UNIVERSITY OF CALIFORNIA, SAN DIEGO

Indicators of Iron Metabolism in Marine Microbial Genomes and Ecosystems

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Oceanography

by

Shane Lahman Hogle

Committee in charge:

Katherine Barbeau, Chair
Eric Allen
Bianca Brahamsha
Christopher Dupont
Brian Palenik
Kit Pogliano

2016

.

The Dissertation of Shane Lahman Hogle is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2016

DEDICATION

Mom, Dad, Joel, and Marie

thank you for everything

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, my greatest thanks to my dissertation advisor, Kathy Barbeau, for providing me with the freedom to explore my ideas and interests throughout my time as a doctoral student. I am most thankful for her scientific insight and guidance, but I also deeply appreciate the flexibility and patience she demonstrated as my scientific interests evolved over the past six years. She has always ensured appropriate funding for my research, provided excellent scientific advice and expertise, and has always welcomed discussion on any variety of subjects. She has been an exceptional mentor to me for which I am grateful.

I am also thankful for the guidance and assistance of my dissertation committee members. Thank you to Bianca Brahamsha for helping turn a young chemist who knew nothing about microbiology or genetics into a now (hopefully) molecular biologist, bacteriologist, and ecologist. I very much appreciate your patience and willingness to share your expertise and your lab space. Thank you to Chris Dupont who really went above and beyond the call of duty as a committee member. Thank you for introducing me to the world of bioinformatics, for always making time to chat about my research, for reviewing and commenting on my manuscripts, for getting me a keycard to the swank J. Craig Venter Institute, and for being an all-around good guy. Thank you to Eric Allen for sharing your lab space, sharing your expertise in sequencing, and for letting me haul your spec20 down to Sverdrup Hall for late night growth curves. Chapter four would not have been possible without your help. Thank you to Brian Palenik for sharing your lab space and equipment. Thank you for the insights you have provided and for teaching me about phytoplankton diversity. Finally, I am honored to have had Kit Pogliano on my

committee. Thank you for sharing your unique insights into microbiology and for being generous with your time.

Thank you to my collaborators and past/present members of the Barbeau and Brahamsha/Palenik labs. Thank you to Cameron Thrash at Louisiana State University for helping with chapter two, for sharing your knowledge about SAR11, and for your kindness and help during my postdoc search. Thank you to Martha Gledhill at GEOMAR Helmholtz Center for Ocean Research, Kiel for helping with heme measurements, HPLC troubleshooting, and for collaborating on chapter one. The field of chemical oceanography is fortunate to have such a talented trace metal chemist. A huge thank you to Javi Paz-Yepes who really helped to initially familiarize me with molecular biology and genetics techniques when I was starting out. Thank you to Emy Daniels and Maitreyi Nagarkar for help with molecular techniques and for showing me the general organization of the Brahamsha/Palenik lab. Thank you to Randie Bundy for being a great lab friend and excellent scientist - for a while we were the only two students! Also thank you to Kelly Roe who taught me how to do cell culturing and other general lab techniques. Thanks for fielding endless questions about the locations of reagents in Sverdrup hall 1237. Thanks to Angel for always keeping an even keel and for being a good friend. You are a champion electrochemist and now a proper boat captain! Thanks to Lauren and Kiefer - you both are so talented and smart, and I'm excited to see where you guys will go with your theses. Lauren, thank you so much for banging your head against the HPLC with me and for your help with the horseradish peroxidase assay. I am thankful I had you around for troubleshooting, and I hope the Barbeau HPLC forever remains on your good side. Finally, I want to thank Leila Whitener and Aarani Arulmoli

for being wonderful undergraduate mentees and for continuously teaching me how to explain things more clearly. Aarani, I am so happy you decided to pursue a career in science, and Leila, I am very inspired to see the good you are doing through the US Peace Corp.

The SIO community as a whole is such a great group of people and it's not possible to thank everyone who deserves it here. Thank you to the graduate office for their support and to the GRD business office for help with administrative odds and ends. Thanks to the people in Sverdrup Hall and Hubbs Hall for sharing missing reagents and for technical advice - particularly Raffie Burke for resupplying me with phytoplankton when I needed them and Jess Blanton for being a DNA sequencing guru. Thank you Alain de Verneil, James Connors, Christian Briseño, James Fischer, Jess Blanton, Angel Ruacho, Stephanie Snyder, Amanda Netburn and many others for your friendship.

Finally, thank you to my mother, father and brother for your support. Thank you for being interested in my science, for asking me questions, for flying across the country to visit me, and for your unwavering love. Also thanks to my grandparents and extended family for the support you have always given me. Thank you to Marie for your patience, kindness, and generosity. I honestly do not know how I would have gotten through this PhD without you... Jag älskar dig.

Chapter 2, in full, is a reprint of the material as it appears in Hogle SL, Barbeau KA, Gledhill MA. (2014). Heme in the marine environment: from cells to the iron cycle. *Metallomics* 6:1107-1120. doi: 10.1039/C4MT00031E. The dissertation author was the primary investigator and first author of this paper.

Chapter 3 is a reprint of the material as it appears in Hogle SL, Thrash JC, Dupont CL, Barbeau KA. (2016). Trace metal acquisition by heterotrophic bacterioplankton with contrasting trophic strategies. *Appl. Environ. Microbiol.* 82(5): 1613-1624. The dissertation author was the primary investigator and first author of this paper.

Chapter 4 is currently being prepared for publication by Shane L. Hogle, Bianca Brahamsha, and Katherine A. Barbeau. The dissertation author was the primary investigator and first author of this paper.

Chapter 5 is currently being prepared for publication by Shane L. Hogle, Randelle A. Bundy, Jessica M. Blanton, Eric E. Allen, and Katherine A. Barbeau. The dissertation author was the primary investigator and first author of this paper.

Chapter 6 is currently being prepared for publication by Shane L. Hogle, Chris L. Dupont, and Katherine A. Barbeau. The dissertation author was the primary investigator and first author of this paper.

VITA

2010	B.A. in Chemistry, Highest Honors, Earlham College
2011 – 2014	National Science Foundation Graduate Research Fellow, University of California, San Diego
2014 - 2015	Graduate Teaching Assistant, University of California, San Diego
2016	Ph.D. in Oceanography, University of California, San Diego

PUBLICATIONS

1. Hogle SL, Thrash JC, Dupont CL, Barbeau KA. (2016). Trace metal acquisition by heterotrophic bacterioplankton with contrasting trophic strategies. *Appl. Environ. Microbiol.* 82(5): 1613-1624

2. Dupont CL, McCrow JP, Valas R, Moustafa A, Walworth N, Goodenough U, Roth R, Hogle SL, Bai J, Johnson ZI, Mann E, Palenik B, Barbeau KA, Venter JC, Allen AA. (2015). Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J.* 1(17):1076-1092. doi:10.1038/ismej.2014.198

3. Hogle SL, Barbeau KA, Gledhill MA. (2014). Heme in the marine environment: from cells to the iron cycle. *Metallomics* 6:1107-1120. doi: 10.1039/C4MT00031E

4. Kharbush JJ, Ugalde JA, Hogle SL, Allen EE, Aluwihare LI. (2013). Composite Bacterial Hopanoids and Their Microbial Producers across Oxygen Gradients in the Water Column of the California Current. *Appl. Environ. Microbiol.* 79(23): 7491-7501. doi: 10.1128/AEM.02367-13

5. Roe KL, Hogle SL, Barbeau KA. (2013). Utilization of heme as an iron source by marine Alphaproteobacteria in the Roseobacter clade. *Appl. Environ. Microbiol.* 79(18) 5753-5762. doi: 10.1128/AEM.01562-13

6. Abbriano RM, Carranza MM, Hogle SL, Levin RA, Netburn AN, Seto KL, Snyder SM, SIO280, Franks PJS. (2011). Deepwater Horizon Oil Spill: A Review

of the Planktonic Response. *Oceanography*. 24(3):294-301. doi: 10.5670/oceanog.2011.80

ABSTRACT OF THE DISSERTATION

Indicators of Iron Metabolism in Marine Microbial Genomes and Ecosystems

by

Shane Lahman Hogle

Doctor of Philosophy in Oceanography

University of California, San Diego, 2016

Professor Katherine Barbeau, Chair

A detailed picture of the chemical speciation of iron in seawater is required to comprehensively understand the marine iron cycle and its downstream effects on the cycles of other nutrient elements. Although much progress has been made in our

understanding of iron-binding ligands in the marine environment, numerous questions remain as to their identity, structure, and biogeochemical sources and sinks. The recent proliferation of "omic" data from cultured organisms and environmental samples now allows us to address some of these chemical questions from a biologically informed perspective.

The research presented in this thesis examines some of the specific molecular mechanisms and underlying genetics that marine heterotrophic bacteria use to acquire iron from the unique chemical conditions of the marine environment. We first examined and reviewed the sources, transformations, and ultimate fate of heme and hemoproteins, iron containing molecules, in seawater. We used comparative genomics to characterize trace metal acquisition strategies in the marine *Roseobacter* and SAR11 clades, and we examined these uptake pathways in light of marine trace metal chemistry, microbial ecology, and genome evolution. We developed a genetic system for targeted mutagenesis and insertional inactivation of a predicted heme uptake TonB dependent outer membrane transporter in a model marine bacterial strain *Ruegeria* sp. TM1040. In a series of iron amended incubations from the California current ecosystem, we examined the heterotrophic bacterial assemblage composition using high throughput sequencing of bacterial marker genes in conjunction with chemical speciation and algal physiology. Finally, we examined putative iron uptake pathways in a metatranscriptome from natural marine phyto- and bacterioplankton communities in the southern California current and integrated patterns in the metatranscriptomic data with biogeochemical field data.

Our results have revealed genes responsible for trace metal metabolisms in marine heterotrophic bacteria and highlight heme uptake as a model heterotrophic iron

assimilation pathway that is potentially important in particulate iron remineralization. In turn, we have provided additional insight to the chemical speciation of iron in seawater and have revealed potential molecular mechanisms by which heterotrophic bacteria shape the marine iron cycle.

Chapter 1

Introduction

Fe is a critical nutrient for all forms of life. It is employed in numerous enzymes that organisms use to perform critical metabolic functions. The large Fe requirements of the enzymes that facilitate these fundamental biogeochemical processes directly link Fe to the carbon and nitrogen cycles of the marine environment. Somewhat paradoxically, dissolved Fe is often vanishingly scarce in seawater due to its unique chemistry. Marine photosynthetic microorganisms have evolved to cope with these low Fe conditions (1) and still generate nearly half of earth's total primary production (2). Fe can be a limiting nutrient for marine phytoplankton in up to 33% of the surface ocean (3), and mesoscale Fe fertilization experiments have unequivocally demonstrated that Fe influences initiation and development of phytoplankton blooms and thus has downstream effects on the cycling of carbon, nitrogen, and silicon as well as effects on trophic dynamics (4). Thus, the availability and abundance of Fe in marine systems are important parameters influencing the global cycling of major nutrient elements.

Fe is extremely scarce in seawater relative to its biological demand due to its thermodynamic propensity to hydrolyze into oxyhydroxide species with very low solubility. These oxyhydroxide species have further tendency to adsorb to sinking particulate matter or other Fe hydroxide species resulting in significant Fe export out of the euphotic zone where, paradoxically, biological demand is highest. The remaining dissolved Fe in the surface ocean is almost exclusively (>99%) bound to largely unknown organic ligands. Traditionally, these ligands have been operationally partitioned into two

classes, L_1 and L_2, based on their conditional stability constants – a measure of the Fe binding affinity for each class (5–8). L_1 ligands, with the highest binding constants, have a greater affinity for Fe than ligands from the L_2 class. These Fe-binding ligands in seawater are hypothesized to be of biological origin, and have been shown to rapidly accumulate after inputs of Fe from mesoscale fertilization experiments (8, 9). Very little is known of the structural diversity of the Fe binding ligand pool, but some constituents have been identified using mass spectrometry techniques (10, 11). To date many of the Fe binding ligands identified from seawater or cultured marine organisms have been siderophores, low molecular weight and high-affinity Fe(III) chelating agents secreted by some bacteria explicitly for the purpose of chelating Fe (12).

Siderophores have been proposed to comprise a major component of the L_1 ligand pool due to the similarity of the binding constants of purified siderophores and natural marine ligands (6, 13). This has led to hypotheses that marine L_1 ligands are biologically derived (9) and may actually be siderophores produced *in situ* by microbial taxa (6). However, for unicellular free living marine organisms in oligotrophic waters, diffusive loss would make siderophore production a risky and metabolically expensive strategy (14). It may be that bacterial siderophore production is confined to particles where diffusion is attenuated and labile nitrogen and carbon are abundant relative to that in bulk seawater (14). More elusive is the weaker L_2 class, which is likely much more diverse with respect to its chemical structure and sources to the water column. Dissolved organic matter is a complex mixture of new and degraded biogenic molecules, some of which have an Fe chelating capacity. It is likely that a proportion of the L_2 class is derived from the decomposition of complex particulate and dissolved organic matter, of which the

altered products have Fe binding capabilities (15). Thus, it is probable that a continuum of Fe-binding molecules exists in seawater. Organisms intentionally produce some of the chelators to extract Fe from their surrounding medium, while other secreted molecules have incidental Fe chelating functional groups. Weaker chelators may be fragments of particulate matter picked over by heterotrophic bacteria (15), polysaccharides (16, 17), humic substances (18), or the liberated prosthetic groups from the metalloproteins of lysed phytoplankton cells. In the generally Fe-deficient conditions of the ocean, it would be advantageous for organisms to employ systems that access and utilize compounds from across the Fe-ligand continuum as well as unbound Fe(II) and Fe(III).

The existence of Fe uptake pathways and homeostasis mechanisms in marine bacterial genomes and metagenomes is bioinformatically inferred from functional information derived from well-studied model bacteria. Relatively little is known about marine microbial Fe transport compared to that for terrestrial and pathogenic microbes, yet the generalized models established in well-characterized pathogens are important guides for unveiling the phenomenon in marine microbes (19). In gram negative bacteria, Fe moves through the outer membrane in three basic forms - inorganic ferric Fe, inorganic ferrous Fe, and ferric Fe bound to organic ligands. Some of these forms are probably able to passively diffuse through porins in the outer membrane while others require the energy dependent transport of tonB dependent outer membrane transporters (TBDTs). Once in the periplasmic space, Fe is transported into the cytoplasm by ferrous Fe transporters or various forms of the ATP binding cassette transport system (ABCT).

Molecules greater than approximately 600 daltons or those that exist in very low concentrations cannot cross the bacterial outer membrane simply by diffusion and require

energized transport from TBDTs to reach the periplasm. Structurally, TBDTs contain a C-terminal membrane spanning β barrel domain and a N-terminal domain that fills the end of the beta barrel forming a plug. The plug domain contains two loops that extend towards the extracellular side of the barrel and form a recognition site for the binding of different substrates. Once a substrate is recognized, conformational changes in the TBDT at the surface of the outer membrane are transferred to the periplasmic facing side, which signals a TonB/ExbB/ExbD complex to somehow transduce energy to the membrane receptor (20). Until recently, tonB dependent transport was shown to only occur with small Fe complexes, such as siderophores, heme, or transferrin as well as the cobalt-containing Vitamin B12. However, recent research has demonstrated TBDT transport of nickel containing compounds, carbohydrates, and organic-phosphorous esters (21–26). Thus, TBDTs may be a far more generalist uptake strategy than was originally thought. TBDT are abundant in sequenced marine heterotrophic bacteria (26, 27) particularly the generalist *Roseobacter* lineage (28) but more modestly distributed in marine metagenomes (27, 29), suggesting that in the environment this capability may be confined to specific niches.

ABCTs are transmembrane, ATP dependent proteins that translocate diverse substrates from the periplasmic space to the cytoplasm by means of substrate-binding, permease, and ATP binding components. Bacteria have been generally assumed to translocate only unbound inorganic Fe(III) through the cellular inner membrane using ABCTs. However, detailed biochemical studies of the system have been limited to only a few organisms, mechanistic details are often conflicting, and microbial ABCTs have been shown to also interact with metal-chelates and ferrous Fe (30–33). TBDTs for

siderophores and heme complexes are often co-localized with ABCTs in bacterial genomes, implying a common transcriptional control. A recent survey found ABCTs to be the most common Fe acquisition system in marine bacterial genomes and metagenomes suggesting it is the most common strategy for Fe acquisition in the marine environment (27). Still, the limited functional characterization of a diversity of ABCTs coupled with the fact that most Fe(III) forms ligand complexes in seawater make the nature of the Fe(III) substrate (free or bound) unclear. In marine bacteria it is likely that both free and complexed ferric Fe are imported to the cytoplasm by ABCT systems.

Fe(II) is diffusible through the outer membrane of gram negative bacteria but requires energy dependent transport across the inner membrane. Three major systems that transport Fe(II) into the cytoplasm have been discovered in model bacteria, which include ABCT-like permeases, natural resistance-associated macrophage proteins (NRAMP), and FeoABC. It appears that the FeoABC system is the dominant pathway for Fe(II) assimilation in most studied pathogens but very little is known of its mechanism of action (34). In the context of the marine environment, Fe(II) is thermodynamically unstable in oxygenated seawater and quickly oxidizes to the ferric oxidation state. However, the assimilation of Fe(II) is likely critical in two step reductive uptake pathways of Fe(III) species such as those employed by many marine phytoplankton (35).

Prior studies have highlighted the importance of meso- and microzooplankton grazers and heterotrophic bacteria in recycling particulate Fe in surface waters (15, 36–38). This pool consists of particulates and colloids in both mineral and organic/detrital forms, and upon remineralization could represent a significant source of Fe to phytoplankton in chronically Fe depleted regions. Very little is known of the molecular

mechanisms through which this recycling occurs. Bacterial enzymatic attack on sinking biogenic particles in the mixed layer could liberate smaller and less complex Fe units such as heme prosthetic groups or Fe-sulfur clusters (39, 40). As heme is a fundamental building block in diverse enzymes across all domains of live, it could be energetically advantageous for microbes to have dedicated systems for assimilating heme directly instead of dismantling it and resynthesizing the porphyrin ring. From this perspective, intact heme molecules could potentially be a component of the bioavailable Fe pool in seawater.

Heme has been detected at picomolar concentrations in particulate organic matter from the subtropical Atlantic Ocean and nanomolar levels in estuarine waters (41–43). It has also been estimated to account for anywhere from 6 - 26% of the total cellular Fe concentration in coccolithophores, diatoms, green algae, and cyanobacteria (41). Heme is a photoreactive compound, and its hydrophobicity suggests it would mostly be surface or particle associated. Thus, it would presumably be most concentrated in low-light and high particulate environments, such as a sinking phytoplankton bloom in the lower part of the mixed layer. It is conceivable that after cell rupture intracellular heme-like complexes are released into the water column and remain structurally intact long enough for the entire molecule to be considered a potential and relevant Fe source to bacterioplankton. Therefore, large cell death events like decaying phytoplankton blooms may potentially provide pulses of heme-type molecules to the water column.

Living and detrital particles are an important habitat in the marine environment, since their high organic content supports abundant microbial populations and may relieve bacteria of macronutrient stress (44–46). This may select for bacteria capable of

producing Fe scavenging (47) molecules and allow for activation of multiple Fe uptake mechanisms targeting any source of bioavailable Fe, including heme. The bioavailability of heme as an Fe source has been well studied in pathogens (48, 49), has received some attention in specific rhizobial bacterial species (50–52), but has only received minor attention with respect to marine organisms (53, 54). The transport mechanism for heme uptake in pathogenic bacteria involves an ATP-dependent TonB complex (ExbB, ExbD, and TonB), an ATP-dependent ABC transport system (HmuV, HmuU, HmuT) and an outer membrane receptor (HmuR). This system closely resembles classical siderophore uptake systems, except that a heme oxygenase and/or chaperone protein (HmuS) located in the cytoplasm appears to be required to successfully utilize the Fe from heme complexes. Identifiable heme outer membrane receptors have been located in genomes from diverse taxonomic groups of nonpathogenic marine bacteria, including *Alphaproteobacteria*, *Gammaproteobacteria*, and *Bacteroidetes* among others (27, 28, 53, 55). The bioavailability of heme to marine bacteria has thus far been shown experimentally for three marine bacterial isolates (27, 53, 56) and heme uptake has been reported in natural microbial assemblages (57).

Some current challenges in the field of marine trace metal chemistry include the structural elucidation of natural marine Fe species, understanding how Fe speciation changes over time and space, and determining how marine microorganisms interact with different Fe species. The goal of my dissertation is to use biological sequence data from marine microbes in order to help reverse engineer some of the chemistry of Fe in seawater and address these existing challenges in a new light. The research presented here examines some of the specific molecular mechanisms and underlying genetics that

marine heterotrophic bacteria use to acquire Fe from the unique chemical conditions of the marine environment, with a particular emphasis on the Fe-containing molecule heme. Ultimately, our 'biologically-informed' perspective may provide a new and useful context for interpreting large-scale geochemical measurement programs like that of the Geotraces.

Chapter two of my dissertation is a literature review where I discuss what is known of the sources, transformations, and ultimate fate of heme and hemoproteins in seawater. The scope of the review spans the cellular physiology of marine organisms to the biogeochemical distributions of heme in ocean basins. Early on it was proposed that Fe porphyrin complexes, such as heme and hemoproteins, could be important weak Fe binding ligands in seawater. Heme-like molecules are probably released into the ocean water column upon the death and rupture of planktonic cells and the degradation of cellular proteins, however the fate of these molecules after cellular release has remained largely unknown. The review covers what is known about heme in the marine environment from the perspectives of how marine microbes acquire, transform, manage, and employ heme-based molecules. It also summarizes recent biogeochemical studies regarding the analysis and distributions of heme in the ocean with an eye as to how these and future studies may inform our current knowledge of the marine biogeochemical Fe cycle.

Chapter three is a comparative genomics study of the trace metal acquisition strategies in the *Roseobacter* and SAR11 lineages, two abundant and cosmopolitan marine heterotrophic bacterial groups. In it we evaluated gradations of uptake capabilities amongst *Roseobacter* and SAR11 genomes and examine these patterns in light of marine

trace metal chemistry, microbial ecology, and lineage evolutionary history. Our results suggest that many marine *Roseobacter* strains are tuned for extracting Fe ligand complexes from their environment while SAR11 strains are not. The patterns of metal uptake genes in these genomes is not predictable from lineage evolutionary history or other major lifestyle factors suggesting that trace metal uptake specialization happened relatively recently in these groups.

Chapter four presents work developing a genetic system for targeted mutagenesis in a model marine bacterial strain, *Ruegeria* sp. TM1040, and the insertional inactivation of a predicted TonB dependent outer membrane receptor in a putative heme uptake gene cluster. After identifying a putative heme uptake locus in the TM1040 genome, we have shown that genes in this locus are co-transcribed as an operon and are upregulated under Fe stress. We have further shown that the insertional inactivation of a predicted TonB dependent outer membrane receptor in the heme uptake operon results in the inability of TM1040 mutant strains to grow on heme as an Fe source. Using these strains, we have demonstrated that TM1040 can utilize Fe from soluble and insoluble diatom lysate while the mutant strain cannot which indicates that heme is an Fe form that TM1040 can acquire from structurally complex particulate organic matter. We have also shown in a closely related bacterial strain that its heme uptake genes are differentially regulated in response to a diatom symbiont. These results provide insights to the role of heme as a nutritional Fe source for marine heterotrophic bacteria, and suggest that heme may be an important molecular currency in the global marine Fe cycle.

In chapter five, we examined the production of different classes of Fe-binding ligands, Fe concentrations, and macronutrient concentrations in Fe amended microcosm

incubations conducted using oligotrophic water collected from the southern California Bight. We examined the phytoplankton and bacterioplankton assemblage composition at the termination of the incubation while monitoring chlorophyll and macronutrients over a period of six days. The phytoplankton assemblage composition was assessed using microscopy while the bacterial/archaeal assemblage structure was assessed using large scale marker gene surveys. These data allowed us to explore relationships between chemical parameters and the abundances of individual microbial taxa and highlighted a potential role for typically rare but transiently abundant microbial taxa in broadly influencing the type and strength of Fe-binding ligands in Fe-amended versus control incubations.

Finally, in chapter six, we identified putative Fe uptake and stress response systems from metatranscriptomes collected from surface waters and waters from the deep chlorophyll maximum in the eastern subtropical Pacific Ocean. The taxonomically annotated metatranscriptomic data suggests that phytoplankton genes related to Fe stress were highly expressed in a handful of eukaryotic groups in the deep chlorophyll maximum. Further, the ratio of dissolved silica to nitrogen provided an independent geochemical indicator of Fe limitation of the overall biological community. Taken together these results suggest that the ecosystem of the California current can be highly dynamic with regards to nutrient limitation and that molecular biomarkers of Fe stress are useful for pinpointing this limitation in situ.

1.1 References

1. Morel FMM, Price NM (2003) The biogeochemical cycles of trace metals in the oceans. *Science* 300:944–947.

2. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281(5374):237–240.

3. Moore JK, Doney SC, Glover DM, Fung IY (2001) Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep Sea Res Part 2 Top Stud Oceanogr* 49:463–507.

4. Boyd PW, Jickells T, Law CS, Blain S, Boyle EA, Buesseler KO, Coale KH, Cullen JJ, de Baar HJW, Follows M, Harvey M, Lancelot C, Levasseur M, Owens NPJ, Pollard R, Rivkin RB, Sarmiento J, Schoemann V, Smetacek V, Takeda S, Tsuda A, Turner S, Watson AJ (2007) Mesoscale Iron Enrichment Experiments 1993-2005: Synthesis and Future Directions. Science 315(5812):612–617.

5. Gledhill M, van den Berg CMG (1994) Determination of complexation of iron(III) with natural organic complexing ligands in seawater using cathodic stripping voltammetry. *Mar Chem* 47(1):41–54.

6. Rue EL, Bruland KW (1995) Complexation of iron(III) by natural organic ligands in the Central North Pacific as determined by a new competitive ligand equilibration/adsorptive cathodic stripping voltammetric method. *Mar Chem* 50:117–138.

7. Wu J, Luther GW (1994) Size-fractionated iron concentrations in the water column of the western North Atlantic Ocean. *Limnol Oceanogr* 39(5):1119–1129.

8. Buck KN, Bruland KW (2007) The physicochemical speciation of dissolved iron in the Bering Sea, Alaska. *Limnol Oceanogr* 52(5):1800–1808.

9. Rue EL, Bruland KW (1997) The role of organic complexation on ambient iron chemistry in the equatorial Pacific Ocean and the response of a mesoscale iron addition experiment. *Limnol Oceanogr* (Iii):901–910.

10. Mawji E, Gledhill M, Milton J a., Tarran G a., Ussher S, Thompson A, Wolff G a., Worsfold PJ, Achterberg EP (2008) Hydroxamate siderophores: occurrence and importance in the Atlantic Ocean. Environ Sci Technol 42(23):8675–8680.

11. Velasquez I, Nunn BL, Ibisanmi E, Goodlett DR, Hunter K a., Sander SG (2011) Detection of hydroxamate siderophores in coastal and Sub-Antarctic waters off the South Eastern Coast of New Zealand. Mar Chem 126(1-4):97–107.

12. Vraspir JM, Butler A (2009) Chemistry of Marine Ligands and Siderophores. *Ann*

Rev Mar Sci 1(1):43–63.

13. Wu J, Luther GW (1995) Complexation of Fe (III) by natural organic ligands in the Northwest Atlantic Ocean by a competitive ligand equilibration method and a kinetic approach. *Mar Chem* 50:159–177.

14. Hopkinson BM, Morel FMM (2009) The role of siderophores in iron acquisition by photosynthetic marine microorganisms. *Biometals* 22(4):659–669.

15. Boyd PW, Ibisanmi E, Sander SG, Hunter K a., Jackson G a. (2010) Remineralization of upper ocean particles: Implications for iron biogeochemistry. *Limnol Oceanogr* 55(3):1271–1288.

16. Hassler C, Schoemann V, Nichols C, Butler E, Boyd P (2011) Saccharides enhance iron bioavailability to Southern Ocean phytoplankton. *Proc Natl Acad Sci U S A* 108(3):1076–1081.

17. Norman L, Worms IAM, Angles E, Bowie AR, Nichols CM, Ninh Pham A, Slaveykova VI, Townsend AT, David Waite T, Hassler CS (2015) The role of bacterial and algal exopolymeric substances in iron chemistry. Mar Chem 173:148–161.

18. Laglera LM, van den Berg CMG (2009) Evidence for geochemical control of iron by humic substances in seawater. *Limnol Oceanogr* 54(2):610–619.

19. Andrews SC, Robinson AK, Rodri-guez-Quiones F (2003) Bacterial iron homeostasis. *FEMS Microbiol Rev* 27(2-3):215–237.

20. Noinaj N, Guillier M, Barnard TJ, Buchanan SK (2010) TonB-dependent transporters: regulation, structure, and function. *Annu Rev Microbiol* 64:43–60.

21. Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* 284(37):24673–24677.

22. González JM, Fernández-Gómez B, Fernàndez-Guerra A, Gómez-Consarnau L, Sánchez O, Coll-Lladó M, Del Campo J, Escudero L, Rodríguez-Martínez R, Alonso-Sáez L, Latasa M, Paulsen I, Nedashkovskaya O, Lekunberri I, Pinhassi J, Pedrós-Alió C (2008) Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). Proc Natl Acad Sci U S A 105(25):8724–8729.

23. Schauer K, Rodionov D, de Reuse H (2008) New substrates for TonB-dependent transport: do we only see the "tip of the iceberg"? *Trends Biochem Sci* 33(7):330–338.

24. Mirus O, Strauss S, Nicolaisen K, von Haeseler A, Schleiff E (2009) TonB-

dependent transporters and their occurrence in cyanobacteria. *BMC Biol* 7(68):1–25.

25. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, Vasse J, Lauber E, Arlat M (2007) Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS One 2(2):e224.

26. Tang K, Jiao N, Liu K, Zhang Y, Li S (2012) Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. *PLoS One* 7(7):e41204.

27. Hopkinson B, Barbeau K (2012) Iron transporters in marine prokaryotic genomes and metagenomes. *Environ Microbiol* 14(1):114–128.

28. Hogle SL, Cameron Thrash J, Dupont CL, Barbeau KA (2016) Trace metal acquisition by marine heterotrophic bacterioplankton with contrasting trophic strategies. *Appl Environ Microbiol*. doi:10.1128/AEM.03128-15.

29. Toulza E, Tagliabue A, Blain S, Piganeau G (2012) Analysis of the global ocean sampling (GOS) project for trends in iron uptake by surface ocean microbes. *PLoS One* 7(2):e30931.

30. Koropatkin N, Randich AM, Bhattacharyya-Pakrasi M, Pakrasi HB, Smith TJ (2007) The structure of the iron-binding protein, FutA1, from Synechocystis 6803. *J Biol Chem* 282(37):27468–27477.

31. Anderson DS, Adhikari P (2004) The hFbpABC transporter from Haemophilus influenzae functions as a binding-protein-dependent ABC transporter with high specificity and affinity for ferric iron. *J Bacteriol* 186(18):6220–6229.

32. Vergauwen B, Elegheert J, Dansercoer A, Devreese B, Savvides S (2010) Glutathione import in Haemophilus influenzae Rd is primed by the periplasmic heme-binding protein HbpA. *Proc Natl Acad Sci U S A* 107(30):13270–13275.

33. Vajrala N, Sayavedra-Soto L a., Bottomley PJ, Arp DJ (2010) Role of Nitrosomonas europaea NitABC iron transporter in the uptake of Fe3+-siderophore complexes. *Arch Microbiol* 192(11):899–908.

34. Cartron ML, Maddocks S, Gillingham P, Craven CJ, Andrews SC (2006) Feo--transport of ferrous iron into bacteria. *Biometals* 19(2):143–157.

35. Morrissey J, Bowler C (2012) Iron utilization in marine cyanobacteria and eukaryotic algae. *Front Microbiol* 3(March):43.

36. Hutchins DA, Bruland KW (1994) Grazer-mediated regeneration and assimilation of Fe, Zn and Mn from planktonic prey. *Mar Ecol Prog Ser* 110(1993):259–269.

37. Barbeau K, Moffett JW, Caron DA, Croot PL, Erdner DL (1996) Role of protozoan grazing in relieving iron limitation of phytoplankton. *Nature* 380:61–64.

38. Sarthou G, Vincent D, Christaki U, Obernosterer I, Timmermans KR, Brussaard CPD (2008) The fate of biogenic iron during a phytoplankton bloom induced by natural fertilisation: Impact of copepod grazing. Deep Sea Res Part 2 Top Stud Oceanogr 55(5-7):734–751.

39. Cho BC, Azam F (1988) Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature* 332(31):441–443.

40. Bidle KD, Azam F (1999) Accelerated dissolution of diatom silica by marine bacterial assemblages. *Nature* 397:508–512.

41. Honey D, Gledhill M, Bibby T, Legiret F, Pratt N, Hickman A, Lawson T, Achterberg E (2013) Heme b in marine phytoplankton and particulate material from the North Atlantic Ocean. Mar Ecol Prog Ser 483:1–17.

42. Gledhill M, Achterberg EP, Honey DJ, Nielsdottir MC, Rijkenberg MJ a. (2013) Distributions of particulate Heme b in the Atlantic and Southern Oceans-Implications for electron transport in phytoplankton. *Global Biogeochem Cycles* 27(4):1072–1082.

43. Vong L, Laës A, Blain S (2007) Determination of iron-porphyrin-like complexes at nanomolar levels in seawater. *Anal Chim Acta* 588(2):237–244.

44. Alldredge AL, Silver MW (1988) Characteristics, dynamics and significance of marine snow. *Prog Oceanogr* 20:41–82.

45. Buchan A, LeCleir GR, Gulvik C a., González JM (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* 12(10):686–698.

46. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R (2012) Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. Science 336(2006):608–611.

47. Cordero OX, Ventouras L-A, DeLong EF, Polz MF (2012) Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc Natl Acad Sci U S A* 109(49):20059–20064.

48. Anzaldi LL, Skaar EP (2010) Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. *Infect Immun* 78(12):4977–4989.

49. Wandersman C, Delepelaire P (2004) Bacterial iron sources: from siderophores to hemophores. *Annu Rev Microbiol* 58:611–647.

50. Noya F, Arias A, Fabiano E (1997) Heme compounds as iron sources for nonpathogenic Rhizobium bacteria. *J Bacteriol* 179(9):3076–3078.

51. Puri S, O'Brian MR (2006) The hmuQ and hmuD genes from Bradyrhizobium japonicum encode heme-degrading enzymes. *J Bacteriol* 188(18):6476–6482.

52. Nienaber A, Hennecke H, Fischer HM (2001) Discovery of a haem uptake system in the soil bacterium Bradyrhizobium japonicum. *Mol Microbiol* 41(4):787–800.

53. Hopkinson BM, Roe KL, Barbeau K (2008) Heme uptake by Microscilla marina and evidence for heme uptake systems in the genomes of diverse marine bacteria. *Appl Environ Microbiol* 74(20):6263–6270.

54. Roe KL, Hogle SL, Barbeau K (2013) Utilization of heme as an iron source by marine alphaproteobacteria in the roseobacter clade. *Appl Environ Microbiol* 79(18):5753–5762.

55. Hogle SL, Barbeau KA, Gledhill M (2014) Heme in the marine environment: from cells to the iron cycle. *Metallomics* 6(6):1107–1120.

56. Septer AN, Wang Y, Ruby EG, Stabb EV, Dunn AK (2011) The haem-uptake gene cluster in Vibrio fischeri is regulated by Fur and contributes to symbiotic colonization. *Environ Microbiol* 13(11):2855–2864.

57. Weaver R, Kirchman D, Hutchins D (2003) Utilization of iron/organic ligand complexes by marine bacterioplankton. *Aquat Microb Ecol* 31:227–239.

Chapter 2

Heme in the marine environment: From cells to the iron cycle

2.1 Abstract

Hemes are iron containing heterocyclic molecules important in many cellular processes. In the marine environment, hemes participate as enzymatic cofactors in biogeochemically significant processes like photosynthesis, respiration, and nitrate assimilation. Further, hemoproteins, hemes, and their analogs appear to be iron sources for some marine bacterioplankton under certain conditions. Current oceanographic analytical methodologies allow for the extraction and measurement of heme *b* from marine material, and a handful of studies have begun to examine the distribution of heme *b* in ocean basins. The study of heme in the marine environment is still in its infancy, but some trends can be gleaned from the work that has been published so far. Here we review what is known about heme in the marine environment from the perspectives of how marine microbes acquire, transform, manage, and employ heme-based molecules. We also summarize recent biogeochemical studies regarding the sources and fates of heme in the ocean with an eye as to how these and future studies may inform our current knowledge of the marine biogeochemical iron cycle. We conclude by presenting some future questions and challenges for the field.

2.2 Intracellular roles for heme in marine microbes

Hemes, iron-porphyrin complexes, function as prosthetic groups in numerous proteins that perform diverse biological functions across all domains of life. The ubiquity of the heme group in enzymes is due to its abilities to readily function as an electron source/sink and small molecule binding site. Some of the roles for hemoproteins include facilitating mitochondrial and chloroplast electron transfer reactions, divalent gas transport and storage, organic substrate oxygenation, peroxide reduction, cellular signal transduction, and regulation of gene expression (1). The heme molecule is a critical component in essential cellular processes that respond both directly and indirectly to the chemical and physical environments of marine phytoplankton and bacterioplankton. From this perspective, heme is a functional molecular link between the cellular level biology of phytoplankton and bacterioplankton and their resulting ecology. Here we summarize some of the major intracellular roles for heme and hemoproteins in the context of marine phytoplankton and bacterioplankton (Figure 2.1).

