removed using the Ribo-Zero Magnetic Kit for bacteria (Illumina, San Diego, CA, USA) as per manufacturer instructions. Purified prokaryotic mRNA was concentrated using the RNeasy MinElute Cleanup Kit (Qiagen) according to the directions provided by the manufacturer. The mRNA concentration and quality was assessed with a BioAnalyzer using the RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA). The Illumina TruSeq RNA sample preparation kit was used by the JP Sulzberger Genome Center at Columbia University (CUGC). Sequencing of 60 million paired end reads from each sample was performed on an Illumina HiSeq at the CUGC. Sequences were deposited in the NCBI SRA under accession number PRJNA381915.

4.3.3 Metatranscriptome analysis

Sequenced reads were trimmed, normalized, and assembled following the Eel Pond Protocol for mRNAseq (Brown et al., 2013a) yielding 1,809,695 contigs total from across the 16 time points. To create one master assembly to which reads from all 16 time points could be mapped, individual assemblies were pooled together and clustered at 98% identity

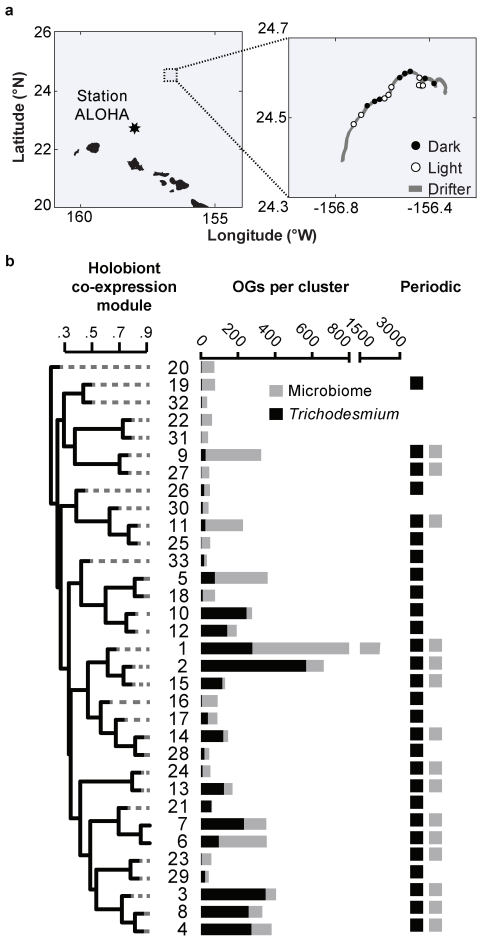

Figure 4.1: Sampling location and holobiont co-expression modules. (a) Lagrangian sampling was carried out Northeast of Station ALOHA in the North Pacific Subtropical Gyre with an approximate location indicated by a dashed box in the left panel, with an expanded view in the right panel. Sampling followed a drifter track (gray line) with dots demarcating where samples were taken and the light conditions at the time of collection, either night (filled symbols) or day (open symbols). (b) A dendrogram showing the 33 co-expression modules generated from *Trichodesmium* and microbiome metatranscriptomes.

using CD-Hit to combine highly similar sequences across the samples (Alexander et al., 2015; Frischkorn et al., 2014; Li and Godzik, 2006; Varaljay et al., 2010). This clustering yielded 1,247,416 contigs total. The merged assembly was then filtered to remove sequences shorter than 210 nucleotides and translated into corresponding amino acid sequences using Prodigal's metagenomics setting (Hyatt et al., 2010).

Taxonomic affiliation of contigs into the *Trichodesmium* and microbiome subsets was determined using DIAMOND (Buchfink et al., 2015) against the NCBI nr database and analyzed using MEGAN5 software (Huson et al., 2013). Functional annotations were obtained by DIAMOND against the UniRef90 database (Suzek et al., 2007) as well as the Kyoto Encyclopedia of Genes and Genomes (KEGG) with the online Automatic Annotation Server using the single-directional best-hit method targeted to prokaryotes and with the metagenomic option selected. Consensus annotations for orthologous groups (OGs) were determined by taking the most abundant UniRef or KEGG annotation for all proteins within that group. In some instances, similar KEGG submodule categories were combined to simplify figures. The amino acid metabolism category includes arginine and proline metabolism, aromatic amino acid metabolism, branched-chain amino acid metabolism, cysteine and methionine metabolism, histidine metabolism and serine and threonine metabolism KEGG submodules. The carbohydrate metabolism category includes central carbohydrate metabolism and other carbohydrate metabolism KEGG submodules. The sugar transport and metabolism includes saccharide and polyol transport system and sugar metabolism KEGG submodules. The transport category includes ABC-2 type and other transport systems, metallic cation, iron-siderophore and vitamin B12 transport system, mineral and organic ion transport system, and peptide and nickel transport system KEGG submodules.

Read mapping to clustered and size-filtered contigs was carried out with RSEM and the default settings with the exception of using the paired end option and the bowtie2 option (Li and Dewey, 2011). On average, over 98% of reads from each time point mapped to this clustered assembly. Read counts were summed across OGs separately for *Trichodesmium* and heterotrophic bacterial epibiont-identified contigs. The OGs were generated by performing a reciprocal comparison with DIAMOND followed by MCL (Markov cluster algorithm) set to an inflation parameter of 1.4, as described elsewhere (Bertrand et al., 2015;

Frischkorn et al., 2017), yielding 462,229 OGs total. To keep downstream analyses conservative, only those OGs with read coverage of greater than 100 and 200 reads total for the *Trichodesmium* and microbiome subsets, respectively, across all time points were used in downstream analyses. This approach is consistent with previous work in other systems (Thaben and Westermark, 2014; Wilson et al., 2017). With this conservative approach, a total of 3,188 *Trichodesmium* OGs and 4,827 microbiome OGs recruited sufficient read coverage across the 16 samples for subsequent analyses. Read counts of these abundant OGs in the *Trichodesmium* and microbiome subsets were normalized using the Variance Stabilizing Transformation (VST) in DESeq (Anders and Huber, 2010). Significant periodicity in normalized OG expression was determined using Rhythmicity Analysis Incorporating Nonparametric Methods (RAIN) (Thaben and Westermark, 2014) in R. OGs with p-values less than 0.1 after false-discovery rate correction (Benjamini and Hochberg, 1995) were considered to have significant periodicity, a cutoff value that has been used previously (Wilson et al., 2017).

To cluster consortia OGs into co-expression modules, counts from *Trichodesmium* and microbiome subsets were first pooled. The combined subsets were then normalized as a whole using VST in DESeq (Anders and Huber, 2010). Normalized read counts were then clustered using the R package WGCNA (Langfelder and Horvath, 2008) with a soft-threshold of 6 selected after a scale-free network topology test and the "blockwiseModules" command set with a minimum module size of 30 OGs and a cut height of 0.25, as previously determined (Wilson et al., 2017). A simplified visualization of the cluster dendrogram produced by WGCNA was generated by hierarchically clustering the pooled, normalized read counts in each module using the hclust command in R.

4.4 Results and discussion

4.4.1 Coordinated expression in the *Trichodesmium* holobiont

High frequency Lagrangian sampling of *Trichodesmium* colonies generated a total of 16 metatranscriptome samples spread over four day-night transitions in the NPSG (Figure 4.1a). After mapping to a merged de novo assembly, read counts were partitioned between

Trichodesmium and all heterotrophic bacteria orthologous groups (OGs) as determined by MEGAN (Huson et al., 2013) and summed for each OG. The microbiome OG subset was dominated by the genera *Pseudoalteromonas* and *Alteromonas*. Other abundant genera included *Fulvivirga* (Bacteroidetes) and *Nisaea* (Alphaproteobacteria). These genera are common members of the *Trichodesmium* microbiome based on previous 16S and metagenomic surveys of Trichodesmium epibiont diversity (Frischkorn et al., 2017; Gradoville et al., 2017; Lee et al., 2017; Rouco et al., 2016a).

Expression profiles from the *Trichodesmium* and microbiome OGs were analyzed with a weighted correlation network analysis (WGCNA) (Aylward et al., 2015; Langfelder and Horvath, 2008; Wilson et al., 2017) to ascertain if there was significant co-expression between *Trichodesmium* and the microbiome. Expression patterns were significantly coordinated between *Trichodesmium* and the microbiome, with 3,140 *Trichodesmium* and 4,730 epibiont OGs clustering into 33 WGCNA co-expression modules (Figure 4.1b). The presence of OGs from both *Trichodesmium* and the microbiome within every significant co-expression module indicates gene expression is tightly integrated within the holobiont. Many co-expression modules were also dominated by microbiome OGs (Figure 4.1b), such as Module 9, which was composed of a suite of different functions including those related to ribosomes, sugar transport and metabolism, ATP synthesis, and phosphate and amino acid transport functions (Supplemental Figure C.1). The presence of microbiome dominated modules suggests that there is significant co-expression between epibionts. Together, these patterns of co-expression suggest coordination between *Trichodesmium* and their microbiome, as well as between epibionts within the microbiome.

Dynamics within the majority of these co-expression modules were defined by diel oscillations in gene expression that had significant periodicity as determined by RAIN (Thaben and Westermark, 2014)(Figure 4.1b). In *Trichodesmium*, these periodic OGs included gene sets related to photosynthesis, phosphorus, iron, N_2, and carbon fixation among others (Supplemental Figure C.2). Global expression patterns in *Trichodesmium* would be expected to be dominated by diel patterns, as photosynthetic and N_2 fixing activities are both tightly regulated across day-night cycles (Bergman et al., 2013). Here, periodic microbiome expression patterns strikingly mirrored those observed in *Trichodesmium* (Supplemental Figure

C.3). In the microbiome, OGs with day-night periodicity included genes related to stress responses, nitrogen metabolism, carbon metabolism, and cofactor and vitamin biosynthesis among others (Supplemental Figure C.2). Similar synchronous diel transcriptional oscillations have also been observed between unicellular marine cyanobacteria and free-living heterotrophic bacteria, as heterotrophic activity appears to be tied to pulses in primary production (Aylward et al., 2015; Ottesen et al., 2014). Overall, microbiome activities are likely tied to both *Trichodesmium* photosynthesis and N_2 fixation as the release of dissolved nitrogen and organic carbon is tied to these processes. Though *Trichodesmium* fixes both carbon and N_2 during daylight (Berthelot et al., 2015), net release of organic carbon, dissolved organic nitrogen, and ammonium oscillate throughout the day and night (Wannicke et al., 2009). The extensive coordination and functional characteristics in gene expression between *Trichodesmium* and their microbiome suggest the physiology of these associated microbes are attuned to the host, and could thus influence the cycling and fate of carbon, nitrogen and other resources in the colony microenvironment.

4.4.2 Coordinated expression in nitrogen and carbon pathways

The OGs related to *Trichodesmium* N_2 fixation, including the genes of the nitrogenase-encoding nif cassette, were clustered into co-expression Module 3. Module 3 contained 348 *Trichodesmium* and 57 microbiome OGs (Figure 4.1b) and was dominated by OGs with maximum or minimum expression just prior to peak photosynthetically active radiation (PAR) and strong day-night periodicity (Figure 4.2a). Strikingly, Module 3 contained a suite of microbiome OGs related to nitrogen processes, which mirrored expression of nitrogenase subunit OGs in Trichodesmium (Figure 4.2b). These nitrogen-related microbiome OGs include aminotransferases, amino acid adenylation proteins, and peptidases with significantly periodic patterns (Figure 4.2b). *Trichodesmium* releases biologically available nitrogen in culture and the environment, even when maintained axenically (Bronk et al., 1994; Mulholland et al., 2006; Mulholland et al., 2004; Wannicke et al., 2009). The extent to which this release facilitates cultivation of a symbiotic heterotrophic bacterial microbiome or whether associated epibionts are simply a product of opportunistic, stochastic colonization remains an outstanding question. However, these expression data, coupled with

recent observations of ubiquitous epibionts across ocean basins (Lee et al., 2017; Rouco et al., 2016a), underscores conserved relationships in these consortia wherein the transfer of *Trichodesmium*-derived fixed N_2 may synchronize microbiome metabolism with host physiology.

Expression patterns between host and microbiome also indicate possible synchrony between photosynthetic carbon fixation and release by *Trichodesmium* and the respiration of that organic carbon by the microbiome. Module 8 had strong day-night periodicity in both *Trichodesmium* and the microbiome and contained a high proportion of *Trichodesmium* photosystem and carbon fixation OGs, including a RuBisCO subunit, three photosystem subunits and a phycoerythrin assembly protein (Figure 4.3a). *Trichodesmium* photosynthetic and carbon fixation related OGs were segregated into two significantly co-expressed patterns, which had peak expression during light and dark periods (Figure 4.3b). The microbiome subset of this module contained a number of functions that suggested synchronization of carbon metabolism with *Trichodesmium* including those related to gluconeogenesis, a sugar binding protein, sugar transporter, and carbohydrate active enzymes, as well as OGs related to ribosome functions and amino acid metabolism (Figure 4.3). The expression patterns in these microbiome OGs exhibited peaks in both light and dark conditions that mirrored those of *Trichodesmium* photosystem and carbon fixation-related gene expression (Figure 4.3b). The microbiome pyruvate carboxyltransferase, an enzyme from gluconeogenesis and a transporter of C4-dicarboxylates like succinate, fumarate, and malate (Janausch et al., 2002) peaked during daylight. Peaking during nighttime were two microbiome glycosyltransferase enzymes, along with *Trichodesmium* RuBisCO (Figure 4.3b). Overall, these microbiome expression patterns are consistent with organic carbon metabolism in the microbiome being tuned to periodic fluxes of *Trichodesmium* organic carbon.

Trichodesmium releases an estimated 50% of fixed carbon (Wannicke et al., 2009) and up to 20% of fixed N_2 from colonies (Berthelot et al., 2015). This loss of resources is paradoxical in the oligotrophic ocean, however it may support microbiome colonization and beneficial syntrophic interactions with the holobiont. Tracing these potential interactions *in situ* is challenging, but they could exert important unrecognized feedbacks on the physiological

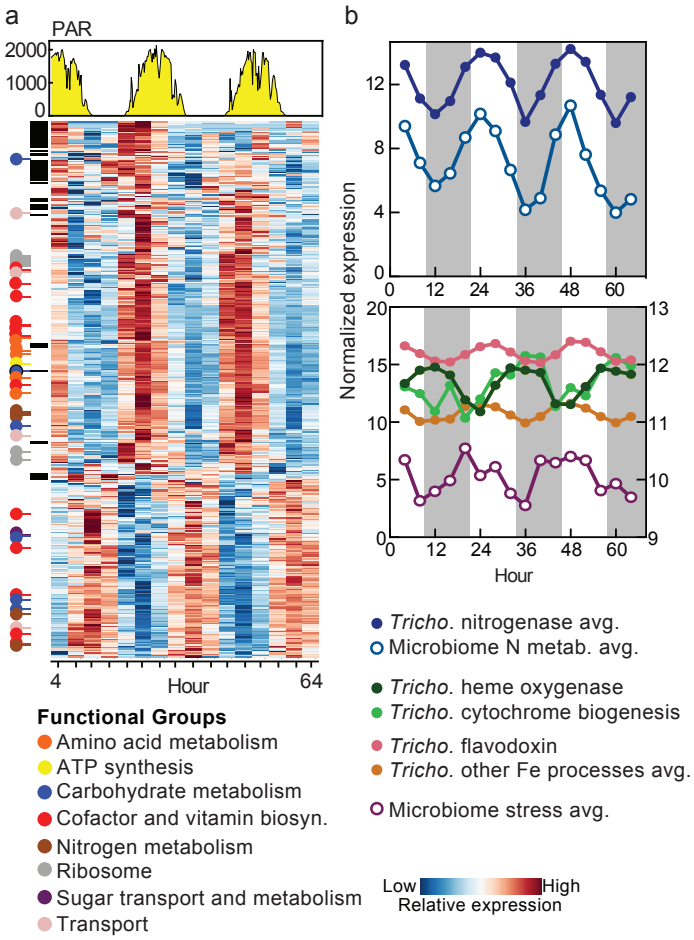

Figure 4.2: Resource-related patterns in the nif-containing co-expression Module 3. (a) Photosynthetically active radiation (PAR) levels are shown over the duration of the Lagrangian sampling period and correspond to vertical bars in the heatmap, which depicts row averaged relative expression of the significantly co-expressed *Trichodesmium* and microbiome orthologous groups (OGs). Black bars denote rows that are microbiome OGs. Colored circles show functional annotations based on KEGG submodules. (b) Details of select OG expression in Module 3 highlighting patterns between *Trichodesmium* and the microbiome related to nitrogen and iron metabolism. Grey bars indicate dark conditions. *Tricho.*, *Trichodesmium*. Nitrogenase avg., average normalized expression for all nif genes clustered in Module 3. Microbiome N metab. avg., average normalized expression of nitrogen-related OGs in the microbiome subset in Module 3, including aminotransferases, amino acid adenylation proteins, and peptidases. *Tricho.* other Fe processes avg., average normalized expression of iron-related OGs including cytochromes, iron-sulfur cluster assembly protein, iron-requiring hydrogenase, ferrochelatase, ferredoxin, and ferredoxin-dependent bilin reductase.

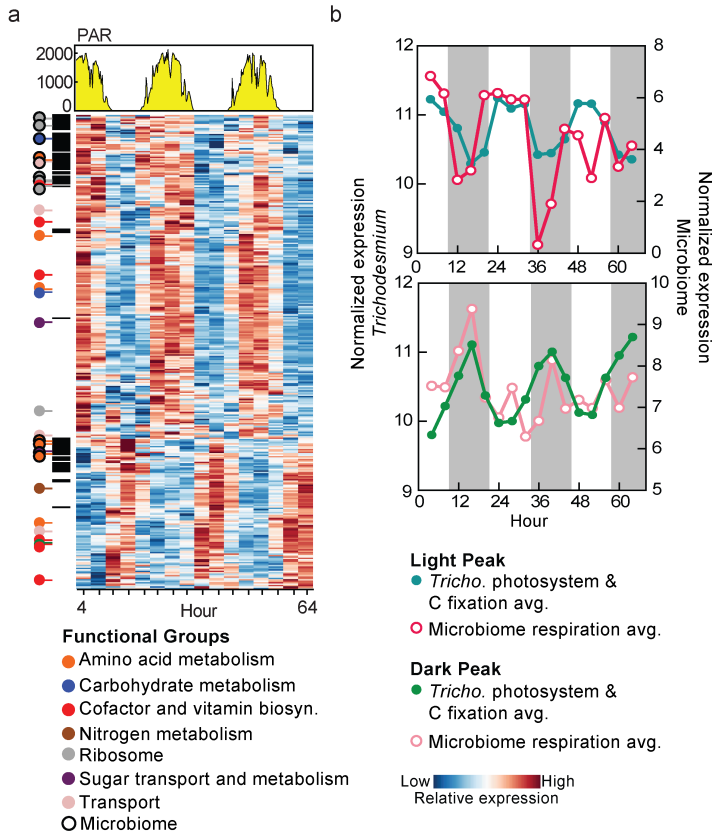

Figure 4.3: Patterns in carbon fixation and respiration in co-expression Module 8. (a) Photosynthetically active radiation (PAR) levels are shown over the duration of the Lagrangian sampling period and correspond to vertical bars in the heatmap, which depicts row averaged relative expression of the significantly co-expressed *Trichodesmium* and microbiome orthologous groups (OGs). Black bars denote rows that are microbiome OGs. Colored circles show functional annotations based on KEGG submodules. (b) Detail of expression patterns in *Trichodesmium* photosystem and carbon fixation OGs and microbiome respiration OGs. Gray bars indicate dark conditions. *Tricho.*, *Trichodesmium*. Microbiome respiration OGs averaged include pyruvate carboxyltransferase and a C4-dicarboxylate transporter in the top panel and two glycosyltransferases in the bottom panel.

ecology of *Trichodesmium* populations. For example, one benefit of the microbiome could be that microbiome respiration of *Trichodesmium* organic carbon alleviates carbon limitation of photosynthesis as well as O_2 inhibition of N_2 fixation (Lee et al., 2017; Paerl et al., 1989). Such a feedback is consistent with the coordinated patterns of gene expression between host and microbiome seen here. If the microbiome modulates the *Trichodesmium* microenvironment in a way that decouples rates of carbon and N_2 fixation from water column geochemistry, or other abiotic drivers of these activities, this would exacerbate ongoing challenges in modeling these processes in the oligotrophic ocean (Capone et al., 2005; McGillicuddy, 2014). Further, the coordinated expression of microbiome nitrogen and carbon metabolism genes with *Trichodesmium* activities, like N_2 fixation and carbon fixation, might allow the microbiome to influence the net flux of resources to the water column.