2.2.1 Electron transfer: Respiration and Photosynthesis

Heme is a central component of electron transport complexes including those participating in the processes of aerobic and anaerobic respiration, some forms of extended anaerobic respiration, and photosynthesis. The terminal portion of the electron transport chain in aerobic and anaerobic respiration utilizes a number of cytochromes, membrane-associated hemoproteins, for the generation of ATP (2, 3). Many of these processes occur at the inner membrane of the mitochondria in eukaryotes and the cytoplasmic membrane of prokaryotes (3, 4). However, extended respiratory electron transport systems have recently been discovered to transport electrons all the way to the

outer membrane of gram negative bacteria thus utilizing extracellular oxidants in respiration (5). These systems are often characterized by an abundance of c-type cytochromes (6, 7). Hemoproteins are also involved in the process of oxygenic photosynthesis with many functional similarities to those in the electron transport chain from respiration (8). Photosystem II contains a heme complex in the D1 reaction center called cytochrome b_{559}, a heterodimer composed of one alpha subunit (PsbE) and one beta (PsbF) subunit (9). Cytochrome b_6f, which participates in the shuttling of electrons from photosystem II to photosystem I, contains a Rieske [2Fe-2S] protein and four heme groups (10). In copper deficient conditions, cyanobacteria and many algae utilize a heme-containing cytochrome c_6 complex instead of plastocyanin to transport electrons between the cytochrome b_6f complex and photosystem I (11).

2.2.2 Electron transfer: Management of Reactive oxygen species

In the marine environment, reactive oxygen species (ROS) are produced from the photochemical oxidation of DOM in waters penetrable by solar radiation, particularly that in the ultraviolet wavelengths (12). However, recent field and laboratory studies have shown that extracellular superoxide is produced directly and light-independently by marine bacteria and algae and may be the dominant source of ROS in some marine waters (13–16). In marine algae it has been inferred that extracellular superoxide is produced by nicotinamide adenine dinucleotide phosphate (NADPH) oxidases (17), which are membrane bound enzyme complexes that produce superoxide through a reactive heme prosthetic group (18). In particular, the genomes of the marine diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* contain putative homologs to

human, plant, and fungal NADPH oxidases, which suggests algal extracellular superoxide production may be due to these hemoproteins (17). A recent study has indicated that marine heterotrophic bacteria may produce extracellular superoxide by way of NADPH oxidases homologous to those in eukaryotes (15). This evidence suggests that both marine heterotrophic bacteria and phytoplankton may produce extracellular superoxide through a heme-dependent reaction.

ROS have negative effects on marine microbes by inhibiting critical metabolic processes such as photosynthesis or by interfering with biological membranes. Hydrogen peroxide is one of the most abundant intracellular and extracellular ROS produced typically through a superoxide intermediate. Excessive intracellular hydrogen peroxide and other ROS are harmful for almost all cell types and thus must be managed rapidly and efficiently. Heme is the prosthetic group in catalases, peroxidases, and catalase-peroxidases that are antioxidant enzymes involved in the degradation of hydrogen peroxide and other organic hydroperoxides (19–21). Catalases and peroxidases both break oxygen–oxygen single bonds, but the nature of the electron donor in the two reaction mechanisms differs between the two enzyme families. Catalases utilize hydrogen peroxide as an electron donor in the final step of the catalytic cycle, while peroxidases can utilize a great variability of organic and inorganic reducing agents including some proteins (22). Catalase-peroxidases show peroxidase activity towards specific organic peroxides while simultaneously retaining activity towards hydrogen peroxides like that seen in monofunctional catalases (23). In higher plants, catalase activity is concentrated in the peroxisome, a single membrane bound organelle that mediates a wide array of biochemical processes including fatty acid β-oxidation, photorespiration, metabolism of

hydrogen peroxide, and synthesis of plant hormones (24). Very little is known of peroxisomes in unicellular marine algae, although some work has been done with the model freshwater alga *Chlamydomonas reinhardtii (25)*. Peroxisomes are predicted to exist in *T. pseudonana* from analyses of its genome (26), but their general distribution in marine algae is largely unknown. Peroxidases in higher plant models are concentrated in the apoplast and vacuole, but little work has been done with model algae. Plant vacuolar peroxidases catalyze the oxidation and polymerization of a variety of phenolic compounds and other secondary metabolites while reducing hydrogen peroxide (27). Extrapolating from these plant models, it is likely that peroxidase activity is also present in algal vacuoles.

2.2.3 Electron transfer: Other redox reactions

Heme-containing proteins in marine bacteria and phytoplankton are also involved in the reduction and oxidation of various endogenous and exogenous compounds. For example, cytochrome P450 monooxygenase and cytochrome b_5 function (often in conjunction) in oxidizing a wide variety of substrates. Cytochrome b_5 functions in the anabolic metabolism of fats and steroids as well as in the catabolism of xenobiotics and compounds of endogenous metabolism (28). Heme-proteins are also involved in processes of inorganic nutrient uptake that often requires reduction of the substrate to a usable form. In marine algae, reduction of nitrate to nitrite is accomplished by a NADPH nitrate reductase. This enzyme contains a molybdenum core as well as heme cofactors that facilitate electron transport between NADPH and nitrate (29). In marine diatoms, the NADPH nitrate reductase enzyme is localized to the cytoplasm (30, 31). Diatom genomes

also encode ferredoxin nitrite reductases that appear to be targeted to the chloroplast (32). Ferredoxin nitrite reductases utilize a siroheme, a heme analog, and a [4Fe–4S] cluster as catalytic cofactors (33). A recent study has convincingly argued that when rapidly relieved of iron stress diatoms appear to partition iron to nitrate and nitrite reductases and other nitrate assimilation proteins while continuing to utilize iron-free photosynthetic proteins (34). In the same study, iron addition also induced a significant upregulation of the porphyrin biosynthesis pathway, further suggesting that diatoms rapidly partition iron into hemoproteins when relieved of iron stress. Sulfite reductases, which conduct the six electron reduction from sulfite to sulfide, also contain heme-like cofactors important in inorganic sulfur assimilation. Much like nitrite reductases, sulfite reductases utilize siroheme and iron-sulfur clusters as cofactors and are targeted to the plastid (35). Finally, heme-containing ferric reductases are employed in the solubilization of extracellular complexed iron. Three types of heme-containing ferric reductases are employed in algae, NADPH oxidases (cytochrome b_{558} containing), cytochrome b_5 reductases, and cytochrome b_{561}, and all appear to be involved in iron mobilization in organelles and at the plasma membrane (36). Heme-based ferric reductase genes have been identified in genomes of P. *tricornutum* (PtFRE1 – PtFRE4)48 and *T. pseudonana* (TpFRE1 and TpFRE2) (26, 37). It is likely some of these putative diatom ferric reductases localize to the outer membrane and are involved in extracellular iron assimilation, although they may be involved in intracellular iron trafficking or other unrelated roles (36).

2.2.4 Signaling and sensing: O_2 and NO

In bacteria another major role for heme-containing proteins is in sensing. At the molecular level, heme sensor proteins act as bistable switches by binding to molecular oxygen (O_2), nitric oxide (NO), or carbon monoxide (CO) and inducing a conformational change in the sensor. The activated sensor protein domain then interacts with a protein domain capable of a response, which leads to modulation of expression levels of specific proteins (38). No biochemical studies have been performed on signaling pathways in planktonic marine bacteria, but the findings uncovered in model terrestrial strains are likely applicable to these organisms. Recently, it has been shown that a globin-coupled diguanylate cyclase induces the synthesis of the second messenger bis-(3′–5′)-cyclic diguanosine monophosphate (cyclic-di-GMP) upon O_2 binding at the heme prosthetic group of the globin. It is still unclear as to the systematic effect of the messenger, but it has been shown to influence biofilm formation in *Bordatella pertussis* (39). In certain rhizobial bacteria a heme-based sensor, FixL, regulates metabolic processes under aerobic and microaerobic conditions (40). The ultimate downstream effect of the protein is to prevent the expression of nitrogen fixation and denitrification systems when O_2 concentrations are high enough to become deleterious. The FixL protein contains an N-terminal heme binding domain as well as C-terminal histadine kinase domain, which is inhibited under the presence of O_2. Kinase activity is restored when O_2 concentrations decrease, which catalyzes the phosphorylation of a transcription factor that interacts with denitrification and nitrogen fixation genes (41).

NO signaling has been extensively studied in vertebrate systems, but has only recently been acknowledged as an important signaling system in bacteria and plants (42, 43). Both NO binding proteins as well as NO synthases utilize heme cofactors. Recently,

a functional heme-nitric oxide/oxygen binding domain was identified in *Shewanella oneidensis* and was shown to influence biofilm formation by modulating cyclic-di-GMP metabolism (44). Many of these systems have been shown to exclusively bind NO, but in some bacterial strains they have equal affinity for both O_2 and NO (42). Regardless, they appear to be important in regulating communal bacterial behavior such as in biofilm formation, dispersal, motility and symbiosis. A charismatic example in the marine environment is the symbiotic colonization of the light organs of the Hawaiian bobtail squid by *Vibrio fischeri*, a bioluminescent marine bacterium. NO production in the light organs of the squid allows for the selective recruitment of planktonic *V. fischeri* to these organs and the subsequent bacterial biofilm formation necessary for bioluminescence (45). NO signaling has also been implicated in regulatory mechanisms in plants, although numerous details as to the production of NO *in vivo* have yet to be elucidated (46). NO production seems to be localized to the cytoplasm in higher plant cells, but cellular localization in algae remains unknown. Further, it appears that there are numerous NO enzymatic sources other than NO synthase in plants (43). Recently, a NO synthase with significant homology to those in vertebrates has been identified in the marine green alga, *Ostreococcus tauri* (47). The NO synthase identified in *O. tauri* was shown to be functional when heterologously expressed in *E. coli*, and its expression *in vivo* was dependent upon light irradiance and growth phase on the alga. This suggests a link between NO production and algal physiology and points to a potential role for NO in regulating cellular processes in *O. tauri.* NO synthases may facilitate the development and demise of algal blooms in the marine environment as NO has been linked to reduced growth and photosynthesis and increased cell death in marine diatoms (48).

2.3 Methods for heme acquisition in marine microbes

As outlined above, heme is required for the functioning of many essential enzymes. To fulfill their needs, marine organisms may synthesize their own heme from starting materials as well as utilize exogenous heme as an iron/heme source.

2.3.1 Heme biosynthesis

The biosynthesis of heme is a fundamental metabolic capability common to both prokaryotes and eukaryotes. Not only is the heme biosynthetic pathway necessary for the function of essential enzymes and proteins, but it is also involved in synthesizing other porphyrin-based molecules such as chlorophylls, bacteriochlorophylls, phycobillins, and the corrin center of vitamin B_{12}. The essential roles of porphyrins in photosynthesis ensures that the pathway is conserved in marine phytoplankton although in eukaryotic algae the different enzymes in the pathway appear to have multiple evolutionary origins including from cyanobacteria, alphaproteobacteria, and heterotrophic eukaryotes (49). The majority of all sequenced bacterial genomes also contain the necessary components for heme biosynthesis (50), but some notable exceptions lack genes necessary in the canonical heme biosynthetic pathway suggesting that they do not use heme, produce it through an unknown pathway, or rely strictly on heme from the external environment (51, 52). Of the genomes of marine organisms surveyed in this review (see methods), all contained the majority of components required for a full heme biosynthetic pathway.

Although, the heme biosynthesis pathway is generally well conserved in prokaryotes and eukaryotes (Figure 2.2) the initial part of the pathway differs between taxonomic groups (50, 53). In the first step the first universal heme precursor synthesized

is δ-aminolevulinic acid (ALA). In marine photosynthetic eukaryotes as well as all prokaryotes excluding the *Alphaproteobacteria*, ALA is synthesized by glutamyl-tRNA synthase through the C5 pathway using glutamate as the starting material. Marine *Alphaproteobacteria* synthesize ALA by condensation of succinyl-CoA with glycine by means of aminolevulinate synthase. After the synthesis of ALA, the remainder of the heme biosynthesis pathway is generally the same in all organisms (Figure 2.2). However, there are two different forms of the enzymes converting coproporphyrinogen III to protoporphyrinogen IX (HemF or HemN) and protoporphyrinogen IX into protoporphyrin IX (HemY or HemG). The difference between the two is that of oxygen-dependence, and the oxygen-independent versions of the enzymes (HemN and HemG) appear to be restricted to prokaryotes. Ultimately, eight molecules of ALA are converted to protoporphyrin IX in a series of six enzymatic steps. In the final enzymatic reaction, a ferrochelatase (HemH) inserts iron into the porphyrin ring generating the final heme molecule which can be used directly or modified further before insertion into hemoproteins (50, 53, 54).

In marine photoautotrophs, the terminus of heme biosynthesis pathway merges with the chlorophyll biosynthesis pathway where magnesium is inserted into the porphyrin ring instead of iron. In marine phototrophic eukaryotes heme biosynthesis and catabolism are compartmentalized to the mitochondria and chloroplasts, while in marine prokaryotes these processes take place in the cytoplasm and/or periplasm. In the chloroplasts heme is utilized in the assembly of the cytochrome b_6f complex of the photosynthetic electron transport chain, while heme catabolic products are utilized as precursors in phycobilin pigments in phytoplankton with phycobilisomes and in the

synthesis of the photoregulatory phytochromobilin apoprotein (55). It appears now that in model phototrophic eukaryotes all of the genes coding for enzymes common to both heme and chlorophyll biosynthesis are expressed exclusively in the chloroplast (Figure 2.1) (54, 56). However, isoforms of protoporphyrinogen oxidase (HemY) and ferrochelatase (HemH) have been shown to be dually targeted to the plastidal inner membrane and the mitochondrial inner membrane and may be involved in heme biosynthesis (57, 58).

2.3.2 Exogenous heme uptake

Marine phytoplankton and bacterioplankton must extract iron from seawater to satisfy their metabolic requirements. The chemical speciation of iron in seawater is complex, and effectively all iron in the ocean is complexed to organic ligands of some kind. Marine microbes must be able to acquire this iron by either utilizing the iron-ligand complexes directly or modifying them at the cell surface to a new chemical form that can be more easily assimilated. Marine photoautotrophic microbes generally have less-well characterized systems for iron uptake, and it has generally been assumed their strategy for iron acquisition involves modification of external compounds to more bioavailable forms, typically ferrous iron (59). An early study reported that heme was more bioavailable than siderophore iron-ligand complexes for two species of marine diatoms as well as for natural algal assemblages. In contrast, the study also reported that heme bioavailability was drastically lower than that of siderophore complexed iron for two marine *Synechococcus* species, suggesting that different groups of phytoplankton specialize in utilizing different classes of ligand-bound iron (60). This work, as well as subsequent

similar studies which employed radiolabelled iron porphyrin complexes presumed to be formed via seawater equilibration (61–65), is difficult to interpret mechanistically due to uncertainty regarding the chemical speciation of iron radiotracers added in association with porphyrins, which do not appear to effectively form complexes with iron under seawater conditions (66).

While results of radiotracer uptake studies can be ambiguous, the existence of iron uptake pathways and homeostasis mechanisms in marine microbial genomes and metagenomes can also be bioinformatically inferred from functional information derived from well-studied model organisms. Relatively little is known about iron transport in marine bacteria compared to terrestrial and pathogenic microbes, yet the generalized models established in well-characterized pathogens are important guides for unveiling the phenomenon in marine microbes. Recent evidence indicates that a number of marine heterotrophic bacteria can directly uptake intact heme from seawater and many have putative systems for heme uptake encoded in their genomes (67, 68). Although the ecological and biogeochemical impacts of this capability have yet to be explored, it does suggest that in certain marine microenvironments heme may be a relatively abundant form of iron. Heme acquisition systems that directly acquire heme or extract it from hemoproteins have been identified in both terrestrial pathogenic Gram-negative and Gram-positive bacteria. In Gram-negative bacteria, heme transport systems are classified as either direct uptake, bipartite uptake, or hemophore-mediated uptake. Less is known about heme uptake in Gram-positive bacteria but both direct uptake and hemophore-mediated uptake systems have been identified (52). As of this date, the direct heme uptake system (67–69) and hemophore-mediated uptake systems are the only forms that

have been identified in the genomes of sequenced marine bacteria. Further, no genomes of Gram-positive bacteria isolated from the marine environment contained homologs to any heme uptake system, although they are generally less-well represented in sequence databases. Here we will review the mechanisms of uptake for the two heme uptake systems identified in Gram-negative marine bacterial genomes.

Direct heme uptake systems bind heme or hemoproteins to a TonB dependent transporter (TBDT) at the outer membrane (OM) and transport heme to the periplasm (Figure 2.3). Once a heme substrate is recognized at the extracellular surface of the TBDT, conformational changes in the protein transmit a signal to a complex consisting of TonB, ExbB, and ExbD proteins located in the periplasmic space (51, 52). The TonB complex energizes the TBDT to unidirectionally shuttle heme through the outer membrane. Once in the periplasm, heme is intercepted by a periplasmic binding protein and shuttled to an ATP binding cassette (ABC) transporter at the inner membrane (IM) after which it is moved to the cytoplasm. Heme ABC transporters consist of IM-spanning permease domains that are energized by ATP binding domains which catalyze ATP hydrolysis on the cytoplasmic side of the IM. Once in the cytoplasm free heme must be dismantled or sequestered due to its reactivity. In many gram negative bacteria, cytoplasmic heme is degraded by heme oxygenases (HO, PFAM01126) with structural similarities to mammalian HOs (70, 71). However, some bacterial genomes lack any homologs to mammalian HOs, and in a pathogenic *Escherichia coli* strain, heme degrading activity is accomplished by a protein, ChuS, lacking structural similarity to any known HOs (72). However, another study has shown that a ChuS homolog in *Pseudomonas aeruginosa* does not have HO activity and is instead responsible for

delivering heme to an already identified heme oxygenase (73). Further, the ChuS family (PFAM05171) contains sequence similarity to a different protein family implicated in heme utilization, HutX (PFAM06228). Even though the exact function of ChuS (PFAM05171) and HutX (PFAM06228) remains unclear, most putative direct heme uptake operons contain a gene encoding one of the two protein families.

Direct heme uptake systems have been identified by homology and conservation of gene order in a number of sequenced marine bacteria. Hopkinson and colleagues (67) showed that *Microscilla marina*, a member of the marine *Cytophagia* group known to be associated with particulate organic matter, can sustain growth on heme as the sole iron source. They also demonstrated that heme uptake genes identified by homology to those in human pathogens were upregulated under iron stress and during growth on heme. A subsequent study (68) demonstrated that *Ruegeria* sp. TrichCH4B, a member of the Roseobacter clade isolated from the marine nitrogen fixing cyanobacterium *Trichodesmium erythraeum*, could sustain growth on heme and had a putative heme uptake genomic locus similar to that found in *M. marina.* Further, *R.* sp. TrichCH4B could utilize a variety of other iron-porphyrin complexes in addition to heme, and its putative heme uptake locus was upregulated under iron stress. The authors of this study also demonstrated the presence of putative heme uptake loci in roughly half of all Roseobacter genomes sequenced to date, suggesting that this capability might be generally common in the clade. They were also able to amplify genes homologous to the putative cytoplasmic heme utilization gene (ChuS, PFAM05171) of *R.* sp. TrichCH4B from a variety of coastal and open ocean waters demonstrating its presence in diverse marine environments. Of all the marine bacterial genomes in the Integrated Microbial

Genomes (IMG) database (74), nearly 24% have at least one type of complete putative heme uptake system with the vast majority being direct uptake systems. This abundance of putative uptake systems may be indicative of the significance of heme as an iron source for marine bacteria. The marine direct uptake systems appear to utilize heme cytoplasmic proteins of both ChuS and HutX families and in roughly equal proportion (Figure 2.4). Interestingly, the HutX protein family is exclusive to marine *Gammaproteobacteria,* mostly of the family *Vibrionaceae,* while ChuS has a broader taxonomic spread including *Alphaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Flavobacteriia, Cytophagia,* and *Sphingobacteri*ia. The most recent published study (69) examining general iron uptake systems in marine bacterial genomes found heme uptake systems to be common in isolate marine bacterial genomes but uncommon in the Global Ocean Sampling (75) (GOS) marine metagenomes. Two additional studies examining GOS reported a similar lack of heme uptake systems (76, 77). However, the lack of heme uptake genes in the GOS dataset is likely due to the dominance in the dataset of picocyanobacteria and *Pelagibacter* species (75) whose genomes almost entirely lack heme uptake systems (69). Targeted metagenomics of marine microzones, such as particles, where heme is more likely to be a component of the bioavailable iron pool may yield a greater diversity and abundance of heme uptake genes.

Hemophore-mediated heme uptake systems utilize a protein secreted outside the cell to bind and mobilize heme to a TBDT at the OM. Generally, the hemophore-mediated heme uptake operon encodes a heme-binding hemophore (HasA), an inner membrane complex for exporting apo-hemophores outside the cell (HasDEF), a TBDT hemophore receptor (HasR), two regulatory proteins (HasI and HasS), a TonB-like

protein, but no IM heme ABC transporter (Figure 2.3) (51). In addition to binding HasA, HasR has affinity for heme and hemoglobin but is most efficient when a hemophore is used. The HasA hemophore-mediated heme uptake system has been identified in *Serratia marcescens*, *P. aeruginosa*, *Pseudomonas fluorescens*, and *Yersinia pestis* (52). In this study we have identified previously unreported putative HasA-like hemophore-mediated heme uptake systems in the genomes of the Gram-negative marine bacteria *Pseudoalteromonas luteoviolacea* 2ta16 (isolated from a tropical coral), *Thalassospira profundimaris* WP0211 (from deep sea sediment), *Thalassospira xiamenensis* M-5 (from oil-contaminated surface seawater), *Alcanivorax dieselolei* B5 (oil-contaminated seawater), and *Pseudovibrio* sp. JE062 (from a Caribbean marine sponge) (Fig 2.3B). Although, direct heme uptake systems appear to be the most common in the marine environment, it appears that hemophore-mediated uptake may be useful under certain conditions. Further sequenced marine bacterial genomes will aid in assessing the prevalence of HasA-like heme uptake in the marine environment.

Even though little is known of heme uptake in marine heterotrophic bacteria, even less is known in marine phytoplankton. The genomes of *Synechococcus* sp. PCC 7002, a strain from brackish water, and *Prochlorococcus marinus* str. MIT9202, isolated from the tropical south Pacific Ocean, are the only marine cyanobacterial genomes that have TBDT with significant homology to those in well-characterized heme uptake operons (69). However, the genomic regions around the TBDT in each strain have little synteny to classical heme uptake operons and lack the presence of genes coding HOs or the ChuS/HutX protein. To our knowledge heme uptake experiments have not been performed with either strain. Recently, a transcriptomic study of the marine diatom *P.*

tricornutum grown under iron limitation reported an increase in the number of transcripts coding for a putative HO (PFAM01126) (78). Although the cellular location of this HO was not determined, the authors postulated that if localized to the outer membrane it could be responsible for the apparent diatom heme utilization reported in by Hutchins *et al.* (60). However, the genomes of many marine cyanobacteria also contain HOs of the same family as in *P. tricornutum*. Thus, if the observed disparity of heme uptake between diatoms and cyanobacteria observed by Hutchins *et al.* is due to known HOs, it must be due to differential regulation and/or cellular localization of a similar gene product in the two phytoplankton classes. Greater insight into heme utilization by marine phytoplankton will likely be gained from further genome sequencing and the genetic manipulation of model strains (79).

2.4 Heme in the marine environment

Undoubtedly, heme-like molecules and hemoproteins are abundant in marine organisms and play important roles in their cellular biology. However, marine scientists are only just beginning to apply what is known of heme at the cellular/molecular level to the scales of ecology and biogeochemistry. Converging evidence is beginning to suggest that heme and hemoproteins persist long enough in seawater and on marine particulates to be considered relevant players in iron biogeochemical cycling. Further, field measurements of intracellular heme and hemoprotein content in marine phytoplankton may be indicative of nutritional status as well as community function. The paucity of measurements in these areas leaves much to be explored in heme biogeochemistry. Here

we review what is currently known about the aqueous chemistry and distributions of extracellular heme and hemoproteins in the ocean and in marine phytoplankton.

2.4.1 Aqueous chemistry of heme and hemoproteins

The aqueous chemistries of hemes are strongly influenced by their tetrapyrroles structure (Figure 2.5). The porphyrin ring makes the complexes inherently insoluble in water (80), however this insolubility is tempered by the presence of different substituted side chains. Consequently, the long hydrophobic side chain of heme *a* decreases this compound's solubility in water at neutral pH, while the presence of the multiple carboxylic acid side chains increases the solubility of siroheme. However, when bound to a protein, the solubility of hemes is controlled by the protein structure, so that heme *b* proteins such as peroxidases, catalases and cytochrome b_5 are readily dissolved in water, while cytochromes such as b_6 and b_{559}, which are membrane associated, are much less soluble. Unmetallated porphyrins are unlikely to form complexes with iron(III) in seawater because incorporation of iron(III) into porphyrin rings is generally not favored as a result of steric hindrance (80). The most likely mechanism for early reports of iron(III) complexation by porphyrins in seawater (81, 82) is perhaps via prior reduction of iron(III) to iron(II) (66). However even incorporation of iron(II) is kinetically slow, with reported yield of only 10% heme after one hour in ideal (reducing) conditions (83), and added porphyrins have been found to have little effect on the solubility of Fe(III) in seawater (84). Once formed, iron porphyrins are relatively stable complexes, and the most common processes for removing iron require the breakdown of the porphyrin ring via oxidation (85). Consequently, an equilibrium stability constant determined under classical conditions has yet to be reported for heme or any other iron porphyrin (to the

best of our knowledge). Hemes are rapidly oxidized in aqueous solution at pH 8, and the coupled oxidation of hemes results in breakdown of the methene bridges between the porphyrin rings producing biliverdins (86). The process appears to involve coordination between oxygen and the unoccupied iron ligand sites, so that oxidation is slowed by the presence of ions that can compete with or shield iron from oxygen (86, 87). In contrast, iron(III) porphyrins are chemically less reactive, although they form insoluble μ-oxo bridged dimers (80, 88). Iron porphyrins are known to have a rich photochemistry due to their characteristic ring structure of conjugated double bonds (89), and have been commonly employed in chemical studies as photosensitizers for their ability to absorb light and transfer energy to desired reactants (90). In seawater, ferrous complexes of protoporphyrin IX have been hypothesized to act as photosensitizing producers of superoxide, increasing the dark production of iron(II) from iron(III) following irradiation (66).

To date, only one study has reported concentrations of iron porphyrins in natural waters (91). Nanomolar equivalents of iron(III) protophorphyrin IX like compounds, which could include hemes and hemoproteins, were detected in estuarine waters. The aqueous chemistry of hemes outlined above suggests that, if released intact from cells in the reduced form into oxygenated seawater at a pH of around 8, oxidation and dimerization processes will dominate the marine chemistry of heme and concentrations of heme in solution would be negligible. However, it is possible that more heavily substituted hemes such as siroheme and soluble heme proteins like peroxidase, catalase and soluble cytochrome c may survive cell lysis and remain in solution, potentially contributing to the dissolved iron pool. The limited investigations into hemes in natural

waters coupled with the lack of knowledge of the intracellular abundance of siroheme and the soluble hemoproteins in marine microbes means that the significance of a dissolved iron porphyrin pool in seawater is still largely unknown.

2.4.2 Methods for the chemical analysis of heme in particulate marine samples

The most common and efficient method for extracting heme from plants and other biological materials is via acidified acetone (92). Heme has traditionally been quantified using the pyridine hemochrome method using the strong secondary adsorption bands resulting from coordination of pyridine to the fifth and sixth ligand binding sites of the reduced iron porphyrin (93). However, this method requires dilution or manipulation of samples to bring them into aqueous solution (92), which increases both the risks of sample degradation and the quantity of starting material required for analysis. Furthermore, direct spectrophotometric determination of heme in marine samples is likely to be problematic as result of the presence of many potential interfering algal pigments. Gledhill (2007) reported an extraction technique for marine samples that utilizes detergents and ammonia rather than acidified acetone (94). Gledhill (2007) was also able to successfully separate extracts by high performance chromatography and then detect heme *b* spectrophotometrically. Ammoniacal detergent standards of iron(III) protoporphyrin IX are stable for at least six months if kept in the dark at 4°C, but heme in extracted samples is not so stable and should be kept at 4°C and analyzed within 24 hours (94). The loss of heme from extracts on prolonged storage possibly results from degradation via oxidation. To date losses of heme in samples have been successfully minimized by reducing the time and temperature between extraction and analysis (95,

96). Analysis of phytoplankton cultures indicates that the relative standard deviation in heme concentrations between experimental triplicates is typically of the order of 25 ± 20 % (n = 21) (95). The detergent extraction technique does not completely liberate heme from all proteins (95, 96). Therefore, the extraction technique is operational rather than fully quantitative, and expressed concentrations of heme *b* are thus underestimates of total cellular heme *b* content. Comparison of acid acetone extraction with ammoniacal detergent extraction suggests that approximately 80 % of heme *b* is extracted using the latter method from two species of marine phytoplankton (95) and recovery of a major part of the heme *b* fraction present in phytoplankton appears to be consistent with comparisons with the cellular iron concentrations.

The determination of heme *b* is not contamination prone as is the determination of particulate iron and sampling is carried out using the same filtration equipment as for chlorophyll *a* with no specialist trace metal free sampling equipment required. Samples can therefore be obtained on cruises where such facilities are unavailable (95), potentially shedding light on an important fraction of the iron pool over a broader temporal and spatial scale than is currently possible. Sensitive chemiluminescence and fluorescence techniques for the detection of heme *b* in plants and algae have also been described (91, 97, 98) and these offer the potential for ship-board analysis or even sensor development, although, as with spectrophotometry, there may be issues relating to the specificity of the analysis (91). Recent work in Gledhill's laboratory has applied electrospray ionization-mass spectrometry (ESI-MS) to the detection of heme *b* after separation by HPLC (99). The use of characteristic collision induced fragmentation patterns resulted in a highly specific detection method that confirmed the identity of heme b in marine particulate

samples, supporting the previously published data. Detection by mass spectrometry was found to be more sensitive and overcame the potential for interferences that can be an issue in the spectrophotometric determination of heme b (94). Furthermore, comparison of heme b concentrations determined by the two methods indicated good agreement between spectrophotometric results and those obtained by ESI-MS, in the absence of interferences (99). A selective mass spectrometric detection technique will clearly be useful for comparison with any future developments of ship-board techniques. However further analytical work is clearly required on optimizing suitable extraction protocols and the determination of hemes other than heme *b*. The tendency for heme in extracts to degrade has so far frustrated the development of such a protocol as many rigorous extraction and digestion techniques require incubation at temperatures higher than 4°C. Further efforts in this regard are still ongoing, in particular with respect to understanding the exact mechanism of heme degradation. A robust total extraction method would enhance our understanding of the overall significance of this iron pool, and possibly lead, amongst other things, to a useful biogenic iron proxy. A further strategy for investigation of hemes in marine organisms is through determination of individual proteins via proteomics (100). Such approaches have been used to gain detailed insight into the metabolic responses of marine organisms to particular environmental stresses (101, 102). However, analysis of proteomes is not trivial (103) and such approaches have yet to be applied to the open ocean environment.

2.4.3 Distributions of particulate heme in the marine environment

The usefulness of heme as a prosthetic group in proteins has made it an abundant component of the biogenic iron pool and in theory, the total heme pool in phytoplankton could be of the order of 40 % of the total "active" iron pool (94). Picomolar concentrations of heme *b* in particulate material have been reported for several regions in the Atlantic Ocean, and the distributions of heme b were found to be oceanographically consistent (95, 96). Reported concentrations are considerably lower than particulate iron concentrations reported for the same regions (Table 1.1). However, this likely results from the non-biogenic origin of much of the particulate iron observed in the ocean (104–106). Biogenic iron concentrations in the ocean are difficult to estimate, although progress has been made with the introduction of washing techniques designed to eliminate non-biogenic particulate iron (107–110), and in the determination of the abundance of iron in individual cells (111–114). Heme *b* has been found to make up between 6 and 26 % of the total cell associated iron for phytoplankton in nutrient and light replete laboratory conditions. Further, extrapolating from dissolved iron content and heme b concentrations at stationary phase for iron limited cultures (0.5 nmol L^{-1} - similar to oceanic dissolved iron concentrations) resulted in 14 – 26% of the total iron inventory being incorporated as heme b (95). This suggests that concentrations of heme *b* determined using ammoniacal detergent extraction represent a significant component of the biogenic iron pool in the marine environment, although the use of heme *b* as a proxy for biogenic iron is limited by the wide range of heme *b*: particulate iron values. The range of heme b: particulate iron values observed between species and growth conditions could arise from variability in the proportion of iron allocated to hemoproteins between individual species. For example, theoretical calculations indicate that the heme b content

of the electron transport chain can vary between 14 and 23 % of the total iron content (94). However, further uncertainty is introduced as a result of the operational nature of the extraction protocol, as interspecies variability in extraction efficiencies has received only limited attention to date (95). Nevertheless, the consistent oceanographic trends observed for heme b concentrations coupled with the broad compatibility observed with biogenic iron concentrations in the Southern Ocean (Table 1) suggest that any errors are likely to be systematic rather than random, increasing the potential interpretative power of the analysis. Thus, while caution should be exercised when interpreting heme b distributions due to uncertainties discussed above and the operational nature of the analysis, determination of heme b has the potential to provide valuable information on iron biogeochemistry and its impacts on microbes in the ocean.

The primary control on heme *b* distributions in the ocean appears to be biomass (95, 96). Heme *b* to particulate organic carbon (POC) ratios in particulate material reported to date vary by an order of magnitude 0.06 - 1.0 μmol mol^{-1}, with the lowest values observed in the iron limited regions of the high latitude North Atlantic (HLNA). However, this variability is not due solely to the intracellular heme b contents of marine phytoplankton because bulk POC contains carbon from heterotrophs as well as detritus. Indeed, comparison of carbon content from POC and that estimated from phytoplankton cell counts indicates that heterotrophic and detrital carbon contribute more to total POC in the HLNA (115) than in other lower latitude areas (116). This disparity in POC source composition may partially explain low heme *b*:POC ratios in the HLNA, although low heme *b*:POC ratios obtained exclusively from phytoplankton biomass in the HLNA suggest that POC source composition cannot account for all of the variability in heme

b:POC ratios (96). Thus, variability in heme *b*:POC observed within biogeographically similar regions may reflect variability in overall nutrient concentrations, as seen in phytoplankton laboratory cultures (95, 96). Still, further data points that also account for phytoplankton community composition are needed to support such a conclusion. Ratios of heme *b* to phytoplankton carbon were reported to vary between 0.07 and 0.78 μmol mol^{-1} in the Celtic Sea and the high latitude North Atlantic (95, 96). Interestingly, the range of these ratios appears to be consistent with the lower end of total Fe:C quotas reported for some field data (114, 117, 118).

Although heme *b* distributions are primarily controlled by biomass, overall correlations with chlorophyll are quite poor, especially considering that the two compounds are both strongly associated with photosynthesis. Thus, heme *b* concentrations do not appear to increase with depth to the same extent as chlorophyll *a* concentrations (95, 96). The differences in the abundance of heme *b* and chlorophyll *a* may relate to the way in which the tetrapyrrole biosynthetic pathway is regulated (53, 56, 119, 120). Low heme content has been shown to increase the production of tetrapyrroles (121) while magnesium chelatase has a higher affinity for protoporphyrin IX than ferrochelatase (122). Consequently, low heme content could potentially increase protoporphyrin IX production, while lack of iron would inhibit ferrochelatase activity, leading to a higher production of chlorophyll relative to heme. Unfortunately, published studies have not been able to differentiate between intracellular heme pools. Different hemoproteins have different lability towards the extraction method (96), and thus, particularly labile intracellular heme pools may be preferentially extracted over others. Decreases in the total cellular abundance of heme *b* and chlorophyll *a* would, therefore,

disproportionately be influenced by reductions in abundance of specific hemoproteins with particularly labile heme *b* prosthetic groups. Further research and optimization in this area have the potential to improve our knowledge of intracellular heme b resource allocation under various nutrient limiting conditions.

2.5 Conclusions and future directions

Currently, one of the grand challenges in chemical oceanography is integrating and reconciling measurements at the molecular/mechanistic level with large scale bulk chemical measurements taken in the field. A comprehensive perspective of the marine iron biogeochemical cycle should ideally integrate geochemical, biochemical, physiological, and genomic/transcriptomic/proteomic information. In this review we have presented the current state of heme biogeochemistry science and what is known or can be inferred about the physiological functions of heme in marine phytoplankton and bacterioplankton. Numerous challenges still remain in the field.

At the cellular scale, we need a better understanding of the allocation of hemes between intracellular pools related to photosynthesis, respiration, nutrient acquisition, ROS management, and cellular signaling and how those allocations may or may not vary with respect to ecological conditions. Proteomics approaches are likely the most promising techniques in this area (100). Although some progress has been made in marine heterotrophic bacteria, we need a better understanding of direct heme uptake including the genes involved, their regulation, and mechanisms. The same inroads need to be made looking at heme bioavailability in marine phytoplankton. In the analytical realm, improved extraction techniques for heme (ideally a universal extraction) are

required from marine material as well as advances in separation and detection methodologies. Chemiluminescence and fluorescence techniques are promising technologies in this area as well as the potential for utilizing genetically modified marine organisms as bioreporters (123). In the field, more data is needed to tease out trends in heme *b* concentrations with respect to biogenic iron and other bulk biogenic properties. The chemical research into heme abundance in marine phytoplankton currently suggests that heme *b* makes up a significant component of the particulate biogenic iron pool, and that heme *b* is depleted relative to chlorophyll *a* and POC in nutrient, particularly iron, limited regions of the ocean. However, further ocean basin-scale measurements will provide new insights into heme biogeochemistry, and the development of simple, field-ready measurement techniques will expedite this process.