4.4.3 Transcriptional evidence of shared phosphorus and iron demand

In addition to the *nif* enzyme subunits and several OGs related to photosynthesis, Module 3 also contained a number of additional *Trichodesmium* OGs encoding iron-requiring proteins (Figure 4.2b), suggesting complex coordination of resources across these processes. In *Crocosphaera*, a marine diazotroph that decouples oxygen evolving photosynthesis and N_2 fixation between day and night respectively, the necessary iron quotas are met by degradation of metalloenzymes that liberate and subsequently repurpose iron across different processes over day-night cycles (Saito et al., 2011). Unlike *Crocosphaera*, *Trichodesmium* cannot shuttle iron between photosynthetic and N_2 fixing processes because both occur during the day. The majority of *Trichodesmium* iron-requiring OGs in Module 3 had peak daytime expression that mirrored *nif* and photosystem expression (Figure 4.2b; Figure 4.3b), and included cytochromes, ferredoxin, ferrochelatase, ferredoxin-dependent bilin reductase, an iron-sulfur cluster assembly protein, and an iron-requiring hydrogenase. Also peaking during daylight in this module was an OG identified as flavodoxin (Figure 4.3b), an electron transport protein that does not require iron and can substitute for ferredoxin during low iron conditions (Erdner and Anderson, 1999). In contrast, the expression of an OG identified as heme oxygenase, an enzyme involved in liberating iron from organic

complexes (Saito et al., 2011), was opposite of nif expression in *Trichodesmium* (Figure 4.2b). In *Crocosphaera*, heme oxygenase also showed diel variability with peak expression opposite that of nitrogenase (Saito et al., 2011), suggesting that this enzyme could be used to liberate iron for cellular processes that are not related to N_2 fixation. In Module 3, a cytochrome biogenesis gene was the only iron-requiring OG that had expression dynamics similar to heme oxygenase (Figure 4.2b), indicating that iron liberated at night could be repurposed for use in electron transport processes. Overall, the dynamics of gene expression in iron-related OGs in *Trichodesmium* underscore high daytime iron requirements that likely modulate competition for this critical resource within the holobiont over day-night cycles.

In addition to nitrogen and iron processes, *Trichodesmium* also expressed OGs for phosphate uptake in Module 3, with significantly periodic daytime expression peaks that were slightly offset from expression of the OGs for phosphonate hydrolysis, which also had significant periodic expression but at a later peak time (Supplemental Figure C.4). These patterns likely indicate increased daytime demand for phosphorus to support growth and N_2 fixation in *Trichodesmium*. Coincident with the expression of these *Trichodesmium* genes in Module 3 was a suite of microbiome OGs including the chaperone *htpG*, one general stress protein, and a *uspA*-like universal stress protein (Figure 4.2b). In other species of bacteria, these protein families have been implicated in a variety of functions which are modulated in response to stressors such as limitation for key resources like phosphorus and iron as well as exposure to antibiotics and oxidative agents (Liu et al., 2007; Nachin et al., 2005). These patterns suggest that relative to nitrogen, there may be an offset in the peak bioavailability of resources like phosphorus or iron to the microbiome. This offset could result in resource limitation scenarios that oscillate within the holobiont microenvironment. These data also highlight the complexity of interactions within the holobiont that span multiple resources. For example, time periods of peak N_2 fixation and photosynthesis, when biologically available resources should be highest for the microbiome, are concurrent with microbiome cellular stress. In conjunction with co-expression in nitrogen and carbon functions, this observation could indicate a system where modulation of *Trichodesmium*-derived resources maintains a stable, tightly-regulated relationship between the host and microbiome. In contrast, at

the end of a bloom *Trichodesmium* releases an abundance of organic material causing a change in community composition of the associated bacteria hypothesized to be related to the growth of opportunistic copiotrophs (Spungin et al., 2016). A similar phenomenon has also been observed in bacteria that colonize corals, where an abundance of labile carbon resulted in a shift towards more virulent activities (Cárdenas et al., 2018). During non-bloom conditions in *Trichodesmium*, avoiding such a resource-rich environment could help maintain a beneficial microbiome and prevent the growth of copiotrophs that could skew potential interactions towards those that are deleterious or parasitic.

4.4.4 Microbiome dependence on *Trichodesmium* cobalamin

Despite the critical importance of the B12 vitamin cobalamin to all organisms, only select bacteria and archaea are capable of *de novo* cobalamin biosynthesis, making exchange between producers and non-producers critical (Bertrand et al., 2015; Sañudo-Wilhelmy et al., 2014; Warren et al., 2002). Based on genome sequence analysis, *Trichodesmium* should possess the capability for *de novo* cobalamin biosynthesis (Helliwell et al., 2016), and OGs related to cobalamin production were detected and expressed in *Trichodesmium* across several co-expression modules largely without significant diel periodicity (Figure 4.4a; Supplemental Figure C.5). Only one OG in the cobalamin biosynthesis pathway (*cobW*) was detected in the microbiome, which indicates that these genes were either not expressed at a high enough level to pass our analysis threshold, or that the microbiome members do not possess the capability to synthesize cobalamin. Together with the recent finding that a conserved *Trichodesmium* microbiome member is a cobalamin auxotroph (Lee et al., 2017) and that marine diazotrophs are known to secrete large amounts of this vitamin (Bonnet et al., 2010), these results suggest that *Trichodesmium* supplies cobalamin to the microbiome.

Recent findings have shown the form of cobalamin produced by cyanobacteria, pseudocobalamin, is not readily bioavailable to heterotrophic organisms (Heal et al., 2017; Helliwell et al., 2016). Prior to use by heterotrophic bacteria, an adenine ligand on cyanobacterial pseudocobalamin must be exchanged with 5,6-dimethylbenzimidizole (DMB)—a process that relies on an exogenous supply of DMB, *de novo* biosynthesis of DMB with the *bluB* gene or the alternative oxygen-sensitive *bzaABCDE* pathway (Hazra et al., 2015)—and fi-

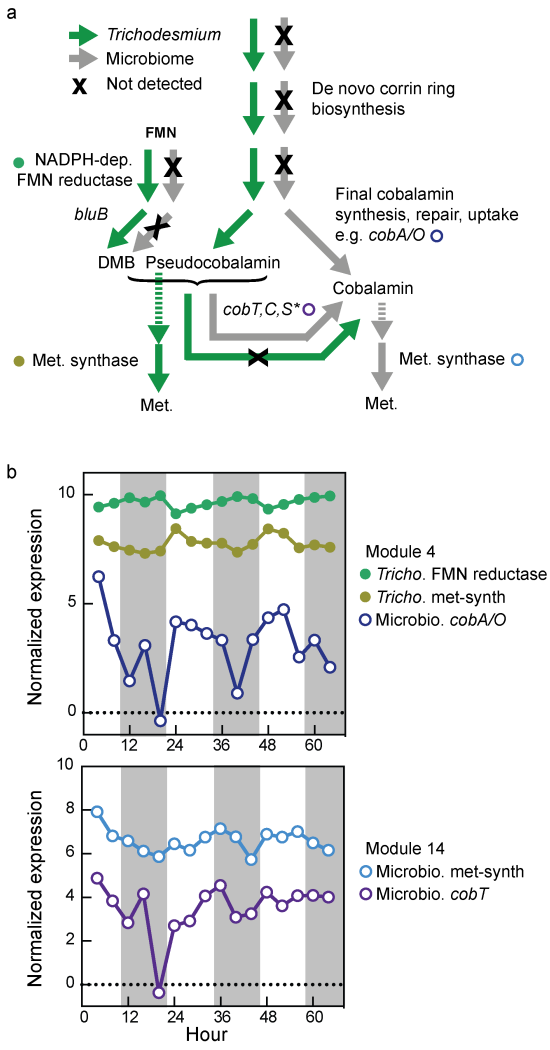

Figure 4.4: Detection of cobalamin pathway orthologous groups (OGs) and related transcriptional patterns. (a) Presence and absence of OGs from the transcriptome data in the *de novo* cobalamin biosynthesis, uptake and modification, and utilization pathways in *Trichodesmium* and the microbiome. *Only *cobT* was detected in the microbiome. Colored circles indicate corresponding expression patterns that are depicted in panel b. (b) Detail of expression patterns in key cobalamin modification and utilization OGs in *Trichodesmium* and the microbiome. FMN, flavin mononucleotide; DMB, 5,6-dimethylbenzimidizole; Met., methionine. *Tricho.*, *Trichodesmium*. Grey bars indicate dark conditions.

nally the activation of DMB with the enzyme nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase (*cobT*) (Heal et al., 2017; Helliwell et al., 2016; Taga et al., 2007). Within the microbiome, the OG for cobalamin uptake (*cobA/O*) was detected in Module 4, a module that also contained a suite of significantly co-expressed *Trichodesmium* OGs for cobalamin synthesis and cobalamin-dependent methionine synthase (Figure 4.4b; Supplemental Figure C.1; Supplemental Figure C.5). Supporting the hypothesis that the microbiome is using *Trichodesmium*-derived cobalamin, the microbiome methionine synthase was clustered in a co-expression module that also contained microbiome *cobT* (Figure 4.4b). However, despite this transcriptional evidence for microbiome conversion of *Trichodesmium*-derived pseudocobalamin and subsequent use, homologs of *bluB* and *bzaABCDE* were not detected in the microbiome subset, indicating that essential DMB is likely supplied from a source outside the microbiome.

Contrary to the overwhelming majority of cyanobacteria that do not possess the capability to synthesize or activate DMB, the *Trichodesmium* IMS101 genome encodes the *bluB* gene for DMB production (Helliwell et al., 2016). Within our dataset, an OG identified as *bluB* was detected and expressed in *Trichodesmium* Module 26 (Supplemental Figure C.5). Furthermore, an OG identified as flavin mononucleotide (*FMN*) reductase, an enzyme that generates the precursor to DMB and acts as the substrate of the BluB enzyme (Taga et al., 2007), was detected in *Trichodesmium* Module 4 (Supplemental Figure C.1) along with *Trichodesmium* OGs from the cobalamin biosynthesis pathway, methionine synthase, and the microbiome *cobA/O* (Figure 4.4b). Strikingly, expression patterns of *Trichodesmium FMN* reductase and cobalamin-dependent methionine synthase had significant, anti-correlated diel periodicity (Figure 4.4b). One interpretation of the presence of *bluB* could be that *Trichodesmium* uses DMB to convert its own pseudocobalamin prior to downstream use of this vitamin. However, the anti-correlated expression dynamics between DMB production and cobalamin-dependent processes (Figure 4.4b) seem to suggest that *Trichodesmium* is not converting its own pseudocobalamin prior to use. Instead, these data suggest that *Trichodesmium* both supplies pseudocobalamin to the microbiome and controls the timing of subsequent ligand conversion within the microbiome, regulating this process to be out of phase with their own cobalamin-dependent enzyme gene expression. These synchronized

transcriptional patterns support evidence from metagenomic data that suggests that in *Trichodesmium* consortia the direction of cobalamin transfer is opposite that of the canonical pathway (Lee et al., 2017), where heterotrophic bacterial producers typically share this vitamin with photosynthetic algae in exchange for fixed carbon (Bertrand et al., 2015; Croft et al., 2005). The processes that govern recruitment and maintenance of microbiome community structure are largely unknown, but the lack of evidence for microbiome cobalamin biosynthesis is consistent with the hypotheses from other research that this vitamin could exert selective pressure on microbiome community structure (Lee et al., 2017). If *Trichodesmium* DMB production does control subsequent microbiome cobalamin use, then this would be an example of a microbiome dependency that drives the community structure of the *Trichodesmium* holobiont. This tightly controlled structure both differentiates *Trichodesmium* colonies from transient microbial consortia on particles that are dominated by remineralization activities (Fontanez et al., 2015), and reinforces the apparent microbiome influence on the flux of important resources, like cobalamin, to the oligotrophic environment.

4.5 Conclusions

The future oceans are predicted to be warmer, higher in CO_2, and have expanded oligotrophic regions (Hutchins et al., 2015; McGillicuddy, 2014; Riebesell et al., 2009). It is expected that these conditions will increase the rate of *Trichodesmium* carbon and N_2 fixation (Hutchins et al., 2007, 2015) and potentially magnify the relative importance of *Trichodesmium* activities within expanded regions of oligotrophy. In other marine systems, these future ocean conditions are known to destabilize the microbiome with resultant deleterious effects on the host (Ainsworth et al., 2010; Bourne et al., 2016; Lesser et al., 2016). Here we show that, like these other host-microbiome systems, the *Trichodesmium* microbiome is closely synchronized with the host, and that host-microbiome interactions likely influence the cycling of carbon, nitrogen, iron, and cobalamin. To fully understand and forecast future ocean dynamics, studies of *Trichodesmium* must consider the role of the microbiome.

Chapter 5

The *Trichodesmium* microbiome can modulate host N_2 fixation

A modified version of this chapter has been submitted as Frischkorn, K.R., Rouco, M., Van Mooy, B.A.S., and Dyhrman, S.T. (2018). The *Trichodesmium* microbiome can modulate host N_2 fixation. *Limnology and Oceanography Letters*, (In review).

5.1 Abstract

Trichodesmium is a marine, diazotrophic cyanobacterium that plays a central role in the biogeochemical cycling of carbon and nitrogen. Colonies ubiquitously co-occur with a diverse microbiome of heterotrophic bacteria. We show that manipulation of the microbiome with quorum sensing acyl homoserine lactones (AHLs) significantly modulated rates of N_2 fixation by *Trichodesmium* collected from the western North Atlantic, with positive and negative effects of varied magnitude. Changes in *Trichodesmium* N_2 fixation did not correlate with changes in microbiome composition or geochemistry. Metatranscriptome sequencing revealed significant changes in the abundance of microbiome transcripts encoding metabolic functions consistent with quorum sensing responses in model bacteria. There was little overlap in specific microbiome transcriptional responses to AHL addition between stations, and this variability in microbiome gene expression may underpin the heterogeneous changes in *Trichodesmium* N_2 fixation. Overall, host-microbiome interactions are a complex interplay of biotic and environmental factors that together form an overlooked mechanism modulating *Trichodesmium* N_2 fixation.

5.2 Introduction

Trichodesmium is a keystone member of marine environments because of its ability to provide fixed N_2 that fuels primary productivity in otherwise nutrient poor regions (Capone et al., 1997). Some estimates predict *Trichodesmium* accounts for approximately half of the total oceanic fixed N_2 (Bergman et al., 2013). In oligotrophic regions, *Trichodesmium* N_2 fixation is strongly affected by the availability of nutrients in the water column (Sohm et al., 2011). With the high iron quotas associated with N_2 fixation and photosynthesis, iron limitation is the canonical constraint on diazotrophy in *Trichodesmium* (Berman-Frank et al., 2001). In the oligotrophic western North Atlantic however, high iron concentrations relative to phosphorus lead to phosphorus depletion, and *Trichodesmium* distribution and N_2 fixation is thought to be more strongly influenced by phosphorus availability (Moore et al., 2013; Rouco et al., 2014; Sañudo-Wilhelmy et al., 2001). Despite the fact that resource controls on *Trichodesmium* eco-physiology are well established, using geochemistry

to predict and model the distribution and activities of this organism remains challenging (Capone et al., 2005; McGillicuddy, 2014; Snow et al., 2015a). Recent studies have suggested that some of the challenges associated with modeling *Trichodesmium* dynamics is in part due to the fact that its physiology is tightly linked to that of its microbiome (Frischkorn et al., 2017; Lee et al., 2017).

Trichodesmium ubiquitously co-occurs with a microbiome of epibiotic microorganisms (Frischkorn et al., 2017; Hewson et al., 2009b; Hmelo et al., 2012; Lee et al., 2017; Paerl et al., 1989; Rouco et al., 2016a), yet the role of this microbiome in modulating *Trichodesmium* physiological ecology is still poorly understood. These epibionts are tightly associated with colonies of *Trichodesmium* filaments, and together make up a conserved community that is unique from planktonic microbes in the surrounding water column (Hmelo et al., 2012; Rouco et al., 2016a). In other systems, such communities of bacteria are known to regulate physiological activities on a population-wide scale through cell-cell signaling called quorum sensing (Miller and Bassler, 2001). Quorum sensing molecules in the acyl homoserine lactone (AHL) family modulate a range of activities in bacteria by altering gene expression and subsequently behavior (Waters and Bassler, 2005), thus altering physiology through a mechanism that is decoupled from external environmental stimuli. These molecules have been detected in *Trichodesmium* colonies collected from the environment (Van Mooy et al., 2012), however *Trichodesmium* does not possess the capacity to produce or respond to quorum sensing molecules (Patankar and Gonzalez, 2009; Vannini et al., 2002), suggesting that AHL-driven cell-cell communication takes place solely between microbiome members.

Artificial manipulation of quorum sensing circuits provides a unique opportunity to selectively alter microbiome activity within *Trichodesmium* colonies. In previous field studies, when *Trichodesmium* colonies were artificially amended with a cocktail of AHLs, phosphorus acquisition was altered by the induction of an organic phosphorus hydrolyzing enzyme, providing evidence that phosphorus physiology of the entire consortia could be influenced by solely modulating the microbiome (Van Mooy et al., 2012). The microbiome of *Trichodesmium* contains diverse and abundant metabolic potential (Frischkorn et al., 2017; Lee et al., 2017) that likely influences interactions between the *Trichodesmium* host and the environment, and could modulate other aspects of host physiology like N_2 fixation.

Here we assess if microbiome activities can modulate *Trichodesmium* N_2 fixation in the western North Atlantic by amending freshly collected colonies with AHLs and monitoring for changes in *Trichodesmium* N_2 fixation and microbiome gene expression.

5.3 Materials and methods

We collected *Trichodesmium* colonies from surface water along a cruise transect in the western North Atlantic (Figure 5.1a) aboard the R/V *Atlantic Explorer* (AE1409) during May 2014, as previously described (Frischkorn et al., 2017). We focused on field studies with freshly isolated colonies because the clade of the cultured strain (*Trichodesmium erythraeum* IMS101) is not dominant in the environment (Rouco et al., 2014), and it likely has a derived microbiome. Briefly, we conducted six hauls of a net tow (mesh size of 130 μm) at approximately 7 am in the upper 20 m to collect samples for experiments. We isolated *Trichodesmium* colonies from the net tow and washed them three times with fresh 0.2 μm sterile-filtered local surface seawater, a method that prevents contamination by planktonic microbe carryover and has reproducibly yielded the stable *Trichodesmium* microbiome (Frischkorn et al., 2017, 2018; Rouco et al., 2016a). After washing, we transferred approximately 30 cleaned colonies of similar sizes and morphologies into acid-clean, polycarbonate bottles filled with 30 mL of sterile-filtered seawater. Efforts were made to place colonies in bottles to mimic colony morphology distribution found in net tow samples. To minimize handling effects, time from initial sampling to incubation was less than 15 minutes.

The experiments were designed to selectively alter microbiome activity, rather than isolate quorum sensing pathways. We spiked the six experimental bottles (+AHL) with a cocktail of three di-deuterated AHLs (N-(decanoyl)homoserinelactone, N-(dodecanoyl)homoserinelactone, and N-(tetradecanoyl)homoserinelactone) in dimethylsulfoxide (DMSO) to a final concentration of 500 nmol L^{-1}, and the six control bottles with only DMSO (Krupke et al., 2016). Previous field incubations showed microbiome activity was specifically altered by the three AHL compounds used here and not by N-3-oxooctanoyl homoserine lactone (Van Mooy et al., 2012). Selective manipulation of the microbiome was possible because natural, non-deuterated forms of these molecules have been observed in *Trichodesmium* colonies,

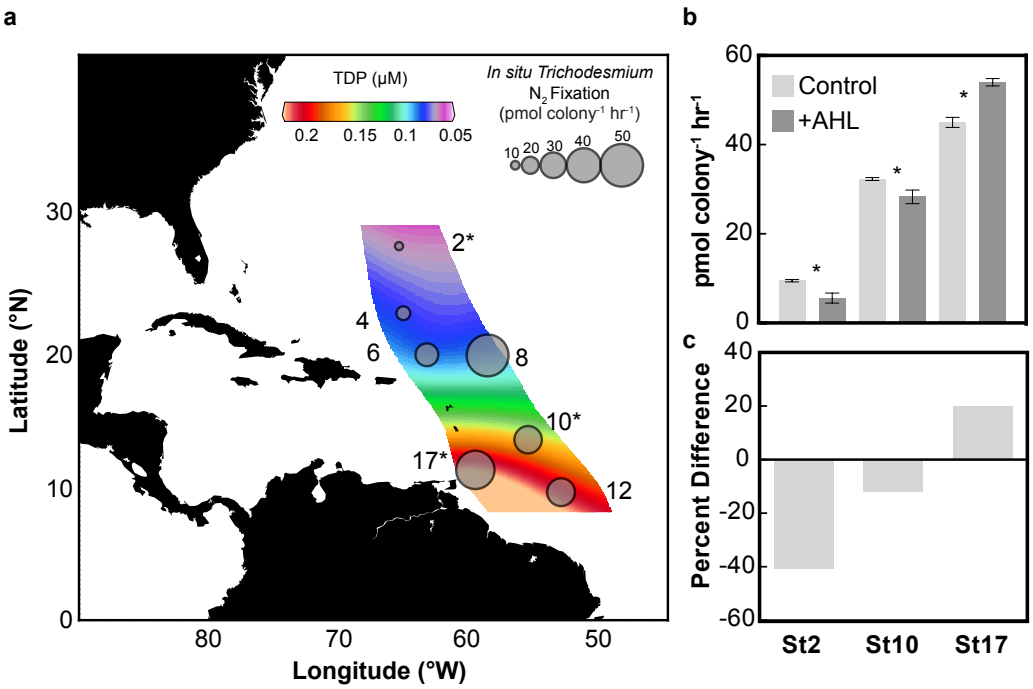

Figure 5.1: Sampling locations of transect stations, and *Trichodesmium* N_2 fixation rate measurements. (a) Circles denote stations along the transect where *Trichodesmium* colony N_2 fixation rate measurements were performed; size is proportional to rate. Stations where +AHL amendment experiments were carried out are denoted with a star. Total dissolved phosphorus (TDP) was contoured using Ocean Data View and the data intensive visualization and analysis (DIVA) grid method. (b) N_2 fixation rates in triplicate control and +AHL experiments. * Indicates significant difference between control and +AHL after student's t-test (<0.05). (c) Percent change in N_2 fixation between control and +AHL.

but *Trichodesmium* itself cannot produce or respond to AHLs (Van Mooy et al., 2012). We incubated the control and +AHL bottles for 4 hours in on-deck flow-through incubators shaded with blue film to mimic *in situ* conditions. A 4 hour incubation was chosen as it was short enough to capture +AHL induced transcriptional changes in the microbiome, and long enough to have induced N_2 fixation changes in *Trichodesmium*. Biomass limitations and destructive sampling precluded time series analysis.

For gene expression analysis, we filtered each of the remaining three bottles of each treatment onto 5 μm pore size polycarbonate filters after the 4 hour incubation period and immediately stored them in liquid nitrogen until sample processing. Three bottles set up as controls as previously described were used to measure in situ N_2 fixation rates at stations where metatranscriptome sequencing experiments were not carried out. Total dissolved phosphorus (TDP) was determined on 0.2 μm filtrates of surface water (approximately 5 m depth) samples collected via CTD into acid-clean polycarbonate bottles. Samples were processed at the SOEST Laboratory for Analytical Biogeochemistry at the University of Hawaii, according to facility protocols.

We extracted prokaryotic RNA from triplicate control and +AHL samples, pooling together triplicate samples, sequencing 60 million paired end reads, normalizing and assembling as previously described (Frischkorn et al., 2018). To obtain read counts for each sample, we mapped cleaned forward and reverse reads to metagenome assemblies from the same sampling locations that were previously characterized and clustered into orthologous groups (OGs) (Frischkorn et al., 2017). We carried out mapping using RSEM with the paired-end and Bowtie2 parameters (Li and Dewey, 2011) and tabulated counts across OGs for *Trichodesmium* and epibiont genome bins separately. We determined significant changes in OG relative abundance between control and +AHL samples using a stringent empirical Bayes approach called Analysis of Sequence Counts (ASC) (Wu et al., 2010). This approach evaluates the posterior probability associated with a given fold change across the pooled triplicates, and performs similarly, but conservatively, on replicated and unreplicated sample datasets (Wu et al., 2010). OGs were considered significantly higher or lower if they had a 95% or higher posterior probability of a fold change greater than 2 between treatment

and control, as previously described (Dyhrman et al., 2012). Taxonomic relative abundance estimates for metagenome samples were previously calculated (Frischkorn et al., 2017).

5.4 Results and discussion

5.4.1 Biological interactions are a driver of *Trichodesmium* N_2 fixation

Trichodesmium in situ N_2 fixation rates increased from north to south, with stations north of 20° latitude having significantly lower rates of N_2 fixation than those in the south (P=0.02, one-way ANOVA; Figure 5.1a; Supplemental Table D.1). Surface total dissolved phosphorus (TDP) concentration in the stations north of 20° latitude differed significantly from those collected to the south (P<0.03, one-way ANOVA; Figure 5.1a; Supplemental Table D.1). Phosphorus is known to be a limiting nutrient for N_2 fixation for *Trichodesmium* in this region (Sañudo-Wilhelmy et al., 2001; Sohm et al., 2011). Although the changes in N_2 fixation and TDP between the northern and southern stations are consistent with phosphorus being a strong driver of N_2 fixation, across the transect the *in situ* rates of N_2 fixation were not significantly correlated with TDP ($R^2 < 0.319$, P=0.186, one-way ANOVA; Figure 5.1a). This suggests that factors other than phosphorus concentration might also influence *Trichodesmium* N_2 fixation.

At all stations where we performed experiments (Figure 5.1a), N_2 fixation was significantly changed in response to microbiome manipulation with the +AHL amendment, with significant decreases in rate at station 2 and 10 (P=0.04 and 0.03 respectively), and a significant increase at station 17 (P=0.01) (Figure 5.1b). Although *Trichodesmium* has co-occurred with a N_2 fixing epibiont in the North Pacific Ocean (Gradoville et al., 2017; Momper et al., 2014), herein *Trichodesmium* colonies did not show evidence of non-*Trichodesmium nif* genes (Frischkorn et al., 2017), indicating that *Trichodesmium* was the only diazotrophic organism within these samples. The magnitude and direction of microbiome-induced changes in host N_2 fixation ranged from significantly decreased at the northernmost station to significantly increased at the southern-most station, despite similarities in TDP between proximal stations (Figure 5.1a,c). This suggests that biotic inter-

actions within colonies can act independently of geochemistry to influence Trichodesmium physiology.

Extrapolating the range of changes we observed in +AHL treatments to *in situ* N_2 fixation rates, we illustrate the potential influence of the microbiome on host physiology (Figure 5.2). This theoretical N_2 fixation range was determined from the maximum (+20%) and minimum (-41%) percent change after +AHL amendment to contextualize the potential for biological interactions to alter observed N_2 fixation rates (Figure 5.2). Although this visualization should be interpreted with caution, the hypothetical 61% range of variation reflects a scenario where biological interactions drive subsequent N_2 fixation higher or lower than would otherwise be expected, even given noted uncertainty in these measurements (coefficient of variation = 10.2%). Such changes are modulated on rapid time scales (<4 hours) and are comparable to the range of variability in community N_2 fixation induced by longer nutrient amendment experiments (Mills et al., 2004), as well as that observed across an Atlantic meridional phosphorus gradient (Moore et al., 2009). The variability we observed suggests that interactions in the microbiome should be considered drivers of N_2 fixation in addition to well-studied constraints like iron, phosphorus, and CO_2 concentration (Hutchins et al., 2015; Paerl, 1994; Sañudo-Wilhelmy et al., 2001). Overall, it may be important to consider these biotic factors in addition to light and surface ocean nutrients when predicting *Trichodesmium* physiological ecology.

5.4.2 Microbiome transcriptional patterns varied with shifts in *Trichodesmium* N_2 fixation

There are 8 core epibiont genomes in samples from this transect (Frischkorn et al., 2017) and they all had OGs with significantly increased or decreased relative abundance in +AHL treatments compared to the controls at all stations tested (Table 5.1). This result suggests all members of the microbiome can respond to AHLs and is consistent with the fact that all genome bins were found to possess putative *LuxR* genes from canonical quorum sensing operons (Frischkorn et al., 2017). The Gammaproteobacteria (bin 7) contributed to 60% of the significantly responsive functional categories across all three stations tested (Figure 5.3). Notably, such Gammaproteobacteria are ubiquitously found in association

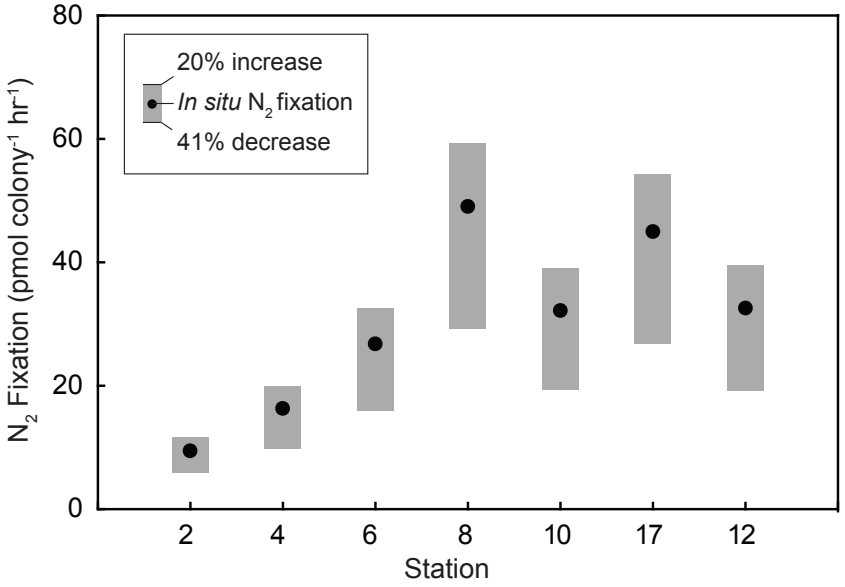

Figure 5.2: Illustration of potential microbiome influence on *in situ Trichodesmium* N_2 fixation across the western North Atlantic transect. Black circles represent *in situ* N_2 fixation rates observed at each station. To highlight the potential magnitude of the effects that biological interactions can have on *Trichodesmium* N_2 fixation, the maximum and minimum percent changes we observed after +AHL amendment of colonies in this study were extrapolated to in situ colony N_2 fixation rates. The grey bars reflect this theoretical range (61%) of variability in N_2 fixation that could occur on rapid (<4 hour) time scales as a result of changes in microbiome activity. Average coefficient of variation of N_2 fixation rates in replicate incubations was 10.2%. Stations are organized in order from north to south.

with colonies across multiple environments and in culture (Lee et al., 2017) and biologically relevant concentrations of AHLs have been found in *Trichodesmium* colonies (Van Mooy et al., 2012). Taken together, it is likely that the microbiome could ubiquitously influence *Trichodesmium* N_2 fixation via quorum sensing pathways.

Table 5.1: Number of OGs from individual microbiome members that had significantly differential relative abundance (SDRA) after +AHL amendment. Counts were tabulated for each gene in a microbiome genome bin that belongs to an OG that was found to be significantly higher or lower in response to +AHL amendment. OGs total = total number of OGs found in each genome bin

Bin	Identity	OGs total	St2: Contribution to SDRA OGs	St10: Contribution to SDRA OGs	St17: Contribution to SDRA OGs	Avg. % of SDRA OGs per Genome
4	Bacteroidetes	8354	121	61	247	1.71
5	Rhodospirillales	2787	26	32	48	1.27
6	Bacteroidetes	6016	41	26	155	1.23
7	Gammaproteobacteria	4195	105	55	163	2.57
8	Rhodobacterales	2693	75	15	102	2.38
10	Rhodospirillales	4206	53	34	106	1.53
11	Rhodospirillales	1766	51	13	60	2.34
12	Rhodospirillales	1884	29	24	34	1.54

Amendment with AHLs induced significant changes in OG relative abundance at each station (Figure 5.3). While the responses were distinct (Supplemental Figure D.1), the OGs encoded a range of functions that aligned at the broadest level with those expected to respond in *Pseudomonas aeruginosa* (Schuster and Greenberg, 2006; Wagner et al., 2003, 2004). The relative abundance of OGs encoding carbohydrate metabolism, cell cycle control and environmental information processing functions were significantly increased in the microbiome at all three stations (Figure 5.3). A suite of OGs encoding environmental information processing functions also had significantly decreased relative abundance at all three stations, which, in addition to transposon and transcription functions, were the only shared responses with decreased relative abundance across the experiments (Figure 5.3). This shared functional response in the expression of environmental information processing and transcription functions may be driven by a hierarchical genetic response induced by

quorum sensing. In model bacteria, AHLs activate a cascade of transcription factors that go on to increase or decrease expression of the genes under their control (Latifi et al., 1996; Whiteley et al., 1999). AHLs are also known to stimulate gene transfer agents and increase transposon mobility (Auchtung et al., 2005; Schaefer et al., 2002), and many quorum sensing genes are adjacent to transposons or encoded within them (Thomson et al., 2000; Wei et al., 2006). Station 17, the only station where +AHL amendment resulted in a significant increase in *Trichodesmium* N_2 fixation, exhibited the broadest range of functions with significantly different expression (Figure 5.3). The increase in host N_2 fixation observed in response could be related to the decreased expression at Station 17 of microbiome transporters for iron and phosphate, resources that are critical to or affected by N_2 fixation, thus ensuring these resources are available to *Trichodesmium*.

Although there were similarities across broad metabolic categories at the stations tested, at the gene family level there was little overlap between specific OGs that contributed to those categories across the three stations (Supplemental Figure D.1). All core microbiome members were consistently responsive to +AHL amendment, but only 5 microbiome OGs were significantly increased or decreased across all stations (Supplemental Figure D.1). These OGs were all annotated as putative, uncharacterized proteins. Sampling a time course in AHL addition experiments would help identify quorum sensing pathways and if there is an initial conserved response in the microbiome that was unique by the 4 hour time point observed here. Regardless, the variability in the OG transcriptional responses here is consistent with the disparate N_2 fixation responses to AHL addition.