2.6 Methods

Marine microbial genomes with the habitat metadata tag "Marine" were searched using the Joint Genome Institute Integrated Microbial Genomes (IMG) database (74). Heme biosynthesis pathways were considered present if genomes contained the components specific to heme biosynthesis in the KEGG metabolic pathway "Porphyrin and chlorophyll metabolism" map00860 (124). PFAM (125) protein families mentioned in the text were searched using the IMG function search. Marine bacterial genomes were considered to have a full heme uptake operon based on the co-localization (within a 10 gene neighborhood) of genes coding for a ChuS or HutX protein, a heme outer membrane TBDT, and the three heme ABC IM transporter components (periplasmic binding protein, permease, ATPase). See Figure 2.3D for an example. The marine bacterial 16S

rRNA phylogeny was constructed from IMG marine genomes that had a corresponding 16S rRNA sequence available in the Greengenes database (126). These 16S rRNA genes were extracted from the total Greengenes alignment and used with RAxML (127) v8.0.0 to generate a maximum likelihood tree under the gamma distribution, using the general time-reversible model for DNA evolution and utilizing 500 random sampling bootstraps.

2.7 Acknowledgements

Chapter 2, in full, is a reprint of the material as it appears in Hogle SL, Barbeau KA, Gledhill MA. (2014). Heme in the marine environment: from cells to the iron cycle. *Metallomics* 6:1107-1120. doi: 10.1039/C4MT00031E. The dissertation author was the primary investigator and first author of this paper. This work was funded by NSF GRFP grant DGE-144086 to S.L.H. and NSF grant OCE-1061068 to K.A.B.

Table 2.1 Biogenic iron and heme *b* in the Southern Ocean. Range of values reported for biogenic iron and heme *b* in the Southern Ocean

Region	Biogenic iron (pM)	Heme *b* (pM)	Size fraction	Method/reference
SE New Zealand (46.24 S, 178.72 E)	40 - 310		> 0.2 µm	Oxalate wash* (104)
S Australia (46-60 S, 139-140 E)	100 - 380		> 0.2 µm	Oxalate wash# (108)
Scotia Sea (52-60 S, 38-45 E)		0.6 - 21	> 0.7 µm	Heme *b* direct determination (96)

*only samples where unwashed and washed fractions were obtained are used
estimated from Figure 4 of reference 108

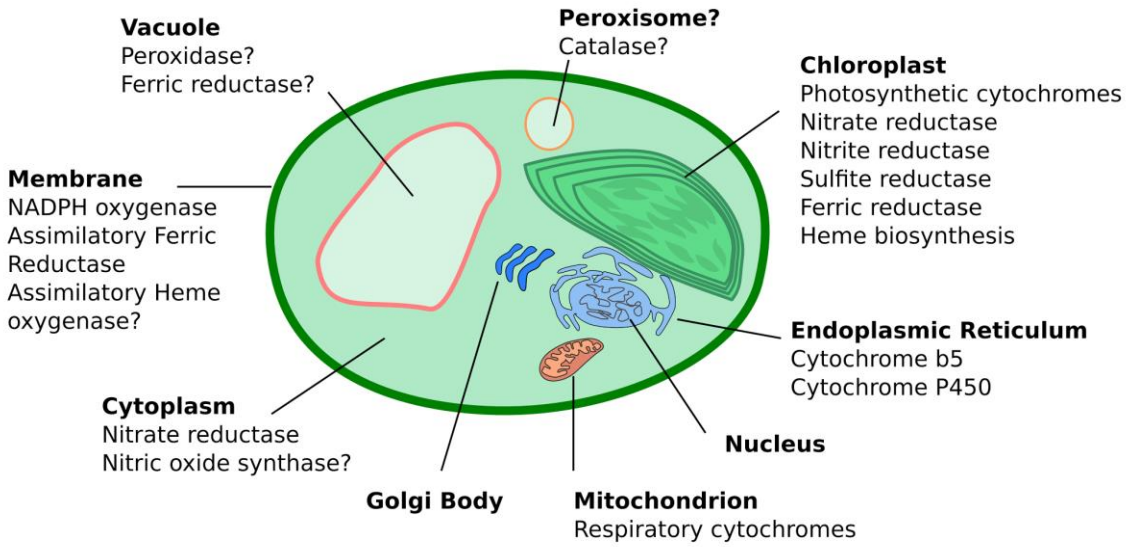


Figure 2.1 Schematic showing the distribution of the major proteins containing heme-like cofactors in a hypothetical diatom. Question marks indicate hypothetical/unconfirmed presence in the cell. The major cellular organelles are listed in boldface and the proteins with heme or siroheme cofactors (and the heme biosynthesis pathway) are listed beneath them. In bacteria these proteins may be either in the cytosol, periplasm, or inner and outer membranes. In cyanobacteria, photosynthetic and respiratory cytochromes reside in the thylakoid membrane.

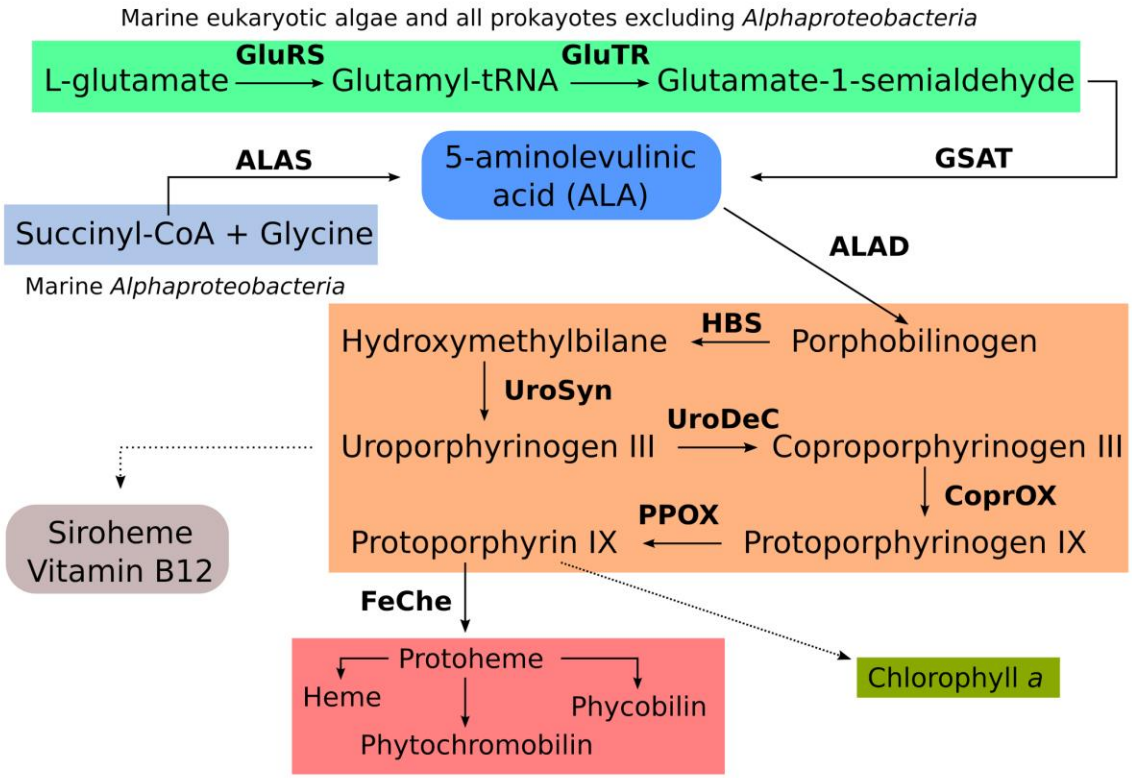


Figure 2.2 Biosynthetic pathway for heme and other tetrapyrroles in marine eukaryotic algae and bacteria. 5-aminolevulinic acid (ALA) is the universal precursor in all organisms. Marine Alphaproteobacteria use Glycine as a starting product while all other marine bacteria and eukaryotic algae use L-glutamate. Protoporphyrin IX is synthesized from ALA and then can be further modified into heme, phytochromobilin, phycobilins or Chlorophyll a. Siroheme, a heme-like cofactor, and Vitamin B12 are synthesized from Uroporphyrinogen III. Dashed lines indicate multiple enzymatic steps. ALAD, ALA dehydratase: ALAS, ALA synthase, CoprOX, coproporphyrinogen oxidase; FeChe, ferrochelatase; GluRS, glutamyl-tRNA synthetase; GluTR, glutamyl-tRNA reductase; GSAT, glutamate-1-semialdehyde aminotransferase; HBS, hydroxymethylbilane synthase; PPOX, protoporphyrinogen IX oxidase; UroDeC, uroporphyrinogen decarboxylase; UroSyn, uroporphyrinogen III synthase.

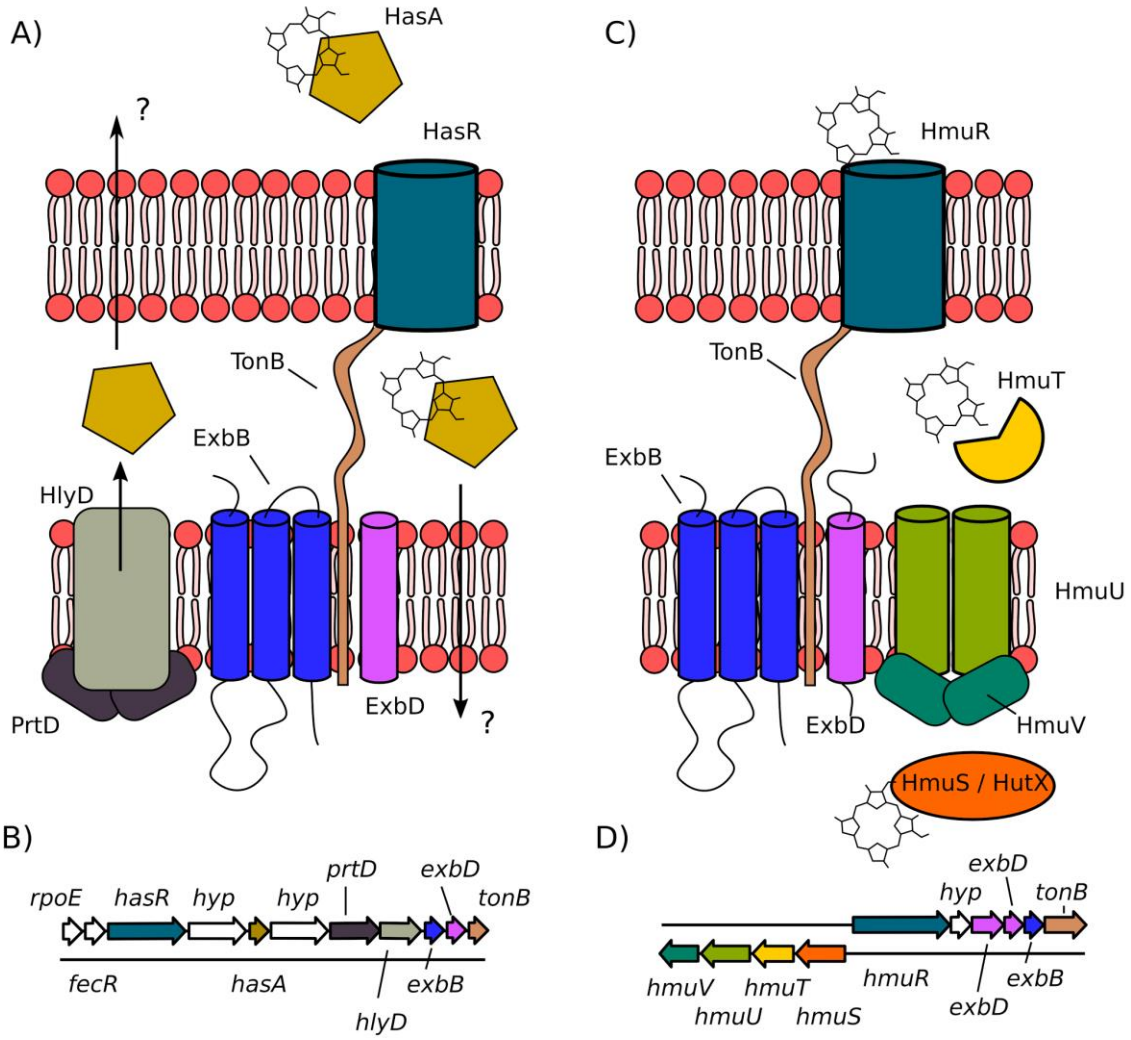


Figure 2.3 Representative schematic of encoding genes and protein machinery utilized in hemophore mediated heme uptake and direct heme uptake. Organization of putative hemophore uptake operon B) in *Pseudovibrio* sp. JE062 and direct heme uptake operon in *Ruegeria* sp. TM1040. Gene symbols are colored as the corresponding proteins in A) and C). ExbB, TBDT energy transduction component; ExbD, TBDT energy transduction component; *fecR*, Iron senstive regulatory element; HasA, Hemophore; HasR, TBDT - heme/hemophore; HlyD, HasA secretion protein; HmuR, TBDT - heme; HmuS, Heme utilization protein – ChuS family; HmuT, Periplasmic heme binding protein; HmuU, IM-spanning ABC permease; HmuV, ABC ATPase; hyp, Hypothetical; HrtD, HasA secretion protein – ATPase; *rpoE*, Iron senstive regulatory element; TonB, TBDT energy transduction component. Question marks indicate hypothetical/unconfirmed pathways for export/import.

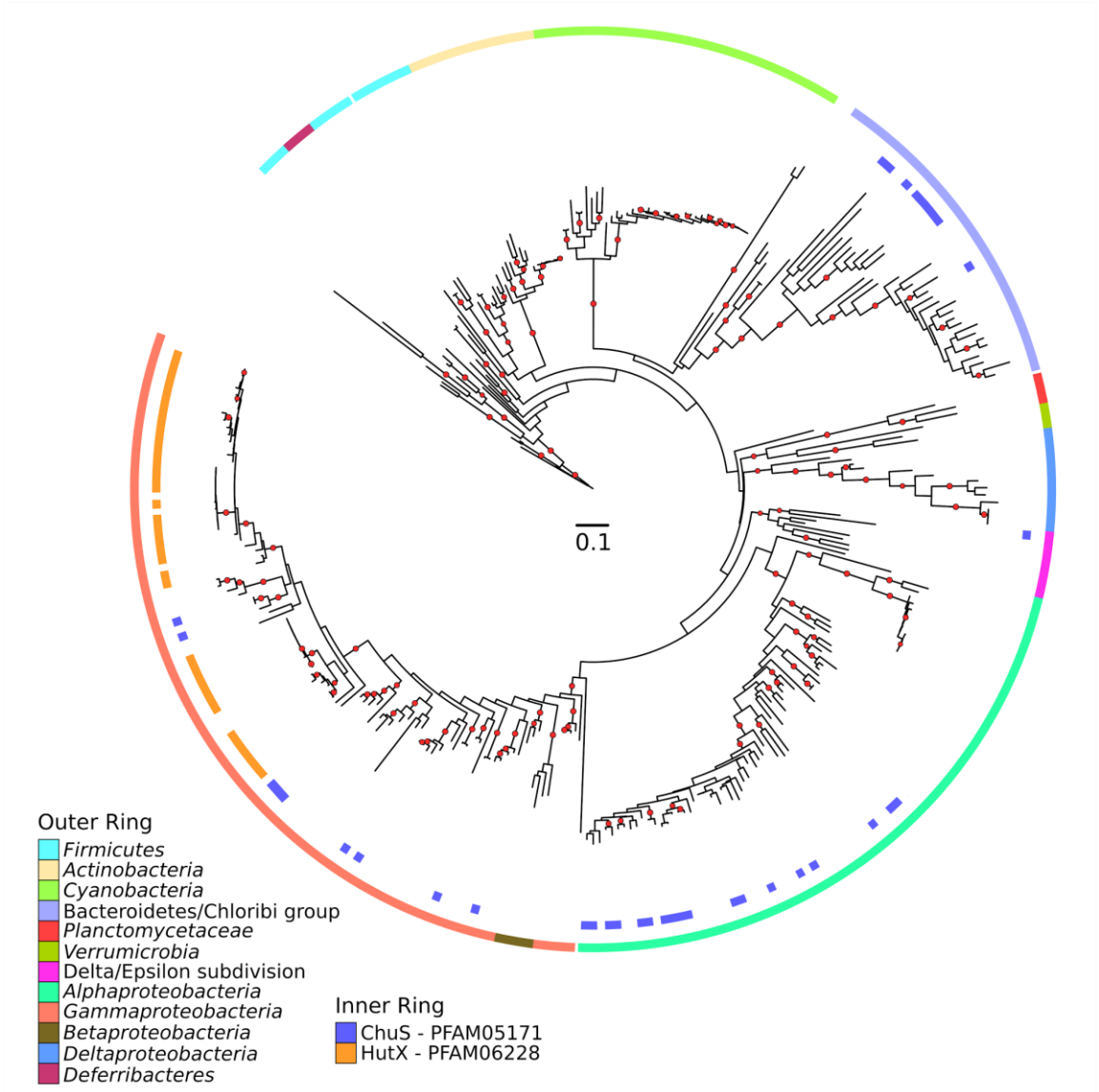
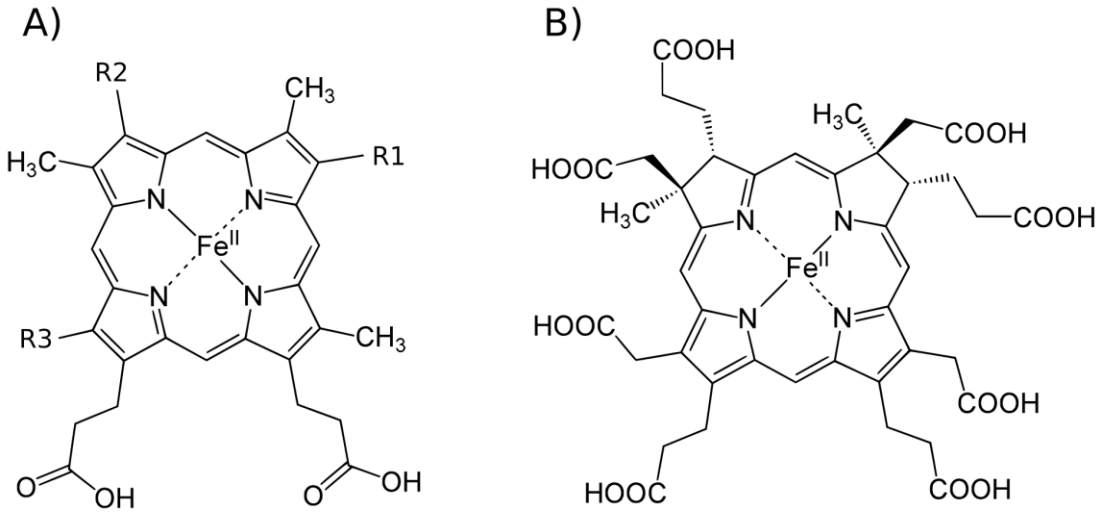


Figure 2.4 Marine bacterial genomes containing heme uptake genes. An unrooted 16S rRNA maximum likelihood phylogeny of marine bacterial genomes indexed in the Integrated Microbial Genomes (IMG) database. Bootstrap values of >80% are shown as red symbols for 500 resamplings. Scale bar equals 0.1 substitution per site. The outer ring is colored by bacterial taxonomic grouping (Phylum or lower). The inner ring represents genomes that contain a full direct heme uptake locus and is colored by the family of the heme utilization protein present in each locus. See methods for further description.



Heme	R1	R2	R3
a	-CH=CH_2	-HFAR	-CHO
b	-CH=CH_2	-CH=CH_2	-CH_3
c	-CH(C-S-yl)CH_3	-CH(C-S-yl)CH_3	-CH_3

HFAR - hydroxyethylfarnesyl group
C-S-yl - Thioether linkage to apoprotein

Figure 2.5 Structures of hemes *a*-*c* and siroheme. Structures of A) the hydrophobic hemes *a*-*c* and B), the more soluble siroheme

2.8 References

1. Chapman S, Daff S, Munro A (1997) Heme: The most versatile redox centre in biology? *Metal Sites in Proteins and Models: Structure and Bonding*, eds Sadler PJ, Thomson AJ (Berlin Heidelberg), pp 39–70. Volume 88.

2. Richter O, Ludwig B (2009) Electron transfer and energy transduction in the terminal part of the respiratory chain - lessons from bacterial model systems. *Biochim Biophys Acta* 1787(6):626–634.

3. Richardson DJ (2000) Bacterial respiration: a flexible process for a changing environment. *Microbiology* 146:551–571.

4. Raven J, Beardall J (2005) Respiration in aquatic photolithotrophs. *Respiration in Aquatic Ecosystems*, eds Giorgio PAD, Williams P (Oxford University Press), pp 36–46.

5. Hartshorne RS, Reardon CL, Ross D, Nuester J, Clarke T, Gates AJ, Mills PC, Fredrickson JK, Zachara JM, Shi L, Beliaev AS, Marshall MJ, Tien M, Brantley S, Butt JN, Richardson DJ. (2009) Characterization of an electron conduit between bacteria and the extracellular environment. *Proc Natl Acad Sci U S A* 106(52):22169–22174.

6. Myers CR, Myers JM (1997) Cloning and sequence of cymA, a gene encoding a tetraheme cytochrome c required for reduction of iron (III), fumarate, and nitrate by Shewanella putrefaciens MR-1. *J Bacteriol* 179(4):1143.

7. Richter K, Schicklberger M, Gescher J (2012) Dissimilatory reduction of extracellular electron acceptors in anaerobic respiration. *Appl Environ Microbiol* 78(4):913–921.

8. Hohmann-Marriott MF, Blankenship RE (2011) Evolution of photosynthesis. *Annu Rev Plant Biol* 62:515–548.

9. Guskov A, Kern J, Gabdulkhakov A, Broser M, Zouni A, Saenger W. (2009) Cyanobacterial photosystem II at 2.9-A resolution and the role of quinones, lipids, channels and chloride. *Nat Struct Mol Biol* 16(3):334–342.

10. Kurisu G, Zhang H, Smith JL, Cramer WA (2003) Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science* 302(5647):1009–1014.

11. Howe CJ, Schlarb-Ridley BG, Wastl J, Purton S, Bendall DS (2006) The novel cytochrome c6 of chloroplasts: a case of evolutionary bricolage? *J Exp Bot* 57(1):13–22.

12. Lesser MP (2006) Oxidative stress in marine environments: biochemistry and physiological ecology. *Annu Rev Physiol* 68:253–278.

13. Rose AL, Webb E a., Waite TD, Moffett JW (2008) Measurement and implications of nonphotochemically generated superoxide in the equatorial Pacific Ocean. *Environ Sci Technol* 42(7):2387–2393.

14. Learman DR, Voelker BM, Vazquez-Rodriguez a. I, Hansel CM (2011) Formation of manganese oxides by bacterially generated superoxide. *Nat Geosci* 4(2):95–98.

15. Diaz JM, Hansel CM, Voelker BM, Mendes CM, Andeer PF, Zhang T. (2013) Widespread production of extracellular superoxide by heterotrophic bacteria. *Science* 340(6137):1223–1226.

16. Kustka AB, Shaked Y, Milligan AJ, King DW, Morel FMM (2005) Extracellular production of superoxide by marine diatoms: Contrasting effects on iron redox chemistry and bioavailability. *Limnol Oceanogr* 50(4):1172–1180.

17. Anderson A, Bothwell JH, Laohavisit A, Smith AG, Davies JM (2011) NOX or not? Evidence for algal NADPH oxidases. *Trends Plant Sci* 16(11):579–581.

18. Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F. (2011) ROS signaling: the new wave? *Trends Plant Sci* 16(6):300–309.

19. Zámocký M, Furtmüller P, Obinger C (2008) Evolution of catalases from bacteria to humans. *Antioxid Redox Signal* 10(9):1527–1548.

20. Zámocký M, Furtmüller PG, Obinger C (2010) Evolution of structure and function of Class I peroxidases. *Arch Biochem Biophys* 500(1):45–57.

21. Bernroitner M, Zamocky M, Furtmüller PG, Peschek G a., Obinger C (2009) Occurrence, phylogeny, structure, and function of catalases and peroxidases in cyanobacteria. *J Exp Bot* 60(2):423–440.

22. Alfonso-Prieto M, Biarnés X, Vidossich P, Rovira C (2009) The molecular mechanism of the catalase reaction. *J Am Chem Soc* 131(33):11751–11761.

23. Smulevich G, Jakopitsch C, Droghetti E, Obinger C (2006) Probing the structure and bifunctionality of catalase-peroxidase (KatG). *J Inorg Biochem* 100(4):568–585.

24. Hu J, Baker A, Bartel B, Linka N, Mullen RT, Reumann S, Zolman BK. (2012) Plant peroxisomes: biogenesis and function. *Plant Cell* 24(6):2279–2303.

25. Gabaldón T (2010) Peroxisome diversity and evolution. *Philos Trans R Soc Lond B Biol Sci* 365(1541):765–773.

26. Armbrust EV, Berges J, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski M, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kröger N, Lau WW, Lane TW, Larimer

FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson T, Saito M, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS. (2004) The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science* 306(5693):79–86.

27. Zipor G, Oren-Shamir M (2013) Do vacuolar peroxidases act as plant caretakers? *Plant Sci* 199-200:41–47.

28. Schenkman JB, Jansson I (2003) The many roles of cytochrome b5. *Pharmacol Ther* 97(2):139–152.

29. Fischer K, Barbier GG, Hecht HJ (2005) Structural basis of eukaryotic nitrate reduction: crystal structures of the nitrate reductase active site. *Plant Cell* 17(April):1167–1179.

30. Gao Y, Smith GJ, Alberte RS (1993) Nitrate Reductase from the Marine Diatom Skeletonema costatum (Biochemical and Immunological Characterization). *Plant Physiol* 103(4):1437–1445.

31. Allen AE, Ward BB, Song B (2005) Characterization of Diatom (Bacillariophyceae) Nitrate Reductase Genes and Their Detection in Marine Phytoplankton Communities. *J Phycol* 41(1):95–104.

32. Bowler C, Vardi A, Allen AE (2010) Oceanographic and Biogeochemical Insights from Diatom Genomes. *Ann Rev Mar Sci* 2(1):333–365.

33. Nakano S, Takahashi M, Sakamoto A, Morikawa H, Katayanagi K (2012) Structure-function relationship of assimilatory nitrite reductases from the leaf and root of tobacco based on high-resolution structures. *Protein Sci* 21(3):383–395.

34. Marchetti A, Schruth DM, Durkin C, Parker MS, Kodner RB, Berthiaume CT, Morales R, Allen AE, Armbrust EV. (2012) Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc Natl Acad Sci U S A* 109(6):E317–25.

35. Takahashi H, Kopriva S, Giordano M, Saito K, Hell R (2011) Sulfur assimilation in photosynthetic organisms: molecular functions and regulations of transporters and assimilatory enzymes. *Annu Rev Plant Biol* 62:157–184.

36. Blaby-Haas CE, Merchant SS (2012) The ins and outs of algal metal transport. *Biochim Biophys Acta* 1823(9):1531–1552.

37. Kustka AB, Allen AE, Morel FMM (2007) Sequence Analysis and Transcriptional Regulation of Iron Acquisition Genes in Two Marine Diatoms. *J Phycol* 43(4):715–729.

38. Liebl U, Lambry J-C, Vos MH (2013) Primary processes in heme-based sensor

proteins. *Biochim Biophys Acta* 1834(9):1684–1692.

39. Wan X, Tuckerman JR, Saito J, Freitas TA, Newhouse JS, Denery JR, Galperin MY, Gonzalez G, Gilles-Gonzalez MA, Alam M. (2009) Globins synthesize the second messenger bis-(3’-5')-cyclic diguanosine monophosphate in bacteria. *J Mol Biol* 388(2):262–270.

40. Gilles-Gonzalez MA, Ditta GS, Helinski DR (1991) A haemoprotein with kinase activity encoded by the oxygen sensor of Rhizobium meliloti. *Nature* 350(14):170–172.

41. Gilles-Gonzalez M-A, Gonzalez G (2005) Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. *J Inorg Biochem* 99(1):1–22.

42. Plate L, Marletta M a. (2013) Nitric oxide-sensing H-NOX proteins govern bacterial communal behavior. *Trends Biochem Sci* 38(11):566–575.

43. Mur L, Mandon J, Persijn S, Cristescu SM, Moshkov IE, Novikova GV, Hall M, Harren FJ, Hebelstrup KH, Gupta KJ. (2013) Nitric oxide in plants: an assessment of the current state of knowledge. *AoB Plants* 5:ls052.

44. Plate L, Marletta M a. (2012) Nitric oxide modulates bacterial biofilm formation through a multicomponent cyclic-di-GMP signaling network. *Mol Cell* 46(4):449–460.

45. Davidson SK, Koropatnick T a., Kossmehl R, Sycuro L, McFall-Ngai MJ (2004) NO means “yes” in the squid-vibrio symbiosis: nitric oxide (NO) during the initial stages of a beneficial association. *Cell Microbiol* 6(12):1139–1151.

46. Besson-Bard A, Pugin A, Wendehenne D (2008) New insights into nitric oxide signaling in plants. *Annu Rev Plant Biol* 59:21–39.

47. Foresi N, Correa-Aragunde N, Parisi G, Caló G, Salerno G, Lamattina L. (2010) Characterization of a nitric oxide synthase from the plant kingdom: NO generation from the green alga Ostreococcus tauri is light irradiance and growth phase dependent. *Plant Cell* 22(11):3816–3830.

48. Vardi A, Bidle KD, Kwityn C, Hirsh DJ, Thompson SM, Callow J, Falkowski P, Bowler C. (2008) A diatom gene regulating nitric-oxide signaling and susceptibility to diatom-derived aldehydes. *Curr Biol* 18(12):895–899.

49. Oborník M, Green BR (2005) Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol* 22(12):2343–2353.

50. Panek H, O’Brian MR (2002) A whole genome view of prokaryotic haem biosynthesis. *Microbiology* 148:2273–2282.

51. Tong Y, Guo M (2009) Bacterial heme-transport proteins and their heme-coordination modes. *Arch Biochem Biophys* 481(1):1–15.

52. Anzaldi LL, Skaar EP (2010) Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. *Infect Immun* 78(12):4977–4989.

53. Tanaka R, Tanaka A (2007) Tetrapyrrole biosynthesis in higher plants. *Annu Rev Plant Biol* 58:321–346.

54. Czarnecki O, Grimm B (2012) Post-translational control of tetrapyrrole biosynthesis in plants, algae, and cyanobacteria. *J Exp Bot* 63(4):1675–1687.

55. Rhie G, Beale SI (1992) Biosynthesis of phycobilins. *J Biol Chem* 267(23):16088–16093.

56. Cornah JE, Terry MJ, Smith AG (2003) Green or red: what stops the traffic in the tetrapyrrole pathway? *Trends Plant Sci* 8(5):224–230.

57. Suzuki T, Masuda T, Singh DP, Tan FC, Tsuchiya T, Shimada H, Ohta H, Smith AG, Takamiya KI. (2002) Two types of ferrochelatase in photosynthetic and nonphotosynthetic tissues of cucumber: their difference in phylogeny, gene expression, and localization. *J Biol Chem* 277(7):4731–4737.

58. Watanabe N, Che FS, Iwano M, Takayama S, Yoshida S, Isogai A. (2001) Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and chloroplasts by alternative use of two in-frame initiation codons. *J Biol Chem* 276(23):20474–20481.

59. Morrissey J, Bowler C (2012) Iron utilization in marine cyanobacteria and eukaryotic algae. *Front Microbiol* 3(March):43.

60. Hutchins DA, Witter AE, Butler A, Luther GW (1999) Competition among marine phytoplankton for different chelated iron species. *Nature* 400(6747):858–861.

61. Weaver R, Kirchman D, Hutchins D (2003) Utilization of iron/organic ligand complexes by marine bacterioplankton. *Aquat Microb Ecol* 31:227–239.

62. Achilles KM, Church TM, Wilhelm SW, Luther GWI, Hutchins D a. (2003) Bioavailability of iron to Trichodesmium colonies in the western subtropical Atlantic Ocean. *Limnol Oceanogr* 48(6):2250–2255.

63. Hassler CS, Schoemann V (2009) Bioavailability of organically bound Fe to model phytoplankton of the Southern Ocean. *Biogeosciences* 6(10):2281–2296.

64. Wells ML, Trick CG, Cochlan WP, Beall B (2009) Persistence of iron limitation in the western subarctic Pacific SEEDS II mesoscale fertilization experiment. *Deep Sea Res Part 2 Top Stud Oceanogr* 56(26):2810–2821.

65. Kondo Y, Takeda S, Nishioka J, Sato M, Saito H, Suzuki K, Furuya K. (2012) Growth stimulation and inhibition of natural phytoplankton communities by model organic ligands in the western subarctic Pacific. *J Oceanogr* 69(1):97–115.

66. Rijkenberg MJ a., Gerringa LJ a., Carolus VE, Velzeboer I, de Baar HJW (2006) Enhancement and inhibition of iron photoreduction by individual ligands in open ocean seawater. *Geochim Cosmochim Acta* 70(11):2790–2805.

67. Hopkinson BM, Roe KL, Barbeau K (2008) Heme uptake by Microscilla marina and evidence for heme uptake systems in the genomes of diverse marine bacteria. *Appl Environ Microbiol* 74(20):6263–6270.

68. Roe KL, Hogle SL, Barbeau K (2013) Utilization of heme as an iron source by marine alphaproteobacteria in the roseobacter clade. *Appl Environ Microbiol* 79(18):5753–5762.

69. Hopkinson B, Barbeau K (2012) Iron transporters in marine prokaryotic genomes and metagenomes. *Environ Microbiol* 14(1):114–128.

70. Frankenberg-Dinkel N (2004) Bacterial heme oxygenases. *Antioxid Redox Signal* 6(5):825–834.

71. Zhu W, Wilks A, Stojiljkovic I (2000) Degradation of Heme in Gram-Negative Bacteria : the Product of the hemO Gene of Neisseriae Is a Heme Oxygenase. *J Bacteriol* 182(23):6783.

72. Suits MD, Pal GP, Nakatsu K, Matte A, Cygler M, Jia Z. (2005) Identification of an Escherichia coli O157:H7 heme oxygenase with tandem functional repeats. *Proc Natl Acad Sci U S A* 102(47):16955–16960.

73. O'Neill MJ, Bhakta MN, Fleming KG, Wilks A (2012) Induced fit on heme binding to the Pseudomonas aeruginosa cytoplasmic protein (PhuS) drives interaction with heme oxygenase (HemO). *Proc Natl Acad Sci U S A* 109(15):5639–5644.

74. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40(Database issue):D115–22.

75. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen J, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcón LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5(3):e77.

76. Toulza E, Tagliabue A, Blain S, Piganeau G (2012) Analysis of the global ocean sampling (GOS) project for trends in iron uptake by surface ocean microbes. *PLoS One* 7(2):e30931.

77. Desai D, Desai F, Laroche J (2012) Factors influencing the diversity of iron uptake systems in aquatic microorganisms. *Front Microbiol* 3(October):362.

78. Allen AE, Laroche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C. (2008) Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc Natl Acad Sci U S A* 105(30):10438–10443.

79. Kröger N, Poulsen N (2008) Diatoms-from cell wall biogenesis to nanotechnology. *Annu Rev Genet* 42:83–107.

80. Hambright P (1975) Dynamic coordination chemistry of metalloporphyrins. *Porphyrins and Metalloporphyrins*, ed Smith K (Elsevier Scientific Pub. Co, Amsterdam, New York), pp 233–271.

81. Witter AE, Hutchins DA, Butler A, Luther GW (2000) Determination of conditional stability constants and kinetic constants for strong model Fe-binding ligands in seawater. *Mar Chem* 69:1–17.

82. Rue EL, Bruland KW (1995) Complexation of iron(III) by natural organic ligands in the Central North Pacific as determined by a new competitive ligand equilibration/adsorptive cathodic stripping voltammetric method. *Mar Chem* 50:117–138.

83. Kassner RJ, Howard W (1973) Heme formation from Fe (II) and porphyrin in the absence of ferrochelatase activity. *Biochim Biophys Acta* 304:294–303.

84. Chen M, Wang W-X, Guo L (2004) Phase partitioning and solubility of iron in natural seawater controlled by dissolved organic matter. *Global Biogeochem Cycles* 18(4):GB4013.

85. Buchler J (1975) Static Coordination Chemistry of Metalloporphyrins. *Porphyrins and Metalloporphyrins*, ed Smith K (Elsevier Scientific Pub. Co, Amsterdam, New York), pp 157–224.

86. Brown S (1976) Stereospecific Haem Cleavage. A Model for the Formation of Bile-Pigment Isomers in vivo and in vitro. *Biochem J* 159(1):23–27.

87. Brown S, King R (1976) The structure of haem in pyridine/water mixtures and its implication in studies of haem catabolism. *Biochem J* 153(2):479–483.

88. Brown S, Dean T, Jones P (1970) Aggregation of Ferrihaems. Dimerization and Protolytic Equilibria of Protoferrihaem and Deutoeroferrihaem in Aqueous Solution. *Biochem J* 117(4):733–739.

89. Horváth O, Huszánk R, Valicsek Z, Lendvay G (2006) Photophysics and photochemistry of kinetically labile, water-soluble porphyrin complexes. *Coord Chem Rev* 250(13-14):1792–1803.

90. Whitten DG (1978) Photochemistry of porphyrins and their metal complexes in solution and organized media. *Res Chem Intermed* 2(2):107–138.

91. Vong L, Laës A, Blain S (2007) Determination of iron-porphyrin-like complexes at nanomolar levels in seawater. *Anal Chim Acta* 588(2):237–244.

92. Espinas N a., Kobayashi K, Takahashi S, Mochizuki N, Masuda T (2012) Evaluation of unbound free heme in plant cells by differential acetone extraction. *Plant Cell Physiol* 53(7):1344–1354.

93. Furhop J, Smith K (1975) Laboratory Methods. *Porphyrins and Metalloporphyrins*, ed Smith K (Elsevier Scientific Pub. Co, Amsterdam, New York), pp 757–862.

94. Gledhill M (2007) The determination of heme b in marine phyto- and bacterioplankton. *Mar Chem* 103(3-4):393–403.

95. Honey D, Gledhill M, Bibby T, Legiret F, Pratt N, Hickman A, Lawson T, Achterberg E. (2013) Heme b in marine phytoplankton and particulate material from the North Atlantic Ocean. *Mar Ecol Prog Ser* 483:1–17.

96. Gledhill M, Achterberg EP, Honey DJ, Nielsdottir MC, Rijkenberg MJ a. (2013) Distributions of particulate Heme b in the Atlantic and Southern Oceans-Implications for electron transport in phytoplankton. *Global Biogeochem Cycles* 27(4):1072–1082.

97. Masuda T, Takahashi S (2006) Chemiluminescent-based method for heme determination by reconstitution with horseradish peroxidase apo-enzyme. *Anal Biochem* 355(2):307–309.

98. Koga S, Yoshihara S, Bando H, Yamasaki K, Higashimoto Y, Noguchi M, Sueda S, Komatsu H, Sakamoto H. (2013) Development of a heme sensor using fluorescently labeled heme oxygenase-1. *Anal Biochem* 433(1):2–9.

99. Gledhill M (2014) The detection of iron protoporphyrin (heme b) in phytoplankton and marine particulate material by electrospray ionisation mass spectrometry – comparison with diode array detection. *Anal Chim Acta* 841:33–43.

100. Saito MA, Bertrand EM, Dutkiewicz S, Bulygin VV, Moran DM, Monteiro FM, Follows MJ, Valois FW, Waterbury JB. (2011) Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph Crocosphaera watsonii. *Proc Natl Acad Sci U S A* 108(6):2184–2189.

101. Bertrand EM, Allen AE, Dupont CL, Norden-Krichmar TM, Bai J, Valas RE, Saito M. (2012) Influence of cobalamin scarcity on diatom molecular physiology and

identification of a cobalamin acquisition protein. *Proc Natl Acad Sci U S A* 109(26):E1762–71.

102. Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska M, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, Laroche J. (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol* 13(7):R66.

103. Slattery M, Ankisetty S, Corrales J, Marsh-Hunkin KE, Gochfeld DJ, Willett KL, Rimoldi JM. (2012) Marine proteomics: a critical assessment of an emerging technology. *J Nat Prod* 75(10):1833–1877.

104. Frew RD, Hutchins D, Nodder S, Sanudo-Wilhelmy S, Tovar-Sanchez A, Leblanc K, Hare CE, Boyd PW. (2006) Particulate iron dynamics during FeCycle in subantarctic waters southeast of New Zealand. *Global Biogeochem Cycles* 20(1). doi:10.1029/2005GB002558.

105. von der Heyden BP, Roychoudhury a. N, Mtshali TN, Tyliszczak T, Myneni SCB (2012) Chemically and geographically distinct solid-phase iron pools in the Southern Ocean. *Science* 338(6111):1199–1201.

106. Slemons L, Paul B, Resing J, Murray JW (2012) Particulate iron, aluminum, and manganese in the Pacific equatorial undercurrent and low latitude western boundary current sources. *Mar Chem* 142-144:54–67.

107. Hudson R, Morel F (1989) Distinguishing between extra-and intracellular iron in marine phytoplankton. *Limnol Oceanogr* 34(6):1113–1120.

108. Tovar-Sanchez A, Sañudo-Wilhelmy S, Garcia-Vargas M, Weaver RS, Popels LC, Hutchins D. (2003) A trace metal clean reagent to remove surface-bound iron from marine phytoplankton. *Mar Chem* 82(1-2):91–99.

109. Tang D, Morel FMM (2006) Distinguishing between cellular and Fe-oxide-associated trace elements in phytoplankton. *Mar Chem* 98(1):18–30.

110. Hassler CS, Schoemann V (2009) Discriminating between intra- and extracellular metals using chemical extractions : an update on the case of iron. *Limnol Oceanogr Methods* 7:479–489.

111. Twining BS, Baines SB, Fisher NS, Landry MR (2004) Cellular iron contents of plankton during the Southern Ocean Iron Experiment (SOFeX). *Deep Sea Res Part I* 51(12):1827–1850.

112. Twining BS, Baines SB, Bozard JB, Vogt S, Walker E, Nelson DM. (2011) Metal quotas of plankton in the equatorial Pacific Ocean. *Deep Sea Res Part 2 Top Stud Oceanogr* 58(3-4):325–341.

113. Nuester J, Vogt S, Newville M, Kustka AB, Twining BS (2012) The unique

biogeochemical signature of the marine diazotroph trichodesmium. *Front Microbiol* 3(April):150.

114. Twining BS, Baines SB (2013) The trace metal composition of marine phytoplankton. *Ann Rev Mar Sci* 5:191–215.

115. Poulton AJ, Charalampopoulou A, Young JR, Tarran G, Lucas MI, Quartly GD. (2010) Coccolithophore dynamics in non-bloom conditions during late summer in the central Iceland Basin (July-August 2007). *Limnol Oceanogr* 55(4):1601–1613.

116. Hickman AE, Holligan PM, Moore CM, Sharples J, Krivtsov V, Palmer MR. (2009) Distribution and chromatic adaptation of phytoplankton within a shelf sea thermocline. *Limnol Oceanogr* 54(2):525–536.

117. King A, Sañudo-Wilhelmy S, Boyd P, Twining B, Wilhelm S, Breene C, Ellwood M, Hutchins D. (2012) A comparison of biogenic iron quotas during a diatom spring bloom using multiple approaches. *Biogeosciences* 9(2):667–687.

118. Hopkinson BM, Seegers B, Hatta M, Measures CI, Greg Mitchell B, Barbeau K. (2013) Planktonic C:Fe ratios and carrying capacity in the southern Drake Passage. *Deep Sea Res Part 2 Top Stud Oceanogr* 90:102–111.

119. Mochizuki N, Tanaka R, Grimm B, Masuda T, Moulin M, Smith AG, Tanaka A, Terry MJ. (2010) The cell biology of tetrapyrroles: a life and death struggle. *Trends Plant Sci* 15(9):488–498.

120. Yin L, Bauer CE (2013) Controlling the delicate balance of tetrapyrrole biosynthesis. *Philos Trans R Soc Lond B Biol Sci* 368(1622):20120262.

121. Thomas J, Weinstein J (1992) Free heme in isolated chloroplasts : an improved method of assay and its physiological importance. *Plant Physiol Biochem* 30(3):285–292.

122. Guo R, Luo M, Weinstein JD (1998) Magnesium-Chelatase from Developing Pea Leaves. Characterization of a Soluble Extract from Chloroplasts and Resolution into Three Required Protein Fractions. *Plant Physiol* 116:605–615.

123. Boyanapalli R, Bullerjahn GS, Pohl C, Croot PL, Boyd PW, McKay RM. (2007) Luminescent whole-cell cyanobacterial bioreporter for measuring Fe availability in diverse marine environments. *Appl Environ Microbiol* 73(3):1019–1024.

124. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42(1):D199–205.

125. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. (2012) The Pfam protein families database. *Nucleic Acids Res*

40(Database issue):D290–301.

126. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6(3):610–618.

127. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.

Chapter 3

Trace metal acquisition by marine heterotrophic bacterioplankton with contrasting trophic strategies

3.1 Abstract

Heterotrophic bacteria in the SAR11 and *Roseobacter* lineages shape the marine carbon, nitrogen, phosphorous, and sulfur cycles, yet do so having adopted divergent ecological strategies. Currently, it is unknown whether these globally significant groups partition into specific niches with respect to micronutrients (e.g. trace metals), and how that may affect marine trace metal cycling. Here we use comparative genomics to identify diverse iron, cobalt, nickel, copper, and zinc uptake capabilities in SAR11 and *Roseobacter* genomes and uncover surprising unevenness within and between lineages. The strongest predictors for the extent of the metal uptake gene content are the total number of transporters per genome, genome size, total metal transporters, and GC content, but numerous exceptions exist in both groups. Taken together our results suggest that SAR11 have strongly minimized their trace metal uptake versatility, with high affinity zinc uptake being a unique exception. The larger *Roseobacter* genomes have greater trace metal uptake versatility on average, but also appear to have a greater plasticity resulting in phylogenetically similar genomes having largely different capabilities. Ultimately, phylogeny is predictive of the diversity and extent of 20 to 33% of all metal uptake systems, suggesting that specialization in metal utilization mostly occurred independently from overall lineage diversification in both SAR11 and *Roseobacter*. We interpret these results as reflecting relatively recent trace metal niche

partitioning in both lineages, suggesting that concentrations and chemical forms of metals in the marine environment are important factors shaping the gene content of marine heterotrophic *Alphaproteobacteria* from the SAR11 and *Roseobacter* lineages.

3.2 Introduction

The bioactive trace metals manganese, iron, cobalt, nickel, copper, and zinc are important enzyme cofactors for microbially mediated processes that drive nutrient cycling in the ocean. Marine phytoplankton and heterotrophic bacterioplankton require these metals for important cellular metabolisms (1) with some closely related species having variable cellular metal requirements and metal-induced physiological responses (2). At larger scales the spatial and temporal distributions of certain trace metals can have profound ecosystem-wide consequences (3). The concentrations of the bioactive trace elements are generally very low in the open ocean due to the isolation of the pelagic ocean from terrestrial inputs, and in some cases limited solubility (e.g. Fe). The chemical speciation of Fe, Cu, and potentially Co in seawater is highly dependent upon each metal's propensity to interact with heterogeneous organic ligands (4–6). Ni and Zn also interact with ligands in seawater (7, 8), but estimates of the percentage complexed are variable. Natural organic ligands in seawater appear to minimally interact with Mn (1). The dilute concentrations of marine trace metal species, their variable redox states, and vast structural diversity likely provide a spectrum of trace metal niches for marine heterotrophic bacteria and phytoplankton.

There is a large variety of currently identified trace metal uptake pathways (9, 10), and here we briefly introduce the major known pathways for metal uptake in gram

negative bacteria (Table 2.1). It should be noted that these pathways have primarily been characterized in copiotrophic, non-marine organisms, many of which are host pathogens. In the context of the marine environment it is probable that transport systems with little to no functional precedent are employed by certain marine bacteria, especially those from unique, taxonomically underrepresented, and unculturable lineages.

Inorganic Fe^{3+} is transported through the bacterial inner membrane by ATP binding cassette transporters (ABCT). ABCTs are transmembrane, ATP dependent transport proteins comprised of a periplasmic substrate-binding protein, a permease, and an ATP binding component. Inorganic Fe^{2+} is transported by four inner membrane transporter families. NRAMP-like proteins facilitate Fe^{2+} and Mn^{2+} transport in some bacteria (11), while the ZIP family can import Fe^{2+}, Mn^{2+}, Zn^{2+}, and Co^{2+} (12). FTR1-like proteins can function as Fe^{2+} transporters (13), while the *feoAB* system is an Fe-specific bacterial permease (14). Many Fe transporters are regulated by the ferric uptake repressor protein (Fur) (15), a transcription factor that utilizes Fe^{2+} as a corepressor. The Fur protein represses transcription by first binding Fe^{2+} and then binding to a conserved 19 base pair inverted repeat called a Fur-box. Iron regulatory motifs, called the Iron-Rhodo-Box, with a different but related palindromic repeat to the Fur box have previously been predicted to be upstream of most iron transporters in 12 different *Roseobacter* genomes (16).

Siderophores and heme/hemoproteins are two major organic Fe forms utilized by bacteria. Hydroxamate and catecholate functional groups are two structural motifs found in siderophores and both chemical classes are biosynthesized in non-ribosomal peptide synthetase (17) (NRPS) or NRPS-independent pathways (18). TonB dependent

transporters (TBDTs) import Fe-bound siderophores across the bacterial outer membrane, while ABCTs move siderophores through the inner membrane. Periplasmic substrate binding proteins of the *fatB* family transport catecholate siderophores such as enterobactin and anguibactin (19), while *fhuD* substrate binding proteins are specific for hydroxamate siderophores (20). In the cytoplasm, siderophore-bound Fe is reduced by siderophore interacting proteins (SIPs) which release Fe^{2+} to be utilized in downstream cellular processes (21). In gram-negative bacteria, heme is imported by heme-specific TBDTs coupled with heme-specific ABCT systems in a manner analogous to siderophores. Heme uptake ABCTs utilize a characteristic substrate binding protein, *hutB.* A cytoplasmic binding protein, *hmuS,* is also typically encoded within characterized heme uptake operons, although its exact function is unresolved (22).

In model organisms Cu, Zn, Co, and Ni move through porins and TBDTs at the outer membrane and through ABCTs and other transmembrane proteins at the inner membrane. *TroA* family ABCT substrate binding proteins (23) participate in the uptake of Mn^{2+} (24), Zn^{2+} (25), and vitamin B_{12} (26). Ni complex uptake has been demonstrated to occur through a non-TroA ABCT system, *nikA* (27). Other Ni and Co permeases include the secondary transporter family NiCoT (Ni and Co) (28), the related *hupE*/*ureJ (29)* family (Ni), and the *cbiMNQO*/*nikMNQO* systems, which are hypothesized to transport Co and Ni, respectively (30). *CbtA* is predicted to be an inner membrane ion channel specific for Co (31), although this is yet to be experimentally confirmed. *CorA* inner membrane ion channels were initially characterized as Mg^{2+} transporters, but recent work demonstrates the family can be highly selective for Co^{2+} (32). As mentioned before, NRAMP and ZIP transporters can operate as generalized divalent metal transporters. Cu

tolerance in certain bacteria is known to be facilitated by P_{1B}-type ATPases, which include both efflux and import transporters. The Cu P_{1B}-type ATPase system *copA (33)* has been identified as a Cu efflux transporter essential for Cu resistance. The same protein family is also required for the biosynthesis of multiple Cu-containing enzymes in *Rubrivivax gelatinosus*, implicating it in Cu import (34). Some bacteria potentially use metallochaperones of the *copZ* family (35) to manage intracellular Cu levels (36).

Heterotrophic marine bacteria are commonly divided into two ecological categories: those that are streamlined oligotrophs (37), and those that are copiotrophs (38, 39), although a continuum certainly exists between these two extremes. Genome-streamlined bacteria appear to be "background-adapted" and succeed by utilizing the persistent but extremely low background concentrations of nutrients under relatively static conditions, whereas metabolically variable copiotrophs are "patch-adapted" and exploit transient nutrient hotspots and variable microscale habitats. Although these coarse divisions do not capture the full complexity of microbial niche space and evolution, they have been shown to be useful in conceptualizing marine microbial ecosystems (39, 40). Patch-adapted marine bacteria rapidly colonize particles and other surfaces and are thought to be primarily responsible for hydrolyzing and degrading structurally complex organic matter while liberating smaller and more labile molecules (41, 42). In contrast, background-adapted organisms are primarily free-living, do not readily associate with surfaces, and have generally low extracellular enzymatic activity. Both patch and background-adapted organisms use membrane-bound transporters to extract specific molecules from their immediate environment in order to acquire nutrients. These bacterial uptake and degradation processes may alter the microscale trace metal reactivity

landscape by liberating metals from sinking particles, modifying metal speciation in the dissolved phase, or by selectively removing certain metal complexes.

The *Roseobacter* and SAR11 lineages are two diverse and highly abundant groups of marine *Alphaproteobacteria* that generally represent patch-adapted and background-adapted ecological strategies, respectively (43). Although roseobacters do not neatly cluster into ecotypes, many cultured representatives have extensive and diverse gene inventories for carbon and energy acquisition consistent with a patch-adapted lifestyle. Recent evidence suggests that some uncultivated roseobacters have lifestyles more consistent with background-adapted organisms (44, 45). Other roseobacters frequently dominate the bacterial community on particles (46) and have been shown to be highly enzymatically active (47). In contrast, the SAR11 lineage is comprised of different ecotypes (48), its members have small streamlined genomes with low GC content and comparatively limited metabolic capability (49), and they do not associate with particles or surfaces. Combined, SAR11 and *Roseobacter* can comprise up to 40% of total bacteria in marine surface waters (50).

Gene content and diversity has been shown to reflect microbial adaptation at ocean basin scales (51–53) as well as at microscales (54, 55), but these adaptations have largely only been explored for nutrients like nitrogen and phosphorous. Here we analyze genomes from two extremes of the patch-adapted to background-adapted continuum (exemplified by the *Roseobacter* and SAR11 genomes, respectively) in order to explore how generalized ecological strategy shapes the specific genomic capabilities for trace metal uptake. First, we survey the extent and diversity of known Mn, Fe, Co, Ni, Cu, and Zn uptake systems in 42 *Roseobacter* genomes and 22 SAR11 genomes. Second, we

evaluate gradations of uptake capabilities amongst genomes and examine lineage evolutionary history as a structuring factor. Finally, we explore relationships between the genetic potential for trace metal uptake, environmental factors, and genomic features and how these relationships are organized in a patch-adapted versus background-adapted framework. *Roseobacter* and SAR11 are not the only bacterial lineages representing this ecological paradigm (e.g. the background-adapted SAR86 lineage (56) and patch-adapted *Alteromonadales* (57)), and caution should be taken when extrapolating the trends from this work to other marine bacterial groups. However, this is the first study, to our knowledge, to contextualize marine bacterial trace metal transporters within evolutionary and ecological frameworks. Although the phylogenetic representation here is limited to two dominant marine groups from the *Alphaproteobacteria*, we expect our conclusions may have broader implications for other marine heterotrophic bacterial groups.

3.3 Results

Genome features: The strains in this study were isolated from a variety of ocean basins, both coastal and pelagic waters, and from particles or surfaces and bulk seawater. 56 of the genomes are either closed or in permanent draft status with the remaining 8 in draft status. The genomes range from 1 to 339 scaffolds and are estimated to be between 93.8% to 100% complete. The genome sizes range from 1.11 Mbp (*Candidatus Pelagibacter ubique* HIMB05) to 5.52 Mbp (*Citreicella* sp. SE45) and code for between 1269 (*Candidatus Pelagibacter ubique* HIMB058) to 5519 (*Pelagibaca bermudensis* HTCC2601) genes. GC content ranges from 29% (*Candidatus Pelagibacter* sp. HTCC7211) to 70% (*Oceanicola granulosus* HTCC2516). The single cell SAR11

AAA240-E13 genome is least complete (93.8%) but is included here to increase the phylogenetic diversity of SAR11 genomes in our analyses.

3.3.1 Uptake systems for inorganic Fe

We used the solute-binding protein, the most divergent and informative component of the system (58), to discriminate the particular metal substrate of ABCTs. In the case of Fe^{3+}, there are five different families of ferric solute binding proteins across the genomes (Figure 3.1, Table 3.1). All *Roseobacter* genomes contain Fe^{3+} ABCTs and 70% of the genomes have at least two different solute binding families. All SAR11 genomes contain Fe^{3+} ABCTs with the exception of AAA240-E13 (clade Ic, isolated from 770 m at Station ALOHA) and HIMB058 (clade II, isolated from Kāneʻohe Bay, HI) (Figure 3.1). Approximately 80% of SAR11 genomes contain single copies of ferric solute binding protein families, with three genomes in clade Ia and one in IIIa having two copies. The most common type of ABCT solute binding protein in all genomes is the FutA1-like protein. Dedicated systems for Fe^{2+} transport are rare in the *Roseobacter* clade. 37% of the genomes surveyed contain proteins in the ZIP family, but ZIP proteins transport metals other than Fe. The Fe-specific *feoAB* system is only found in the genome of *Roseobacter* sp. R2A57, NRAMP homologs are present in the genomes of *Citreicella* sp. 357 and *Ruegeria* sp. TrichCH4B, and no *Roseobacter* genome contains orthologs to FTR1 (Figure 3.1). Fe^{2+} uptake systems are also rare in SAR11 genomes. No SAR11 genome contains ZIP, *feoAB,* or NRAMP uptake systems, but the deeply branching genomes of HIMB058 and HIMB114 both have FTR1 permeases (Figure 3.1).

3.3.2 Fur proteins and binding sites

We predict approximately 90 new Iron-Rhodo-Box transcription binding sites in *Roseobacter* genomes using the previously identified Iron-Rhodo-Box palindromic motif (16). In this expanded work, many but not all *Roseobacter* iron uptake genes are preceded by the Iron-Rhodo-Box motif. 70% of *Roseobacter* genomes appear to have at least one Fe^{3+} ABCT under Fur regulation and many *Roseobacter* TBDTs are preceded by Iron-Rhodo-Boxes (Table 2.2). In the case of SAR11, Fur-box motifs generally match a 15 base pair (7-1-7) Fur-box inverted repeat (59). Only the HTCC7217, HTCC7211, HIMB59, and AA240-E13 genomes are missing apparent Fur-boxes, and Fur-boxes in the remaining genomes always precede components of the Fe^{3+} ABCT uptake system as well as FTR1 permeases. Regardless, it appears that all *Roseobacter* and SAR11 genomes have at least one Fur-like protein, with most *Roseobacter* genomes containing multiple copies.

3.3.3 Siderophore and heme uptake

We used gene neighborhood analysis for the presence of *fatB, fhuD*, siderophore interacting proteins, or putative siderophore biosynthesis genes to predict a role in siderophore-like molecule uptake for 34 TBDTs identified in *Roseobacter* genomes (Table S2, Dataset 5). We also used genome neighborhood analysis to identify 18 genetic loci involved in heme uptake based on the co-localization of *hmuS, hutB*, and TBDTs. Many *Roseobacter* genomes contain multiple different siderophore uptake genetic loci, whereas none contain multiple copies of heme uptake systems. No SAR11 genomes surveyed here contain systems for siderophore or heme uptake. In total it appears that

45% of *Roseobacter* genomes contain the potential for exogenous heme utilization, while 40% contain the potential for siderophore uptake.

The 34 putative siderophore TBDTs cluster by sequence similarity into three different groupings (MCL2, 3, 4) largely consistent with the family of substrate binding protein present nearby. Significant enrichment of *fatB* within ±10 genes of MCL2 suggest MCL2 is involved in the uptake of catecholate-like siderophores, while enrichment of *fhuD* and siderophore interacting proteins within ±10 genes of MCL3 suggest this cluster participates in hydroxamate-like siderophore uptake (Table 2.2). MCL4 gene neighborhoods are enriched in siderophore interacting proteins, substrate binding proteins only assignable at the superfamily level, and occasionally with either *fatB* or *fhuD* suggesting it is involved with the uptake of siderophores of an unclear structural class. Iron-Rhodo-Box transcription binding sites are upstream of many but not all of these siderophore TBDTs, and the catecholate and mixed siderophore clusters (MCL4 and MCLnull) have the lowest percentage of TBDTs with these sites (Table 2.2). In addition to the 34 siderophore TBDTs identified by gene neighborhood alone, eight other TBDTs in ambiguous neighborhoods also cluster with MCL2, MCL3, and MCL4. These eight TBDTs all have upstream Iron-Rhodo-Boxes suggesting they may have a role in siderophore transport or the transport of unknown Fe complexes.

3.3.4 Siderophore biosynthesis

Putative siderophore biosynthesis clusters are present in four *Roseobacter* genomes and no SAR11 genomes (Figure 3.1, Table 3.1 “Sidero. biosyn. + proc.” category). Both the *P. gallaeciensis* 2.10 and *P. gallaeciensis* 17395 genomes contain

clusters with NRPS-independent type siderophore biosynthesis genes, which are exclusive to the biosynthesis of siderophore secondary metabolites (18). In addition, the genomes of *Citreicella* sp. SE45 and *Oceanicola* sp. S124 have TBDTs co-localized with genes containing domain-specific hits to enterobactin synthase subunits E and F (NRPS) and other genes involved in siderophore uptake, biosynthesis, and regulation. Interestingly, TRipartite ATP-independent Periplasmic (TRAP) transporters are present within two putative siderophore uptake clusters (Figure 3.2). TRAP transporters utilize a substrate binding protein to shuttle substrates in the periplasm, but the substrate is moved through the inner membrane by the cotransport of a counter ion (usually Na^+ or H^+) in the opposite direction (60), rather than the hydrolysis of ATP. The use of TRAP transporters in conjunction with TBDTs (particularly siderophore transporters) has thus far been unexplored (61). The co-localization of TRAP transporters with seven different putative Fe TBDTs and siderophore biosynthesis genes (Figure 3.2) is unique, and to our knowledge has not been reported before. It has been postulated that using two counter Na^+ ions would impart a significant energy savings compared to the use of an ATP binding cassette transporter (ABCT) system (62). In the marine environment where substrates are highly dilute, the energy savings of utilizing TRAP transporters over ABCTs may be significant, especially when used in concert with ion gradients generated by proteorhodopsins (63). Every Roseobacter genome contained at least eight different substrate binding protein components of TRAP transporters with *Citreicella* sp. SE45 containing over 40 different homologs. TRAP uptake is constrained by the presence of at least one carboxylic acid group in the substrate (60). Carboxylic acid groups are frequently involved in the chelation of metal species (64) making metal-ligand complexes

a potential substrate for TRAP transporters. The identification of TRAP transporters next to siderophore biosynthesis genes and siderophore TBDTs implies a previously unrecognized role in Fe chelate transport for the TRAP family.

3.3.5 Uptake systems for Mn, Zn, Co, Cu, Ni

The number of experimentally characterized TBDT systems for trace metals other than Fe is small, and we observe no TBDTs unambiguously related to Ni, Co, or Cu transport in *Roseobacter* and SAR11 genomes. We therefore focus our searches on relatively better known inner membrane Ni, Co, Cu, Mn, and Zn transporters (Figure 3.1, Table 2.1). No SAR11 genomes but all 42 *Roseobacter* genomes contain at least one helical backbone solute binding protein assignable only at the *troA* superfamily level. The *troA* subfamilies present in *Roseobacter* genomes are *psaA* (38%), *troA*-a (38%), *hemV2* (10%), *troA*-f (2%), and *znuA* (2%). 83% of *Roseobacter* genomes have at least one of the troA subfamilies, NRAMP proteins, or ZIP family proteins (Figure 3.1), and roughly half have copies of multiple families. In contrast, SAR11 genomes have reduced diversity of Mn and Zn transport systems compared to *Roseobacter* genomes. The only troA subfamily present in SAR11 appears to be the high affinity Zn transporter *znuA*, present in 50% of the genomes (Figure 3.1). In these genomes the synteny of the *znuA* uptake system is highly conserved, with all regions containing a Fur-family protein as well as the additional ABCT components. These putative znuA amino acid sequences have approximately 43% sequence identity to curated *znuA* sequences in UniProt, supporting their assignment as Zn transporters.

Nickel and cobalt transporters vary between *Roseobacter* and SAR11 genomes (Figure 3.1). 90% of *Roseobacter* genomes contain homologous proteins to *hupE/ureJ* nickel transporters, while only *Citreicella* sp. SE45 has the well-characterized nickel transporter *nikA*. 45% of SAR11 genomes contain putative *hupE/ureJ* systems. In both SAR11 and *Roseobacter hupE/ureJ* gene neighborhoods are highly variable, although in both groups many *hupE/ureJ* genes are co-localized with ABCTs suggesting a role in transport. 76% of *Roseobacter* genomes have at least one of the *cbiMQ*, *cbtAB*, or *corA* cobalt uptake systems, and half have more than one kind of cobalt transporter. The most common *Roseobacter* cobalt transporter is *corA* found in half of the genomes. In contrast, corA is the only potential cobalt ion transporter in SAR11 and is only present in the earliest diverging genome, HIMB59. 98% of *Roseobacter* genomes contain *copA* systems, while 50% contain *copZ* (Figure 3.1). In contrast, SAR11 appears to have completely eschewed the use of *copA*, but 54% of genomes contain *copZ* metallochaperones.

3.3.6 Unknown tonB dependent transporters and other secondary transporters

TonB dependent transporters (TBDTs) have a broad range of substrates, but interestingly, 77% of Roseobacter TBDTs appear to be involved in the uptake of small Fe complexes based on gene neighborhood, sequence similarity, and predicted FUR regulation. This is in contrast to TBDTs in other marine groups such as SAR86 (56) and the phylum *Bacteroidetes (65),* which appear to be mostly dedicated to the uptake of high molecular weight carbon compounds. TBDTs in clusters MCL6 and MCLnull were the most difficult to assign substrates from gene context and sequence similarity, but were

located next to substrate binding proteins for peptide and sugar transport as well as non-specific metal transport.

3.3.7 Phylogenetic conservation of metal uptake genes

Genes coding for microbial cellular functions can be differentially conserved with respect to a phylogeny derived from a universal and conserved reference gene such as the 16S rRNA gene. The degree of correspondence of a particular trait with an organism's phylogeny can be thought of in terms of phylogenetic trait depth. We used the τ_D statistic (trait depth) from the consenTRAIT algorithm (66) and Fritz and Purvis' D for phylogenetic dispersion (67) to predict the extent to which *Roseobacter* and SAR11 phylogeny (Figure 3.1) explains the distribution of metal uptake categories in each genome. The use of two independent approaches also allowed us to qualitatively assess the degree of uncertainty for our phylogenetic conclusions. The τ_D statistic is a continuous metric that corresponds to the mean branch length between root nodes and leaves in a phylogenetic tree where 90% of the leaves have a particular trait. When trait depth (τ_D) is large, a metal uptake trait will be shared among members of deeply branching clades, suggesting that these traits are consistently passed on to daughter lineages. When trait depth (τ_D) is small, the trait will more likely be found in small, dispersed clades in a phylogeny. Such dispersion could suggest the trait is evolutionarily labile, having been lost and gained multiple times during the evolutionary trajectory of a microbial lineage. Unlike D, trait depth is not normalized to phylogeny size and represents direct phylogenetic distances with larger values indicating deeper clades. Fritz and Purvis' D is calculated by the sum of changes in nodal values of a binary trait along

edges in a phylogeny. The metric is robust for phylogeny sizes down to approximately 25 leaves and where trait prevalence is greater than 0.2 (see methods for details).

We hypothesized that patch-adapted and background-adapted genomes would have differing signatures of heritability for trace metal uptake due to the constraints of genome size. Whereas larger genomes contain more genetic material upon which processes of recombination and lateral transfer can operate, background-adapted genomes would have streamlined to such an extent that essential trace metal uptake pathways would have been largely phylogenetically fixed across a lineage. 57% of the estimates of phylogenetic dispersion in *Roseobacter* are less than 0.5, but only 35% of traits have probabilities of matching a stochastic distribution of less than 5% ($P(\mathrm{D})_{\mathrm{random}} < 0.05$) and have prevalence values greater than 0.2 (*Prevalence* > 0.2). In SAR11, 55% of D values are less than 0.5, but only 33% of traits have *Prevalence* > 0.2 and $P(\mathrm{D})_{\mathrm{random}} < 0.05$. This indicates that the remaining metal uptake traits are indistinguishable from patterns of random or convergent evolution based on the current number of genomes surveyed in each group. Generally, there is poor agreement between τ_D and the independent Fritz and Purvis' D metric as to which traits are non-randomly distributed. However, both metrics estimate that a similar proportion (20% to 35%) of traits are non-randomly distributed. Trait depth values for trace metal transport categories shared between *Roseobacter* and SAR11 are directly comparable because their phylogenies are based on the same protein families (Table 3.3, Figure 3.3). Four of the six shared categories have significantly greater trait depths in SAR11 than *Roseobacter*.

3.3.8 Relationships between patterns of metal uptake genes, habitat, and genome features

As metals are essential cofactors in a variety of ecologically relevant metabolic processes, we tested for associations between the abundance of metal transport systems, microbial habitat, and genome features. A principle components analysis (PCA) based on diversity and abundance of trace metal transporters indicates that SAR11 genomes cluster tightly in ordination space while *Roseobacter* genomes are highly dispersed (Figure 3.4). The multivariate homogeneity of group dispersions for *Roseobacter* has a greater average distance to the median (4.227) compared to SAR11 (1.087) and this difference is significant ($P < 0.001$). The larger dispersion value indicates that *Roseobacter* genomes have greater within-group variability in uptake capabilities than SAR11. For example, the *Roseobacter* HTCC2255 genome is more similar to a SAR11 genome (distance to SAR11 mean centroid of 1.16) than it is to the average *Roseobacter* genome (distance to *Roseobacter* mean centroid of 3.39). Multivariate homogeneity determined using Bray-Curtis dissimilarity (with non-metric multidimensional scaling) also indicates that the *Roseobacter* dispersion is significantly larger than SAR11 demonstrating the robustness of the trend.

Factors corresponding to ocean basin of isolation and coastal versus pelagic isolation are not significantly associated with the PCA ordination. We classified genomes as having a “planktonic” or “surface associated” lifestyle based on whether a strain was described to associate with particles experimentally (68) or was isolated from some kind of biotic or abiotic surface. Strains where lifestyle data could not be clearly determined were omitted from statistical analyses. As a result, surface associated and planktonic

factors are significantly associated ($R^2 = 0.33$, $P < 0.001$) with the reduced PCA ordination (Figure 3.4). The total number of metal transporters, the total number of transporters in each genome mapping to the Transporter Classification Database (TCDB) (69), the predicted number of biosynthetic gene clusters per genome, GC content, and number of genes per genome are all strongly correlated with the ordination. The number of predicted laterally transferred genes is not significantly correlated.

In the combined dataset, many of the correlations between genome features and metal transport categories are driven by the inclusion of SAR11. For example, the strong negative correlation between *znuA* and many genome features is a result of its prevalence in SAR11 genomes. In *Roseobacter* (Figure 3.5B), the number of metal transporters per genome, the total transporters per genome, and GC content are strongly positively correlated with TBDT categories. The strongest correlations when considering both SAR11 and Roseobacter genomes (Figure 3.5A) are between genome features (grey diamonds) and nonspecific metals (blue) as well as between transporters predicted to be in the same uptake pathway (e.g. TBDT MCL1, hemS, and hutB). The large number of *troA* superfamily proteins, siderophore transporters, and heme transporters in *Roseobacter* are positively correlated with genome features such as increasing genome size and increasing number of total transporters. Ultimately, metal pathways related to uptake of small defined iron complexes (siderophores and heme) are best correlated with genome features such as increasing genome size, increasing number of total transporters per genome. A previous study reported a statistically significant negative correlation between high-affinity Fe^{2+} transporters and Fe^{3+} transporters suggesting that marine bacteria tend to rely on either *feoB* or Fe^{3+} ABCTs but not both (70). We do not see this

negative correlation in our results, neither do we observe a positive correlation between ZIP divalent metal transporters and Fe^{3+} ABCTs as was also reported earlier (70). However, the previous study included genomes from many different taxonomic groups, and those trends may have been driven mostly by specific bacterial taxa not included in our study. In the case of the *Alphaproteobacteria* genomes examined here it appears that neither Fe^{3+} or Fe^{2+} uptake capabilities are exclusive of one another and indeed many strains have the capabilities for both reduced and oxidized Fe uptake. Interestingly, the number of genes per genome predicted to be laterally transferred was positively correlated with genes related to siderophore biosynthesis and processing, suggesting a potential linkage between lateral gene transfer and the ability to synthesize siderophores.

The stark difference between the trace metal uptake inventory in SAR11 and *Roseobacter* generates a number of correlations that are not observed when examining either group individually. Effectively no significant correlations exist between genome features and metal uptake pathways in SAR11 genomes because they are lacking most transporters observed in this study. When considering only Roseobacter genomes (Figure 3.5B), the total number of metal transporters per genome are correlated with 14 metal transport categories, most having to do with siderophore and heme uptake. The total number of Transporter Classification Database (TCDB) (69) transporters per genome is correlated with four metal transport categories (MCL6, MCLnull, ZIP, and P_{1B} ATPases) as well as the total number of metal transporters. GC content is also correlated with the total number of transport genes per genome, ZIP transporters, and P_{1B} ATPases. Genome size, the number of predicted biosynthetic clusters, and the total number of genes with a

functional prediction only correlated with the TCDB count and no specific metal uptake pathway.

In the combined *Roseobacter*/SAR11 dataset (Figure 3.5A) and *Roseobacter* alone (Figure 3.5B) dataset there is significant positive correlation between metal transporter abundance and transporters predicted to be in the same pathway based on synteny and sequence homology. Ultimately, metal uptake pathways for small, defined iron complexes (siderophores and heme) are best correlated with genome features such as increasing genome size, increasing number of total transporters per genome, and increasing metal transporters per genome. Indeed, it does generally appear that siderophore and heme uptake are largely biased towards the largest genomes with the most transporters, but they are also unevenly distributed in *Roseobacter* (Figure 3.6). The strongest positive correlations for the combined and individual lineages are between the total number of transporters per genome and the genome size (Figure 3.7A). However, the number of metal transporters per genome has the poorest correlation with genome size (Figure 3.7C) and marginally better correlation with the total number of transporters (Figure 3.7D).

3.4 Discussion

Consistent with streamlining theory (37), the background-adapted SAR11 lineage has relatively few trace metal transporters and an apparently limited regulatory capacity for Fe uptake. In contrast, the mostly patch-adapted *Roseobacter* genomes investigated here have multiple diverse pathways for the acquisition of both organically complexed and inorganic metals. This suggests that roseobacters are able to adapt to and occupy a

range of trace metal niches in the marine environment and that the availability of trace metal resources may influence *Roseobacter* genome diversification. The variable inventories of trace metal transporters in *Roseobacte*r and SAR11 may ultimately reflect variable metabolic demands for metals as enzymatic cofactors. However, no studies, to our knowledge, have specifically examined metal quotas for SAR11 and *Roseobacter* strains. A handful of *Roseobacter* genomes, for example HTCC2255, appear closer in gene content to background-adapted SAR11 genomes (44, 50), suggesting that trace metal streamlining is also a valuable ecological strategy for some roseobacters. Indeed, background-adaptation is probably a more prominent strategy in the *Roseobacter* lineage than currently available isolates would suggest (44). Thus the results presented here for *Roseobacter* are likely biased towards the portions of the lineage that are patch-adapted.