Little is known about the quorum sensing and quenching pathways in the *Trichodesmium* microbiome, but in model systems these pathways are complex, operating in circuits that can influence each other in distinct ways and lead to a cascade of unique transcriptional responses and resulting shifts in activity (Schuster and Greenberg, 2006; Wagner et al., 2004). Previous work found that community composition of the microbiome varied significantly between northern stations (e.g. station 2) and the southern stations (e.g. station 10 and 17), but that the *Trichodesmium* host did not (Frischkorn et al., 2017). However, in the southern stations where community composition and geochemistry were similar, transcriptional responses in the microbiome were dissimilar and yielded different effects on N_2 fixation. For

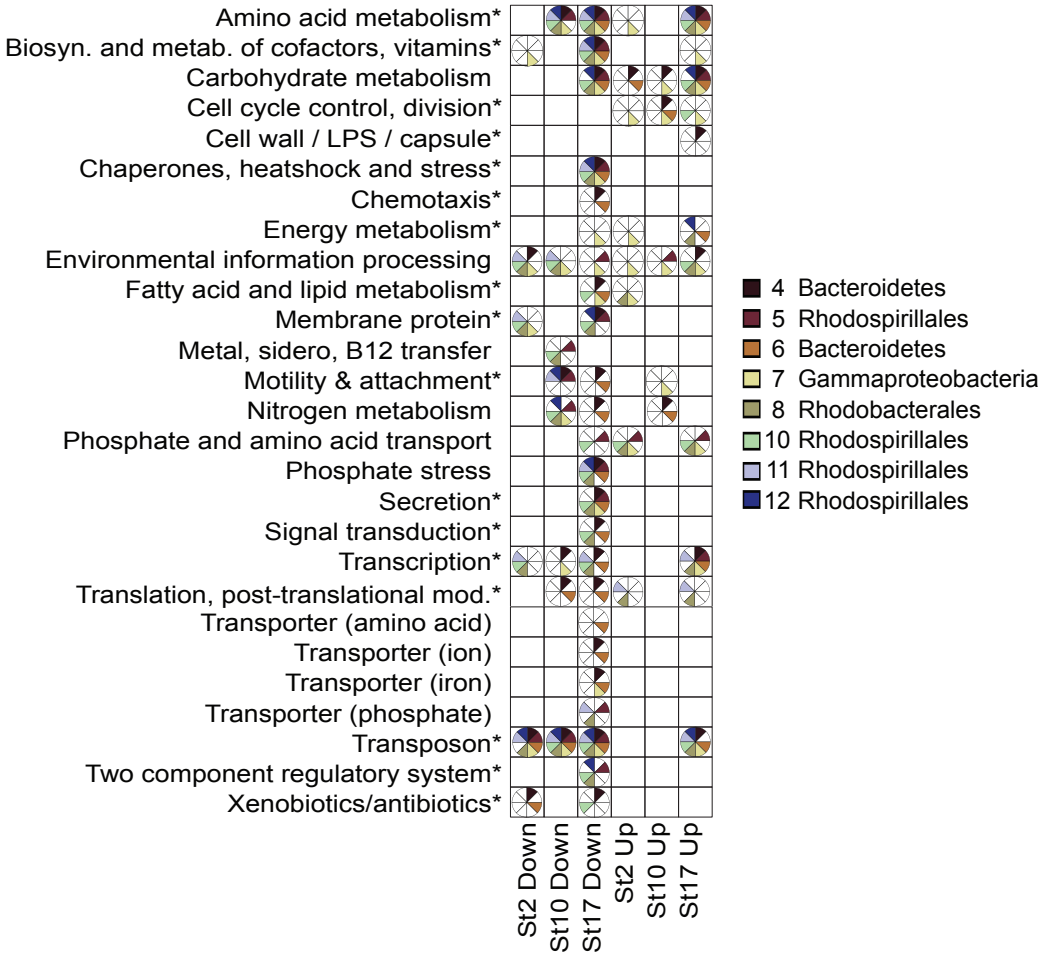

Figure 5.3: Microbiome functional gene categories with significant changes in relative abundance between the control and the +AHL treatments. Pie wedges are filled if that epibiont's genome bin contributed to an OG with significantly increased or decreased relative abundance that fell into a functional category as determined by annotation against the Kyoto Encyclopedia of Genes and Genomes (KEGG). Starred functional categories indicate categories that are known to have significantly increased or decreased relative abundance after stimulating quorum sensing in the model bacterium *Pseudomonas aeruginosa* (Schuster and Greenberg, 2006; Wagner et al., 2003).

example, although relative abundances of the core microbiome members between Stations 10 and 17 were not significantly different (Frischkorn et al., 2017), and TDP was similar, variable epibiont transcription forced by quorum sensing elicited different *Trichodesmium* N_2 fixation responses. In sum, N_2 fixation responses are not predictable from microbiome community composition alone. Variability in AHL responses in model systems can vary as much as 100-fold due to nutrient status, oxygen availability, and whether cultures were planktonic or growing as a biofilm (Duan and Surette, 2007; Schuster and Greenberg, 2006). Furthermore, the transcriptional regulators that are activated by quorum sensing molecules do not exist in isolation, but rather interact with a web of regulators and quorum quenching molecules that affect physiology on a genome-wide scale—a finding that has been used as an explanation for the rapid adaptability of bacteria to fluctuating environments (Schuster and Greenberg, 2006). The concentration of the AHLs experienced by the epibionts could also affect the direction, magnitude, and characteristics of gene expression, as hydrolytic enzyme activity in marine particulate matter is strongly affected by the concentration of AHLs added (Krupke et al., 2016). Similarly, in *P. aeruginosa*, different concentrations of the same quorum sensing signaling molecule result in different responses that subsequently elicit opposite host physiological responses (Williams and Cámara, 2009). In the *Trichodesmium* holobiont the effects of the microbiome on N_2 fixation likely reflect a complex interplay of environment, community composition, chemical signaling, and functional response, and a more mechanistic understanding of activities in the microbiome is needed to model how biological interactions modulate *Trichodesmium* N_2 fixation.

5.4.3 Conclusions

Here we show that selective manipulation of microbiome activities can alter the N_2 fixation rate of the *Trichodesmium* host over short time scales, expanding the suite of factors that are known drivers of marine N_2 fixation. If the observed interplay between host and microbiome holds true across the full range of oligotrophic environments *Trichodesmium* inhabits, then these interactions are likely an overlooked factor that influences *Trichodesmium* N_2 fixation, and future ecological studies of *Trichodesmium* should take into account the activities of the microbiome.

Chapter 6

Conclusion

6.1 Dissertation summary

Trichodesmium is a keystone marine cyanobacterium: a diazotroph thought to supply nearly half of the total biologically fixed nitrogen in the ocean (Bergman et al., 2013). The occurrence of this organism in the natural environment has been a source of wonder for millennia. Thousands of years before the first microorganism was discovered, the sudden blood red darkening of the sea—now thought to be blooms of *Trichodesmium*—were given supernatural import: they were a cosmological phenomenon or a portentous omen (Ehrenberg, 1830). *Trichodesmium* was described by Jules Verne in *20,000 Leagues Under the Sea*[1], noted by Charles Darwin as he sailed on the *Beagle*[2], and cataloged by Captain Cook during his expeditions on the H.M.S *Endeavour*[3]. In the age of modern scientific inquiry, it was found that *Trichodesmium* was capable of turning inert N_2 gas into biologically available ammonium (Dugdale et al., 1961). *Trichodesmium* was subsequently propelled into planetary biogeochemical relevance with the discovery that its global input to "new" nitrogen was underestimated by approximately 15 fold (Capone et al., 1997; Mahaffey et al., 2005; Westberry and Siegel, 2006).

Today, the availability of high throughput sequencing technologies and the bioinformatic tools to analyze "big data" have ushered in a new age of *Trichodesmium* discoveries pioneered by molecular microbial oceanographers. For example, genome sequencing revealed an expanding genetic capacity in *Trichodesmium* that contrasted paradoxically with other marine bacteria that exhibit genomic streamlining to survive in the oligotrophic ocean (Walworth et al., 2015). Metagenomics and 16S amplicon sequencing have been used to show in the field and in laboratory isolates the presence of ubiquitous microbiome members, the metabolic capabilities they possesses, and how they are altered across environments (Lee et al., 2017; Rouco et al., 2016a). Transcriptome sequencing and epigenetics have revealed the mechanisms of how *Trichodesmium* N_2 fixation could be expected to change under future ocean conditions (Hutchins et al., 2015; Walworth et al., 2017a). Metatranscriptomics in

[1]Verne, Jules, 1828-1905. *20,000 Leagues Under the Sea.* New York; Bantam Dell, 1962. Print.

[2]Darwin, Charles, 1809-1882. *The Voyage of the Beagle.* New York; Modern Library, 2001. Print.

[3]J. C. Beaglehole, Ed., *The* Endeavor *Journals of Joseph Banks 1768–1771.* Cambridge; Cambridge University Press, 1955. Print.

the field have even been wielded to understand the molecular and ecological underpinnings of a *Trichodesmium* bloom's demise (Spungin et al., 2016). Such technologies continue become cheaper, more efficient, and increasingly innovative. There are no doubt more *Trichodesmium* discoveries to be made—oceanographers will be limited only by the ability to parse through staggeringly large datasets and interpret them using databases that are not optimized for marine microorganisms.

In this thesis I set out to codify the importance of the microbiome to *Trichodesmium's* ecology and physiology. This line of inquiry was a critical next step, considering the importance of this keystone organism to global elemental cycles as well as growing appreciation of the influence of a microbiome on its host. Broadly, the studies I carried out delved into unanswered questions about the physiological capabilities of the *Trichodesmium* microbiome, how those capabilities respond across diverse environments and are intertwined with the host physiology, and the influence of previously unappreciated biological interactions in ecologically and biogeochemically important processes like N_2 fixation. Encompassing four field expeditions across three ocean basins, my research has implemented cutting edge nucleic acid sequencing approaches coupled with traditional oceanographic and biochemical measurements to investigate this host microbiome system.

To first establish the importance and implications of the relationship between *Trichodesmium* and the microbiome, Chapter 2 used metagenomic sequencing to show that the microbiome possesses a large amount of metabolic functional potential and is altered across a geochemical gradient in the North Atlantic ocean. In Chapter 3, I showed that this metabolic potential was indeed wielded across the oligotrophic environment of the western tropical South Pacific, uncovering the mechanisms of how *Trichodesmium* and their microbiome influence the reduced phosphorus cycle. The coordinated gene expression patterns within consortia that I demonstrated in Chapter 4 indicated metabolic interactions between *Trichodesmium* and the microbiome encompass carbon, nitrogen and vitamin pathways in addition to phosphorus, interactions that might short circuit the transfer of these resources to the water column at large. Finally, in Chapter 5, I demonstrated that interactions with the microbiome could impact *Trichodesmium's* physiology by showing that selective manipulation of the microbiome can alter host N_2 fixation.

In short, I believe that this research creates a frame shift in how we think about *Trichodesmium's* physiology and ecology: this organism cannot be considered in isolation, but rather as an integrated microscopic community in which the interactions have global relevance. Coupled with traditional geochemical measurements that are canonical in biological oceanography, considering biotic interactions will strengthen numerical models of ocean processes, predictions, and experimental efforts pertaining to *Trichodesmium*. With these approaches together, we might come closer to being able to answer a persistent question in oceanography: what forces determine the structure of microbial communities in aquatic habitats, and subsequently, how do they influence the planet?

6.2 Future directions

Throughout the course of my dissertation research I have generated several terabytes worth of sequence data, genome and transcriptome assemblies, and ancillary analysis files. Stated more simply: that's a lot of data. With each sample encompassing a yield of up to hundreds of thousands of individual genes—many of which have only hypothetical annotations of function—delving into a new project can seem like a daunting task. With new bioinformatic programs published frequently, embarking on big data analysis involves some measure of reinventing the wheel. Despite the many technological advances that have made collecting all these data points possible, the results in this thesis are on the precipice of the unknown: databases to annotate genes, the bread and butter of my research, are not optimized for marine microbes. Thus, a large proportion of those hundreds of thousands of genes have only hypothetical or unknown functions. Still, the more data generated, the more these knowledge gaps can be filled in. Six years ago, it would have been impossible to partition metatranscriptomic reads between *Trichodesmium* and their microbiome. Today, I can do this with a few keystrokes. Coupling all of these molecular data with more traditional experimental approaches—like the geochemical, enzyme activity, and uptake rate measurements I used in my research—ground truth functional inferences and provide environmental context. Applied to the *Trichodesmium* system, the ultimate yield is a better

appreciation of the integral associations within these consortia, the metabolic currencies exchanged, and the potential geochemical implications of these interactions.

Even though these novel approaches to understanding *Trichodesmium* ecology push the envelope of what is possible, it is still critical to think towards the next frontier. With sequence data from across the globe, it is possible to delve into questions of how microbial communities are structured using the relatively simple *Trichodesmium* consortium as a model system (Morris and Hmelo, 2014). Based on existing research, it is clear that the bacterial associations with *Trichodesmium* are not stochastic. Similar species cohabitate with *Trichodesmium* across ocean basins (Lee et al., 2017; Rouco et al., 2016a), but are these associations driven by key physiological capabilities that these individual epibiont species bring to the holobiont? Such questions are a key next step that could be addressed with existing data.

In a different vein, gaps left by the unknown and the hypothetical genes could be addressed by co-opting new molecular techniques that enable targeted genome modification. Ideally, such investigations should bridge targeted laboratory experiments with environmental surveys, and ultimately modeling efforts. With existing cultures of *Trichodesmium*, and its macroscopic colonial morphology that renders it relatively easy to detect in field samples, not to mention the wealth of archived genetic data, this system is well suited to these efforts. I am intrigued by the prospect of applying genome editing technology like CRISPR (Nymark et al., 2016; Rastogi et al., 2016) in *Trichodesmium* and the microbiome to test the function of specific genes known or unknown, in solo or in combination. For example, targeting some of the pathways I found to be implicated in phosphorus or iron exchange between *Trichodesmium* and microbiome for CRISPR deletion would enable delving into the physiological mechanisms that structure and maintain a beneficial relationship in the holobiont and could be coupled with the aforementioned questions of the principles of community structure (Morris and Hmelo, 2014).

The wealth of global marine microbial sequence data is also primed for bridging laboratory experiments with environmental surveys in a way that targets specific mechanisms. Boiteau and colleagues recently investigated microbial iron physiology the low iron Pacific Ocean using one such tiered approach (Boiteau et al., 2016). Their approach used mass

spectrometry to detect iron chelating siderophore compounds across a transect, coupled with searches of the Tara Ocean Expedition gene catalog (Sunagawa et al., 2015) for the presence of siderophore biosynthesis genes, as well as a phylogenetic analysis of the molecular history of these genes (Boiteau et al., 2016). Together, their results highlight a previously mysterious ocean mechanism: siderophore-based adaptations to low iron conditions that have been horizontally transferred between bacteria to a point where they are detectable across the global ocean (Boiteau et al., 2016). Considering the global relevance of *Trichodesmium* to nitrogen and carbon cycles, applying similar approaches could yield fruitful insights. Detailed investigations of specific processes on a global level could be particularly informative inputs to biogeochemical models (Coles et al., 2017; Follows et al., 2007), especially considering the noted challenges to accurately modeling *Trichodesmium* (Capone et al., 2005; McGillicuddy, 2014).

Ultimately, the microscopic organisms that reside in the ocean form the microbiome of Earth. I predict that much like the human microbiome is the arbiter of health and disease in the body, the microbial systems of the natural environment are integral to the health of the planet (Blaser, 2014; Sampson and Mazmanian, 2015). As my research and others' have demonstrated, these microbes interact with one another and the environment, and such interactions and transformations influence geochemical cycles: through primary productivity, the sequestration of the greenhouse gas CO_2, and the overall habitability of the planet (Karl, 1999). Building a holistic understanding of microbial processes in the environment will be essential to forecast the ocean's role in the future climate.

With unmitigated climate change, the future oceans are forecasted to have higher temperatures, more acidic waters, and more expansive oligotrophic regions (Hutchins et al., 2015; McGillicuddy, 2014; Riebesell et al., 2009). Previous research has demonstrated that such conditions could expand the geographic range of *Trichodesmium* (Hutchins et al., 2007), increase its overall fitness (Walworth et al., 2016), and lead to elevated carbon and N_2 fixation rates (Hutchins et al., 2015). As of yet, these studies have performed experiments, modeled, and made predictions about *Trichodesmium* without taking the microbiome into account. Rising temperatures and CO_2 concentrations, however, are known to alter marine host-microbiome relationships with deleterious consequences (Ainsworth et al., 2010;

Bourne et al., 2016; Lesser et al., 2016). Whatever the next phase of *Trichodesmium* research entails, as the research in this thesis has demonstrated, it is critical that future investigations take into account biological interactions and the influence of the microbiome.

Bibliography

Achilles, K. M. et al. (2003). Bioavailability of iron to *Trichodesmium* colonies in the western subtropical Atlantic Ocean. *Limnol. Oceanogr.* 48(6):2250–2255. DOI: 10.4319/lo.2003.48.6.2250.

Ainsworth, T. D. et al. (2010). The future of coral reefs: a microbial perspective. *Trends Ecol. Evol.* 25(4):233–240. DOI: 10.1016/j.tree.2009.11.001.

Alexander, H. et al. (2015). Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc. Natl. Acad. Sci.* 112(17):E2182–E2190. DOI: 10.1073/pnas.1421993112.

Anders, S. and W. Huber (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11(10):R106.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26:32–46.

Arrigo, K. R. (2005). Marine microorganisms and global nutrient cycles. *Nature* 437(7057):343–348. DOI: 10.1038/nature04158.

Auchtung, J. M. et al. (2005). Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci.* 102(35):12554–12559. DOI: 10.1073/pnas.0505835102.

Aylward, F. O. et al. (2014). Convergent bacterial microbiotas in the fungal agricultural systems of insects. *MBio* 5(6):e02077. DOI: 10.1128/mBio.02077-14.

Aylward, F. O. et al. (2015). Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc. Natl. Acad. Sci.* 112(17):5443–5448. DOI: 10.1073/pnas.1502883112.

Azam, F. et al. (1983). The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* 10:257–263.

Aziz, R. K. et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9(1):75. DOI: 10.1186/1471-2164-9-75.

Bassler, B. L. (2012). Microbes as Menaces, Mates & Marvels. *Daedalus* 141(3):67–76. DOI: 10.1162/DAED_a_00162.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57(1):289–300.

Bergman, B. et al. (2013). *Trichodesmium*—a widespread marine cyanobacterium with unusual nitrogen fixation properties. *FEMS Microbiol. Rev.* 37(3):286–302. DOI: 10.1111/j.1574-6976.2012.00352.x.

Berman-Frank, I. et al. (2001). Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. *Limnol. Oceanogr.* 46(6):1249–1260. DOI: 10.4319/lo.2001.46.6.1249.

Berthelot, H. et al. (2015). Assessment of the dinitrogen released as ammonium and dissolved organic nitrogen by unicellular and filamentous marine diazotrophic cyanobacteria grown in culture. *Front. Mar. Sci.* 2:80. DOI: 10.3389/fmars.2015.00080.

Bertrand, E. M. et al. (2015). Phytoplanktonbacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci.* 112(32):9938–9943.

Blain, S. et al. (2008). Dissolved iron distribution in the tropical and subtropical South Eastern Pacific. *Biogeosciences* 5(1):269–280. DOI: 10.5194/bg-5-269-2008.

Blaser, M. J. (2014). The microbiome revolution. *J. Clin. Invest.* 124(10):8–11. DOI: 10.1172/JCI78366.

Bock, N. et al. (2018). Microbial community structure along trophic and N_2 fixation gradients in the Western Tropical South Pacific. *Biogeosciences* 15(12):3909–3925. DOI: 10.5194/bg-15-3909-2018.

Boiteau, R. M. et al. (2016). Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific Ocean. *Proc. Natl. Acad. Sci.* 113(50):14237–14242. DOI: 10.1073/pnas.1608594113.

Bonnet, S. et al. (2010). Vitamin B12 excretion by cultures of the marine cyanobacteria *Crocosphaera* and *Synechococcus*. *Limnol. Oceanogr.* 55(5):1959–1964. DOI: 10.4319/lo.2010.55.5.1959.

Bonnet, S. et al. (2017). Hot spot of N_2 fixation in the western tropical South Pacific pleads for a spatial decoupling between N_2 fixation and denitrification. *Proc. Natl. Acad. Sci.* 114(14):E2800–E2801. DOI: 10.1073/pnas.1619514114.

Bonnet, S. et al. (2018). In depth characterization of diazotroph activity across the Western Tropical South Pacific hot spot of N_2 fixation. *Biogeosciences* 15(13):4215–4232. DOI: 10.5194/bg-2017-567.

Bordenstein, S. R. and K. R. Theis (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLOS Biol.* 13(8):e1002226. DOI: 10.1371/journal.pbio.1002226.

Bourne, D. G. et al. (2016). Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annu. Rev. Microbiol.* 70(1):317–340. DOI: 10.1146/annurev-micro-102215-095440.

Bronk, D. A. et al. (1994). Nitrogen uptake, dissolved organic nitrogen release, and new production. *Science* 265(5180):1843–1846. DOI: 10.1126/science.265.5180.1843.

Brown, C. T. et al. (2013a). khmer protocols documentation. https://khmer-protocols.readthedocs.io/en/latest/mrnaseq/index.html.

Brown, J. M. et al. (2013b). Characterization of *Trichodesmium*-associated viral communities in the eastern Gulf of Mexico. *FEMS Microbiol. Ecol.* 84(3):603–613. DOI: 10.1111/1574-6941.12088.

Brussaard, C. P.D. et al. (2016). The interactive microbial ocean. *Nat. Microbiol.* 2:16255. DOI: 10.1038/nmicrobiol.2016.255.

Buchfink, B. et al. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12(1):59–60. DOI: 10.1038/nmeth.3176.

Caffin, M. et al. (2018a). N_2 fixation as a dominant new N source in the western tropical South Pacific Ocean (OUTPACE cruise). *Biogeosciences* 15(8):2565–2585. DOI: 10.5194/bg-15-2565-2018.

Caffin, M. et al. (2018b). Transfer of diazotroph-derived nitrogen to the planktonic food web across gradients of N_2 fixation activity and diversity in the Western Tropical South Pacific. *Biogeosciences* 15(12):3795–3810. DOI: 10.5194/bg-15-3795-2018.

Capone, D. G. (1993). Determination of nitrogenase activity in aquatic samples using the acetylene reduction procedure. *Handbook of Methods in Aquatic Microbial Ecology*:ed. by E. B. Sherr.

Capone, D. G. (2001). Marine nitrogen fixation: what's the fuss? *Curr. Opin. Microbiol.* 4(3):341–348.

Capone, D. G. et al. (1997). *Trichodesmium*, a globally significant marine cyanobacterium. *Science* 276(5316):1221–1229.

Capone, D. G. et al. (1998). An extensive bloom of the N_2-fixing cyanobacterium *Trichodesmium erythraeum* in the central Arabian Sea. *Mar. Ecol. Prog. Ser.* 172:281–292.

Capone, D. G. et al. (2005). Nitrogen fixation by *Trichodesmium* spp.: an important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochem. Cycles* 19(2):GB2024. DOI: 10.1029/2004GB002331.

Cárdenas, A. et al. (2018). Excess labile carbon promotes the expression of virulence factors in coral reef bacterioplankton. *ISME J.* 12(1):59–76. DOI: 10.1038/ismej.2017.142.

Case, R. J. et al. (2008). AHL-driven quorum-sensing circuits: their frequency and function among the Proteobacteria. *ISME J.* 2:345–349. DOI: 10.1038/ismej.2008.13.

Chappell, P. D. and E. A. Webb (2010). A molecular assessment of the iron stress response in the two phylogenetic clades of *Trichodesmium*. *Environ. Microbiol.* 12(1):13–27. DOI: 10.1111/j.1462-2920.2009.02026.x.

Chappell, P. D. et al. (2012). Molecular evidence of iron limitation and availability in the global diazotroph *Trichodesmium*. *ISME J.* 6(9):1728–1739. DOI: 10.1038/ismej.2012.13.

Chen, C. C. H. et al. (2006). Structure and kinetics of phosphonopyruvate hydrolase from *Voriovorax* sp. Pal2: New insight into the divergence of catalysis within the pep mutase/isocitrate lyase superfamily. *Biochemistry* 45(38):11491–11504. DOI: 10.1021/bi061208l.

Coles, V. J. et al. (2017). Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* 358(6367):1149–1154. DOI: 10.1126/science.aan5712.

Croft, M. T. et al. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438(7064):90–93. DOI: 10.1038/nature04056.

De Verneil, A. et al. (2017). The fate of a southwest Pacific bloom: gauging the impact of submesoscale vs. mesoscale circulation on biological gradients in the subtropics. *Biogeosciences* 14(14):3471–3486. DOI: 10.5194/bg-14-3471-2017.

Deutsch, C et al. (2007). Spatial coupling of nitrogen inputs and losses in the ocean. *Nature* 445(7124):163–167. DOI: 10.1038/nature05392.

Dombrowski, N. et al. (2016). Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nat. Microbiol.* 1(7):16057. DOI: 10.1038/NMICROBIOL.2016.57.

Duan, K. and M. G. Surette (2007). Environmental regulation of *Pseudomonas aeruginosa* PAO1 *las* and *Rhl* quorum-sensing systems. *J. Bacteriol.* 189(13):4827–4836. DOI: 10.1128/JB.00043-07.

Ducklow, H. W. et al. (2001). Upper ocean carbon export and the biological pump. *Oceanography* 14(4):50–58. DOI: 10.5670/oceanog.2001.06.

Dugdale, R. C. et al. (1961). Nitrogen fixation in the Sargasso Sea. *Deep-Sea Res.* 7:298–300.

Dyhrman, S. T. and K. C. Ruttenberg (2006). Presence and regulation of alkaline phosphatase activity in eukaryotic phytoplankton from the coastal ocean: implications for dissolved organic phosphorus remineralization. *Limnol. Oceanogr.* 51(3):1381–1390.

Dyhrman, S. T. et al. (2002). Cell-specific detection of phosphorus stress in *Trichodesmium* from the western North Atlantic. *Limnol. Oceanogr.* 47(6):1832–1836.

Dyhrman, S. T. et al. (2006). Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* 439(7072):68–71. DOI: 10.1038/nature04203.

Dyhrman, S. T. et al. (2007). Microbes and the marine phosphorus cycle. *Oceanogr.* 20(2):110–116. DOI: 10.5670/oceanog.2007.54.

Dyhrman, S. T. et al. (2009). A microbial source of phosphonates in oligotrophic marine systems. *Nat. Geosci.* 2(10):696–699. DOI: 10.1038/ngeo639.

Dyhrman, S. T. et al. (2012). The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS One* 7(3):e33768. DOI: 10.1371/journal.pone.0033768.

Eddy, S. R. (1995). Multiple alignment using hidden Markov models. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 3:114–120.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792–1797. DOI: 10.1093/nar/gkh340.

Ehrenberg, C. G. (1830). Neue Beobachlungen uber blutartige Erscheinungen in Aegypten, Arabien und Sibirien, nebst einer Uebersicht und Kritik der fruher bekannten. *Ann. der Phys. und Chemie* 94(4):477–514.

Engel, P. et al. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci.* 109(27):11002–11007. DOI: 10.1073/pnas.1202970109.

Eppley, R. W. and B. J. Peterson (1979). Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* 282:677–680.

Erdner, D. L. and D. M. Anderson (1999). Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. *Limnol. Oceanogr.* 44(7):1609–1615. DOI: 10.4319/lo.1999.44.7.1609.

Falkowski, P. G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. *Nature* 387(6630):272–275. DOI: 10.1038/387272a0.

Falkowski, P. G. et al. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* 281(5374):200–207. DOI: 10.1126/science.281.5374.200.

Fetzner, S. (2014). Quorum quenching enzymes. *J. Biotechnol.* 201:2–14. DOI: 10.1016/j.jbiotec.2014.09.001.

Follows, M. J. et al. (2007). Emergent biogeography of microbial communities in a model ocean. *Science* 315(5820):1843–1846. DOI: 10.1126/science.1138544.

Fontanez, K. M. et al. (2015). Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. *Front. Microbiol.* 6:469. DOI: 10.3389/fmicb.2015.00469.

Frischkorn, K. R. et al. (2014). *De novo* assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. *Front. Microbiol.* 5:375. DOI: 10.3389/fmicb.2014.00375.

Frischkorn, K. R. et al. (2017). Epibionts dominate metabolic functional potential of of *Trichodesmium* colonies from the oligotrophic ocean. *ISME J.* 11(9):2090–2101. DOI: 10.1038/ismej.2017.74.

Frischkorn, K. R. et al. (2018). Coordinated gene expression between *Trichodesmium* and its microbiome over day-night cycles in the North Pacific Subtropical Gyre. *ISME J.* 12(4):997–1007. DOI: 10.1038/s41396-017-0041-5.

Gradoville, M. R. et al. (2017). Microbiome of *Trichodesmium* colonies from the North Pacific Subtropical Gyre. *Front. Microbiol.* 8:1122. DOI: 10.3389/fmicb.2017.01122.

Guieu, C. et al. (2018). Iron from a submarine source impacts the productive layer of the Western Tropical South Pacific. *Sci. Rep.* 8:9075. DOI: 10.1038/s41598-018-27407-z.

Handley, K. M. et al. (2012). Biostimulation induces syntrophic interactions that impact C, S and N cycling in a sediment microbial community. *ISME J.* 7(4):800–816. DOI: 10.1038/ismej.2012.148.

Hazra, A. B. et al. (2015). Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proc. Natl. Acad. Sci.* 112(34):10792–10797. DOI: 10.1073/pnas.1509132112.

Heal, K. R. et al. (2017). Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *Proc. Natl. Acad. Sci.* 114(2):364–369. DOI: 10.1073/pnas.1608462114.

Helliwell, K. E. et al. (2016). Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B12. *Curr. Biol.* 26(8):999–1008. DOI: 10.1016/j.cub.2016.02.041.

Hentschel, U. et al. (2012). Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* 10(9):641–654. DOI: 10.1038/nrmicro2839.

Herbst, V. and J Overbeck (1978). Metabolic coupling between the alga *Oscillatoria redekei* and accompanying bacteria. *Naturwissenschaften* 65(11):598–599. DOI: 10.1007/BF00364919.

Hewson, I. et al. (2009a). *In situ* transcriptomic analysis of the globally important keystone N_2-fixing taxon *Crocosphaera watsonii*. *ISME J.* 3(5):618–631. DOI: 10.1038/ismej.2009.8.

Hewson, I. et al. (2009b). Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *ISME J.* 3(11):1286–1300. DOI: 10.1038/ismej.2009.75.

Hmelo, L. R. et al. (2011). Possible influence of bacterial quorum sensing on the hydrolysis of sinking particulate organic carbon in marine environments. *Environ. Microbiol. Rep.* 3(6):682–688. DOI: 10.1111/j.1758-2229.2011.00281.x.

Hmelo, L. et al. (2012). Characterization of bacterial epibionts on the cyanobacterium *Trichodesmium*. *Aquat. Microb. Ecol.*1–14. DOI: 10.3354/ame01571.

Hong, K. W. et al. (2012). Quorum quenching revisited-from signal decays to signalling confusion. *Sensors* 12(4):4661–4696. DOI: 10.3390/s120404661.

Hopkinson, B. M. and François M M Morel (2009). The role of siderophores in iron acquisition by photosynthetic marine microorganisms. *Biometals* 22(4):659–69. DOI: 10.1007/s10534-009-9235-2.

Howarth, R. W. and R. Marino (1988). Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 2. Biogeochemical controls. *Limnol. Oceanogr.* 33:688–701.

Huson, D. H. et al. (2013). Improved metagenome analysis using MEGAN5:DOI: http://ab.inf.uni-tuebingen.de/software/megan5/.

Hutchins, D. A. et al. (2007). CO_2 control of *Trichodesmium* N_2 fixation, photosynthesis, growth rates, and elemental ratios: implications for past, present, and future ocean biogeochemistry. *Limnol. Oceanogr.* 52(4):1293–1304. DOI: 10.4319/lo.2007.52.4.1293.

Hutchins, D. A. et al. (2015). Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. *Nat. Commun.* 6:8155. DOI: 10.1038/ncomms9155.

Hyatt, D. et al. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. DOI: 10.1186/1471-2105-11-119.

Hynes, A. M. et al. (2009). Cross-basin comparison of phosphorus stress and nitrogen fixation in *Trichodesmium*. *Limnol. Oceanogr.* 54(5):1438–1448. DOI: 10.4319/lo.2009.54.5.1438.

Hynes, A. M. et al. (2012). Comparison of cultured *Trichodesmium* (Cyanophyceae) with species characterized from the field. *J. Phycol.* 48(1):196–210. DOI: 10.1111/j.1529-8817.2011.01096.x.

Janausch, I. G. et al. (2002). C4-dicarboxylate carriers and sensors in bacteria. *Biochim. Biophys. Acta* 1553:39–56. DOI: 10.1016/S0005-2728(01)00233-X.

Karl, D. M. (1999). A sea of change: biogeochemical variability in the North Pacific subtropical gyre. *Ecosystems* 2(3):181–214.

Karl, D. M. (2007). Microbial oceanography: paradigms, processes and promise. *Nat. Rev. Microbiol.* 5(10):759–769. DOI: 10.1038/nrmicro1749.

Karl, D. M. (2014). Microbially mediated transformations of phosphorus in the sea: mew views of an old cycle. *Ann. Rev. Mar. Sci.* 6(1):279–337. DOI: 10.1146/annurev-marine-010213-135046.

Karl, D. M. et al. (2008). Aerobic production of methane in the sea. *Nat. Geosci.* 1(7):473–478. DOI: 10.1038/ngeo234.

Kearse, M. et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649. DOI: 10.1093/bioinformatics/bts199.

Krishnamurthy, A. et al. (2007). Effects of atmospheric inorganic nitrogen deposition on ocean biogeochemistry. *J. Geophys. Res.* 112(G2):G02019. DOI: 10.1029/2006JG000334.

Krupke, A. et al. (2016). Quorum sensing plays a complex role in regulating the enzyme hydrolysis activity of microbes associated with sinking particles in the ocean. *Front. Mar. Sci.* 3:55. DOI: 10.3389/fmars.2016.00055.

Langfelder, P. and S. Horvath (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. DOI: 10.1186/1471-2105-9-559.

Latifi, A. et al. (1996). A hierarchical quorum-sensing cascade in *Pseudomonas aeruginosa* links the transcriptional activators LasR and RhlR (VsmR) to expression of the stationary-phase sigma factor RpoS. *Mol. Microbiol.* 21(6):1137–1146. DOI: 10.1046/j.1365-2958.1996.00063.x.

Lee, M. D. et al. (2017). The *Trichodesmium* consortium: conserved heterotrophic co-occurrence and genomic signatures of potential interactions. *ISME J.* 11(8):1813–1824. DOI: 10.1038/ismej.2017.49.

Lee, M. D. et al. (2018). Transcriptional activities of the microbial consortium living with the marine nitrogen-fixing cyanobacterium *Trichodesmium* reveal potential roles in community-level nitrogen cycling. *Appl. Environ. Microbiol.* 84(1):e02026–17.

Lesser, M. P. et al. (2016). Climate change stressors destabilize the microbiome of the Caribbean barrel sponge, *Xestospongia muta*. *J. Exp. Mar. Bio. Ecol.* 475:11–18. DOI: 10.1016/j.jembe.2015.11.004.

Letunic, I. and P. Bork (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44(W1):W242–W245. DOI: 10.1093/nar/gkw290.

Li, B. and C. N. Dewey (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. DOI: 10.1186/1471-2105-12-323.

Li, H. et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079. DOI: 10.1093/bioinformatics/btp352.

Li, W. and A. Godzik (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659. DOI: 10.1093/bioinformatics/btl158.

Liu, W. T. et al. (2007). Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. *Microb. Pathog.* 42(1):2–10. DOI: 10.1016/j.micpath.2006.09.002.

Lomas, M. W. et al. (2010). Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences* 7(2):695–710. DOI: 10.5194/bg-7-695-2010.

Luo, H. et al. (2009). Subcellular localization of marine bacterial alkaline phosphatases. *Proc. Natl. Acad. Sci.* 106(50):21219–21223. DOI: 10.1073/pnas.0907586106.

Luo, Y.-W. et al. (2012). Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data* 4(1):47–73. DOI: 10.5194/essd-4-47-2012.

Mahaffey, C. et al. (2005). The conundrum of marine N_2 fixation. *Am. J. Sci.* 305:546–595. DOI: 10.2475/ajs.305.6-8.546.

Mahaffey, C. et al. (2014). Alkaline phosphatase activity in the subtropical ocean: insights from nutrient, dust and trace metal addition experiments. *Front. Mar. Sci.* 1:73. DOI: 10.3389/fmars.2014.00073.

McGillicuddy, Dennis J. (2014). Do *Trichodesmium* spp. populations in the North Atlantic export most of the nitrogen they fix? *Global Biogeochem. Cycles* 28(2):103–114. DOI: 10.1002/2013GB004652.

McGrath, J. W. et al. (2013). Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. *Nat. Rev. Microbiol.* 11(6):412–9. DOI: 10.1038/nrmicro3011.

Metcalf, W. W. and W. A. van der Donk (2009). Biosynthesis of phosphonic and phosphinic acid natural products. *Annu. Rev. Biochem.* 78:65–94. DOI: 10.1146/annurev.biochem.78.091707.100215.Biosynthesis.

Metcalf, W. W. and R. S. Wolfe (1998). Molecular genetic analysis of phosphite and hypophosphite oxidation by *Pseudomonas stutzeri* WM88. *J. Bacteriol.* 180(21):5547–5558.

Metcalf, W. W. et al. (2012). Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science* 337(6098):1104–1107. DOI: 10.1126/science.1219875.

Miller, M. B. and B. L. Bassler (2001). Quorum sensing in bacteria. *Annu. Rev. Microbiol.* 55:165–199. DOI: 10.1146/annurev.micro.55.1.165.

Mills, M. M. et al. (2004). Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* 429(6989):292–294. DOI: 10.1038/nature02550.

Momper, L. M. et al. (2014). A novel cohabitation between two diazotrophic cyanobacteria in the oligotrophic ocean. *ISME J.* (4):882–893. DOI: 10.1038/ismej.2014.186.

Moore, C. M. et al. (2008). Relative influence of nitrogen and phosphorus availability on phytoplankton physiology and productivity in the oligotrophic sub-tropical North Atlantic Ocean. *Limnol. Oceanogr.* 53(1):291–305. DOI: 10.4319/lo.2008.53.1.0291.

Moore, C. M. et al. (2009). Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nat. Geosci.* 2(12):867–871. DOI: 10.1038/ngeo667.

Moore, C. M. et al. (2013). Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* 6(9):701–710. DOI: 10.1038/ngeo1765.

Moran, M. et al. (2016). Deciphering ocean carbon in a changing world. *Proc. Natl. Acad. Sci.* 113(12):3143–3151. DOI: 10.1073/pnas.1514645113.

Morris, J. J. and L. R. Hmelo (2014). Organizing principles for marine microbial consortia. *Eco-DAS IX Symposium Proceedings* 1:1–15.

Moutin, T. et al. (2005). Phosphate availability controls *Trichodesmium* spp. biomass in the SW Pacific Ocean. *Mar. Ecol. Prog. Ser.* 297:15–21. DOI: 10.3354/meps297015.

Moutin, T. et al. (2008). Phosphate availability and the ultimate control of new nitrogen input by nitrogen fixation in the tropical Pacific Ocean. *Biogeosciences* 5(1):95–109. DOI: 10.5194/bg-5-95-2008.