3.4.1 Similarities between SAR11 and Roseobacter genomes

Fe^{3+} ABCTs are the most abundant Fe transporter identified in this study, suggesting that free Fe^{3+} is the most common form of Fe in the periplasm in both *Roseobacter* and SAR11. All investigated genomes on either end of the background-adapted to patch-adapted spectrum contain ABCT transporters for Fe^{3+} uptake, with the exception of the SAR11 genomes AAA240-E13 and HIMB058, although both these genomes may be missing these transporters due to genome incompleteness. Nevertheless, it appears that ABCT uptake is likely the default mechanism for Fe^{3+} uptake across the bacterial inner membrane for both the SAR11 and *Roseobacter* lineages and may be an essential system. Fe^{2+} is often at very low concentrations in the marine environment,

although at times it can accumulate to significant proportions of the total Fe pool (71). However, Fe^{2+} transporters are rare in both SAR11 and *Roseobacter*.

3.4.2 Differences between SAR11 and Roseobacter genomes

Roseobacter and SAR11 genomes are mostly different with respect to trace metal transporter inventory. Even though Fe^{3+} ABCTs are basically present in all genomes, 30 out of 42 *Roseobacter* genomes have multiple copies of Fe^{3+} ABCTs from at least two different domain families (TrichCH4B has five), while only four of the 22 SAR11 genomes have multiple copies. This suggests a nuanced distinction between patch-adapted *Roseobacter* and background-adapted SAR11 that may be reflective of Fe niches. For example, multiple versions of substrate transporters have been invoked to explain multiphasic kinetics for glucose uptake in bacterial isolates and natural assemblages (72, 73).

The Fur protein family has a large diversity of metal selectivity, making it challenging to assign specific metal cofactors using sequence homology alone (74). Regardless, *Roseobacter* genomes generally have multiple copies of Fur-like regulatory proteins and many of their iron uptake genes are downstream of Fur-box regulatory motifs. Past work suggests that Fur proteins are present only in the genomes of SAR11 clade Ia members (49), but our results indicate that all SAR11 genomes have some form of a metal-dependent transcription factor. It appears that SAR11 has greatly downsized its metal-dependent regulatory networks compared with patch-adapted roseobacters, which is consistent with prior observations of its overall regulatory complexity (37). This suggests that patch-adapted roseobacters have the capability to sense and react to a wide

variety of trace metal forms, while the background-adapted SAR11 may respond to only a limited portion of the overall trace metal chemical diversity in seawater.

Many of the Fe-binding ligands identified from cultured marine bacteria (75) have been siderophores, low molecular weight and high-affinity Fe^{3+} chelating agents secreted by some bacteria explicitly for the purpose of chelating Fe. Siderophores have also been detected in bulk seawater, are predicted to shape bacterial social interactions (76), may reflect degrees of habitat structure (77), and are hypothesized to comprise a significant component of the marine strong iron-binding ligand pool (78). Therefore, it is plausible that siderophore and perhaps other undiscovered strongly-bound iron complexes are important iron sources for some marine bacteria. 40% of roseobacters can probably acquire at least one type of siderophore or other small molecule Fe chelator, and ~10% have the potential of producing siderophores. Indeed, both *Phaeobacter* strains included in this study have been shown to produce siderophores in culture (79).

Even though ~40% of Roseobacter genomes appear to have siderophore uptake systems, very few strains appear able to biosynthesize siderophores. This suggests that most Roseobacters with putative organic iron-complex TBDTs rely on siderophores or small Fe chelating ligands of similar structure that are not endogenously produced as secondary metabolites. This observation is consistent with a "public goods" dynamic that has been demonstrated for a large and deeply sampled grouping of particle associated *Vibrio* strains (76). In this case, siderophore "cheaters" found on large particles do not contribute to the community pool of siderophores and are able to access the public goods generated by siderophore producers (76). Marine particles are often rich in Roseobacters (80, 81), and these same public goods dynamics apparent in Vibrio communities may

have driven the evolution of siderophore biosynthesis and uptake capabilities in Roseobacter communities. The apparent lack of siderophore biosynthesis in the Roseobacter clade may be due to an oversampling of siderophore cheater genomes. An alternative explanation is that putative siderophore uptake genes may in fact be targeted to unknown strong marine ligands with similar chemical moieties to siderophores.

In contrast, no SAR11 genomes have either capability, and siderophores have been used to experimentally Fe-limit *P. ubique* HTCC1062 in culture (82). Here our focus is on direct transport of siderophore complexes and is not meant to address the possibility of extracellular processing of organically bound Fe into bioavailable forms, although this is certainly an important, yet understudied possibility. Our results suggest that *direct* uptake of small intact organic-Fe complexes via TBDTs is an important uptake strategy for some, but not all, roseobacters, while it appears to be an expendable strategy for the background-adapted SAR11.

Heme *b*, another form of organic Fe, is a dynamic and significant component of the marine Fe cycle (22). Like siderophores, no heme uptake systems were detected in SAR11 genomes. Interestingly, about equal proportions of *Roseobacter* genomes from this study have heme and siderophore uptake systems (45% and 40% respectively) and nine strains have only a single TBDT, which is specific for heme. The *Roseobacter* heme uptake gene locus is also highly conserved with respect to synteny and TBDT sequence similarity suggesting a tightly controlled evolution of this gene cluster. The proportion of heme uptake systems in *Roseobacter* identified here is consistent with what has been described earlier (83), but it is interesting that the prevalence of heme uptake is roughly equivalent to that of siderophore uptake. This is significant in that siderophores are the

dominantly researched Fe-ligand complex in marine systems. In the case of the *Roseobacter* clade, it appears that heme may be equally important as siderophores, which may be due to the frequent association of *Roseobacter* with hemoprotein-rich phytoplankton (22, 46). Supporting this hypothesis, many roseobacters with heme uptake systems were isolated from phytoplankton or are known to associate with other organisms (38).

How do background-adapted SAR11, one of the most abundant organisms on earth, manage to satisfy their iron requirements by apparently only utilizing inorganic Fe^{3+}, the scarcest form of oxidized iron in the oceans (4)? One hypothesis is that SAR11 cells directly modify refractory extracellular dissolved organic, colloidal, or particulate Fe species into usable forms. Another hypothesis is that SAR11 relies on the activity of external agents in microbial ecosystems to produce enough labile Fe for its survival, analogous to how *Prochlorococcus* cells appear to rely on the activity of microbial community members for hydrogen peroxide oxidation (84). Background-adapted organisms like SAR11 often co-exist in food webs strongly controlled by micro-grazers and viruses whereby regular biomass turnover may produce significant and regularly occurring labile Fe sources. For example, if there is a strong diel structuring of *Prochlorococcus* mortality as has been observed previously (85), co-occurring heterotrophic bacteria may in turn synchronize the expression of trace metal transporters to daily periods of increased *Prochlorococcus* lysis. In support of this idea, many *Roseobacter*, SAR11, and SAR116 transporter transcripts displayed strong diel periodicity in a recent field study (86). Ultimately, our results indicate that SAR11 cells do not have the ability to directly transport intact organic iron complexes suggesting they

utilize Fe ions or very small charged complexes moved through the outer membrane probably by passive transport. SAR11 appears to produce large amounts of ABCT solute binding proteins under Fe stress (82) and solute binding proteins are highly abundant in natural populations (87). It may be that highly expressed solute binding proteins efficiently intercept all periplasmic Fe^{3+} and drive a gradient inward toward the cytoplasm and/or that unknown cell surface binding proteins operate to locally concentrate Fe at the outer membrane.

Mn, Zn, Ni, Co, and Cu are cofactors in many biogeochemically significant metabolic pathways and are important micronutrients for marine heterotrophic bacteria. Most *Roseobacter* genomes contain dedicated inner membrane transporters for transition metals other than Fe and many have apparently redundant systems for some metals. For example, two *troA* superfamily substrate binding proteins identified in this analysis from R. sp. TM1040 and R. sp. AzwK-3b have been shown to be highly expressed under Mn limiting conditions (88). As with organic Fe transporters, Mn, Zn, Ni, Co, and Cu transporters were largely absent in background-adapted SAR11 genomes. However, an intriguing exception suggest that some metal uptake traits have escaped the purging effect of genome streamlining in SAR11. Early diverging SAR11 lineages and some SAR11 group Ia members appear to occupy niches where high-affinity zinc uptake (*znuA*) is useful, while roseobacters appear to have mostly rejected this trait. Although the SAR11 *znuA* sequences are quite similar to characterized *znuA* proteins from model bacteria, there is a possibility that they are involved in the uptake of other metals. Physiological experiments are needed to confirm substrate specificity, but based on the level of sequence similarity we anticipate our bioinformatic predictions here are robust.

Currently, it is unknown as to whether the presence of *znuA* in these SAR11 strains represents greater absolute Zn requirements relative to other trace metals, reduced Zn concentrations in specific niches, or some other factor. Genomes with *znuA* do not contain more annotated Zn-binding domains than other SAR11 genomes, nor are they significantly connected by isolation source as it is defined in this study. However, it is intriguing that most SAR11 strains isolated from pelagic surface waters have *znuA*, which may be related to the extremely low concentrations of Zn and other metals in the pelagic surface ocean.

3.4.3 Phylogenetic signal in metal uptake traits

To our knowledge, this is the first study to contextualize trace metal transporters within a phylogenetic framework in marine microbes in order to explore combined patterns of heritability, gene loss, and lateral transfer. Our results indicate that the majority of metal uptake traits are not significantly associated with phylogeny in either SAR11 or *Roseobacter*, and it appears that in both groups trace metal niche adaptation has occurred through evolutionary mechanisms indistinguishable from stochastic processes. A higher percentage of metal transporters do appear to be nonrandomly distributed in SAR11 than *Roseobacter* potentially suggesting an overall greater role for vertical heritability in SAR11. Furthermore, certain metal uptake traits shared between groups cluster at significantly different clade depths, which implies that distributions of these families across the SAR11 lineage are more likely to be fixed across fine scale phylogenetic diversity than they are in the *Roseobacter* lineage.

Recent surveys of carbon utilization traits (66) and extracellular enzymes (89) using the consenTRAIT metric suggest that these traits mostly exhibit non-random phylogenetic distribution despite their generally shallow clade depth. We show here that most metal uptake traits in both *Roseobacter* and SAR11 also have shallow clade depths (small τ_D), but appear to be randomly associated with their reference phylogenies. We interpret these results as potentially reflecting differing degrees of selective pressure with respect to specific metals in the *Roseobacter* and SAR11 lineages. Microbial niche exploration, changing trace metal availability in existing niches, or altered absolute metal requirements may have resulted in gene-specific selective sweeps in both patch-adapted and background-adapted *Roseobacter* and SAR11 lineages. These processes may also have resulted in selective loss of capabilities as well. Although the τ_D and D metrics cannot distinguish between lateral gene transfer or selective gene loss events, the end result is that phylogenetically similar strains within each lineage have strongly ecologically differentiated with respect to trace metal uptake genes. We do not explore mechanisms of selective gene gain versus loss in *Roseobacter* and SAR11 here, but our results are consistent with the hypothesis that marine trace metal resources influence genome content in both patch-adapted and background-adapted genomes from the *Roseobacter* and SAR11 lineages.

3.4.4 Metal transporters, habitat, and genome features

The four factors most strongly correlating with the metal uptake ordination are metal transporters per genome, the GC content of each genome, the total transporters per genome, and genome size. An inspection of individual pairwise correlations shows that

genome size is strongly positively correlated with many specific uptake pathways, but many of these specific correlations are lost when only *Roseobacter* genomes are included. In general, it appears that smaller *Roseobacter* genomes have fewer total metal transporters, but metal transporters do not neatly scale with genome size in many other cases. *Octadecabacter articus* 238 has the third largest genome size of all roseobacters, but it has no TBDTs and appears to only have systems for Fe^{3+} and Mn^{2+} uptake. On the contrary, half of the ten smallest *Roseobacter* genomes have transporters for heme and siderophore-like complexes.

Habitat categories (eg. coastal, pelagic, ocean basin isolation) as defined in this study are not correlated with the trace metal uptake inventory in either *Roseobacter* or SAR11. However, significant correlation exists when considering only strains with confirmed particle attachment lifestyles. Supporting these results, recent meta-omic studies (90–92) have demonstrated that prefilter size (particle fractions) is a better predictor of differences in community structure, metabolic capability, and transcriptional activity than both depth and geographical variability. However, many uncultured roseobacters are predicted to be background-adapted and may not associate with particles at all (44, 45), and it is unknown whether uncultured *Roseobacter* genomes contain organic Fe uptake systems. Therefore, it is possible that trace metal uptake inventory may not overlap neatly with a particle-adapted lifestyle in the *Roseobacter* lineage due to a culturing bias and/or incomplete lifestyle assignments. HTCC2255, the only background-adapted *Roseobacter* genome included here, is thought to be only free-living (50) and is similar in trace metal transporter inventory to SAR11. More work is needed to obtain full

genome sequences of background-adapted roseobacters, and to understand the potential for particle attachment in both background and patch-adapted strains.

3.5 Conclusions

Our results indicate that both the patch-adapted and background-adapted genomes included in this study exist along spectrums of trace metal acquisition capability. At one extreme, genomes contain multiple and apparently redundant pathways for the uptake of most metals, while at the other end they lack many transport families. In patch-adapted roseobacters the presence/absence of any particular metal uptake pathway is not predictable from genome size alone, and only a small subset of all metal uptake pathways are present in all genomes. This also indicates that the presence of any one metal acquisition pathway, such as siderophore or heme uptake, is not representative of the capabilities of the *Roseobacter* lineage as a whole. Even though the background-adapted genomes included in this work have reduced metal uptake capabilities when compared with patch-adapted genomes, they also appear to have a surprising degree of variability, for example the *hupE/ureJ* family and the patchy distribution of *znuA*. We interpret this as reflecting degrees of trace metal niche differentiation, whereby marine trace metal concentrations and chemical speciation influence genomic content at fine levels of phylogenetic differentiation in both background-strategists and patch-strategists. Thus, trace metal niches should be considered an important factor in shaping the genomic content of marine heterotrophic bacteria and should be considered when examining microbial roles in broader marine ecosystems and biogeochemical cycles.

3.6 Materials and Methods

3.6.1 Genomic sequence data and genome classification schemes

All microbial genomes and associated metadata were obtained from the IMG database (93) (April 2015). *Roseobacter* genomes were selected from IMG to reflect the content of Roseobase, which is a comprehensive genomic resource for marine *Roseobacter* strains. All publically available SAR11 genomes available in IMG with a genome completeness greater than 90% were included in this study. Genome completeness and integrity of all isolates were assessed using the CheckM pipeline (94). Lifestyle (surface associated, free-living), isolation location (Atlantic, Pacific, Indian, Polar), and isolation land proximity (coastal, pelagic) were assigned when the data were available in IMG or the primary literature. Genome size, total gene counts, GC content, the number of genes with a functional prediction, the number of genes assignable to the Transporter Classification Database, the estimated number of genes predicted to be horizontally transferred, and the estimated number of biosynthetic clusters per genome were obtained from IMG database annotations. Isolation data (geographical and lifestyle) for bacterial strains used in this study are compiled from the primary literature. If isolation data could not be reliably determined from the literature, then these data were omitted from subsequent statistical analyses.

3.6.2 Functional prediction and annotation

The metal transport systems used for searches in *Roseobacter* and SAR11 genomes (Table 2.1) include the majority of currently characterized/predicted metal transport systems (9, 10). Metal transporters were identified using the NCBI conserved

domain database (95). Orthologs were identified by rpsblast hits (evalue < 10^{-5}) to conserved domain database models and only bidirectional reciprocal "specific hits" to domain models (those above significance threshold values) were retained. In the case of the *troA* superfamily, sequences matching the "TroA helical backbone superfamily domain" (cl00262) but not matching other specific subfamilies above a bitscore threshold were identified as hits to "*troA*-superfam" which we use to represent generic metal-interacting solute binding proteins. Nonribosomal peptide synthetase independent siderophore (NIS) biosynthesis capabilities were assigned if genomes contained NIS synthetases and NIS acetyl transferases. Nonribosomal peptide synthetase (NRPS) pathways synthesize many secondary metabolites including siderophores (17), so the presence of NRPS pathways cannot be simply attributed to siderophore production. To identify NRPS siderophore biosynthesis, all NRPS-like domains were identified then manually annotated using ferrichrome synthetase (96), enterbactin synthetase subunits E and F, the phosphopantetheinyl transferase component of enterobactin synthetase, and enterochelin esterase (17). Abundances of multi-gene pathways or enzymes known to utilize a particular metal, for example [NiFe] hydrogenases or Vitamin B_{12} biosynthesis, were identified by orthology to KEGG through the IMG web service.

3.6.3 Identification of Fur box DNA regulatory elements

The *Rhodobacterales* have previously been described to employ DNA regulatory motifs divergent from that of the canonical Fur box (16). As such, searches for the 19 DNA base pair inverted repeat consensus fur box (59) generally yielded poor matches in the Roseobacter genomes. To address this, a collection of 99 inverted repeat sections of

the binding sequence for Fur proteins (Fur box) was obtained from published sequences from a previous study utilizing ten different Roseobacter genomes (16). These sequences were used in a standard blastn search against the genomes of all 42 Roseobacter used in this study. This resulted in approximately 170 potential Roseobacter Fur boxes. These 170 sequences were collected and used in another blastn search against Roseobacter genomes using lenient search parameters including a match reward = +1, match penalty = -3, gapopen = +5, gapextend = +2, word size = 7, and no masking or filtering for low complexity. Regions were discarded if they contained less than 15 base pairs of exact similarity to an existing sequence or if they were not within an intergenic region. The regions of all new hits were manually searched for candidate Fur-regulated genes based upon whether the gene's predicted function was related to iron uptake or homeostasis. If the newly identified Fur box was near a TBDT this information was used to update the potential substrate if necessary. In contrast, SAR11 genomes had readily identifiable Fur-box sequences based on a blastn search using the 15 basepair (7-1-7) inverted repeat identified in prior work (59). SAR11 genomes were searched with the 15 base pair motif using the same blastn parameters as described for the Roseobacters and candidate iron-regulated genes were identified. The identified.

3.6.4 Markov Clustering of TonB Dependent Transporters (TBDT) and protein family enrichment

To classify TBDTs by sequence similarity, all TBDTs identified by rpsblast were bi-directionally evaluated using BLASTp using an Evalue cutoff of 1.00E-15. TBDTs were clustered using the MCL algorithm (97) of the clusterMaker (98) plugin in

Cytoscape. An edge weight conversion of -Log(Evalue), edge weight cutoff of 1, a pruning threshold of 1.00E-015, 16 iterations, and a maximum residual value of 0.001 were used in the Markov Clustering of an 'all vs. all' blastp analysis of identified TBDTs. The number of clusters was explored using granularity parameters of 1.4, 2.0, 2.2, 2.5. Ultimately, a granularity parameter of 2.0 was chosen as it best partitioned TBDTs into clusters with neighboring genes of coherent function. Substrate specificity for TonB Dependent Transporters (TBDT) was assigned by manually examining the genome neighborhood (10 genes upstream and downstream) for the presence siderophore SBPs, SIPs, siderophore biosynthesis genes, heme uptake genes, other metal uptake systems, and the Fur box DNA regulatory motif. Fisher's exact test was then used to test if protein families were enriched in TBDT gene neighborhoods (10 genes upstream and downstream of TBDT). Briefly, protein neighborhoods were collected from the UniProt database and COG/Pfam/CDD/TIGRFam domains existing in more than 20% of TBDT neighborhoods of a particular MCL cluster were enumerated inside and outside of the TBDT neighborhood and used in Fisher's exact test. *P* values from the Fisher's exact test were corrected using the Benjamini Hochberg method for controling the false discovery rate. Protein families were considered enriched if $P < 10^{-5}$ and the Odds Ratio was > 1.

3.6.5 Phylogenetic tree inference and phylogenetic conservation of functional traits

26 of 28 composition-homogenous orthologous protein families (99) (excluding COG0238 and COG0522) were identified in the 64 genomes. Amino acid sequences in each orthologous set were aligned using MUSCLE (100), culled using Gblocks (101) with the following settings: -b1=(n/2)+1 -b2= (n/2)+1 -b3= (n/2) -b4=2 -b5=h, where

n=number of sequences in the alignment (42 for *Roseobacter* and 22 for SAR11), and concatenated. Phylogenetic inference was performed with RAxML HPC v7.7.6 using the gamma model for rate heterogeneity with optimized substitution rates, the LG amino acid substitution matrix (102), and 1000 bootstrap resamplings. Phylogenetic trees were rooted at HTCC2255 for *Roseobacter* and HIMB59 for SAR11.

3.6.6 Phylogenomic structures of biological traits

Phylogenetic signal' or 'phylogenetic structuring' are used in this study to refer to traits (in this case the ability to acquire various metals) whose presence among taxa in a phylogeny are autocorrelated with the structure of the phylogeny. We used the τ_D statistic (trait depth) from the consenTRAIT algorithm and Fritz and Purvis' D for phylogenetic dispersion to predict the extent with which Roseobacter and SAR11 phylogeny explains the distribution of metal uptake categories in each genome. D was calculated using the CAPER package in R with 1000 random permutations under a model of Brownian motion for discrete trait evolution, where $D < 0$ indicates strong trait clustering, $D = 0$ indicates Brownian evolution, $D = 1$ indicates random evolution, and $D > 1$ indicates over-dispersion. D was calculated in the R package 'caper' using the highest scoring maximum likelihood tree from Roseobacter and SAR11. R Command: phylo.d(caper_object, binvar=TRACEMETAL_TRAIT). D values were considered significant if the probability of matching a random distribution ($P(D)_{random}$) was smaller than 5% ($P(D)_{random} < 0.05$).

The τ_D statistic was estimated as the average amino acid distance in substitutions between leaves and root node of a clade carrying a particular trait, whereas clades are

defined by 90% of leaves possessing a metal uptake trait. Traits without any neighbors were scored using half the distance to the nearest internal node. We considered a particular τ_D value as significant ($\alpha = 0.05$) when fewer than 5% of τ_D values resulting from 1000 random permutation of taxa across the tree (10 permutations per bootstrap) were greater than or equal to the true τ_D value. We did not calculate D and τ_D for metal uptake traits present in all genomes because they are completely phylogenetically conserved. Trait depth (τ_D) was calculated using a custom R script provided with the original description of the consenTRAIT algorithm (66). The script was modified to include a random permutation procedure for significance testing. Briefly, for each phylogenetic tree in the multi-tree bootstrap file the distributions of binary traits were randomly shuffled ten times for each tree resulting in 10X the individual trait depth calculation performed in the original script. The proportion of those calculations greater than the value calculated for the best maximum likelihood tree is then equal to the *P* value (proportion of values greater than that observed by chance alone).

3.6.7 Multivariate statistics and correlation analysis

Multivariate statistics were performed using the VEGAN (103) package in R. Patterns of metal uptake genes among genomes were explored using both Principal Coordinates Analysis (PCA) and nonmetric multidimensional scaling (NMDS) using the Bray-Curtis dissimilarity index. To quantify the spread of SAR11 and *Roseobacter* genomes in the ordination, the permdisp2 procedure from VEGAN was used to look at multivariate homogeneity of group dispersions for each taxonomic class. Significance of group dispersion magnitude was assessed using permutation analysis, and the test was

performed to account for unequal sample size. Effect sizes of differences in dispersion between *Roseobacter* and SAR11 were assessed using Cohen's D, while significance was assessed using a Mann-Whitney *U* test. Classification schemes and genome features were fit as vectors to the ordination using the envit VEGAN function using 999 permutations to assess significance. Spearman's rank correlation coefficient was used to test for correlations between transporter abundances and Pearson's chi-squared test was used for categorical data. All multiple testings were corrected using the Benjamini Hochberg method for controlling the false discovery rate.

3.7 Acknowledgements

Chapter 3 is a reprint of the material as it appears in Hogle SL, Thrash JC, Dupont CL, Barbeau KA. (2016). Trace metal acquisition by heterotrophic bacterioplankton with contrasting trophic strategies. *Appl. Environ. Microbiol.* 82(5): 1613-1624. The dissertation author was the primary investigator and first author of this paper. This work was funded by NSF GRFP grant DGE-144086 to S.L.H., NSF grant OCE-1061068 to K.A.B and S.L.H. and the National Aeronautics and Space Administration through the NASA Astrobiology Institute under Cooperative Agreement No. NNA15BB03A to C.L.D.

Table 3.1 Metal transport components in *Roseobacter* and SAR11 genomes

Name	Function	Metal(s)
hupE/ureJ	Predicted secondary inner membrane Ni transporter, associated with hydrogenase/urease	Ni^{2+}
nikA	Type II solute binding protein	Ni^{2+}
corA	Inner membrane ion channel	Co^{2+}, Ni^{2+}, Mg
cbiM	Inner membrane permease	Co^{2+}
cbiQ	Inner membrane permease	Co^{2+}
cbtA	Predicted inner membrane channel	Co^{2+}
cbtB	Predicted inner membrane channel	Co^{2+}
btuF	TroA family solute binding protein	Co^{2+} as Vitamin B_{12}
copZ	Heavy metal binding domain with N-terminal Cu-interacting ATPase	$Cu^{+/2+}$ and other heavy metals
copA	P-type IB ATPase - Cu import and efflux	$Cu^{+/2+}$ and other heavy metals
NRAMP	Inner membrane permease	Fe^{2+}, Mn^{2+}
ZIP	Inner membrane permease	Fe^{2+}, Zn^{2+}
psaA	troA family solute binding protein	Mn^{2+}
hemV2	troA family solute binding protein	Mn^{2+}, Zn^{2+}
troA-a	troA family solute binding protein	Mn^{2+}, Zn^{2+}
troA-f	troA family solute binding protein	Mn^{2+}, Zn^{2+}
znuA	troA family solute binding protein	Zn^{2+}
afuA	Type II periplasmic binding family	Fe^{3+}
PBP2 Fbp-like 1	Type II periplasmic binding family	Fe^{3+}
PBP2 Fbp-like 2	Type II periplasmic binding family	Fe^{3+}
PBP2 Fbp-like 3	Type II periplasmic binding family	Fe^{3+}
PBP2 futA-like	Type II periplasmic binding family	Fe^{3+}
feoB	Inner membrane permease	Fe^{2+}
FTR1	Inner membrane permease	Fe^{2+}
hemS	Cytoplasmic Oxygenase/Chaperone	Heme/Hemoglobin
hutB	troA family solute binding protein	Heme/Hemoglobin
TBDT MCL1	TonB-dependent transporter	Heme/Hemoglobin
fatB	troA family solute binding protein	Catecholate Siderophores
TBDT MCL2	TonB-dependent transporter	Catecholate Siderophores
fhuD	troA family solute binding protein	Hydroxamate Siderophores
TBDT MCL3	TonB-dependent transporter	Hydroxamate Siderophores
fepB	troA family solute binding protein	Mixed / multiple siderophores
TBDT MCL4	TonB-dependent transporter	Mixed / multiple siderophores
entE	Siderophore biosynthesis	Siderophore biosynthesis (NIS or NRPS)
entF	Siderophore biosynthesis	Siderophore biosynthesis (NIS or NRPS)
rbhC	Siderophore biosynthesis	Siderophore biosynthesis (NIS or NRPS)

Table 3.1 Metal transport components in *Roseobacter* and SAR11 genomes, continued

Name	Function	Metal(s)
SIP	Cytoplasmic ferric reductase	Siderophore processing
TBDT MCL5	TonB-dependent transporter	Unknown
TBDT MCL6	TonB-dependent transporter	Unknown
TBDT MCLnull	TonB-dependent transporter	Unknown / siderophores
troA superfamily	Solute binding protein assignable only to troA superfamily	Unknown / Multiple

"Function" designates the molecular role of each gene/family, while "Metal(s)" indicates the metal or metal complex with which each family is known or predicted to interact.

Table 3.2 TBDT Markov Clusters and their occurrence in *Roseobacter* genomes

					Gene enrichment				
Cluster	Substrate Transported	Cluster Abun	Freq	Fur-box Freq	Gene/ PFAM	*P* value	Odds Ratio	Cluster Frac	Avg Gene Dist
MCL1	Heme	19	45%	85%	*hmuS*	4.0×10^{-45}	4225	0.95	1.29
					hutB	1.3×10^{-35}	5973	0.95	2.14
					FecCD	7.6×10^{-32}	94	1.1	4
					ExbD	1.8×10^{-22}	68	0.84	4.2
					TonB_2	3.9×10^{-19}	240	0.53	4.7
					MotA_ExbB	2.5×10^{-17}	54	0.67	3.2
					DUF4178	5.3×10^{-8}	189	0.21	8.5
MCL2	Catecholate Siderophores	18	26%	33%	*fatB*	3.6×10^{-19}	1332	0.56	2.25
					FecCD	2.9×10^{-34}	105	1.2	3.77
					DUF2218	2.7×10^{-15}	511	0.39	3.57
					MarR_2	4.8×10^{-11}	58	0.61	2
MCL3	Hydroxamate Siderophores	18	24%	55%	*fhuD*	1.5×10^{-8}	287	0.44	1
					FecCD	7.1×10^{-16}	51	0.67	2.3
					SIP	1.3×10^{-12}	148	0.33	3
					FAD_binding_9	1.8×10^{-10}	116	0.33	3
					HTH_18	1.5×10^{-8}	10	0.67	1
					ABC_membrane	2.7×10^{-6}	11	0.44	3.83
MCL4	Mixed Siderophores	6	14%	33%	FecCD	5.5×10^{-26}	199	2.33	6.36
					ExbD	1.5×10^{-21}	165	2	3.5
					SIP	2.9×10^{-13}	376	1	5
					FAD_binding_9	3.4×10^{-13}	364	1	5
					Peripla_BP_2	5.5×10^{-12}	110	1.16	7.42
					TonB_2	5.8×10^{-11}	313	0.83	1
					MotA_ExbB	1.4×10^{-9}	75	1	4.5
MCL5	Unknown – Iron?	5	12%	60%	TonB_C	2.5×10^{-10}	1784	0.8	4
					2OG-FeII_Oxy_3	2.3×10^{-8}	206	0.8	1
					FMN_dh	1.9×10^{-6}	63	0.8	5
					hemP	4.6×10^{-6}	138	0.6	3.3

Table 3.2 TBDT Markov Clusters and their occurrence in *Roseobacter* genomes, continued

					Gene enrichment				
Cluster	Substrate Transported	Cluster Abun	Freq	Fur-box Freq	Gene/ PFAM	*P* value	Odds Ratio	Cluster Frac	Avg Gene Dist
MCL6	Unknown	4	10%	0%	Peripla_ BP_2	5.0×10^{-7}	90	1	1
					TatD_D Nase	2.6×10^{-6}	170	0.75	4.7
MCL null	Mostly unknown, siderophore?	11	14%	27%	NA	NA	NA	NA	NA

TBDT sequence similarity clusters, their putatively ascribed function, and gene/PFAM enrichment for each cluster. “Cluster Abun.” is the total number of TBDTs in each cluster. “Freq.” indicates the percentage of Roseobacter genomes that contain at least one of the TBDTs from a given cluster. “Fur-box Freq.” denotes the percentage of TBDT in the cluster with a detectable upstream Fur-box binding motif. “Gene/PFAM” are genes from Table 1 and/or additional PFAMs that are enriched in the neighborhood of each TBDT cluster. Specific gene groups from Table 1 are listed preferentially over their parent PFAM superfamilies when both are significantly enriched (e.g., *hutB* is listed for MCL1 instead of its parent protein family Peripla_BP_2). For enriched genes/PFAMs, “p value” indicates the significance of enrichment (see Materials and Methods), “Odds Ratio” is the frequency of the gene/PFAM in the TBDT neighborhood divided by the frequency outside the neighborhood, “Cluster Frac.” is the total abundance of a gene/PFAM inside all neighborhoods of a particular cluster divided by the total number of TBDTs in that cluster, and “Avg. Gene Dist.” is the average number of genes separating a TBDT of the given cluster and the respective gene/PFAM.

Table 3.3 Trait depth of metal uptake genes with significant phylogenetic signal

		τ_D *Roseobacter*		τ_D SAR11	
Name	Metal(s)	Mean	Std dev	Mean	Std dev
afuA	Fe^{3+}	***0.0638***	***0.0040***		
PBP2 Fbp-like 1 *	Fe^{3+}	**0.0414**	**0.0020**	*0.0407*	*0.0017*
PBP2 FutA1-like ***	Fe^{3+}	***0.0565***	***0.0068***	*0.1781*	*0.0091*
hutB	Heme	*0.0274*	*0.0015*		
TBDT MCL2	Catecholate	*0.0331*	*0.0015*		
fatB	Catecholate	*0.0363*	*0.0016*		
TBDT MCL3	Hydroxamate	*0.0355*	*0.0015*		
fhuD	Hydroxamate	**0.0462**	**0.0021**		
SIP	Sidero. Biosyn. + proc.	**0.0476**	**0.0021**		
ZIP	Fe^{2+}, Zn^{2+}	*0.0386*	*0.0028*		
znuA ***	Zn^{2+}	*0.0271*	*0.0024*	**0.0562**	**0.0021**
psaA	Mn^{2+}	***0.0423***	***0.0056***		
troA-a	Mn^{2+}, Zn^{2+}	*0.0263*	*0.0014*		
cbtA	Co^{2+}	*0.0307*	*0.0018*		
cbtB	Co^{2+}	*0.0346*	*0.0024*		
corA ***	Co^{2+}, Ni^{2+}, Mg	*0.0400*	*0.0025*	**0.2785**	**0.0127**
btuF	Co^{2+} as Vitamin B12			**0.0809**	**0.0046**
copZ ***	$Cu^{+/2+}$, other heavy metals	*0.0268*	*0.0013*	*0.0070*	*0.0010*
TBDT MCL5	Unknown / Multiple	**0.0441**	**0.0025**		
Total TBDT ***	Multiple	***0.0493***	***0.0039***	*0.0652*	*0.0032*

Values in bold denote non-random phylogenetic distribution as assessed by the consenTRAIT algorithm ($P < 0.1$) and those in italics denote non-random phylogenetic distribution from independent D metric for phylogenetic dispersion of Fritz and Purvis ($P(D)_{random} < 0.05$). For all transporters shared between *Roseobacter* and SAR11 differences in trait depth for metal uptake traits are significant (Student's *t*-test, $P < 0.05$). Asterisks after transporter labels indicate effect size calculated using Cohen's D. A single asterisk represents a small effect size ($D < 0.5$), while three asterisks represent a large effect size ($D > 1$).

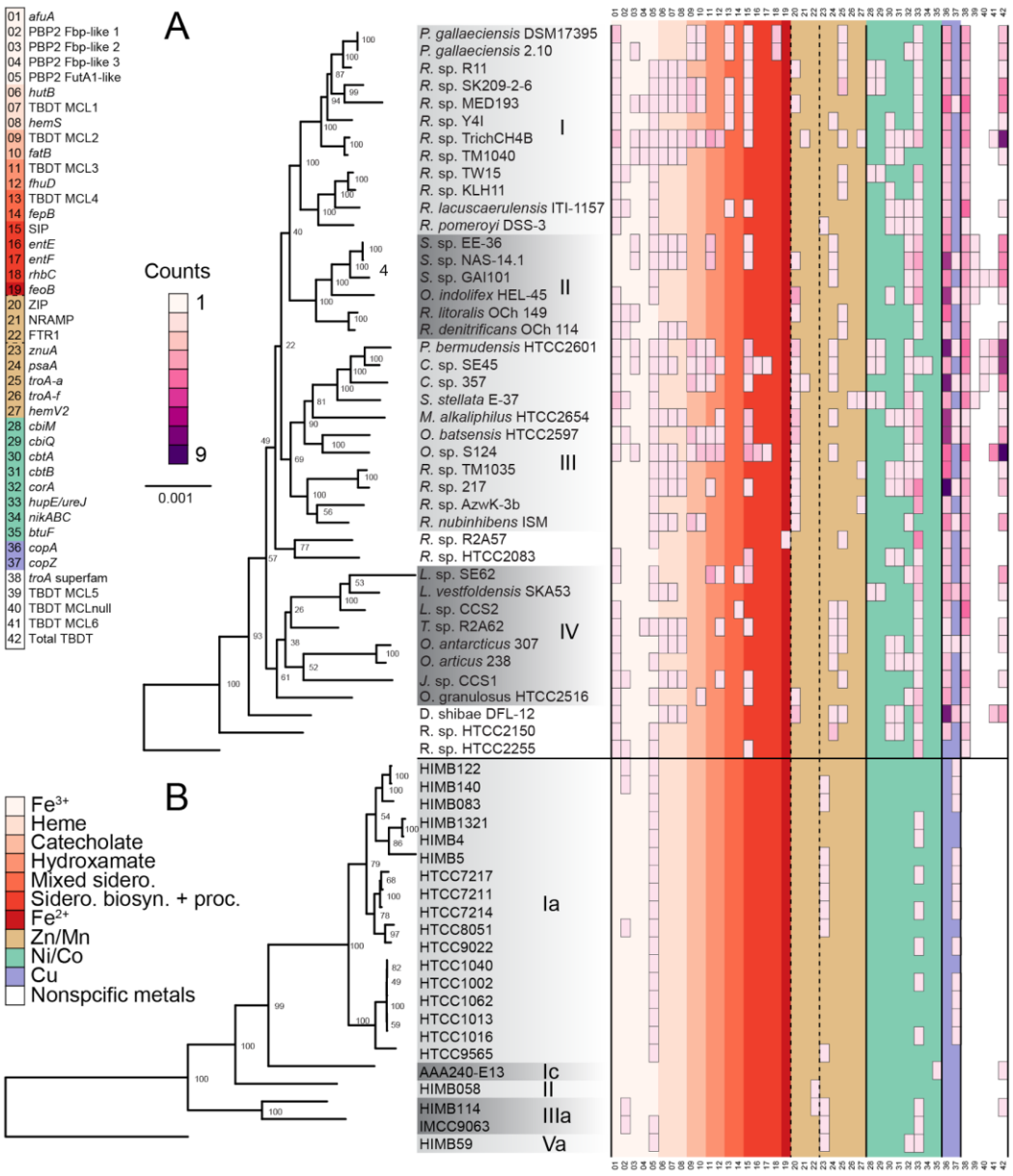


Figure 3.1 Maximum likelihood phylogenies of 42 *Roseobacter* genomes and 22 SAR11 genomes. Node values indicate bootstrap values from 1000 resamplings. Scale bar represents 0.001 substitutions per sequence position in both trees. Shaded boxes beneath *Roseobacter* names in (A) indicate the four major clades as presented in (Luo & Moran 2014), while those in (B) denote five major clades in the SAR11 group. The central grid represents the absolute abundance of each respective metal uptake system per genome, the values of which are represented by color in the scale labeled "Counts." Classes of metal uptake systems are partitioned by color as seen in the lower left corner. Specific metal uptake genes are referenced by number and color on the upper left side of the figure, where numbers correspond to rows in the central grid. NRAMP , FTR1, and ZIP transporters interact with Fe, Zn, and Mn and are indicated with dashed lines.