Moutin, T. et al. (2017). Preface: The Oligotrophy to the UlTra-oligotrophy PACific Experiment (OUTPACE cruise, 18 February to 3 April 2015). *Biogeosciences* 14(13):3207–3220. DOI: 10.5194/bg-14-3207-2017.

Mulholland, M. R. et al. (2006). Nitrogen fixation and release of fixed nitrogen by *Trichodesmium* spp. in the Gulf of Mexico. *Limnol. Oceanogr.* 51(4):1762–1776. DOI: 10.4319/lo.2006.51.4.1762.

Mulholland, Margaret R et al. (2004). Dinitrogen fixation and release of ammonium and dissolved organic nitrogen by *Trichodesmium* IMS101. *Aquat. Microb. Ecol.* 37(1):85–94. DOI: 10.3354/ame037085.

Nachin, L. et al. (2005). Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J. Bacteriol.* 187(18):6265–6272. DOI: 10.1128/JB.187.18.6265.

Nausch, M. (1996). Microbial activities on *Trichodesmium* colonies. *Mar. Ecol. Prog. Ser.* 141:173–181. DOI: 10.3354/meps141173.

Nymark, M. et al. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. DOI: 10.1038/srep24951.

Oksanen, J. et al. (2015). The vegan package. http://vegan.r-forge.r-project.org/.

Orchard, E. D. et al. (2009). Molecular analysis of the phosphorus starvation response in *Trichodesmium* spp. *Environ. Microbiol.* 11(9):2400–2411. DOI: 10.1111/j.1462-2920.2009.01968.x.

Orchard, E. D. et al. (2010). Polyphosphate in *Trichodesmium* from the low-phosphorus Sargasso Sea. *Limnol. Oceanogr.* 55(5):2161–2169. DOI: 10.4319/lo.2010.55.5.2161.

Ottesen, E. A. et al. (2014). Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* 345(6193):207–212. DOI: 10.1126/science.1252476.

Overbeek, R. et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42(D1):206–214. DOI: 10.1093/nar/gkt1226.

Paerl, H. W. (1994). Spatial segregation of CO_2 fixation in *Trichodesmium* spp. - linkage to N_2 fixation potential. *J. Phycol.* 30(5):790–799.

Paerl, H. W. and B. M. Bebout (1988). Direct measurement of O_2-depleted microzones in marine *Oscillatoria*: Relation to N_2 fixation. *Science* 241(4864):442–445.

Paerl, H. W. et al. (1989). Bacterial associations with marine *Oscillatoria* sp. (*Trichodesmium* sp.) populations: ecophysiological implications. *J. Phycol.* 25(4):773–784. DOI: 10.1111/j.0022-3646.1989.00773.x.

Patankar, A. V. and J. E. Gonzalez (2009). Orphan LuxR regulators of quorum sensing. *FEMS Microbiol. Rev.* 33(4):739–756. DOI: 10.1111/j.1574-6976.2009.00163.x.

Peng, Y. et al. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428. DOI: 10.1093/bioinformatics/bts174.

Polyviou, D. et al. (2015). Phosphite utilisation by the globally important marine diazotroph *Trichodesmium*. *Environ. Microbiol. Rep.* 7(6):824–830. DOI: 10.1111/1758-2229.12308.

Polyviou, D. et al. (2018). Desert dust as a source of iron to the globally important diazotroph *Trichodesmium*. *Front. Microbiol.* 8:2683. DOI: 10.3389/fmicb.2017.02683.

Price, M. N. et al. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490. DOI: 10.1371/journal.pone.0009490.

Prufert-Bebout, L. et al. (1993). Growth, nitrogen fixation, and spectral attenuation in cultivated *Trichodesmium* species. *Appl. Environ. Microbiol.* 59(5):1367–1375.

Pulido-Villena, E. et al. (2010). Transient fertilizing effect of dust in P-deficient LNLC surface ocean. *Geophys. Res. Letters* 37:L01603. DOI: 10.1029/2009GL041415.

Rao, M. et al. (2015). Nitric oxide mediates biofilm formation and symbiosis in *Silicibacter*. *MBio* 6(3):e00206–15. DOI: 10.1128/mBio.00206-15.

Rastogi, A. et al. (2016). PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. *BMC Bioinformatics* 17(1):261. DOI: 10.1186/s12859-016-1143-1.

Raymond, J. et al. (2004). The natural history of nitrogen fixation. *Mol. Biol. Evol.* 21(3):541–554. DOI: 10.1093/molbev/msh047.

Repeta, D. J. et al. (2016). Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nat. Geosci.* 9:884–887. DOI: 10.1038/ngeo2837.

Riebesell, U. et al. (2009). Sensitivities of marine carbon fluxes to ocean change. *Proc. Natl. Acad. Sci.* 106(49):20602–20609. DOI: 10.1073/pnas.0813291106.

Rodriguez, F. et al. (2014). Crystal structure of the *Bacillus subtilis* phosphodiesterase PhoD reveals an iron and calcium-containing active site. *J. Biol. Chem.* 289(45):30889–30899. DOI: 10.1074/jbc.M114.604892.

Roe, K. L. et al. (2012). Acquisition of iron by *Trichodesmium* and associated bacteria in culture. *Environ. Microbiol.* 14(7):1681–95. DOI: 10.1111/j.1462-2920.2011.02653.x.

Rouco, M. et al. (2014). *Trichodesmium* sp. clade distributions in the western North Atlantic Ocean. *Limnol. Oceanogr.* 59(6):1899–1909. DOI: 10.4319/lo.2014.59.6.1899.

Rouco, M. et al. (2016a). Microbial diversity within the *Trichodesmium* holobiont. *Environ. Microbiol.* 18(12):5151–5160. DOI: 10.1111/1462-2920.13513.

Rouco, M. et al. (2016b). Variable depth distribution of *Trichodesmium* clades in the North Pacific Ocean. *Environ. Microbiol. Rep.* 8(6):1058–1066. DOI: 10.1111/1758-2229.12488.

Rouco, M. et al. (2018). Transcriptional patterns identify resource controls on the diazotroph *Trichodesmium* in the Atlantic and Pacific oceans. *ISME J.* 12(6):1486–1495. DOI: 10.1038/s41396-018-0087-z.

Rubin, M. et al. (2011). Dust- and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. *Nat. Geosci.* 4(8):529–534. DOI: 10.1038/ngeo1181.

Ryther, J. H. and W. M. Dunstan (1971). Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science* 171(3975):1008–1013.

Saito, M. A. et al. (2011). Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocosphaera watsonii*. *Proc. Natl. Acad. Sci.* 108(6):2184–2189. DOI: 10.1073/pnas.1006943108.

Salomon, D. et al. (2015). Type VI secretion system toxins horizontally shared between marine bacteria. *PLoS Pathog.* 11(8):e1005128. DOI: 10.1371/journal.ppat.1005128.

Sampson, T. R. and S. K. Mazmanian (2015). Control of brain development, function, and behavior by the microbiome. *Cell Host Microbe* 17(5):565–576. DOI: 10.1016/j.chom.2015.04.011.

Sañudo-Wilhelmy, S. A. et al. (2001). Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* 411(6833):66–69. DOI: 10.1038/35075041.

Sañudo-Wilhelmy, S. A. et al. (2014). The role of B vitamins in marine biogeochemistry. *Ann. Rev. Mar. Sci.* 6(1):339–367. DOI: 10.1146/annurev-marine-120710-100912.

Sargent, E. C. et al. (2016). Evidence for polyploidy in the globally important diazotroph *Trichodesmium*. *FEMS Microbiol. Lett.* 363(21):fnw244. DOI: 10.1093/femsle/fnw244.

Schaefer, A. L. et al. (2002). Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J. Bacteriol.* 184(23):6515–6521. DOI: 10.1128/JB.184.23.6515.

Schuster, M. and E. P. Greenberg (2006). A network of networks: quorum-sensing gene regulation in *Pseudomonas aeruginosa*. *Int. J. Med. Microbiol.* 296(2-3):73–81. DOI: 10.1016/j.ijmm.2006.01.036.

Sebastian, M. and J. W. Ammerman (2009). The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME J.* 3(5):563–572. DOI: 10.1038/ismej.2009.10.

Sheridan, C. C. et al. (2002). The microbial and metazoan community associated with colonies of *Trichodesmium* spp.: a quantitative survey. *J. Plankton Res.* 24:913–922. DOI: 10.1093/plankt/24.9.913.

Snow, J. T. et al. (2015a). Environmental controls on the biogeography of diazotrophy and *Trichodesmium* in the Atlantic Ocean. *Global Biogeochem. Cycles* 29:865–884. DOI: 10.1002/2015GB005090.

Snow, J. T. et al. (2015b). Quantifying integrated proteomic responses to iron Stress in the globally important marine diazotroph *Trichodesmium*. *PLoS One* 10(11):e0142626. DOI: 10.1371/journal.pone.0142626.

Sohm, J. A. and D. Capone (2006). Phosphorus dynamics of the tropical and subtropical north Atlantic: *Trichodesmium* spp. versus bulk plankton. *Mar. Ecol. Prog. Ser.* 317:21–28. DOI: 10.3354/meps317021.

Sohm, J. A. et al. (2011). Emerging patterns of marine nitrogen fixation. *Nat. Rev. Microbiol.* 9(7):499–508. DOI: 10.1038/nrmicro2594.

Somers, E. et al. (2005). *Azospirillum brasilense* produces the auxin-like phenylacetic acid by using the key enzyme for indole-3-acetic acid biosynthesis. *Appl. Environ. Microbiolgy* 71(4):1803–1810. DOI: 10.1128/AEM.71.4.1803.

Spaepen, S. and J. Vanderleyden (2011). Auxin and plant-microbe interactions. *Cold Spring Harb. Perspect. Biol.* 3(4):1–13. DOI: 10.1101/cshperspect.a001438.

Spungin, D. et al. (2016). Mechanisms of *Trichodesmium* bloom demise within the New Caledonia Lagoon during the VAHINE mesocosm experiment. *Biogeosciences* 13(14):4187–4203. DOI: 10.5194/bg-13-4187-2016.

Stenegren, M. et al. (2017). Distribution and drivers of symbiotic and free-living diazotrophic cyanobacteria in the Western Tropical South Pacific. *Biogeosciences Discuss.* 15(5):1559–1578. DOI: 10.5194/bg-2017-63.

Subramaniam, A. et al. (2008). Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. *Proc. Natl. Acad. Sci.* 105(30):10460–10465. DOI: 10.1073/pnas.0710279105.

Sunagawa, S. et al. (2015). Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348(6237):1261359. DOI: 10.1126/science.1261359.

Suzek, B. E. et al. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23(10):1282–1288. DOI: 10.1093/bioinformatics/btm098.

Taga, M. E. et al. (2007). BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* 446(7134):449–453. DOI: 10.1038/nature05611.

Tang, K. et al. (2012). Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. *PLoS One* 7(7):e41204. DOI: 10.1371/journal.pone.0041204.

Teeling, H. et al. (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* 336(6081):608–611. DOI: 10.1126/science.1218344.

Thaben, P. F. and P. O. Westermark (2014). Detecting rhythms in time series with rain. *J. Biol. Rhythms* 29(6):391–400. DOI: 10.1177/0748730414553029.

Thomson, N. R. et al. (2000). Biosynthesis of carbapenem antibiotic and prodigiosin pigment in *Serratia* is under quorum sensing control. *Mol. Microbiol.* 36(3):539–556. DOI: 10.1046/j.1365-2958.2000.01872.x.

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

Tzubari, Y. et al. (2018). Iron and phosphorus deprivation induce sociality in the marine bloom-forming cyanobacterium *Trichodesmium*. *ISME J.* 12:1682–1693. DOI: 10.1038/s41396-018-0073-5.

Valdés, V. et al. (2018). Nitrogen and phosphorus recycling mediated by copepods in Western Tropical South Pacific. *Biogeosciences Discuss.*1–28. DOI: 10.5194/bg-2017-563.

Van Mooy, B. A. S. et al. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* 457(7234):69–72. DOI: 10.1038/nature07659.

Van Mooy, B. A. S. et al. (2012). Quorum sensing control of phosphorus acquisition in *Trichodesmium* consortia. *ISME J.* 6(2):422–429. DOI: 10.1038/ismej.2011.115.

Van Mooy, B. A. S. et al. (2015). Major role of planktonic phosphate reduction in the marine phosphorus redox cycle. *Science* 348(6236):783–785. DOI: 10.1126/science.aaa8181.

Vannini, A. et al. (2002). The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J.* 21(17):4393–4401. DOI: 10.1093/emboj/cdf459.

Varaljay, V. A. et al. (2010). Deep sequencing of a dimethylsulfoniopropionate-degrading gene (*dmdA*) by using PCR primer pairs designed on the basis of marine metagenomic data. *Appl. Environ. Microbiol.* 76(2):609–617. DOI: 10.1128/AEM.01258-09.

Villarreal-Chiu, J. F. et al. (2012). The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. *Front. Microbiol.*19. DOI: 10.3389/fmicb.2012.00019.

Wagner, V. E. et al. (2003). Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. *J. Bacteriol.* 185(7):2080–2095. DOI: 10.1128/JB.185.7.2080.

Wagner, V. E. et al. (2004). Transcriptome analysis of quorum-sensing regulation and virulence factor expression in *Pseudomonas aeruginosa*. *Vaccine* 22:S15–S20. DOI: 10.1016/j.vaccine.2004.08.011.

Walworth, N. G. et al. (2015). *Trichodesmium* genome maintains abundant, widespread noncoding DNA in situ, despite oligotrophic lifestyle. *Proc. Natl. Acad. Sci.* 112(14):4251–4256. DOI: 10.1073/pnas.1422332112.

Walworth, N. G. et al. (2016). Mechanisms of increased *Trichodesmium* fitness under iron and phosphorus co-limitation in the present and future ocean. *Nat. Commun.* 7:12081. DOI: 10.1038/ncomms12081.

Walworth, N. G. et al. (2017a). Biogeographic conservation of the cytosine epigenome in the globally important marine, nitrogen-fixing cyanobacterium *Trichodesmium*. *Environ. Microbiol.* 19(11):4700–4713. DOI: 10.1111/1462-2920.13934.

Walworth, N. G. et al. (2017b). Nutrient co-limited *Trichodesmium* as nitrogen source or sink in a future ocean. *Appl. Environ. Microbiol.* 84(3):02137–17. DOI: 10.1128/AEM.02137-17.

Wannicke, N. et al. (2009). Release of fixed N_2 and C as dissolved compounds by *Trichodesmium erythreum* and *Nodularia spumigena* under the influence of high light and high nutrient (P). *Aquat. Microb. Ecol.* 57(2):175–189. DOI: 10.3354/ame01343.

Warren, M. J. et al. (2002). The biosynthesis of adenosylcobalamin (vitamin B12). *Nat. Prod. Rep.* 19(4):390–412. DOI: 10.1039/b108967f.

Waters, C. M. and Bonnie L Bassler (2005). Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* 21:319–346. DOI: 10.1146/annurev.cellbio.21.012704.131001.

Webb, E. A. et al. (2001). Iron stress in open-ocean cyanobacteria (*Synechococcus*, *Trichodesmium*, and *Crocosphaera* spp.): identification of the IdiA protein. *Appl. Environ. Microbiol.* 67(12):5444–5452. DOI: 10.1128/AEM.67.12.5444-5452.2001.

Webb, E. A. et al. (2007). Molecular assessment of phosphorus and iron physiology in *Trichodesmium* populations from the western Central and western South Atlantic. *Limnol. Oceanogr.* 52(5):2221–2232. DOI: 10.4319/lo.2007.52.5.2221.

Wei, J. et al. (2006). A mobile quorum-sensing system in *Serratia marcescens*. *J. Bacteriol.* 188(4):1518–1525. DOI: 10.1128/JB.188.4.1518.

Westberry, T. K. and D. A. Siegel (2006). Spatial and temporal distribution of *Trichodesmium* blooms in the world's oceans. *Global Biogeochem. Cycles* 20(4):GB4016. DOI: 10.1029/2005GB002673.

Whiteley, M. et al. (1999). Identification of genes controlled by quorum sensing in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* 96(24):13904–13909. DOI: 10.1073/pnas.96.24.13904.

Williams, P. and M. Cámara (2009). Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. *Curr. Opin. Microbiol.* 12(2):182–191. DOI: 10.1016/j.mib.2009.01.005.

Wilson, S. T. et al. (2017). Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocosphaera*. *Nat. Microbiol.* 2:17118. DOI: 10.1038/nmicrobiol.2017.118.

Wu, J. et al. (2000). Phosphate depletion in the western North Atlantic Ocean. *Science* 289(5480):759–762.

Wu, Y. et al. (2015). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32(4):605–607. DOI: 10.1093/bioinformatics/btv638.

Wu, Z. et al. (2010). Empirical bayes analysis of sequencing-based transcriptional profiling without replicates. *BMC Bioinformatics* 11(1):564. DOI: 10.1186/1471-2105-11-564.

Yong, S. C. et al. (2014). A complex iron-calcium cofactor catalyzing phosphotransfer chemistry. *Science* 345(6201):1170–1173. DOI: 10.1126/science.1254237.

Yu, X. et al. (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc. Natl. Acad. Sci.* 110(51):20759–20764. DOI: 10.1073/pnas.1315107110.

Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. 19(4):162–173. DOI: 10.1016/j.tim.2010.12.004.

Zehr, J. P. and B. B. Ward (2002). Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl. Environ. Microbiol.* 68(3):1015–1024. DOI: 10.1128/AEM.68.3.1015.

Appendix A

Chapter 2 supplemental material

A.1 Supplemental figures

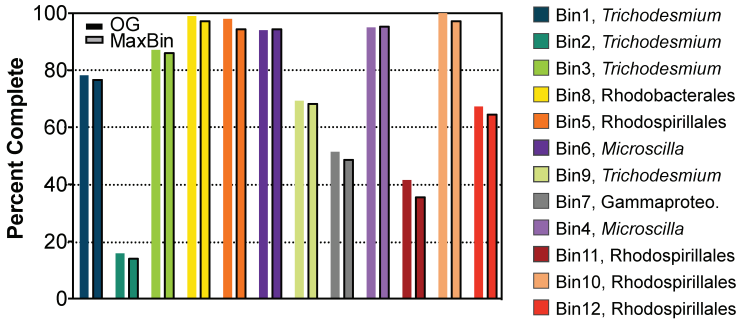

Figure A.1: Completeness of the genome bins. Percent completeness is estimated using abundance of single copy genes as previously described (Handley et al., 2012) and using the MaxBin program (Wu et al., 2015).

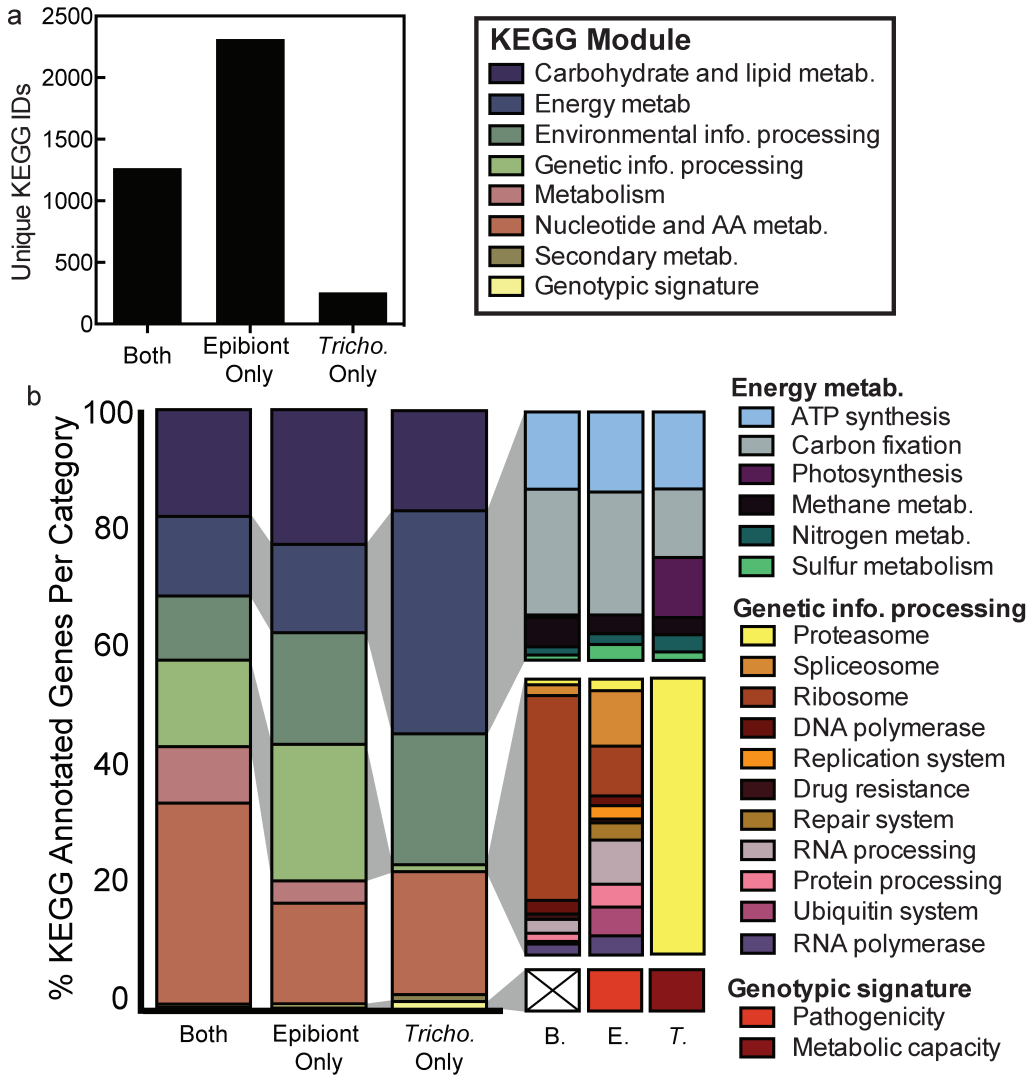

Figure A.2: Distribution of KEGG functional annotations of OGs predicted from the individual assemblies. (a) Functions (as determined by KEGG annotation) that are uniquely found in *Trichodesmium* only OGs (T), epibiont only OGs (E), or shared between *Trichodesmium* and epibiont (Both or B). (b) A finer scale breakdown of functionality in OG category shows the metabolic diversity contained within epibiont only OGs.

A.2 Supplemental tables

Table A.1: Sequencing reads, proteins from the metagenome assembly and environmental metadata from stations in the western North Atlantic.

Station	Lat.	Lon.	Reads	Total prots. binned per station	N_2 fixation rate (pmol colony^{-1} hr^{-1})	PO_4 turnover rate (hr^{-1})	Sal.	Temp. (°C)	Total dissolved P (µM)	*Tricho.* APA (nmol P col^{-1} hr^{-1})
2	27.9	-65	33,044,652	47,177	4.3	0.03	36.7	24	0.07	0.124
5	21.2	-64.9	45,080,408	81,064	8.06	0.07	36.4	26.4	0.09	0.146
9	16.5	-57.3	24,522,732	49,198	39.34	0.06	36.4	26.4	0.17	0.288
10	14	-55.7	9,067,799	17,975	14.64	0.05	34.8	27.1	0.16	0.087
16	9.9	-58.5	18,492,342	25,845	16.67	0.06	34.8	27.5	0.23	0.134
17	11.9	-59.4	26,590,472	42,773	20.43	0.09	35.8	27.4	0.22	0.2

Table A.2: Frequency of key phosphorus and iron genes in the *Trichodesmium* microbiome compared to free-living bacterioplankton communities sampled from the Sargasso Sea region of the western North Atlantic.
[a]Percentage of core epibiont community genome bins (out of 8 total bins) containing the gene family
[b]Percentage of free living community calculated relative to the single copy gene RecA
*Values estimated from data previously calculated (Tang et al., 2012)

Gene Family	% Core Microbiome[a]	% Free Living[b]	% Free Living Reference
Phosphonate CP lyase (*PhnJ*)	38	19	Karl et al., 2008
		26	Martinez et al., 2010
Siderophore/Vitamin transporters	100	~18*	Tang et al., 2012
Heme/Hemophores/Iron(heme) binding protein transporters	100	~2*	Tang et al., 2012

Table A.3: Glycoside hydrolase (GH) orthologous groups (OGs) detected in the metagenome assemblies. Substrate specificity (target) is noted when available.

OG	Bins	Category	GH Family	Target
OG_1244	1,3,9,12	Both	38	Mannose
OG_3154	3,4,6	Both	18	Chitin
OG_11528	4,6	Epibiont	NA	
OG_12010	6	Epibiont	30	Xylan
OG_12197	6	Epibiont	17	1,3;1,4-β-D-glucan
OG_16240	6	Epibiont	16	β-1,4 or β-1,3 bonds in glucans and galactans
OG_1918	4,6,12	Epibiont	31	Broad specificity
OG_20631	Other	Epibiont	NA	
OG_2728	4,6	Epibiont	3	Broad specificity (monosaccharides)
OG_28509	Other	Epibiont	NA	
OG_3117	5,8	Epibiont	4	Broad specificity (phosphorylated sugars)
OG_31908	Other	Epibiont	NA	
OG_3410	7,8,Other	Epibiont	NA	
OG_3693	5,6	Epibiont	43	Arabinose or xylose
OG_3765	4	Epibiont	NA	
OG_43634	Other	Epibiont	NA	
OG_52433	Other	Epibiont	18	Chitin
OG_53917	Other	Epibiont	10	Xylan/cellobiose
OG_6351	6	Epibiont	8	Lichenin, xylan, glucan, chitin
OG_7475	4,6	Epibiont	NA	
OG_7827	6	Epibiont	NA	
OG_8937	8,11	Epibiont	24	Unassigned
OG_9303	Other	Epibiont	88	Double bonded carbons in sugars
OG_10713	3	Tricho	57	Broad specificyt
OG_2844	1,3,9	Tricho	57	Broad specificity
OG_4914	1,2	Tricho	32	Involved in inverting sugars

Appendix B

Chapter 3 supplemental material

B.1 Supplemental figures

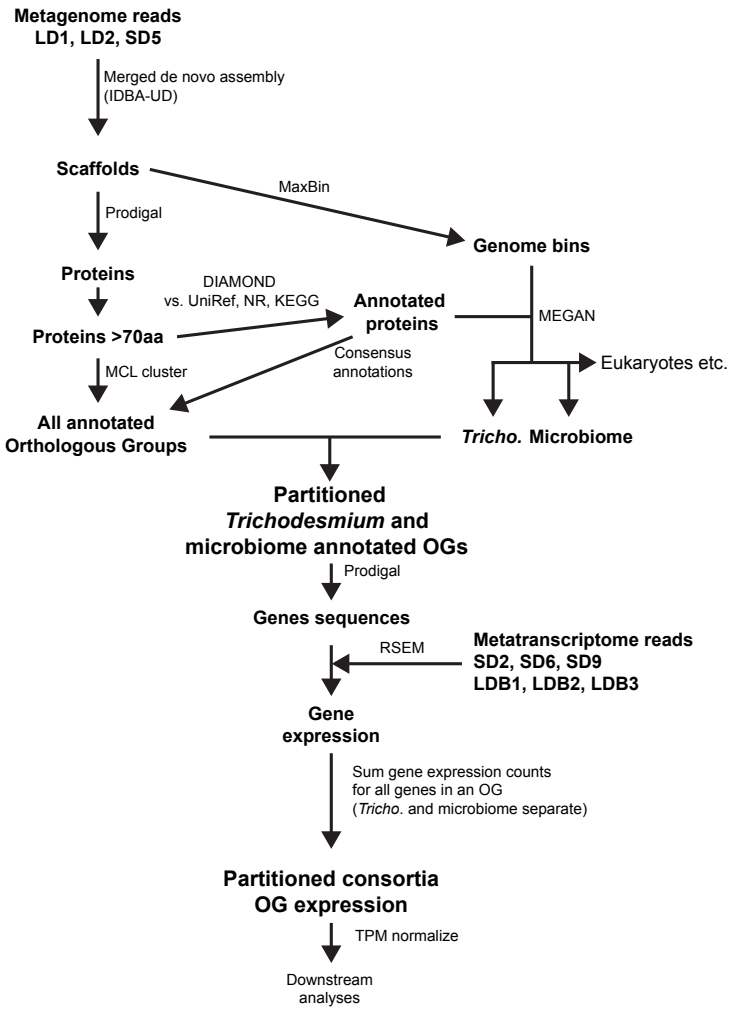

Figure B.1: Diagram detailing approach used to assemble and partition the metagenome between *Trichodesmium* and the microbiome, annotate and cluster protein coding sequences into orthologous groups (OGs), and obtain gene expression values from metatranscriptomes. Eukaryote identified sequences as well as phototrophs other than *Trichodesmium* were excluded from downstream analysis after the genome binning and analysis step. TPM, transcripts per million.

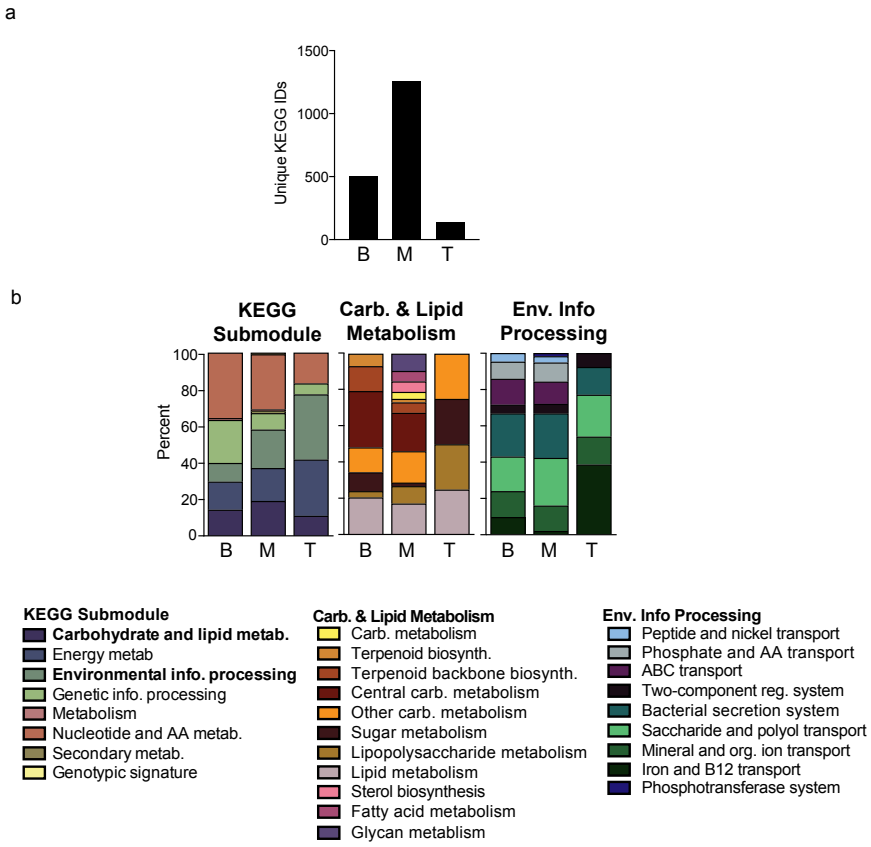

Figure B.2: Distribution and annotations of KEGG functional annotations of OGs found uniquely in the Trichodesmium (T) or microbiome (M) or those composed of both Trichodesmium and microbiome proteins (B). (a) Total number of OGs in each category. (b) A functional breakdown of these annotations at the KEGG module level and detailed annotations from within two KEGG module categories.

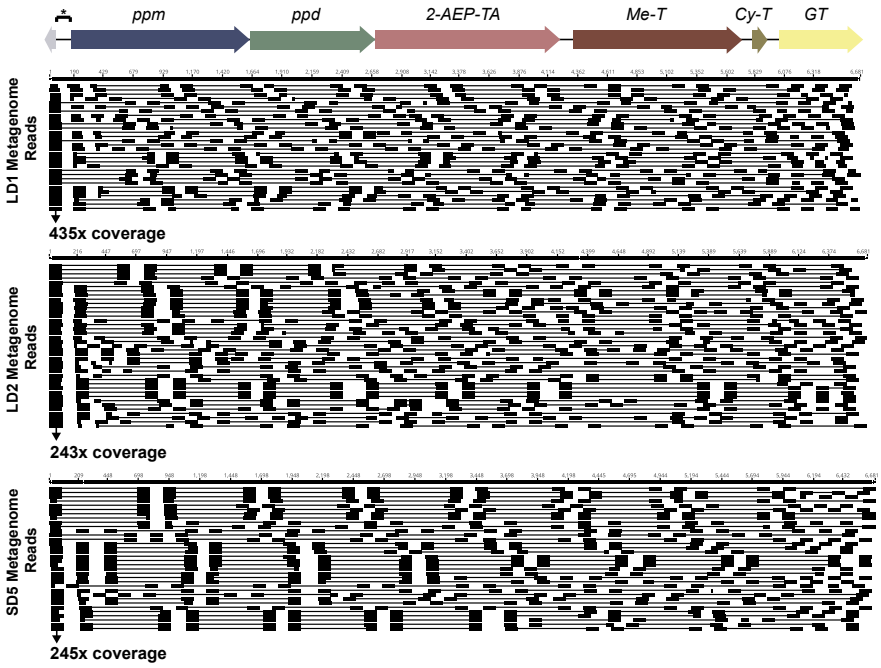

Figure B.3: Metagenome reads aligned to the scaffold containing the *ppm* cassette. Thick black rectangles depict 100 bp reads, connected by thin black lines to their mate paired read.

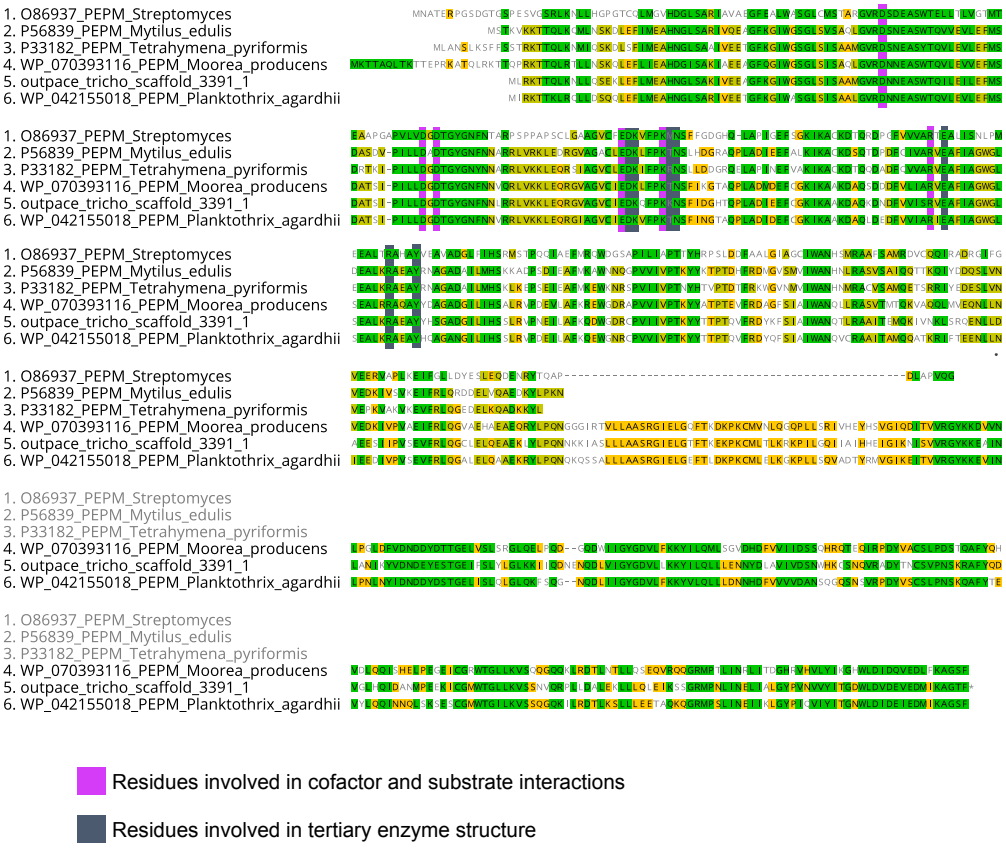

Figure B.4: Amino acid alignment of the Ppm protein recovered from a *Trichodesmium* metagenome bin against experimentally verified Ppm sequences in other organisms. Genes highlighted in green denote regions with 100% amino acid identity across all sequences. Shades of yellow denote conservation across the majority of amino acids in the column. Un-highlighted amino acids indicate divergent residues or regions with little conservation. Purple and grey highlighted columns denote residues that were previously determined to be important to the structure or activity of this enzyme (Chen et al., 2006).