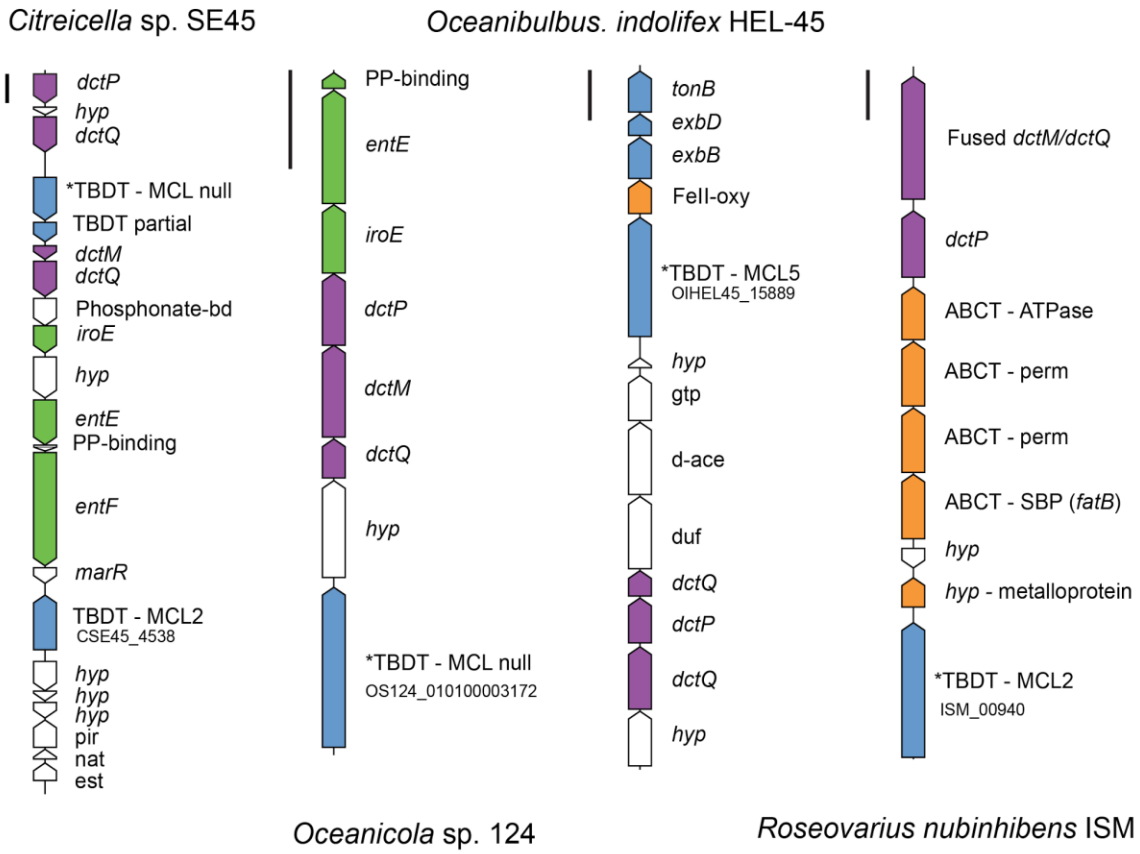


Figure 3.2 Examples of TRAP transporters in genome neighborhoods containing siderophore biosynthesis or siderophore-like uptake genes. Purple arrows are TRAP genes, green arrows represent genes likely involved in siderophore biosynthesis, light blue arrows represent TBDTs and associated energy transduction components, and orange arrows represent other genes related to Fe uptake and metabolism. Genes labeled with an asterisk have an upstream Fur box motif. The locus IDs of four TBDT are provided for reference. Scale bars in the upper left corner of each gene group represent 1000 base pairs. PP-binding, phosphopantetheine attachment site; entE, enterobactin synthase subunit E; entF, enterobactin synthase subunit F; iroE, putative enterobactin esterase; Phosphonate-bd, ABCT phosphonate SBP; dctQ, TRAP large permease component; dctM, TRAP small permease component; dctP, TRAP periplasmic component; marR, transcriptional regulator; pir, pirin-like Fe2+ containing protein; nat, GCN5-related N-acetyl-transferase; est, predicted esterase; tonB, periplasmic protein TonB; exbB, biopolymer transport protein exbD/tolR; exbD, motA/tolQ/exbB proton channel family; FeII-oxy, Fe2+ dependent oxygenase; gtp, putative GTPases (G3E family); d-ace, D-aminoacylases (N-acyl-D-Amino acid amidohydrolases); duf, protein of unknown function (DUF1485); ABCT, ATP Binding Cassette Transporter; perm, IM permease; SBP, inner membrane solute binding protein.

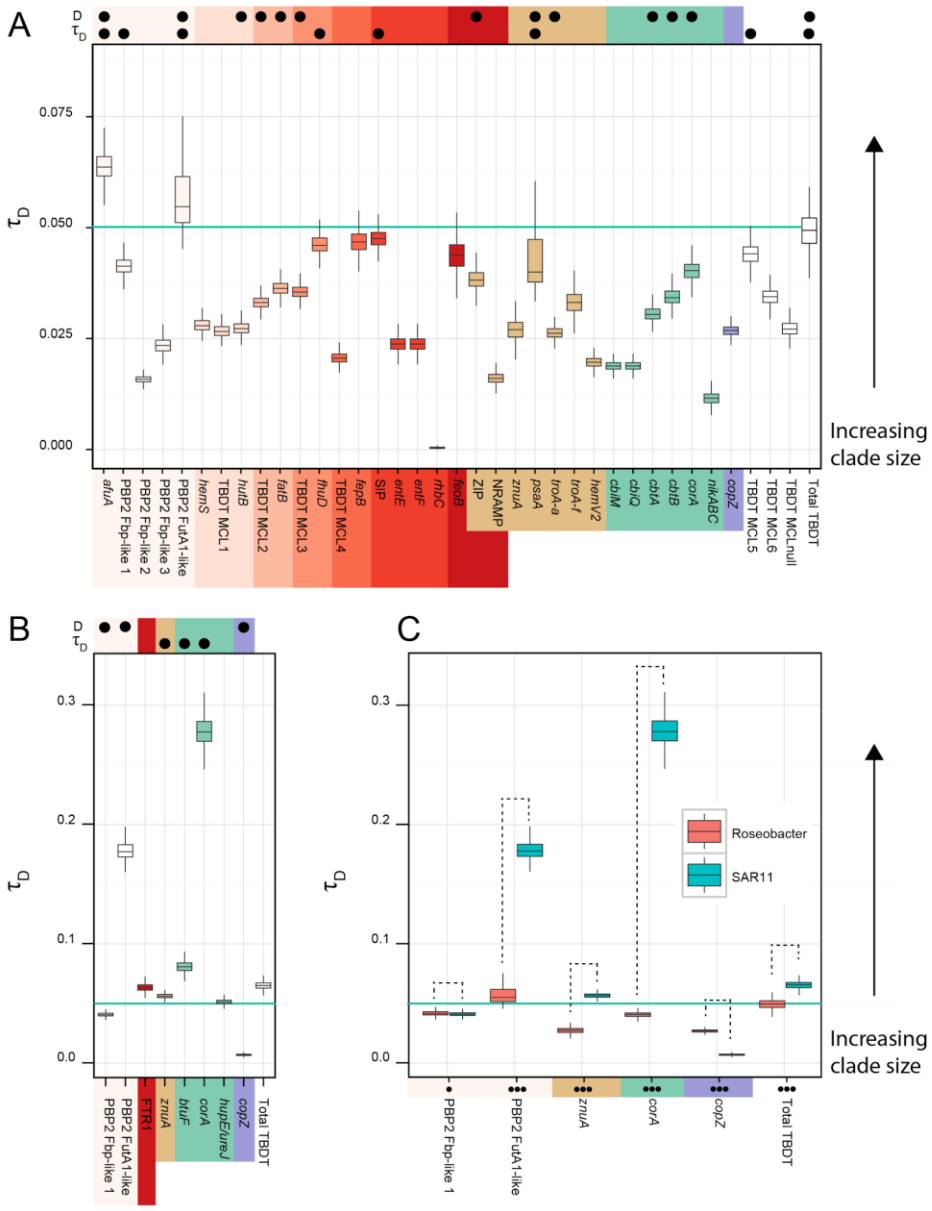


Figure 3.3 Boxplot showing trait depth (τ_D) of metal uptake genes with respect to *Roseobacter* and SAR11 phylogenies. Black circles above plots in (A) and (B) denote non-random phylogenetic distribution as assessed by the consenTRAIT algorithm (P < 0.1) and the independent D metric for phylogenetic dispersion of Fritz and Purvis ($P(D)_{random}$ < 0.05). (C) displays differences in trait depth for metal uptake traits shared between the SAR11 and roseobacter groups. Differences in mean are significant for all comparisons (Student's *t*-test, P < 0.05). Dots above transporter labels indicate effect size calculated using Cohen's D. Single dot represents a small effect size (D < 0.5), while three dots represent a large effect size (D > 1). Transporters are colored as in Figure 3.1. For visual reference, a teal line is marked at τ_D of 0.05 in each panel that corresponds to the maximum value for subtrees marked in Figure 3.1.

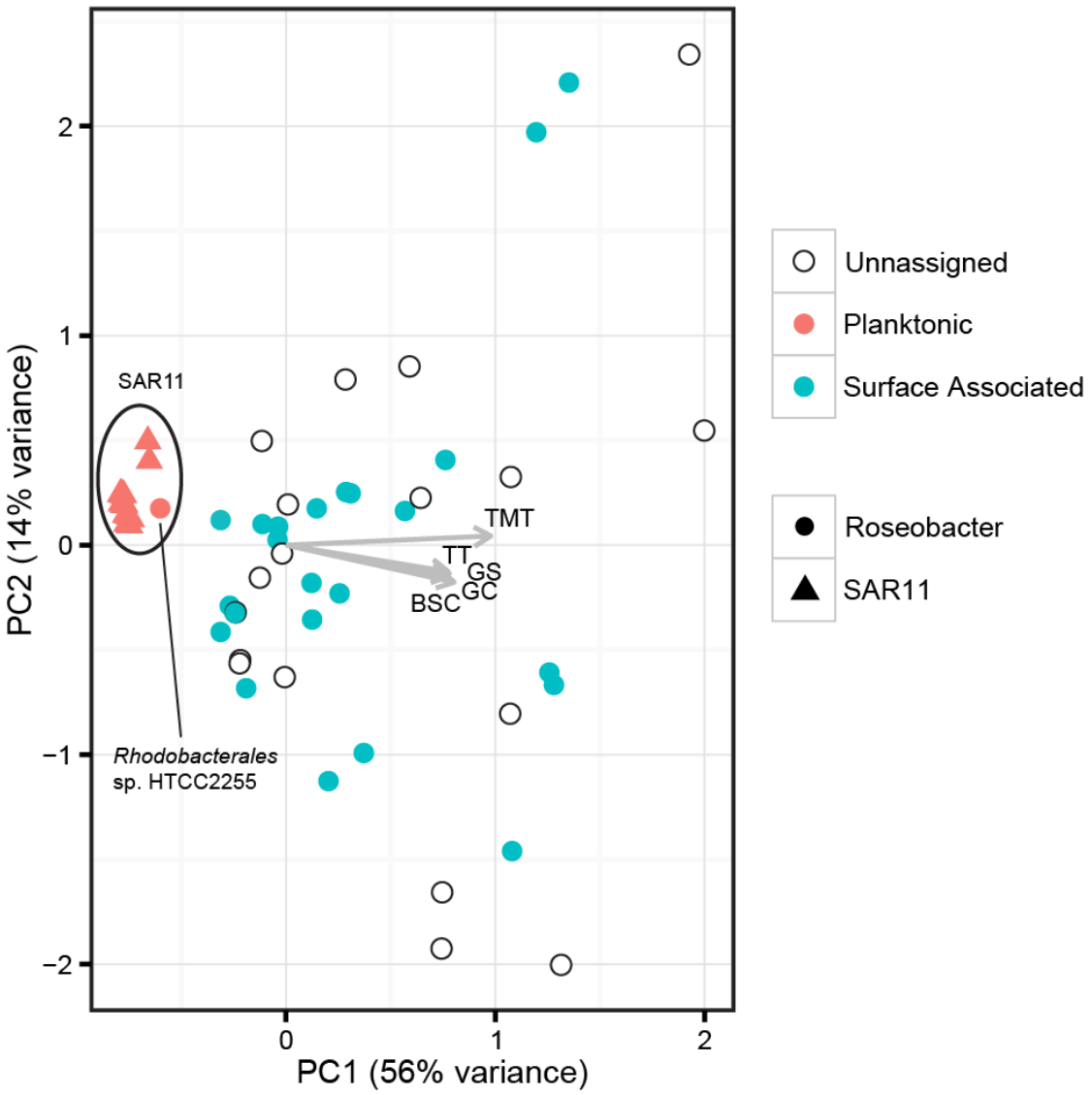


Figure 3.4 Principal Component Analysis (PCA) based on the diversity and abundance of transporters in *Roseobacter* and SAR11 genomes. The 42 *Roseobacter* genomes (circles) and 22 SAR11 genomes (triangles) are plotted with respect to whether the organism has been observed to be surface-associated (blue), planktonic (red), or with insufficient data to assign a lifestyle (white). Arrows represent fitted vectors of continuous associated variables (genome features) and show the direction of the increasing gradient. Arrow length is proportional to the correlation between the variable and ordination. TMT = total number of metal transporters per genome, $R^2 = 0.96$, $P = 0.001$; TT = total number of transporters per genome, $R^2 = 0.71$, $P < 0.001$; BSC = total number of predicted biosynthetic gene clusters per genome, $R^2 = 0.63$, $P < 0.001$; GC = GC content per genome, $R^2 = 0.67$, $P < 0.001$; GS = total number of predicted genes per genome, $R^2 = 0.63$, $P < 0.001$.

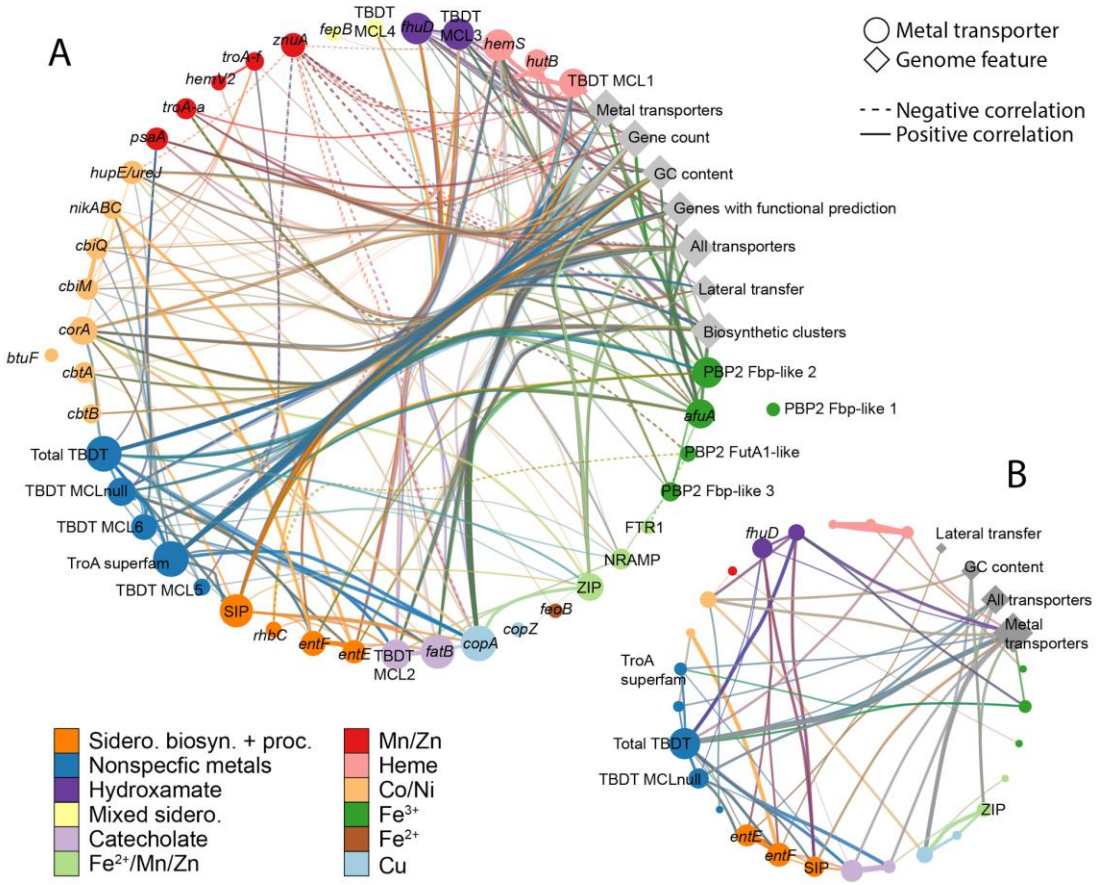


Figure 3.5 Network visualization of pairwise Spearman's rank correlation coefficients between metal uptake genes and genome features. (A) represents ρ values calculated from a combined SAR11 and Roseobacter dataset (N = 64), while (B) is calculated from Roseobacter only (N=42). Nodes are colored based on the category of metal uptake, and their size is proportional to the number of linked edges. Edges represent correlation coefficients between variables with $P < 0.05$. Solid edges represent positive correlations while dashed edges represent negative correlation. Edges are colored corresponding to nodes. Edge thickness and opacity is proportional to the magnitude of the correlation coefficient. Network layout was calculated using the attribute layout in Cytoscape and edges are bundled for clarity. In (B) singleton nodes and labels for some nodes are omitted for clarity.

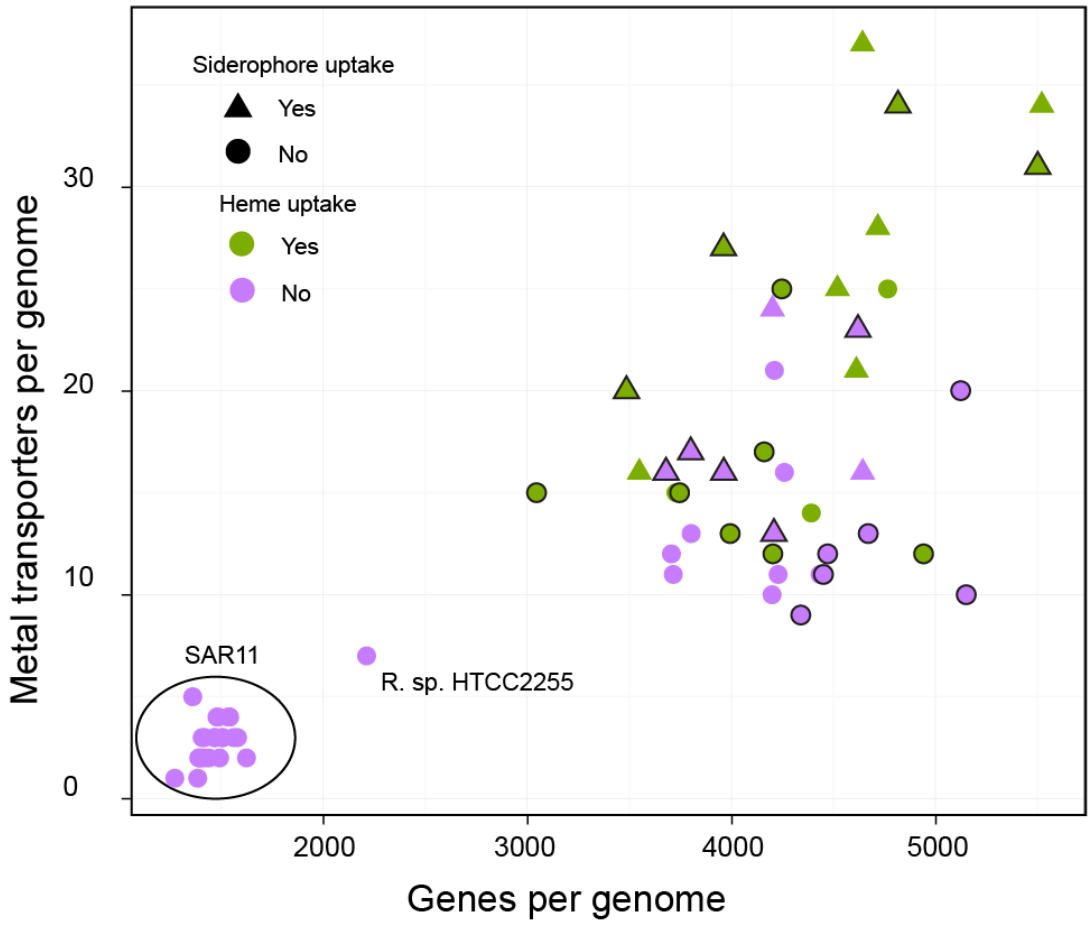


Figure 3.6 Scatterplot of metal transporters per genome versus total genes per genome for both *Roseobacter* and SAR11 groups. Each point represents one genome and is colored according to the genomic potential for heme uptake and shaped according to the genomic potential for siderophore uptake. Points with a black outline have been experimentally confirmed to be particle associated or were isolated from an abiotic or biotic surface. The SAR11 genomes and the HTCC225 genome are the only genomes with confirmed planktonic lifestyles. Spearman's rank correlation coefficient (ρ) for SAR11 genomes (N= 22), $\rho_{SAR11} = 0.37$, $P > 0.05$; for *Roseobacter* genomes (N = 42), $\rho_{Roseobacter} = 0.31$, $P < 0.05$; for combined genomes (N = 64), $\rho_{combined} = 0.78$, $P < 10^{-10}$.

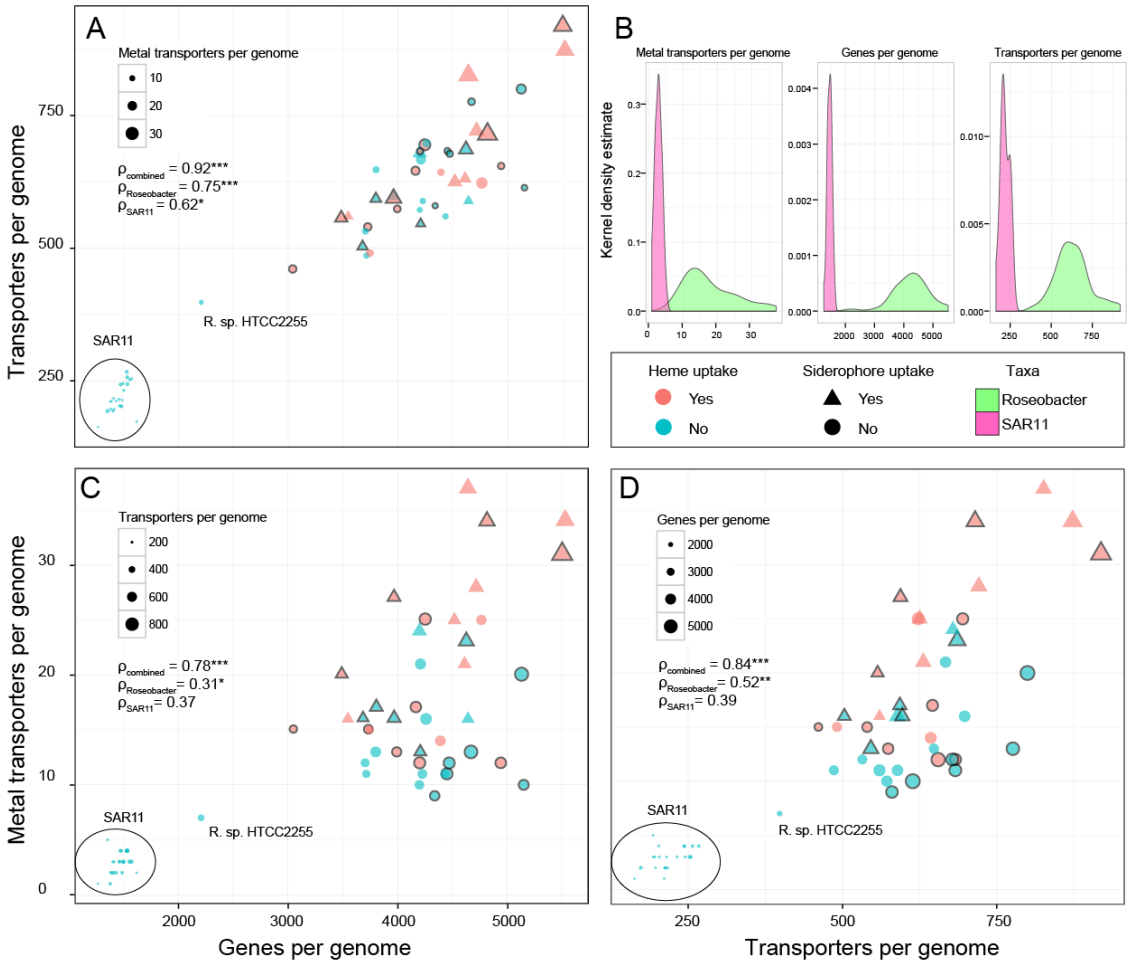


Figure 3.7 Matrix scatterplot of pairwise comparisons between three different genome features in SAR11 and *Roseobacter*. Scatterplots display pairwise comparisons of three different genome features (genes per genome, transporters per genome, and metal transporters per genome) determined from the 62 SAR11 and Roseobacter genomes. In (A), (C), and (D) each point represents one genome and is colored according to the genomic potential for heme uptake, shaped according to the genomic potential for siderophore uptake, and sized based on the genome feature omitted from the bivariate comparison. Points with a black outline have been experimentally confirmed to be particle associated or were isolated from an abiotic or biotic surface. The SAR11 genomes circled in black and the HTCC225 genome are the only genomes with confirmed planktonic lifestyles. In each scatterplot, Spearman's rank correlation coefficients (ρ) are displayed for SAR11 genomes only (N= 22), Roseobacter genomes only (N = 42), and the combined genomes from each group (N=64). Asterisks denote statistical significance; *** $P<1e^{-10}$, ** $P<1e^{-3}$, * $P<0.05$. (B) displays kernel density estimates for each of (A, C and D) and is colored by SAR11 and Roseobacter groups.

3.8 References

1. Morel FMM, Price NM (2003) The biogeochemical cycles of trace metals in the oceans. *Science* 300:944–947.

2. Mackey KR, Post AF, McIlvin MR, Cutter G, John SG, Saito M. (2015) Divergent responses of Atlantic coastal and oceanic Synechococcus to iron limitation. *Proc Natl Acad Sci U S A* 112(32):201509448.

3. Moore JK, Doney SC, Glover DM, Fung IY (2001) Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep Sea Res Part 2 Top Stud Oceanogr* 49:463–507.

4. Rue EL, Bruland KW (1995) Complexation of iron(III) by natural organic ligands in the Central North Pacific as determined by a new competitive ligand equilibration/adsorptive cathodic stripping voltammetric method. *Mar Chem* 50:117–138.

5. Coale KH, Bruland KW (1988) Copper complexation in the Northeast Pacific. *Limnol Oceanogr* 33(5):1084–1101.

6. Saito MA, Moffett JW (2002) Temporal and spatial variability of cobalt in the Atlantic Ocean. *Geochim Cosmochim Acta* 66(11):1943–1953.

7. Xue HB, Jansen S, Prasch A, Sigg L (2001) Nickel Speciation and Complexation Kinetics in Freshwater by Ligand Exchange and DPCSV. *Environ Sci Technol* 35(3):539–546.

8. Kim J-M, Baars O, Morel FMM (2015) Bioavailability and Electroreactivity of Zinc Complexed to Strong and Weak Organic Ligands. *Environ Sci Technol* 49(18):10894–10902.

9. Ma Z, Jacobsen FE, Giedroc DP (2009) Coordination chemistry of bacterial metal transport and sensing. *Chem Rev* 109(10):4644–4681.

10. Nies DH, Silver S (2007) *Molecular Microbiology of Heavy Metals* (Springer Berlin Heidelberg).

11. Forbes JR, Gros P (2001) Divalent-metal transport by NRAMP proteins at the interface of host-pathogen interactions. *Trends Microbiol* 9(8):397–403.

12. Grass G, Franke S, Taudte N, Dietrich H, Kucharski LM, Maguire ME, Rensing C, Nies DH. (2005) The Metal Permease ZupT from Escherichia coli Is a Transporter with a Broad Substrate Spectrum. *J Bacteriol* 187(5):1604–1611.

13. Große C, Scherer J, Koch D, Otto M, Taudte N, Grass G. (2006) A new ferrous iron-uptake transporter, EfeU (YcdN), from Escherichia coli. *Mol Microbiol* 62(1):120–131.

14. Cartron ML, Maddocks S, Gillingham P, Craven CJ, Andrews SC (2006) Feo--transport of ferrous iron into bacteria. *Biometals* 19(2):143–157.

15. Carpenter BM, Whitmire JM, Merrell DS (2009) This is not your mother's repressor: the complex role of fur in pathogenesis. *Infect Immun* 77(7):2590–2601.

16. Rodionov D, Gelfand M, Todd J, Curson A, Johnston A (2006) Computational reconstruction of iron- and manganese-responsive transcriptional networks in alpha-proteobacteria. *PLoS Comput Biol* 2(12):e163.

17. Crosa JH, Walsh CT (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* 66(2):223–249.

18. Oves-Costales D, Kadi N, Challis GL (2009) The long-overlooked enzymology of a nonribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis. *Chem Commun* (43):6530–6541.

19. Zawadzka AM, Abergel RJ, Nichiporuk R, Andersen UN, Raymond KN (2009) Siderophore-mediated iron acquisition systems in Bacillus cereus: Identification of receptors for anthrax virulence-associated petrobactin. *Biochemistry* 48(16):3645–3657.

20. Clarke TE, Braun V, Winkelmann G, Tari LW, Vogel HJ (2002) X-ray crystallographic structures of the Escherichia coli periplasmic protein FhuD bound to hydroxamate-type siderophores and the antibiotic albomycin. *J Biol Chem* 277(16):13966–13972.

21. Miethke M, Hou J, Marahiel M a. (2011) The siderophore-interacting protein YqjH acts as a ferric reductase in different iron assimilation pathways of Escherichia coli. *Biochemistry* 50(50):10951–10964.

22. Hogle SL, Barbeau KA, Gledhill M (2014) Heme in the marine environment: from cells to the iron cycle. *Metallomics* 6(6):1107–1120.

23. Claverys J (2001) A new family of high-affinity ABC manganese and zinc permeases. *Res Microbiol* 152(3-4):231–243.

24. Dintilhac A, Alloing G, Granadel C, Claverys JP (1997) Competence and virulence of Streptococcus pneumoniae: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol* 25(4):727–739.

25. Patzer S, Hantke K (1998) The ZnuABC high-affinity zinc uptake system and its regulator Zur in Escherichia coli. *Mol Microbiol* 28(6):1199–1210.

26. Cadieux N, Bradbeer C (2002) Identification of the periplasmic cobalamin-binding protein BtuF of Escherichia coli. *J Bacteriol* 184(3):706–717.

27. Navarro C, Wu L, Mandrand-Berthelot M (1993) The nik operon of Escherichia coli

encodes a periplasmic binding-protein-dependent transport system for nickel. *Mol Microbiol* 9(6):1181–1191.

28. Eitinger T, Suhr J, Moore L, Smith JAC (2005) Secondary transporters for nickel and cobalt ions: theme and variations. *Biometals* 18(4):399–405.

29. Brito B, Prieto R, Cabrera E, Mandrand-Berthelot M, Imperial J, Ruiz-Argüeso T, Palacios J. (2010) Rhizobium leguminosarum hupE encodes a nickel transporter required for hydrogenase activity. *J Bacteriol* 192(4):925–935.

30. Rodionov DA, Hebbeln P (2006) Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J Bacteriol* 188(1):317–327.

31. Rodionov D, Vitreschak A, Mironov A, Gelfand M (2003) Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem* 278(42):41148–41159.

32. Xia Y, Lundbäck AK, Sahaf N, Nordlund G, Brzezinski P, Eshaghi S. (2011) Co^{2+} selectivity of Thermotoga maritima CorA and its inability to regulate Mg^{2+} homeostasis present a new class of CorA proteins. *J Biol Chem* 286(18):16525–16532.

33. Rensing C, Fan B, Sharma R, Mitra B, Rosen BP (2000) CopA: An Escherichia coli Cu(I)-translocating P-type ATPase. *Proc Natl Acad Sci U S A* 97(2):652–656.

34. Hassani BK, Astier C, Nitschke W, Ouchane S (2010) CtpA, a copper-translocating P-type ATPase involved in the biogenesis of multiple copper-requiring enzymes. *J Biol Chem* 285(25):19330–19337.

35. Radford D, Kihlken M, Borrelly G, Harwood C, Le Brun N, Cavet J. (2003) CopZ from Bacillus subtilis interacts in vivo with a copper exporting CPx-type ATPase CopA. *FEMS Microbiol Lett* 220(1):105–112.

36. Jordan IK, Natale D, Galperin M (2000) Copper chaperones in bacteria: Association with copper-transporting ATPases. *Trends Biochem Sci* 25(10):480–481.

37. Giovannoni SJ, Cameron Thrash J, Temperton B (2014) Implications of streamlining theory for microbial ecology. *ISME J* 8(8):1553–1565.

38. Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, Howard EC, King E, Oakley C, Reisch CR, Rinta-Kanto JM, Sharma S, Sun S, Varaljay V, Vila-Costa M, Westrich JR, Moran MA. (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME J* 4(6):784–798.

39. Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, DeMaere MZ, Ting L, Ertan H, Johnson J, Ferriera S, Lapidus A, Anderson I, Kyrpides N, Munk C, Detter C, Han CS, Brown MV, Robb FT, Kjelleberg S, Cavicchioli R. (2009) The

genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* 106:15527–15533.

40. Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M, Meng J, Sun S, Medeiros PM, Paul JH, Coles VJ, Yager PL, Moran M. (2014) Microspatial gene expression patterns in the Amazon River Plume. *Proc Natl Acad Sci U S A* 111(30):11085–11090.

41. Smith DC, Simon M, Alldredge AL, Azam F (1992) Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. *Nature* 359(10):139–142.

42. DeLong EF, Franks DG, Alldredge AL (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* 38(5):924–934.

43. Luo H, Moran MA (2015) How do divergent ecological strategies emerge among marine bacterioplankton lineages? *Trends Microbiol* 23(9):577–584.

44. Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA (2014) Evolutionary analysis of a streamlined lineage of surface ocean Roseobacters. *ISME J* 8(7):1428–1439.

45. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, Luo H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* 110(28):11463–11468.

46. Buchan A, LeCleir GR, Gulvik C a., González JM (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* 12(10):686–698.

47. Riemann L, Steward GF, Azam F (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol* 66(2):578–587.

48. Vergin KL, Beszteri B, Monier A, Thrash JC, Temperton B, Treusch AH, Kilpert F, Worden AZ, Giovannoni SJ. (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J* 7(7):1322–1332.

49. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS. (2012) Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* 3(5):e00252–12.

50. Luo H, Csuros M, Hughes AL, Moran MA (2013) Evolution of divergent life history

strategies in marine alphaproteobacteria. *MBio* 4(4):1–8.

51. Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC (2010) Characterization of Prochlorococcus clades from iron-depleted oceanic regions. *Proc Natl Acad Sci U S A* 107(37):16184–16189.

52. Carini P, Van Mooy BA, Thrash JC, White A, Zhao Y, Campbell EO, Fredricks HF, Giovannoni SJ. (2015) SAR11 lipid renovation in response to phosphate starvation. *Proc Natl Acad Sci U S A* 112(25):7767–7772.

53. Coleman ML, Chisholm SW (2010) Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci U S A* 107(43):18634–18639.

54. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabo G, Polz MF, Alm EJ. (2012) Population Genomics of Early Events in the Ecological Differentiation of Bacteria. *Science* 336(2012):48–51.

55. Yawata Y, Cordero OX, Menolascina F, Hehemann JH, Polz MF, Stocker R. (2014) Competition-dispersal tradeoff ecologically differentiates recently speciated marine bacterioplankton populations. *Proc Natl Acad Sci U S A* 111(15):5622–5627.

56. Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Alexander Richter R, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS, Nealson K, Friedman R, Craig Venter J. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* 6(6):1186–1199.

57. Pedler BE, Aluwihare LI, Azam F (2014) Single Bacterial Strain Capable of Significant Contribution to Carbon Cycling in the Surface Ocean. *Proc Natl Acad Sci U S A* 111(20):7202–7207.

58. Tam R, Saier MH (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Mol Biol Rev* 57(2):320–346.

59. Baichoo N, Helmann JD (2002) Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence. *J Bacteriol* 184(21):5826–5832.

60. Mulligan C, Fischer M, Thomas GH (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* 35(1):68–86.

61. Schauer K, Rodionov D, de Reuse H (2008) New substrates for TonB-dependent transport: do we only see the “tip of the iceberg”? *Trends Biochem Sci* 33(7):330–338.

62. Mulligan C, Geertsma ER, Severi E, Kelly DJ, Poolman B, Thomas GH. (2009) The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc Natl Acad Sci U S A* 106(6):1778–1783.

63. Yoshizawa S, Kumagai Y, Kim H, Ogura Y, Hayashi T, Iwasaki W, DeLong EF, Kogure K. (2014) Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc Natl Acad Sci U S A* 111(18):6732–6737.

64. Sandy M, Butler A (2009) Microbial iron acquisition: marine and terrestrial siderophores. *Chem Rev* 109(10):4580–4595.

65. Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, Pedrós-Alió C. (2013) Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* 7(5):1026–1037.

66. Martiny AC, Treseder K, Pusch G (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME J* 7(4):830–838.

67. Fritz S a., Purvis A (2010) Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conserv Biol* 24(4):1042–1051.

68. Slightom RN, Buchan A (2009) Surface colonization by marine roseobacters: Integrating genotype and phenotype. *Appl Environ Microbiol* 75(19):6027–6037.

69. Saier M, Reddy V, Tamang D, Västermark Å (2014) The transporter classification database. *Nucleic Acids Res* 42(D1):251–258.

70. Hopkinson B, Barbeau K (2012) Iron transporters in marine prokaryotic genomes and metagenomes. *Environ Microbiol* 14(1):114–128.

71. Roy EG, Wells ML, King DW (2008) Persistence of iron(II) in surface waters of the western subarctic Pacific. *Limnol Oceanogr* 53(1):89–98.

72. Nissen H, Nissen P, Azam F (1984) Multiphasic uptake of D-glucose by an oligotrophic marine bacterium. *Mar Ecol Prog Ser* 16:155–160.

73. Azam F, Hodson R (1981) Multiphasic Kinetics for D-Glucose Uptake by Assemblages of Natural Marine Bacteria. *Mar Ecol Prog Ser* 6:213–222.

74. Lee JW, Helmann JD (2007) Functional specialization within the fur family of metalloregulators. *Biometals* 20(3-4):485–499.

75. Vraspir JM, Butler A (2009) Chemistry of Marine Ligands and Siderophores. *Ann Rev Mar Sci* 1(1):43–63.

76. Cordero OX, Ventouras L-A, DeLong EF, Polz MF (2012) Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc Natl Acad Sci U S A* 109(49):20059–20064.

77. Kümmerli R, Schiessl KT, Waldvogel T, McNeill K, Ackermann M (2014) Habitat structure and the evolution of diffusible siderophores in bacteria. *Ecol Lett*

17(12):1536–1544.

78. Gledhill M, Buck KN (2012) The organic complexation of iron in the marine environment: a review. *Front Microbiol* 3(February):69.

79. Thole S, Kalhoefer D, Voget S, Berger M, Engelhardt T, Liesegang H, Wollherr A, Kjelleberg S, Daniel R, Simon M, Thomas T, Brinkhoff T. (2012) Phaeobacter gallaeciensis genomes from globally opposite locations reveal high similarity of adaptation to surface life. *ISME J* 6(12):2229–2244.

80. Luo H, Löytynoja A, Moran MA (2012) Genome content of uncultivated marine Roseobacters in the surface ocean. *Environ Microbiol* 14(1):41–51.

81. Biers EJ, Sun S, Howard EC (2009) Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Appl Environ Microbiol* 75(7):2221–2229.

82. Smith DP, Kitner JB, Norbeck AD, Clauss TR, Lipton MS, Schwalbach MS, Steindler L, Nicora CD, Smith RD, Giovannoni SJ. (2010) Transcriptional and translational regulatory responses to iron limitation in the globally distributed marine bacterium Candidatus pelagibacter ubique. *PLoS One* 5(5):e10487.

83. Roe KL, Hogle SL, Barbeau K (2013) Utilization of heme as an iron source by marine alphaproteobacteria in the roseobacter clade. *Appl Environ Microbiol* 79(18):5753–5762.

84. Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER (2011) Dependence of the cyanobacterium Prochlorococcus on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PLoS One* 6(2):e16805.

85. Ribalet F, Swalwell J, Clayton S, Jiménez V, Sudek S, Lin Y, Johnson ZI, Worden AZ, Armbrust EV. (2015) Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. *Proc Natl Acad Sci U S A* 112(26):8008–8012.

86. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC, Scholin CA, DeLong EF. (2014) Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* 345(6193):207–212.

87. Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, Barofsky DF, Giovannoni SJ. (2011) Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J* 5(5):856–865.

88. Learman D, Hansel C (2014) Comparative proteomics of Mn(II)-oxidizing and non-oxidizing Roseobacter clade bacteria reveal an operative manganese transport system but minimal Mn(II)-induced expression of manganese oxidation and antioxidant enzymes. *Environ Microbiol Rep* 6(5):501–509.

89. Zimmerman AE, Martiny AC, Allison SD (2013) Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J* 7(6):1187–1199.

90. Dupont CL, McCrow JP, Valas R, Moustafa A, Walworth N, Goodenough U, Roth R, Hogle SL, Bai J, Johnson ZI, Mann E, Palenik B, Barbeau K, Craig Venter J, Allen AE. (2015) Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J* 9(5):1076–1092.

91. Dupont CL, Larsson J, Yooseph S, Ininbergs K, Goll J, Asplund-Samuelsson J, McCrow JP, Celepli N, Allen LZ, Ekman M, Lucas AJ, Hagström Å, Thiagarajan M, Brindefalk B, Richter AR, Andersson AF, Tenney A, Lundin D, Tovchigrechko A, Nylander J, Brami D, Badger JH, Allen AE, Rusch DB, Hoffman J, Norrby E, Friedman R, Pinhassi J, Venter JC, Bergman B. (2014) Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* 9(2):e89549.

92. Ganesh S, Parris DJ, DeLong EF, Stewart FJ (2014) Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J* 8(1):187–211.

93. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40(Database issue):D115–22.

94. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25(7):1043–1055.

95. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita R, Zhang D, Zheng C, Bryant SH. (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(D1):D222–D226.

96. Winterberg B, Uhlmann S, Linne U, Lessing F, Marahiel M, Eichhorn H, Kahmann R, Schirawski J. (2010) Elucidation of the complete ferrichrome A biosynthetic pathway in Ustilago maydis. *Mol Microbiol* 75(5):1260–1271.

97. Enright a. J, Dongen SV, Ouzounis C a. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30(7):1575–1584.

98. Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, Bader GD, Ferrin TE. (2011) clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* 12(1):436.

99. Luo H (2015) Evolutionary origin of a streamlined marine bacterioplankton lineage. *ISME J* 9(6):1423–1433.

100. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.

101. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17(4):540–552.

102. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7):1307–1320.

103. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MH, Wagner H. (2013) vegan: Community Ecology Package. Available at: http://cran.r-project.org/package=vegan.

Chapter 4

Heme as a molecular intermediate in the microbial remineralization of marine biogenic Fe

4.1 Abstract

The degradation and remineralization of biogenic matter is a fundamental process shaping marine biogeochemical cycles, including that of iron. Ecosystem productivity in many regions of the surface ocean is fueled primarily by iron recycled from biomass, but the intermediate molecular compounds involved in this process are poorly understood. Here we characterized a heme uptake system, a microbial iron acquisition mechanism predominantly studied in the context of human infections, in the nonpathogenic marine *Roseobacter* strain *Ruegeria* sp.TM1040. The heme uptake system was strongly upregulated when TM1040 was subjected to Fe stress, and TM1040 was able to fulfill its Fe requirements in a minimal medium by solely utilizing heme or hemoglobin. A mutant TM1040 strain in which the gene encoding the outer membrane TonB dependent transporter *hmuR* was insertionally inactivated lost the ability to grow on heme sources confirming the role of *hmuR* and/or its associated gene cluster in heme uptake. Wild type TM1040 outcompeted the mutant strain when co-cultured in Fe-deficient conditions with diatom lysate as the only provided Fe source. Roughly half of the 153 different marine *Roseobacter* genomes contain heme uptake systems, and one strain significantly downregulated its heme uptake system during symbiotic growth with a marine diatom. We argue that many phytoplankton-associated *Roseobacter* strains likely employ heme uptake systems to scavenge Fe from hemoproteins when their algal partners lyse and die

and that these systems provide them with a competitive advantage while transitioning to a new lifestyle. As a result, heme may be an important chemical currency during the microbial remineralization of marine phytoplankton and may be an important chemical intermediate connecting different biogenic Fe pools in the ocean.

4.2 Introduction

Fe is a critical micronutrient for both marine phytoplankton and heterotrophic bacterioplankton and can limit overall ecosystem productivity in nearly 40% of the surface ocean (1, 2). Over the last three decades much has been learned about the biogeochemical cycle of Fe in the oceans (3), but still relatively little is known about the specific molecules underlying global processes in the marine Fe cycle. Many of the biogeochemical transformations of Fe are facilitated by the activities of the ocean microbiome, which is increasingly being recognized as a microscale structured, dynamic, and heterogeneous landscape occupied by organisms employing an enormous diversity of ecological strategies (4). Recent work has highlighted the role of phytoplankton-bacteria interactions (5) in modulating overall ecosystem primary productivity and has largely focused on mutualistic interactions between heterotrophic bacteria and marine phytoplankton (6, 7). Antagonistic phytoplankton-bacteria interactions are also important in shaping the development of phytoplankton blooms (8, 9), and interactions between marine bacteria and sinking particulate organic matter, much of which is living or dead phytoplankton cells, affect the export of carbon to the deep ocean (10).

Marine phytoplankton are responsible for about half of net global primary productivity, but nearly half of marine primary production is consumed by heterotrophic

marine bacteria through the microbial loop (10). Additionally, heterotrophic marine bacteria can consume a large amount of Fe through the microbial loop, and up to 45% of total biological Fe uptake in Fe-limited waters could be due to heterotrophic bacteria (11). Currently, little is known about the exact molecular units exchanged between biogeochemical Fe pools in the marine environment; for example, the Fe-containing molecules that marine bacteria consume during the remineralization of lysed or decaying phytoplankton. Specific Fe-containing molecules are probably uniquely bioavailable to different fractions of the bacterial communities in seawater (12), but there is a lack of experimental data constraining relationships between microbes and Fe-containing molecules (13). To contribute to addressing this knowledge gap, we characterized a heme uptake system in the marine *Roseobacter* strain *Ruegeria* sp. TM1040. Heme is a heterocyclic enzyme cofactor containing an Fe atom as the reactive center, is biologically ubiquitous, and is a widespread Fe-containing molecule in the marine environment (14). We sought to explore the role of heme as an Fe source for TM1040 in order to understand the ecological conditions under which heme uptake is an advantageous biological trait; to examine the distribution of heme uptake in other marine bacterial genomes; and to gain broader insight into the role of heme in the marine biogeochemical cycle of Fe.

4.3 Results

4.3.1 TM1040 genome contains a putative heme uptake genetic locus

The TM1040 genome contains ten co-localized genes (Figure 4.1) organized similarly to heme uptake genes clusters in other marine bacteria (15). The TM1040 heme uptake clusters have been bioinformatically predicted to be involved in heme uptake (12,

16) but have not been functionally characterized. The ten genes are organized in two divergently transcribed operon-like structures and include putative components of an ATP binding cassette transporter (ABCT) system (*hmuTUV*), a heme oxygenase (*hmuS*), an outer membrane TonB dependent transporter (TBDT) (*hmuR*), and the energy transduction components of the canonical TBDT system (*exbB*, *exbD1*, *tonB*). A duplicate of the *exbD1* gene (*exbD1a*) and a hypothetical protein are also included in the putative TM1040 heme uptake operon (Figure 4.1). The top blast hit of the TM1040 *hmuR* gene to the curated Swiss-Prot database (17) is to *Haemophilus ducreyi* (23.6% amino acid identity, UniProtKB - Q7VNU1) while the top hit of TM1040 *hmuS* is to *Yersinia pestis* (36.8% amino acid identity, UniProtKB: Q56990).

4.3.2 TM1040 heme uptake locus is upregulated under Fe stress and co-transcribed

We examined the expression of the ten putative heme uptake genes under Fe-limiting (no added Fe) and Fe-replete (5 μM added $FeCl_3$) conditions using two-step RT-qPCR. All genes in the putative heme uptake operons were upregulated between approximately 41 and 201-fold under Fe stress relative to Fe-replete conditions (Figure 4.2). The three most highly expressed genes were the putative TBDT (*hmuR*), the solute binding protein of the ABCT system (*hmuT*), and the hypothetical protein (hyp). The putative heme oxygenase (*hmuS*), which has been used as a diagnostic gene for identifying heme uptake gene clusters, was also upregulated more than 140-fold. We also analyzed co-transcription of consecutive genes in the putative heme uptake locus as has been described previously (18). Although the ten components of the heme uptake locus

were not upregulated at uniform levels under Fe stress, the genes in each divergent cluster were co-transcribed. RT-PCR amplification products spanning genes junctions of the six genes in the *hmuR* cluster and four genes in the *hmuS* cluster were observed in all cases (Figure 4.3), indicating co-transcription on the respective strands.

4.3.3 TM1040 can utilize heme or hemoglobin as sole Fe sources

Because the TM1040 genome contains a putative heme uptake gene cluster that we observed to be strongly upregulated under Fe stress, we examined the growth of TM1040 on inorganic $FeCl_3$, hemin chloride (heme), and the model hemoproteins human hemoglobin (Hb) and cytochrome c (cyt c). The specific growth rate of TM1040 in an Fe-depleted medium with 500 nM of added inorganic $FeCl_3$ was 0.31 ± 0.02 hr^{-1} while it was significantly lower ($\mu_{TM1040\ NoFe} = 0.13 \pm 0.004$ hr^{-1}) with no added Fe source. TM1040 grew comparably with $FeCl_3$ conditions when we provided 500 nM heme as the sole Fe source ($\mu_{TM1040\ Heme} = 0.28 \pm 0.01$ hr^{-1}). We then supplied TM1040 with 17 μg/mL Hb and 500 nM cyt c, in order to determine whether TM1040 could utilize more structurally complex hemoproteins. In cyt c the heme prosthetic group is covalently bound to the parent protein while heme is noncovalently embedded in the human Hb tetramer. TM1040 was able to grow equally well with human Hb ($\mu_{TM1040\ Hb} = 0.28 \pm 0.01$ hr^{-1}) as with $FeCl_3$ but not cyt c ($\mu_{TM1040\ cytc} = 0.14 \pm 0.01$ hr^{-1}), suggesting that the structural nature of the parent protein is an important factor in determining heme bioavailability from hemoproteins (Table 1, Figure 4.4).

4.3.4 Partial deletion of heme outer membrane transporter eliminates TM1040 growth on heme

We selected the TM1040 *hmuR* gene for insertional inactivation (Figure 4.5, Figure 4.6) in order to determine whether this particular gene cluster was responsible for TM1040's observed growth on heme and human Hb. We targeted the TM1040 *hmuR* gene because it was the most dramatically upregulated under Fe stress and also encodes the putative outer membrane TBDT, which is the first step in the transfer of heme from the extracellular environment into the bacterial cytoplasm. We deleted a 435 bp portion of the TM1040 *hmuR* gene and substituted the deletion with a kanamycin resistance cassette, generating *Ruegeria* sp. TM1040 strain LH02 (Table 2, Figs. S3, S4). Wild type TM1040 and LH02 both grew equally well when supplied with $FeCl_3$ as an Fe source, but LH02 growth on heme and hemoglobin was greatly reduced (Table 1, Figure 4.4) indicating that the *hmuR* gene cluster is necessary for TM1040 growth on heme. Downstream polar effects on the hypothetical protein and components of the TonB energy transduction system from *hmuR* insertional inactivation may also be responsible for the observed decrease in LH02 growth.

4.3.5 TM1040 outperforms heme uptake mutant when diatom lysate is provided as the sole Fe source

In order to simulate natural conditions where TM1040 cells would be in contact with heterogeneous dissolved and particulate organic matter, we tested whether the wild type TM1040 strain was able to outperform the heme uptake mutant LH02 when soluble and insoluble fractions of lysed diatom cells were supplied as the sole Fe source. We

tested whether the wild type TM1040 or LH02 strain would numerically outcompete the other when co-cultured with 500 nM heme, 17 μg/mL human Hb, 500 nM $FeCl_3$, or cellular lysate obtained from the centric marine diatom *Thalassiosira pseudonana*. Axenic *T. pseudonana* lysate was prepared by first removing extracellular Fe from harvested algal cells in order to prevent confounding effects from extracellular inorganic Fe species (19). Using high performance liquid chromatography (HPLC) we determined the heme content of the original axenic *T. pseudonana* cultures to be approximately 0.72 ± 0.09 nmol/L with a chlorophyll *a* : heme *b* ratio of 264 ± 38 which is in good agreement with heme *b* measurements from other marine phytoplankton (Table S3) (20). After determining the heme concentrations in the lysates, we added *T. pseudonana* lysate to co-cultures under strictly trace metal clean conditions in order to generate a concentration of approximately 11 nM heme equivalents. To distinguish between the two strains in co-culture we developed a qPCR assay with primers targeting the kanamycin resistance cassette and the 435 bp portion of the TM1040 *hmuR* gene that was deleted in LH02 to quantify strain proportions. TM1040 outcompeted LH02 under all conditions, except for when co-cultured with 500 nM $FeCl_3$ or when given no additional Fe (Figure 4.7). $FeCl_3$ supported the largest number of amplicon copies per mL while no Fe addition supported the lowest. We observed the strongest difference in strain proportions when TM1040 and LH02 were co-cultured with heme and hemoglobin. When grown on *T. pseudonana* lysate, the wild type strain outcompeted LH02 in roughly similar proportions to heme and hemoglobin but the lysate supported a lower number of amplicon copies per mL than heme or hemoglobin.

4.3.6 The TM1040 heme uptake outer membrane transporter is similar to those in other *Roseobacter* genomes

To contextualize our previous experiments related to TM1040 physiology, we sought to explore the distribution of potential heme uptake genes in other genomes from the same taxonomic lineage, *Roseobacter*. We examined amino acid homology between the TM1040 heme uptake TBDT (HmuR) and 436 other TBDT sequences from 153 different genomes of the *Roseobacter* lineage. We initially identified TBDTs by sequence homology at the protein family level (Pfam: PF00593), but TBDT sequences are highly divergent and it is challenging to predict substrate specificity with alignment-based methods (21). We instead opted to construct sequence similarity networks as a method for viewing homology between sequences. Sequence similarity networks do not rely on a global sequence alignment, but rather visualize individual pairwise relationships between sequences identified in a reciprocal pairwise BLAST search (22). In these network representations (Figure 4.8) nodes represent sequences while edges represent similarity scores between sequences. For our TBDT network, edges between nodes are drawn if the pairwise similarity score between two sequences is greater than an e-value threshold of 10^{-110}, which corresponds to a mean sequence identity of approximately 34%. If two sequences are not similar at this threshold no edge is drawn which in effect partitions sequences into homology-based clusters. If a sequence does not match any other sequence at the similarity threshold then it is represented by a single, unconnected node.

Most *Roseobacter* TBDTs cluster into six groupings of at least 20 sequences each (Figure 4.8), and the TM1040 HmuR sequence falls into a tight cluster with 44 other sequences (group 4). The most similar TBDT to TM1040 HmuR comes from *Ruegeria*

sp. CH4B, which is also its most closely related relative based on phylogeny (23), while another highly similar sequence to TM1040 HmuR comes from the recently sequenced genome of *Sulfitobacter* sp. SA11 (Figure 4.9) (6). Additionally, most of the TBDT sequences (with the exception of six sequences) in group 5 share a similar synteny to TM1040 *hmuR* even though these TBDTs fall into a different sequence cluster (Figure 4.8). Although sequence identity is useful in aiding annotation, TBDTs are most commonly functionally assigned based on the predicted functions of neighboring genes (12, 21). All of the TBDT sequences in group 4 (Figure 4.8) share a syntenic gene neighborhood (10 genes upstream or downstream) that includes at minimum homologs to *hmuTUV*, *hmuS*, *exbB*, *exbD1*, and *tonB* genes. Generally, the genome neighborhoods of the other *Roseobacter* TBDT clusters are much less regularly ordered than group 4 and group 5. For example, the only consistently common genes in the neighborhoods of groups 1 and 2 encode components of ABCTs, but often are missing genes typically required for their functioning. It appears that 16% of all *Roseobacter* TBDT genes code for heme uptake systems. As a result, 45% of the 153 *Roseobacter* genomes surveyed here appear to have a complete heme uptake system.

4.3.7 Putative *Sulfitobacter* sp. SA11 heme uptake system is downregulated during exponential growth with a marine diatom in co-culture

After identifying the heme uptake system in *Sulfitobacter* sp. SA11, we leveraged the publically available transcriptome from a recent study (6) that identified a growth promoting factor produced by sp. SA11 when co-cultured with the diatom *Pseudo-nitzschia multiseries* PC9. In the work by Amin *et al*. transcriptomes were generated from

SA11 and *P. multiseries* co-cultures harvested at mid-exponential growth (96 h after inoculation). At 96 hours in co-culture with *P. multiseries* SA11 strongly upregulated pathways associated with the production of the plant hormone indole-3-acetic acid relative to axenic controls, which stimulated the growth of *P. multiseries*. We examined in this dataset the expression of 13 putative Fe uptake systems (12) encompassing 42 different genes identified in SA11 (Dataset S2). At mid-exponential growth in co-culture there appeared to be broad shift in the expression of Fe uptake capabilities in SA11 (Figure 4.10). For example, genes encoding a putative heme uptake system, two putative siderophore uptake systems ("Sid1" and "Sid3"), and a putative Fe uptake ABCT system ("Fe^{3+} SBP3") were significantly downregulated during exponential phase co-culture. In contrast, a different siderophore uptake system ("Sid2") and two other putative Fe uptake ABCTs ("Fe^{3+} SBP1 and SBP2") were significantly upregulated. Most of the genes in the six other potential Fe uptake systems ("Other putative Fe uptake") were either not differentially expressed or had small Log_2 fold changes.

4.4 Discussion

Roseobacter is an abundant lineage within the marine *Rhodobacteraceae* (*Alphaproteobacteria*) and members of this group play important roles in the global biogeochemical cycles of carbon and sulfur (7, 24). The *Roseobacter* strain *Ruegeria* sp. TM1040 was originally isolated from the bloom-forming dinoflagellate *Pfiesteria piscicida* (25, 26), is motile, and degrades algal-produced dimethylsulfoniopropionate (DMSP) (27). *Sulfitobacter* sp. SA11, the other *Roseobacter* strain highlighted in this study, was isolated from the marine diatom *Pseudo-nitzschia multiseries* PC9 and has

been shown to be important in promoting the growth of its algal partner via production of a hormone (6). Both of these strains were isolated from the phycosphere, a carbon-concentrated microscale layer surrounding algal cells (5, 28). Other members of the *Roseobacter* clade have frequently been isolated or detected in association with phytoplankton and other biological surfaces (29) and have been documented to increase in abundance during algal blooms (30, 31). The metabolic diversity of the *Roseobacter* clade indicates that these organisms can utilize multiple organic energy sources (23) and are well adapted for a dynamic environment where substrates can vary, such as the rise and demise of an algal bloom (31).

Here we have shown that the putative heme uptake system previously identified in the TM1040 genome is indeed involved in the uptake of heme. This system is upregulated when TM1040 is Fe limited suggesting that it is involved in heme scavenging under Fe-depleted conditions. The elimination of TM1040 growth on heme and hemoglobin after the insertional inactivation of *hmuR* demonstrates that the HmuR gene cluster is required for heme uptake and provides strong evidence that recognition of the heme moiety at the outer membrane TBDT is the first step in membrane translocation. Additionally, it demonstrates that uptake is occurring directly through HmuR rather than by extracellular degradation of the heme cofactor prior to translocation. The ability of TM1040 to grow on hemoglobin but not cytochrome c suggests that there is a bioavailability effect related to source hemoprotein structure. Our data is consistent with the hypothesis that HmuR is targeted to the heme cofactor rather than the parent hemoprotein. For example, the heme prosthetic group of hemoglobin is embedded within each tetrahedral subunit of the parent protein, but it is not covalently

bonded as it is in cytochrome c. This suggests that covalently linked hemoproteins may require additional enzymatic processing or chemical decomposition before they are rendered bioavailable, although it is possible that other, yet to be identified, bacterial strains possess the enzymatic capacity to use them directly.

Previous work has hypothesized that heme uptake is a mechanism by which heterotrophic bacteria remineralize particulate organic Fe in the marine environment (16, 32). Phytoplankton cells contain relatively high levels of hemoproteins mostly associated with the photosynthetic apparatus (33), and lysed or decaying phytoplankton cells could be a significant source of heme or hemoproteins to nearby bacteria. In a prior study in terrestrial systems a heme uptake gene cluster was characterized in the alphaproteobacterial N_2-fixing plant symbiont *Bradyrhizobium japonicum (34)*. The authors hypothesized this system would be advantageous for *B. japonicum* when the plant root nodule that harbored the bacterium lost integrity, began to decay, and the bacterium was forced to adapt to a new free-living lifestyle where leghemoglobin, a hemoprotein found in the nitrogen-fixing root nodules of leguminous plants, was a new and abundant Fe source (34). It seems likely that a similar strategy would be appropriate in a marine context, particularly that of heterotrophic bacteria interacting with sinking marine biogenic particles. Indeed, the results from competition experiments between TM1040 and LH02 suggest that the heme uptake system provides an advantage to TM1040 when these strains are competing for biogenic Fe derived from algal material (Figure 4.7), which in turn suggests that hemoproteins derived from algal material are Fe sources for TM1040 in the wild.

The close association of many *Roseobacter* strains with algal cells (29, 31) and the abundance and conservation of synteny of putative heme uptake genes within the clade suggest that nearly half of cultivated members likely utilize heme as an Fe source. We found that 45% of 153 different *Roseobacter* genomes contain probable heme uptake systems, which is in excellent agreement with what has been reported in earlier studies utilizing smaller collections of *Roseobacter* genomes (12, 16). However, recent evidence suggests that many uncultivated marine roseobacters have streamlined genomes and are likely ecological k-strategists (35, 36). Most of the *Roseobacter* genomes surveyed here contain extensive metabolic capability suggesting they belong to organisms that are copiotrophs or r-strategists. Thus our estimates for the prevalence of heme uptake in the *Roseobacter* clade may be biased against uncultivated k-strategists. Heme uptake systems have also been noted as puzzlingly absent in marine metagenomic datasets (37, 38). We revisited the questions of heme uptake gene distributions in environmental samples here, but we were unable to find significant abundances of heme uptake genes in any of the publically available marine meta-omic datasets in the IMG or NCBI databases. This may reflect a genuine lack of heme uptake genes in the marine environments where these samples were collected, or it may reflect the numerical dominance in most of these datasets of bacterial strains like SAR11 and *Prochlorococcus* which lack heme uptake genes (37). However, bacteria that are typically rare but can be transiently abundant are increasingly being recognized as disproportionate contributors to marine carbon recycling (39) and ecosystem global gene transcription (40). These same strains may disproportionately affect the marine Fe cycle as well.

Because it appears that TM1040 acquires Fe from lysed diatom cells via its heme TBDT, we subsequently sought to investigate the role of HmuR during symbiosis with a living phytoplankton partner. We leveraged a publicly available transcriptome derived from the *Roseobacter* strain *Sulfitobacter* sp. SA11 grown axenically and grown in co-culture with a marine diatom (6). The transcriptomes were collected after 96 hours of co-culture during *P. multiseries* log phase growth, and at this time the putative heme uptake system in SA11 was strongly downregulated relative to axenic culture. Because Fe concentrations in the growth medium were nominally the same in axenic and co-culture conditions, this would suggest that SA11 differentially regulates its heme acquisition system in response to cues from its algal host. Generally, it appears that SA11 shifted its Fe uptake strategy away from organic Fe resources (Heme and two potential siderophores) towards a strategy based on small inorganic Fe^{3+} chemical species during co-culture with exponentially growing *P. multiseries* PC9.

TM1040 has been observed to oscillate between mutualistic and antagonistic lifestyles depending on local concentrations of *p*-coumaric acid (41), which is a byproduct of algal lignin decomposition. In the first stages of phytoplankton-bacteria interaction motile TM1040 colonizes the phytoplankton cell surface and enters a sessile life phase where it utilizes phytoplankton-derived DMSP and other carbon compounds while simultaneously producing a suite of small molecules from which the host benefits. As the algal cell senesces it releases lignin breakdown products, which triggers the biosynthesis in TM1040 of a small infochemical that stimulates TM1040 motility and also causes increased algal lysis. As the host cell lyses and dies, it is plausible that algal hemoproteins become a relatively abundant iron resource for the newly motile TM1040

as it searches out a new algal symbiont. We propose an additional Fe component to the "swim or stick" model (41, 42) where *Roseobacter* symbionts utilize different suites of Fe compounds based on whether they are growing in association with a living algal cell or have adapted a planktonic, motile lifestyle (Figure 4.11). In particular, heme uptake is strongly downregulated during bacterial-algal symbiosis, presumably due to a lack of availability of heme-containing molecules, while it is upregulated during cell decomposition and becomes a major mode of Fe nutrition for the bacteria as they are transitioning to a new host cell.

4.5 Conclusion

Recycled Fe has been estimated to make up between 50% (43) and 90% (44) of the total Fe supply to marine planktonic ecosystems, and nearly 25% of all particulate Fe in some ocean surface waters is estimated to be routed through heterotrophic bacteria (45). More recently, significant concentrations of heme in particulate organic matter have been directly measured in marine waters (46, 47) and the distribution of heme in ocean basins is likely to be controlled by total phytoplankton biomass. Few studies have examined turnover rates of algal Fe in the marine environment, but if we assume an average contribution of 30 pmol L^{-1} diatom-derived particulate organic Fe as has been measured in naturally Fe fertilized regions (13, 43), a turnover rate of 5-17% algal particulate Fe day^{-1} by particle attached bacteria with no differential rates of turnover between intracellular Fe species (45), that 40% of algal Fe is in the form of heme (20), that a conservative 20% of the bacterial community on marine particles is composed of roseobacters (48, 49), and that 45% of roseobacters have the capability for heme uptake

this would translate to a maximum rate of between 0.05 - 0.18 pmol diatom heme L^{-1} day^{-1} that is remineralized by particle or algal associated *Roseobacter* assemblages. This rate is probably a significant underestimate of the total rate of heme remineralized by particle or algal-associated bacteria because many additional marine bacterial lineages that have been observed to associate with diatoms have the capability for heme uptake and diatoms are variably abundant in marine ecosystems. However, during conditions such as diatom blooms the fraction of total bacterial Fe recycling being mediated by *Roseobacter* heme uptake could be significant.

Our findings indicate that some marine bacteria acquire heme through an outer membrane TBDT and use this system to extract heme from particulate organic matter, namely lysed and decomposed phytoplankton cells. The role of heme uptake in phytoplankton-bacteria interactions is less clear, but we present evidence that induction of this system is dependent upon phytoplankton host physiology and is most likely advantageous during the initial stages of host decomposition rather than during symbiosis. The potential for heme uptake appears to be widespread in the often numerically dominant marine *Roseobacter* clade, particularly in the genomes of strains known to be isolated from phytoplankton and other biological surfaces. This raises the intriguing possibility that some marine phytoplankton cells may be "primed' for iron remineralization depending on the composition of their associated bacterial assemblage. Furthermore, the proportion of total algal intracellular Fe partitioned into hemoproteins has been shown to vary substantially depending on environmental and culture conditions (50, 51). This may mean that fluctuating proportions of the total algal intracellular Fe pool are susceptible to rapid remineralization via heterotrophic heme uptake depending

on local oceanographic conditions. If these variables could be parameterized they may be useful in defining Fe remineralization terms in global models of the marine Fe cycle. Here our results highlight heme as a rapidly exchanged "currency" of biogenic Fe and suggest that bacterial heme uptake is likely an important mechanism by which Fe is recycled from biomass in the marine environment.

4.6 Materials and Methods

4.6.1 Bacterial/algal strains and growth conditions

Escherichia coli str. DH5α (New England BioLabs) grown in LB medium and Kanamycin (50 μg/mL) and/or chloramphenicol (10 μg/mL) was used for the selection and maintenance of plasmids. *Ruegeria* sp. TM1040 was grown in either 2216 Marine broth/agar (Difco) or a Heart Infusion Seawater (HISW) based medium (25 g Heart Infusion broth (Difco), 15 g sea salts (Sigma), 1.5% agarose) at either 22°C or 30°C. Kanamycin (120 μg/mL) and chloramphenicol (10 μg/mL) were used for selection in plates and liquid cultures to ensure selection of insertion mutants. Iron limited cultures were grown in a modified PC medium termed PC+ (16). Briefly, the medium consisted of 1 liter of 0.2 μm filtered north Pacific seawater collected from the Scripps Institution of Oceanography pier, 1 g bacteriological peptone, 1 g casein, 10 mM glucose, 4.7 mM NH_4Cl, 600 μM KH_2PO_4, 50 μM Na_2EDTA, 40 nM $ZnSO_4 \cdot 7H_2O$, 230 nM $MnCl_2 \cdot 4H_2O$, 25 nM $CoCl_2 \cdot 6H_2O$, 10 nM $CuSO_4 \cdot 5H_2O$, 100 nM $Na_2MoO_4 \cdot 2H_2O$, 10 nM Na_2SeO_3. All components except for trace metals, KH_2PO_4, and EDTA were mixed, filtered through a 0.2 μm filter, stirred approximately 24 hours with 7% w/v Chelex resin (Biorad), then refiltered through a 0.2 μm filter to remove the chelex. Trace metal stocks were prepared

in 0.1 M HCl, and combined in necessary proportions to generate a mastermix. KH_2PO_4 was prepared in MilliQ water, and EDTA was prepared in MilliQ water at pH 8. All stocks were filter sterilized (0.2 μm) before use, and KH_2PO_4, EDTA, trace metal mastermix, and $FeCl_3 \cdot 6H_2O$ were added to PC+ aliquots immediately before inoculation with TM1040. $FeCl_3 \cdot 6H_2O$ was added to generate final concentrations of 50 nM, 100 nM, 500 nM, and 5 μM. Cultures of *Thalassiosira pseudonana* were made axenic by transferring three times in 1:10000 dilutions through F/2 enriched seawater medium with 100 μg/mL ampicillin, 25 μg/mL streptomycin, and 25 μg/mL neomycin. To test for bacterial contamination, 200 μL aliquots of *T pseudonana* were plated on 2216 marine agar, incubated for 48 hours at 30°C in the dark, and examined visually for bacterial growth.

4.6.2 Culturing strains in Fe depleted conditions

Bacterial colonies were inoculated into 5 mL of 2216 MB or HISW medium and allowed to grow at 30°C in the dark for 12 hours on a platform shaker (190 rpm). 250 μL of this culture was then transferred into PC+ medium with 5 μM $FeCl_3$ or no added iron and grown at room temperature in the dark for 12 hours. Then, 250 μL of TM1040 in PC+ with 5 μM $FeCl_3$ was transferred to a fresh 5 mL of PC+ with 5 μM $FeCl_3$ and allowed to grow for 12 hours at room temperature. The same procedure was performed for the culture with no added iron. Iron limitation was confirmed by spiking 5 μM of $FeCl_3$ into the iron-depleted culture and observing renewed growth. Fresh stocks of heme were prepared as a growth substrate immediately before each experiment by dissolving required amounts of hemin chloride (Sigma) in 0.3 M ammonium hydroxide. The pH of

the solution was adjusted to 8.0 with concentrated HCl, then filter sterilized through a 0.2 μm membrane. Dissolved heme prepared in this fashion has been determined to have an upper limit of free iron of approximately 4 nM per 1 μM of hemin chloride (52). Stocks of hemoproteins were prepared fresh before each experiment by dissolving the required mass of protein in MilliQ water and filter sterilizing through a 0.22 μm membrane. Significant contamination of free Fe in growth media was observed when using lyophilized hemoglobin from commercial sources confounding the growth effect differences between mutant and wild type TM1040. To circumvent this contamination, purified human hemoglobin was prepared in the Skaar lab at Vanderbilt University by anion exchange high-performance liquid chromatography and protein dialysis (53) and used in all subsequent experiments. All experiments were conducted in the dark to prevent photodegradation of heme and hemoproteins.

4.6.3 TM1040 genomic DNA extraction

1.5 mL of TM1040 culture at an OD_{600} of approximately 0.6 in 2216 MB medium was pelleted by centrifugation. DNA was extracted from cell pellets using the DNeasy Blood and Tissue Kit (Qiagen) following manufacturer's protocols. DNA was quantified using a NanoDrop 1000-D Spectrophotometer (Nano-drop Technologies).

4.6.4 TM1040 RNA extraction

TM1040 cultures were grown for 12 hours in PC+ medium supplemented with 5 μM of $FeCl_3$ or no added iron. Cultures were grown in triplicate. After 12 hours, iron-depleted cultures had an average OD_{600} of 0.045 while iron-rich cultures had an OD_{600} of

0.22. 5 mL of triplicate iron-depleted and iron-rich culture were spun at 4000 rpm for 10 minutes at 4 °C. The resulting pellets were resuspended in 1 mL TRI-Reagent (Zymo Research), and total RNA was extracted using a Direct-zol RNA miniprep kit (Zymo Research) following the manufacturer's instructions. Traces of contaminating genomic DNA were removed from total RNA using TURBO DNase (Life Technologies) following manufacturer's instructions. Total RNA concentrations were measured using a NanoDrop 1000-D Spectrophotometer. The absence of contaminating DNA was confirmed by visually inspecting (1% agarose gel) PCR reactions using the primers for *rpoD* (rpoD_F and rpoD_R) on the total RNA.

4.6.5 Reverse transcription of mRNA

Between 0.3 to 1 μg of total extracted RNA was reverse-transcribed using random hexamer primers and SuperScript III First-Strand Synthesis SuperMix (Invitrogen) following the manufacturer protocol. Resulting cDNA concentrations were measured using a NanoDrop 1000-D Spectrophotometer and samples were diluted to final concentrations of 5 ng/μL for use in reverse transcription quantitative PCR (RT-qPCR).