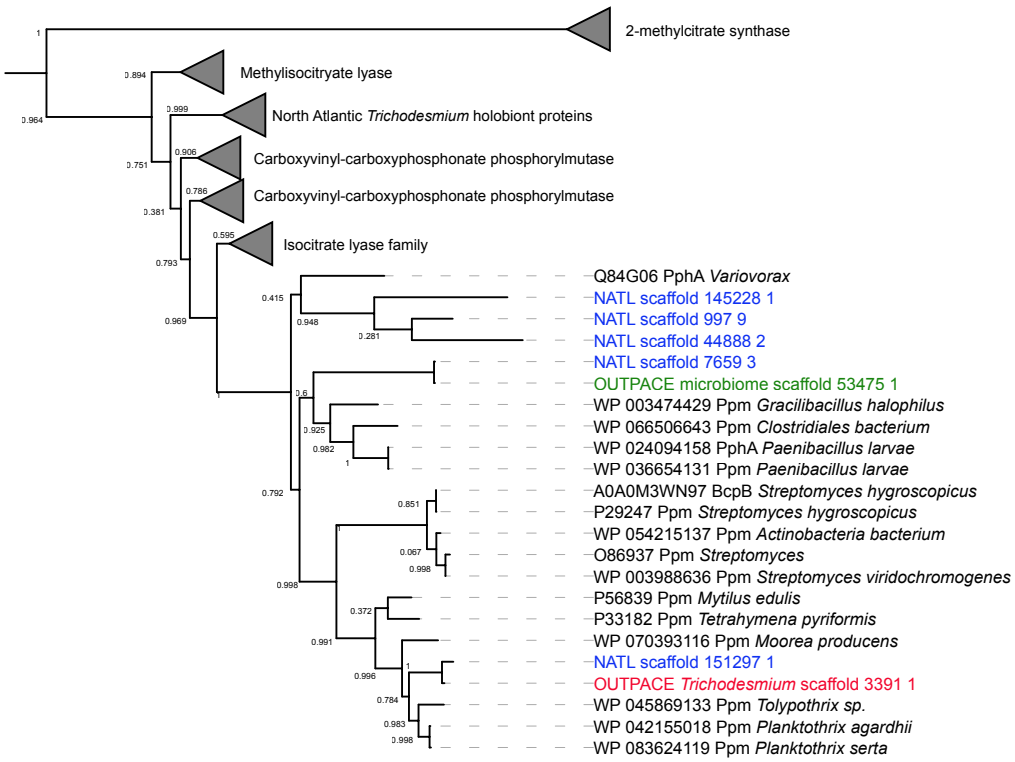

Figure B.5: Phylogenetic tree showing the placement of the *Trichodesmium* Ppm protein (red), a microbiome proteins from this study's metagenome assembly that is similar to Ppm but lacking conservation at key residues (green), as well as homologous proteins from a previously assembled North Atlantic *Trichodesmium* metagenome assembly (blue) (Frischkorn et al., 2017), along with homologous sequences obtained from the NCBI nr database (black). The tree was generated with FastTree using the default settings (Price et al., 2010). Numbers at the branch labels indicate FastTree support percentages for the sequences in that branch.

B.2 Supplemental tables

Table B.1: Summary of water column (10 m) geochemical data collected across the OUT-PACE cruise transect. *Measurement obtained at 30 m.

Station	Lat	Lon	Water colum dissolved inorganic phosphorus (nmol/L) at 10 m		Water column phosphate turnover (1/hr)		Phosphate turnover time (hrs)	Water column phosphate uptake rate (nmol/L/hr)	Water column dissolved iron (nmol/L) at 10 m
			avg.	stdev	avg	stdev			
SD1	-18.0	159.9	11.2	0.5376					0.67
SD2	-18.6	162.1	16.4	0.7872					0.36
SD3	-19.5	165.0	8.3	0.3984	0.0251	0.0013	39.7986	0.209	0.38
LDA	-19.2	164.6	13.8*	0.6624	0.0494	0.0031	20.2525	0.681	0.85
SD4	-20.0	168.0	4.8	0.2304					0.69
SD5	-22.0	170.0	2.3	0.1104	0.0026	0.0004	383.4898	0.006	0.44
SD6	-21.4	172.1	13.4	0.6432					
SD7	-20.8	174.3	10.21	0.49008	0.0029	0.0000	340.9558	0.030	0.21
SD8	-20.7	176.4	8.2	0.3936	0.0012	0.0009	827.7212	0.010	0.38
SD9	-21.0	178.6	6.1	0.2928	0.0389	0.0004	25.7207	0.237	0.22
SD10	-20.5	-178.5	6.6	0.3168					0.97
SD11	-20.0	-175.7	6.6	0.3168	0.0366	0.0002	27.3563	0.241	1.16
SD12	-19.5	-172.8	18.7	0.8976	0.0019	0.0008	519.0585	0.036	0.94*
LDB	-18.2	-170.7	7	0.336	0.0694	0.0044	14.4041	0.486	0.65
SD13	-18.2	-169.1							
LDC	-18.4	-165.9			0.5041	0.2785	1.9839		0.37
SD14	-18.4	-163.0	214	10.272					0.46
SD15	-18.3	-160.0	230	11.04					0.31

Appendix C

Chapter 4 supplemental material

C.1 Supplemental figures

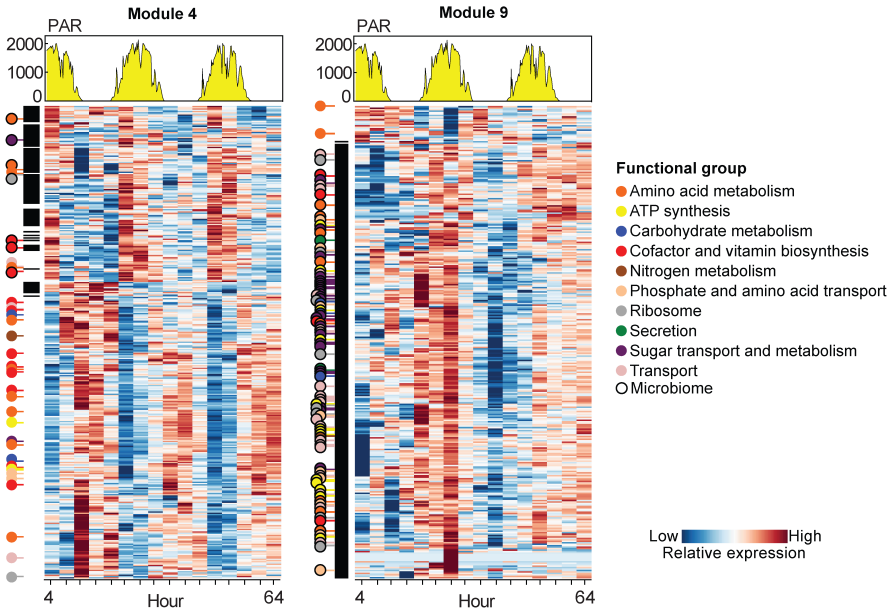

Figure C.1: Expression patterns in a module containing microbiome cobalamin OGs, and a module dominated by microbiome OGs. Photosynthetically active radiation (PAR) levels are shown over the duration of the Lagrangian sampling period and correspond to vertical bars in the heatmap, which depicts row averaged relative expression of the significantly co-expressed *Trichodesmium* and microbiome orthologous groups (OGs) within two modules that exhibited strong diel periodicity and a number of microbiome OGs related to cobalamin (Modules 4) and a module that was the most dominated by microbiome OGs (Module 9). Black bars denote rows that are microbiome OGs. Colored circles indicate functional annotations of *Trichodesmium* or microbiome OGs based on KEGG submodules.

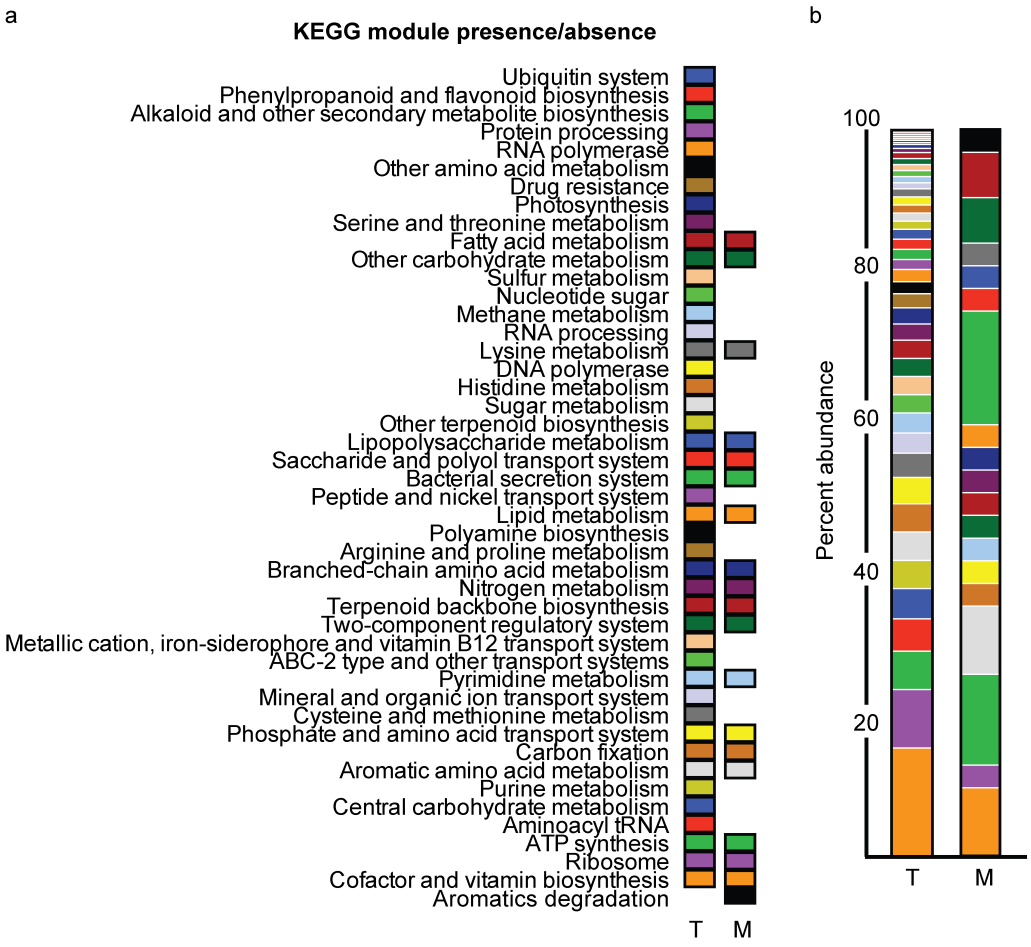

Figure C.2: KEGG profile of significantly periodic orthologous groups (OGs) in *Trichodesmium* and the microbiome. (a) Presence/absence of KEGG modules in significantly periodic OGs in *Trichodesmium* and the microbiome. (b) Relative abundance of significantly periodic OGs for each detected KEGG module in *Trichodesmium* and microbiome data. T, *Trichodesmium*; M, microbiome. Significant periodicity was determined using RAIN (Thaben and Westermark, 2014).

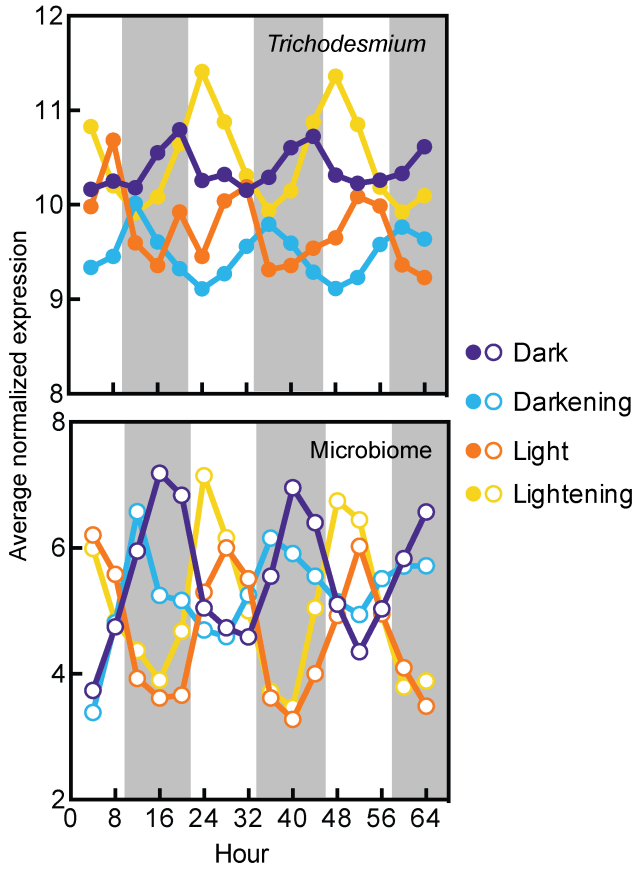

Figure C.3: Day-night trends in significantly periodic holobiont orthologous groups (OGs). Expression patterns of significantly periodic *Trichodesmium* and microbiome OGs averaged across those with peak expression times during four discrete day-night periods: peak dark and peak light periods, and when conditions were transitioning from light to dark (darkening) and dark to light (lightening). Significant periodicity was determined using RAIN (Thaben and Westermark, 2014).

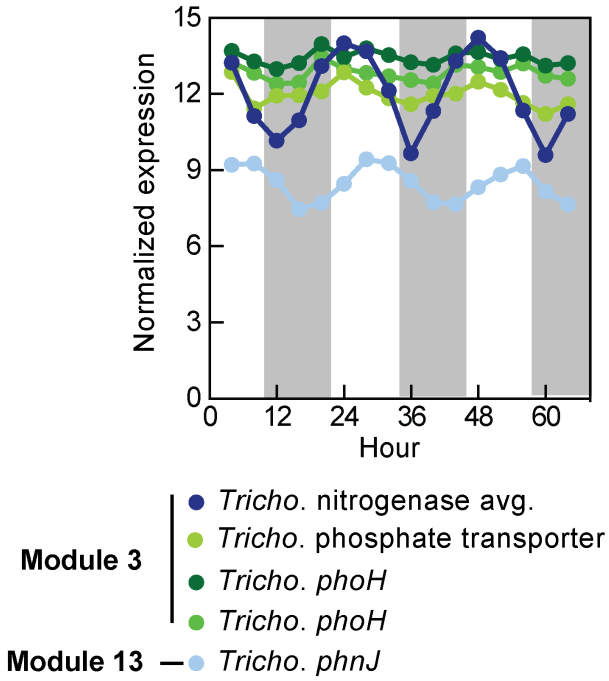

Figure C.4: Expression patterns of N_2 fixation and key phosphorus-related genes in *Trichodesmium*. Expression profiles of *Trichodesmium* phosphorus-related orthologous groups (OGs) from Module 3 as well as the OG for the phosphonate C-P lyase *phnJ* compared to the average expression of *nif* OGs from Module 3.

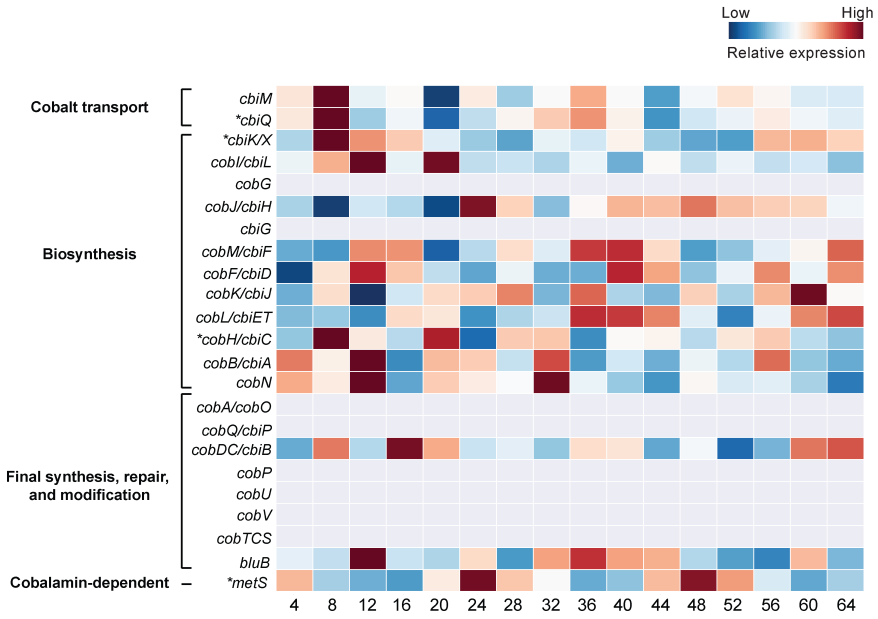

Figure C.5: Expression patterns of all cobalamin-related orthologous groups (OGs) detected in *Trichodesmium*. Heatmap showing presence/absence and row-averaged relative expression of all OGs related to cobalamin biosynthesis and utilization in *Trichodesmium*. Grey rows indicate OGs that were not detected in *Trichodesmium* field data. *Multiple OGs were detected with similar annotations and expression was averaged to yield expression values for that row.

Appendix D

Chapter 5 supplemental material

D.1 Supplemental figures

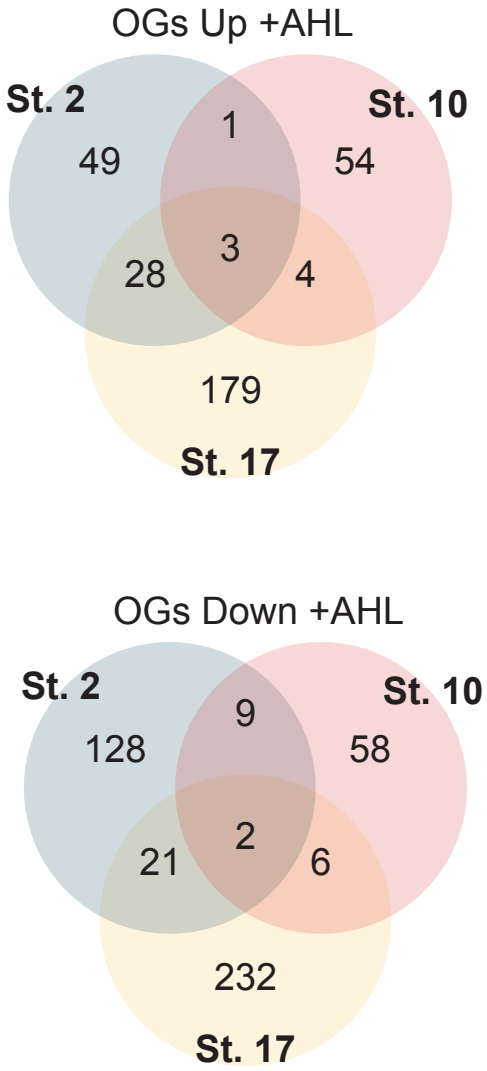

Figure D.1: Venn diagrams showing minimal overlap in microbiome OGs that were significantly increased (top panel) or decreased (bottom panel) in the +AHL treatment relative to the control across the three stations tested.

D.2 Supplemental tables

Table D.1: N_2 fixation rate and total dissolved phosphorus concentration at each station sampled across the transect.

Station	Latitude	Longitude	*In situ* N_2 Fixation Rate (pmol colony^{-1}hr^{-1})	*In situ* N_2 Fixation Rate Standard deviation	+AHL N_2 Fixation Rate (pmol colony^{-1}hr^{-1})	+AHL N_2 Fixation Rate Standard deviation	TDP (µM)
2	27.87	64.99	9.5	0.3	5.6	1.1	0.07
4	23.03	65.01	16.4	6.9			0.08
6	20.01	63.1	26.8	5.4			0.08
8	18.4	58.56	49.1	4			0.1
10	13.97	55.68	32.2	0.31	28.3	1.5	0.16
12	10.09	53.24	32.6	0.8			0.21
17	11.86	59.44	45	1.1	54	0.82	0.22