4.6.6 Reverse Transcription Quantitative PCR (RT-qPCR)

RT-qPCR reactions were carried out on a Qiagen RotorGene-Q (Qiagen) using Promega GoTaq qPCR Mastermix (Promega) in 25 μL total reaction volumes. The diluted cDNA samples (three biological replicates for each of the two iron conditions) were run in triplicate as technical replicates. Quantitative PCR reactions using total purified RNA were run in triplicate for the biological replicates of each iron condition for

the genes *rpoD* and *hmuR* (rpoD_F, rpoD_R and hmuR_F, hmuR_R). Results were below detection limit and never higher than the cycle thresholds for No-Template Controls demonstrating no genomic DNA contamination in samples. Five-point standard curves ranging from 311268 to 31 genome copies per 25 µl reaction in 10-fold dilutions were generated for each gene of interest using TM1040 genomic DNA with the formula $m = (n)1.096 \times 10^{-21}$ where m is the genome mass and n is the size of the genome in base pairs. For each individual calibration curve amplification efficiency as calculated from the slope of the standard curves was always greater than 90% and less than 100%. Cycle threshold values for standard curves were used with the REST 2009 software (54) to generate expression values for each gene relative to iron depleted conditions and normalized to the housekeeping genes *rpoD*, *gyrA*, and *gmkA*. Significance of expression ratios were calculated using the native randomization and bootstrapping algorithms in the REST 2009 software.

4.6.7 Partial deletion and insertional inactivation of *hmuR* in TM1040 genome

A non-replicating plasmid (pLH02) was generated through two iterations of the Gibson assembly technique (55) using the Gibson Assembly Master Kit (New England Biolabs). Two iterations were used in order to circumvent problems caused by high GC% in the *hmuR* gene (TM1040_0347) and earlier problems with assembly possibly due to secondary structure of 5' and 3' overhangs of assembly fragments (Figure 4.5). The first iteration plasmid (pLH01) was derived from pPY17a (18), a pRL271-derived suicide vector containing *sacB* for selection of double recombinants, by PCR amplifying two fragments with overlapping ends (gibfrag_01, gibfrag_02) from plasmid pPY17a using

primers gibfrag01_fwd, gibfrag01_rev, gibfrag02_fwd, and gibfrag02_rev. These fragments were combined with a portion of TM1040_0347 that was 645 bp long and 330 bp upstream of the 5' start site. The TM1040_0347 fragment was amplified from TM1040 genomic DNA with primers (0347US_fwd and 0347US_rev) containing sequence overhangs matching gibfrag_01 and gibfrag_02. The three fragments were assembled through Gibson assembly, electroporated into *E. coli* DH5α, and then recovered by selection using Kanamycin and sucrose sensitivity to isolate the correctly assembled pLH01. An analogous process was performed using pLH01 as a template to produce fragments gibfrag_03 and gibfrag_04 (primers: gibfrag03_fwd, gibfrag03_rev, gibfrag04_fwd, and gibfrag04_rev) and TM1040 genomic DNA to produce a 517 bp fragment 1435 bp upstream of the 5' start site of TM1040_0347 (primers: 0347DS_fwd and 0347DS_rev). These fragments were again assembled through Gibson assembly, electroporated into *E. coli* DH5α and recovered by Kanamycin selection and screening for sucrose sensitivity. The resulting plasmid (pLH02) contained a 1162 bp portion of TM1040_0347 interrupted by a 460 bp deletion into which a Km resistance cassette was inserted (Figure 4.4). All PCR reactions to produce Gibson assembly fragments were performed using Q5 High-Fidelity Polymerase (NEB).

The plasmid pLH02 was introduced to *Ruegeria* sp. TM1040 by electroporation as has been described earlier (42). Briefly, TM1040 was grown in 50 mL HISW at 30°C with shaking at 200 rpm to an OD_{600} of approximately 0.5, chilled on ice for 15 minutes, and centrifuged at 8000 rpm for 10 minutes at 4°C. The supernatant was removed and the cell pellet was gently resuspended in 10 mL of 10% glycerol in MQ water. The cell suspension was then centrifuged at 8000 rpm for 10 minutes at 4°C, and the glycerol

washing procedure was repeated for a total of three washes. After the final wash, the cell pellet was suspended in 0.5 mL of 10% glycerol in MQ water, a 65 µL aliquot was mixed with 50 ng of pLH02, and the mixture was incubated on ice for one minute. Cells were electroporated in a 0.2 cm cuvette at 2500 V, 400 Ω, and 25 µF, then immediately suspended in 1 mL of 30°C HISW medium and incubated at 30°C with shaking (200 rpm) for 2.5 hours. Double recombinant mutants were selected on HISW plates containing 120 µg/mL Kanamycin and 5% sucrose wt/vol, and individual colonies were picked after approximately 14 hours. Transformants exhibiting both Kanamycin and sucrose resistance were expected to have incorporated the mutagenic TM1040_0347 construct into the genome by double cross-over. The double cross-over event was verified by PCR to confirm correct insertion (primers: insrt_cnfirm_fwd, insrt_cnfirm_rev). A 2926 bp insert confirmed a 460 bp deletion in TM104_0347 and an insertion of the 1374 bp Kanamycin resistance cassette in its place (Figure 4.6). The expected PCR product of the wild type gene using these primers is 2013 bp in length.

4.6.8 Preparation of TM-clean algal lysate and removal of extracellular Fe

Axenic *Thalassiosira pseudonana* was grown in 1 L of F/2 enriched seawater medium with 120 µM added EDTA until cells were at a density of 1-2 x 10^6 cells/mL. A 10 mL aliquot of the culture was filtered for chlorophyll *a* measurement, and the remaining volume was gently filtered (~ 5 mmHg) through an acid soaked 3.0 µm nuclepore filter, using an acid cleaned teflon filter rig. The cells were resuspended in an acid cleaned centrifuge bottle in 200 mL of low Fe (~ 2 nM Fe) seawater collected from

the California Current. 20 mL of an oxalate wash (300 mM NaCl, 10 mM KCl, 100 mM Na_2oxalate, 50 mM Na_2EDTA (19)) was added to the resuspended cells in low Fe seawater. The solution was mixed gently and allowed to rest at room temperature for 20 minutes after which it was centrifuged at 6000 rpm for 10 minutes at 4°C. The resulting cell pellet was washed again in low Fe seawater and oxalate solution for 20 minutes, centrifuged, and finally resuspended in 1 mL of trace metal clean MilliQ water. 300 µL of the suspension was immediately sonicated at 4°C using a Diogenode Bioruptor Standard system on high power for 6 minutes, using a 60 second ON + 30 second OFF cycle. The resulting soluble and insoluble lysate fractions were immediately frozen in liquid N_2 until use.

4.6.9 Analytical Measurement of heme *b* and chlorophyll from algal lysate

Aliquots of algal lysate were mixed with two volumes of acidified acetone (80:20 v/v acetone:1.6 M HCl), ultrasonicated in a water bath sonicator (10 minutes at 4°C), and centrifuged at 13000 rpm (10 minutes at 4°C) to extract heme *b* as described before (20, 56). The supernatant was diluted by a factor of 50 in mobile phase A, and heme *b* was measured spectrophotometrically by high performance liquid chromatography (HPLC) using a polymeric reverse phase column (PRP-1, 100 x 2.1 mm, 5 µm, Hamilton Inc.) as described earlier (20, 57). HPLC was performed using binary gradient pumps (Waters 1525) with a microcell diode array spectrophotometer (Waters 2489) with a manual full loop injection (50 µL). Heme separation was completed at 22°C using a gradient of 100% A to 100% B over 10 mins, followed by an isocratic elution of mobile phase B for two minutes at a flow rate of 0.5 mL/min. Mobile phases consisted of (A) 30:70:0.08 v/v

acetonitrile:water:trifluoroacetic acid and (B) 100:0.08 v/v acetonitrile:trifluoroacetic acid. Elution of heme *b* (retention time 6.8 min) was monitored at 400 nm. For chlorophyll analysis, samples of T. pseudonana were filtered onto 0.7 μm GF/F filters (Whatman), stored in liquid N_2 until use, extracted using a 90% acetone extraction/acidification method, and measured using a Turner Designs fluorometer (58).

4.6.10 Study of competition between wild type TM1040 and the LH02 mutant strain

Quantitative PCR (qPCR) was performed on genomic DNA extracted from co-cultures in order to calculate the relative abundance of TM1040 and LH02 strains. Oligonucleotide primers (Table 4) matching the Kanamycin resistance cassette in LH02 (kmR_F_compete, kmR_R_compete) and the 435 bp deleted portion of the TM1040 *hmuR* gene in LH02 (hmuR_F_compete, hmuR_R_compete) were used to distinguish between TM1040 and LH02. Five-point standard curves ranging from 311268 to 31 genome copies per 25 μl reaction in 10-fold dilutions were generated for each primer pair, and reaction efficiencies were always between 95% and 100%. TM1040 and LH02 cultures were iron limited in PC+ marine medium with no added Fe for two subsequent transfers and inoculated in duplicate into 20 mL PC+ medium at the same initial cell density ($\sim 5 \times 10^6$ cells/mL). Co-cultures were supplied with 500 nM $FeCl_3$, 500 nM heme, 10.6 nM heme equivalents of *Thalassiosira pseudonana* cellular lysate, or no additional Fe source. Co-cultures were maintained at 22°C with shaking at 190 rpm and were grown in the dark to prevent photodegradation of heme and hemoproteins. At each experimental time point 2 mL of each co-culture was pelleted by centrifugation and total

DNA was extracted from cell pellets using the DNeasy Blood and Tissue Kit (Qiagen) following manufacturer's protocols. Total genomic DNA from each co-culture was quantified using a NanoDrop 1000-D Spectrophotometer (Nano-drop Technologies) and used in subsequent qPCR reactions.

4.6.11 Construction of TonB Dependent Receptor Sequence Similarity Network

The genomes of 153 different *Roseobacter* strains were searched for the TonB-dependent receptor domain (Pfam00593) using the Integrated Microbial Genomes (IMG) database (59). Peptide sequences matching Pfam00593 were downloaded resulting in 436 sequences total. These sequences were compared in an all-vs-all BLAST v2.2.30 analysis using default parameters with blastp. The resulting blast output was visualized and edited in Cytoscape (v3.2.0). An E-value threshold of 10^{-110} (corresponding to a mean of 34.3% identity) was chosen for network visualization by manually iterating through increasingly stringent thresholds until clusters persisted through three subsequent iterations (22, 60). In order to assist the functional annotation of TBDT SSN clusters, the resulting SSN was submitted to the Genome Neighborhood Network tool at the Enzyme Function Initiative (61) with a gene neighborhood size of 10 and a co-occurrence lower limit of 20% resulting in a collection of PFAMs present in the genome neighborhood (± 10 genes) of at least 20% of the TBDT sequences in each sequence similarity network cluster.

4.6.12 Transcriptome data from *Sulfitobacter* sp. SA11 in co-culture with *Pseudo-nitzschia multiseries* PC9

Transcriptome data from a recent study (6) examining expression changes during a co-culturing experiment with SA11 and *P. multiseries* PC9 was downloaded from the NCBI Gene Expression Omnibus (GEO accession GSE65189) (62). To identify genes that were differentially expressed between co-culture and axenic culture conditions, SA11 count data was processed using the DESeq2 (63) package in R. Deseq2 fits a generalized linear model (GLM) to expression data by modeling read counts under a negative binomial distribution, while shrinking count dispersion factors and fold-change estimation using an empirical Bayesian model that assumes similar dispersion across genes with similar mean expression. The DESeq2 procedure was performed with default parameters, namely using the conventional null hypothesis of zero logarithmic fold change between conditions and a False Discovery Rate (FDR) threshold of 10% ($P_{\text{FDR corrected}} < 0.1$). Potential Fe uptake systems were annotated using the NCBI Conserved Domain Database (64) using cutoffs as described earlier (12).

4.7 Acknowledgements

Chapter 4 is currently being prepared for publication by Shane L. Hogle, Bianca Brahamsha, and Katherine A. Barbeau. We would like to thank Dr. Robert Belas for providing TM1040, Dr. Javi Paz-Yepes for help with Gibson cloning and qPCR, Emy Daniels for assistance with molecular biology protocols, Lauren Manck and Dr. Martha Gledhill for their assistance with HPLC, Dr. Chris Dupont for helpful discussions and assistance with metagenome searches, and Dr. Eric Skaar and his lab for providing purified human hemoglobin. This work was funded by NSF GRFP grant DGE-144086 to S.L.H. and NSF grant OCE-1061068 to K.A.B and B.B.

Table 4.1 Growth rates of TM1040 and LH02 on different Fe sources

Condition	Strain	Growth Rate (hr^{-1})	*P* value
0 nM $FeCl_3$	TM1040	0.13 ± 0.004	ns
0 nM $FeCl_3$	LH02	0.13 ± 0.003	ns
500 nM $FeCl_3$	TM1040	0.31 ± 0.02	ns
500 nM $FeCl_3$	LH02	0.29 ± 0.02	ns
500 nM Heme	TM1040	0.28 ± 0.01	< 0.0002
500 nM Heme	LH02	0.16 ± 0.01	< 0.0002
17 μM Hemoglobin	TM1040	0.27 ± 0.006	< 0.0002
17 μM Hemoglobin	LH02	0.20 ± 0.007	< 0.0002
500 nM Cytochrome C	TM1040	0.14 ± 0.01	ns
500 nM Cytochrome C	LH02	0.15 ± 0.007	ns

P value indicates the results of an independent two tailed Student's t test between strain growth rates at each condition. Values greater than 0.05 are omitted. There were three replicates for each condition.

Table 4.2 Bacterial strains and plasmids

Strain or Plasmid	Description	Source
E. coli str. DH5α	Electrocompetent strain used as host for vector constructs. Catalog # C2989K	New England BioLabs
Ruegeria sp. TM1040	*hmuR*+; wildtype strain	1
Ruegeria sp. LH02	Δ*hmuR*::Km^r; TM1040 electroporated with pLH02 resulting in a recombinant strain (LH02) with a partial deletion of *hmuR* and insertion of a Kanamycin resistance cassette.	This work
pPY17a	Km^r, Em^r, Cm^r, Suc^s; used to construct pLH01 and pLH02	2
pHL01	Suc^s, Δ*hmuR*::Kmr; used to construct pLH02	This work
pHL02	Suc^s, Δ*hmuR*::Kmr; used for transformation	This work

1. Belas R, Horikawa E, Aizawa S-I, Suvanasuthi R (2009) Genetic determinants of *Silicibacter* sp. TM1040 motility. *J Bacteriol* 191(14):4502–4512.
2. Paz-Yepes J, Brahamsha B, Palenik B (2013) Role of a microcin-C-like biosynthetic gene cluster in allelopathic interactions in marine *Synechococcus*. *Proc Natl Acad Sci U S A* 110(29):12030–12035.

Table 4.3 Heme concentrations in soluble and insoluble *T. pseudonana* lysate

Culture	heme *b* in lysate (nM)	heme *b* in original culture (pM)	Chl *a* (nM)	Cells L^{-1}	heme *b* (nmol) $cell^{-1}$	Chl *a* (nmol) $cell^{-1}$	Chl *a* : heme *b*
T. pseudonana 3	761	684	340	[a]1.24×10^6 [b]7.67×10^8	9.92×10^{-7}	2.74×10^{-4}	276.45
T. pseudonana 2	831	748	342	[a]1.30×10^6 [b]7.00×10^8	1.19×10^{-6}	2.63×10^{-4}	221.54
T. pseudonana 1	624	562	335	[a]1.10×10^6 [b]6.05×10^8	1.03×10^{-6}	3.05×10^{-4}	295.03

[a]Cell concentration before washing and concentration step. Used to calculate Chl *a* $cell^{-1}$
[b]Cell concentration after washing and concentration step. Used to calculate heme *b* $cell^{-1}$

Table 4.4 Oligonucleotide primers

Primer name	Sequence (5' to 3')	Description
gibfrag01_fwd	ACACAAAAGAACCCCCAGA	Amplify fragment 01 from pPY17a for first Gibson construction
gibfrag01_rev	CAGGAAAGAACATGTGAGCAAA	Amplify fragment 01 from pPY17a for first Gibson construction
gibfrag02_fwd	CCTTTTGCTGGCCTTTTG	Amplify fragment 02 from pPY17a for first Gibson construction
gibfrag02_rev	GTTAAGGGATGCAGGTCG	Amplify fragment 02 from pPY17a for first Gibson construction
0347US_fwd	CGACCTGCATCCCTTAACAACCAAAAGGTCCAGATCC	Amplify upstream region of *hmuR* from TM1040 genomic DNA and add overhangs. Used in first Gibson construction
0347US_rev	TCTGGGGGTTCTTTTGTGTCTCGATGTATTGATCCGCAT	Amplify upstream region of *hmuR* from TM1040 genomic DNA and add overhangs. Used in first Gibson construction
gibfrag03_fwd	AAATAATAGTGAACGGCAGG	Amplify fragment 03 from pLH01 for second Gibson construction
gibfrag03_rev	AGAGCTTTGTTGTAGGTG	Amplify fragment 03 from pLH01 for second Gibson construction
gibfrag04_fwd	TACGATACACTTCCGCTC	Amplify fragment 04 from pLH01 for second Gibson construction
gibfrag04_rev	ACCCATCACATATACCTG	Amplify fragment 04 from pLH01 for second Gibson construction
0347DS_fwd	CACCTACAACAAAGCTCTCCCAGACTGAAAAAGCGG	Amplify downstream region of *hmuR* from TM1040 genomic DNA and add overhangs. Used in second Gibson construction
0347DS_rev	GAGCGGAAGTGTATCGTATTGCGTCTGATATTGCTTGT	Amplify downstream region of *hmuR* from TM1040 genomic DNA and add overhangs. Used in second Gibson construction
insrt_cnfirm_fwd	AGAGGCCGCAAGAGTGAA	Used to confirm double cross-over event in R. LH02
insrt_cnfirm_rev	GAGGGAGGTGAGGACAGCAA	Used to confirm double cross-over event in R. LH02
hmuV_F	ATCAACGCCTGTCAGAGGTC	(B4) – used in RT-qPCR
hmuV_R	GTTTGCGGCAGAACAAAGAC	(A1) – used in RT-qPCR
hmuU_F	TCATCAGCCGTACCATCATC	(B5) – used in RT-qPCR
hmuU_R	CGTTTCAGCAAGATCCACAG	(A2) – used in RT-qPCR
hmuT_F	TATCCGTGAACCCAAACCTG	(B6) – used in RT-qPCR
hmuT_R	GTTGATCTTGTCGACGATGC	(A3) – used in RT-qPCR

Table 4.4 Oligonucleotide primers, continued

Primer name	Sequence (5' to 3')	Description
hmuS_F	GATTCTGCAGCCAGAAATCG	(B7) – used in RT-qPCR
hmuS_R	TGTCGGTGTCTTCATTCAGG	(A4) – used in RT-qPCR
hmuR_F	AATTGACGACCGATCCTGAG	(A5) – used in RT-qPCR
hmuR_R	CGGAGCCATATTCAAACGAC	(B8) – used in RT-qPCR
hyp_F	GATGCTTCTGACGCTCTTTG	(A6) – used in RT-qPCR
hyp_R	ACACCAGCAACATCGAACTG	(B9) – used in RT-qPCR
exbB_F	GTTCTGGACGGCAATAAAGG	(A7) – used in RT-qPCR
exbB_R	CGATCGACAGTCCAATAAGC	(B10) – used in RT-qPCR
exbD1_F	TGCTCTCTTCGACGTTTTCG	(A8) – used in RT-qPCR
exbD1_R	CTTCACCATTCAGCGACATC	(C1) – used in RT-qPCR
exbD1a_F	GATGGTAGCCTGCATTTTCG	(A9) – used in RT-qPCR
exbD1a_R	TTCCTCTTCCCAGATGATGC	(C2) – used in RT-qPCR
tonB_F	AAGAGACCGCTGTGGAAATC	(A10) – used in RT-qPCR
tonB_R	CTGAACAAGGTCCGCAAAAC	(C3) – used in RT-qPCR
rpoD_F	CGCCAAGAAATACACCAACC	- used in RT-qPCR
rpoD_R	TTATAGCCGCGACGGTATTC	- used in RT-qPCR
gyrA_F	ACGCTCTTTGTGGCGAATAC	- used in RT-qPCR
gyrA_R	GGGATTGGCAGAATGTTGAC	- used in RT-qPCR
gmkA_F	TGCTCTTTGACATCGACTGG	- used in RT-qPCR
gmkA_R	AGATCGACAGCGTGTGTTTG	- used in RT-qPCR
kmR_F_compete	ATTCTCACCGGATTCAGTCG	- used for TM1040 + LH02 competition exp
kmR_R_compete	ATTCCGACTCGTCCAACATC	- used for TM1040 + LH02 competition exp
hmuR_F_compete	AAAGGAACGGTTGAATGCGG	- used for TM1040 + LH02 competition exp
hmuR_R_compete	GATCTCCTGGCGTTCGTAGC	- used for TM1040 + LH02 competition exp

Letters in parentheses in the “Description” column indicate shorthand names for primers as designated in the figures

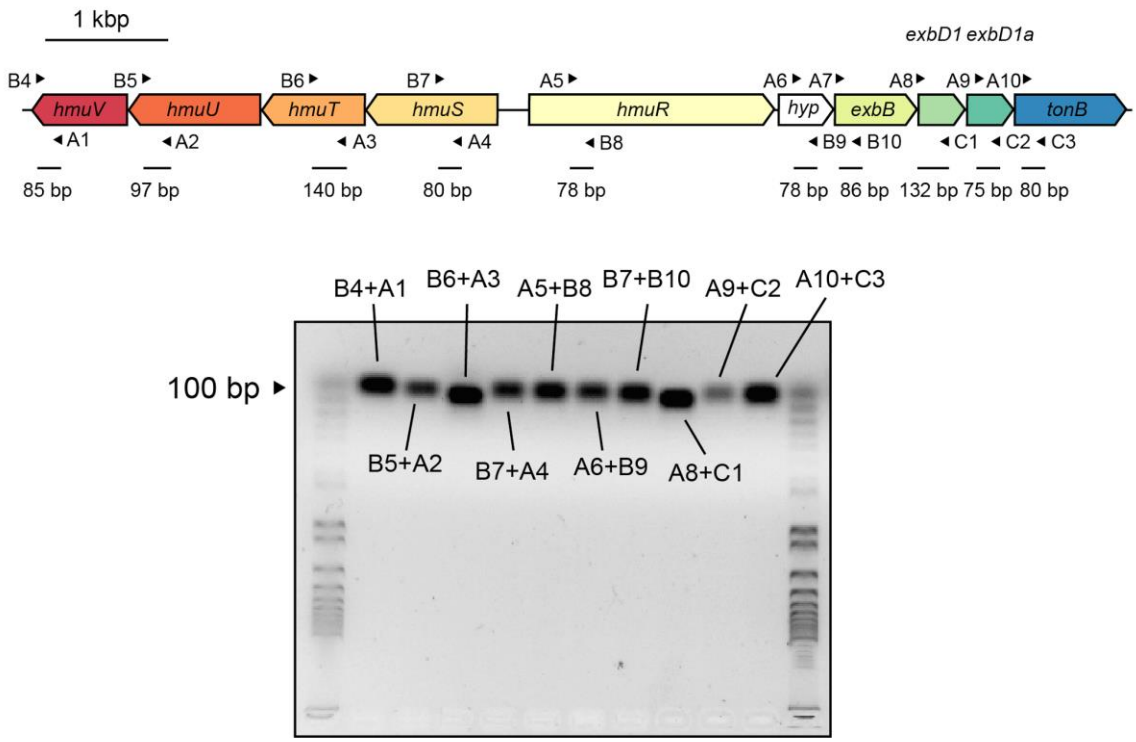


Figure 4.1 Location of qPCR primer pairs in TM1040 heme uptake locus. Primers are identified by shorthand names (Table S2) and product lengths are labeled below. Below - RT-PCR analysis using above primer pairs on cDNA reverse transcribed from RNA collected from an Fe-stressed treatment. First and last lanes contain 1-kb plus ladder (Invitrogen).

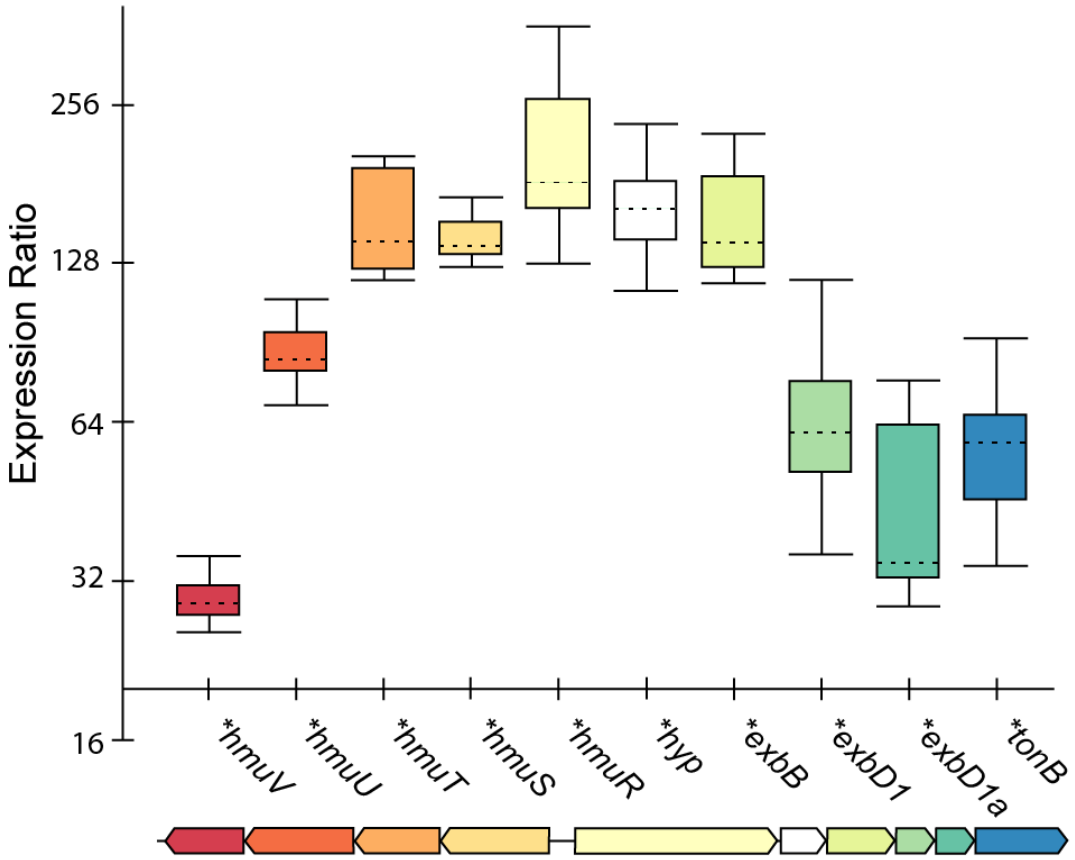


Figure 4.2 Quantitative real-time PCR of expression changes for the TM1040 heme uptake locus in Fe stressed compared with Fe replete culture conditions. The relative expression distribution is displayed as an expression ratio (n = 3 biological replicates per group) normalized against three condition-stable reference genes - *rpoD*, *gyrA*, and *gmkA*. Ratios greater than 1 indicate a higher gene expression in Fe stressed cultures. Asterisks indicate genes determined to be differentially expressed ($P < 0.05$).

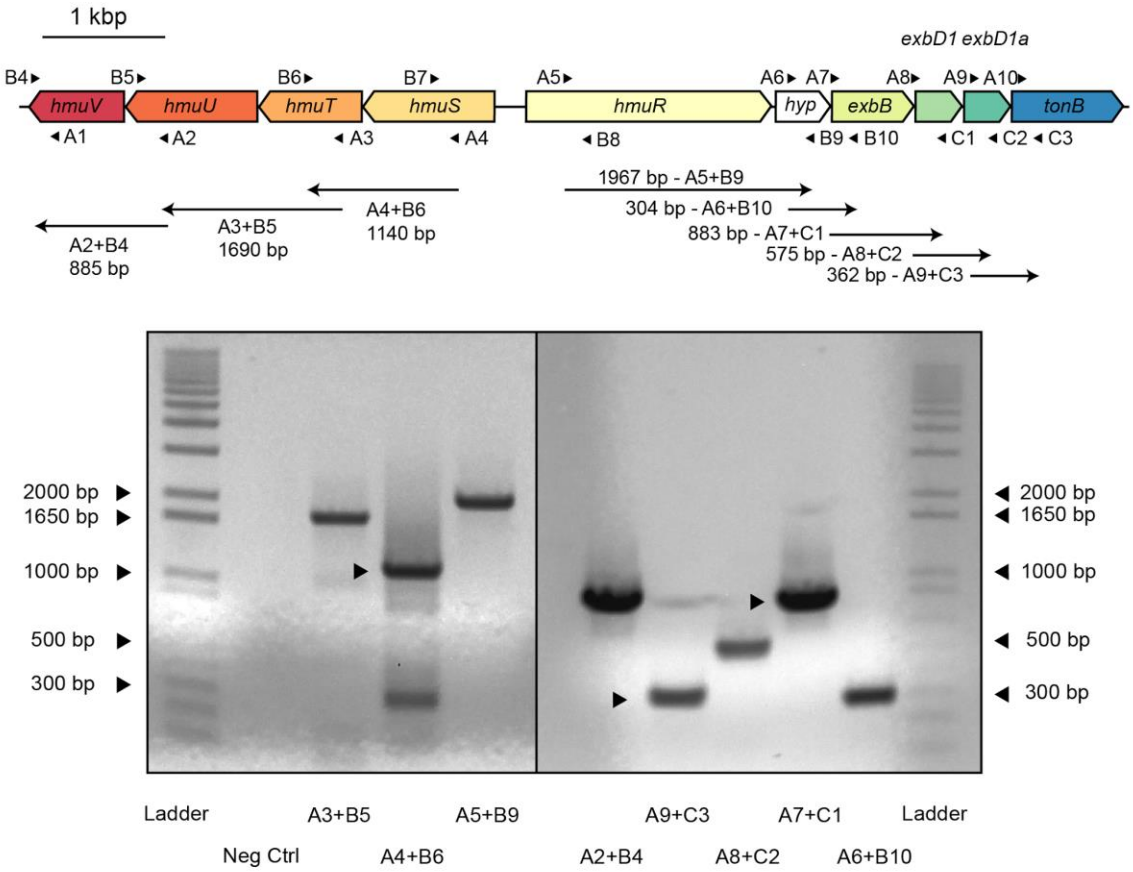


Figure 4.3 Characterization of cotranscribed genes in the two diverging heme uptake gene clusters in the TM1040 genome. The top diagram displays the TM1040 heme uptake genomic region, arrows represent primers used in RT-PCR, and lines denote predicted PCR products for each primer pair, their lengths, and orientation. Lower diagram shows the results of the RT-PCR cotranscription analysis using cDNA reverse transcribed from RNA collected from an Fe-stressed treatment. Each primer pair shown above produced one or more bands after PCR amplification. The band consistent with the expected product length (shown above) is highlighted with an arrowhead in the cases where multiple bands are observed. Primers are listed by their shorthand names (see Table S2). First and last lanes contain 1-kb plus ladder (Invitrogen). Second lane is a negative control containing mRNA before reverse transcription indicating no contaminating genomic DNA.

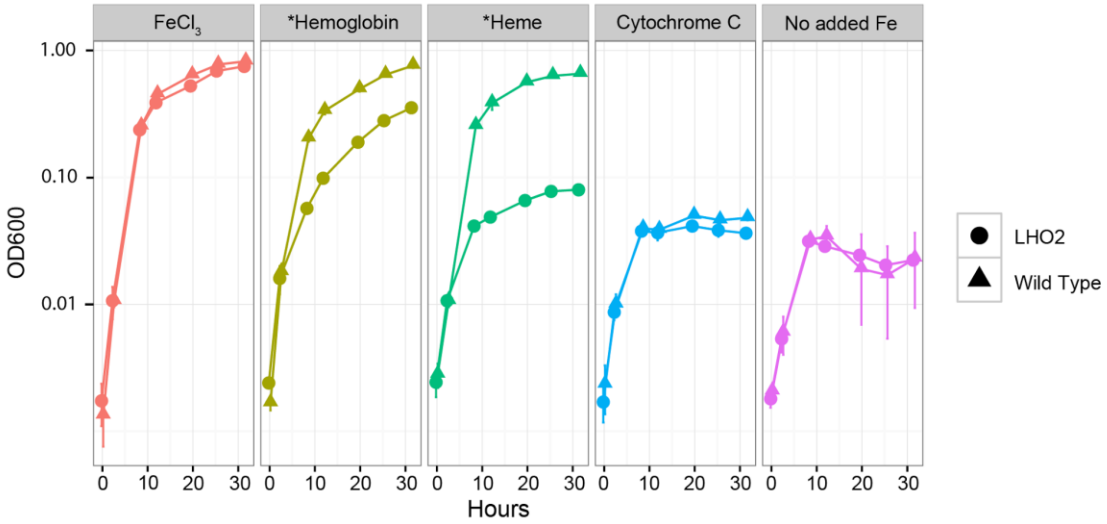


Figure 4.4 Growth of wild type TM1040 and mutant LH02 strains with different added Fe sources. Y axis (log scale) displays optical density measurements at 600 nm with respect to time (X axis). The wild type strain is represented by triangles while the LH02 mutant strain is represented by circles. Conditions preceded with an asterisk have statistically different growth rates between LH02 and TM1040 (independent, two-tailed Student's t-test, $P < 0.05$). All Fe sources were added at 500 nM concentrations except for Hemoglobin, which was added at 17 μg/mL.

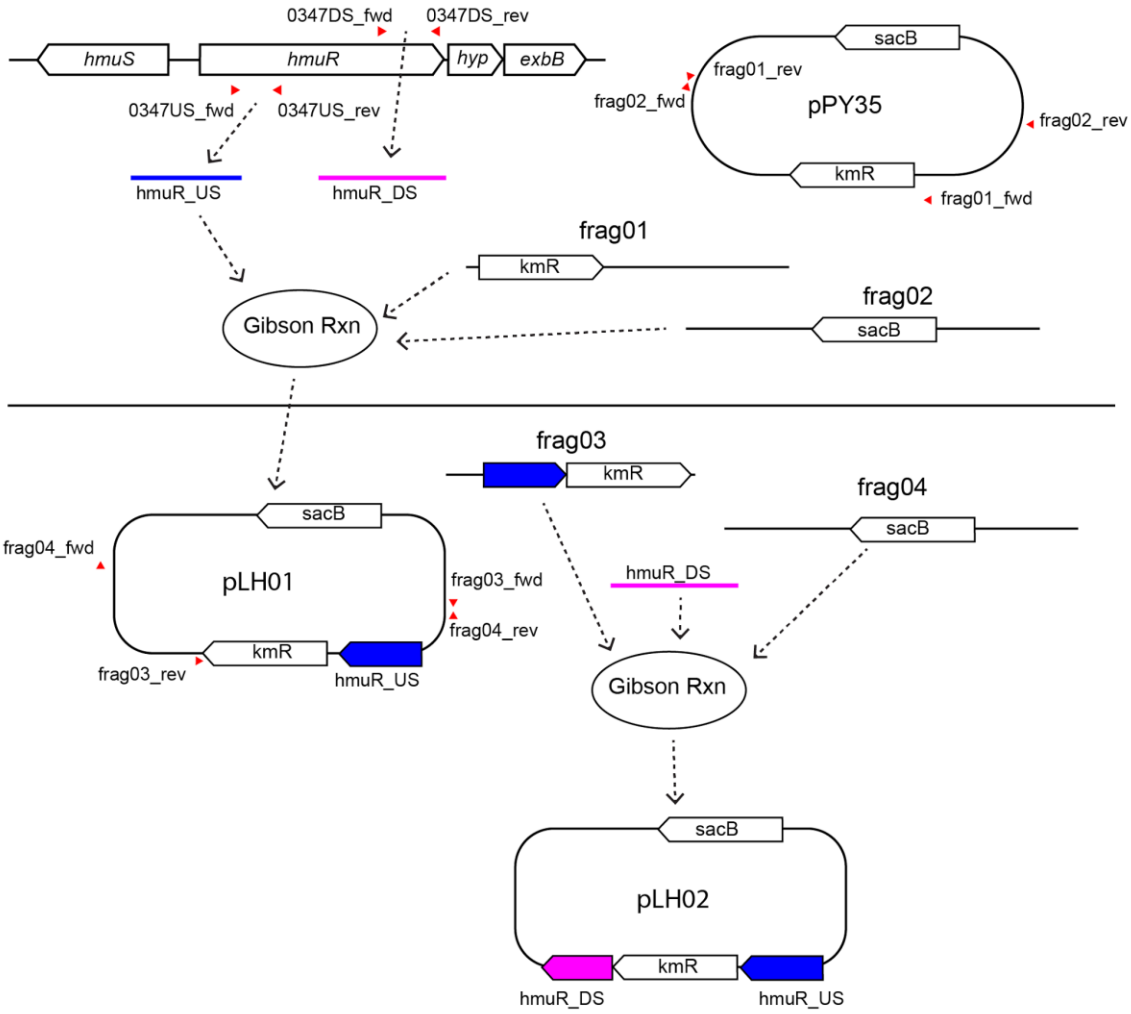


Figure 4.5 pLH02 plasmid generation from two Gibson assembly steps. The first step (above the horizontal line) depicts production of plasmid pLH01 using a Gibson Assembly Kit (New England Biolabs). The second step (below the horizontal line) depicts the final insertion of the kanamycin-resistant cassette between upstream (blue) and downstream (pink) regions of the TM1040 *hmuR* gene again using a Gibson Assembly Kit (New England Biolabs). Red arrowheads denote positions of primers (Table 2) used to produce DNA fragments that were later assembled by Gibson reaction.

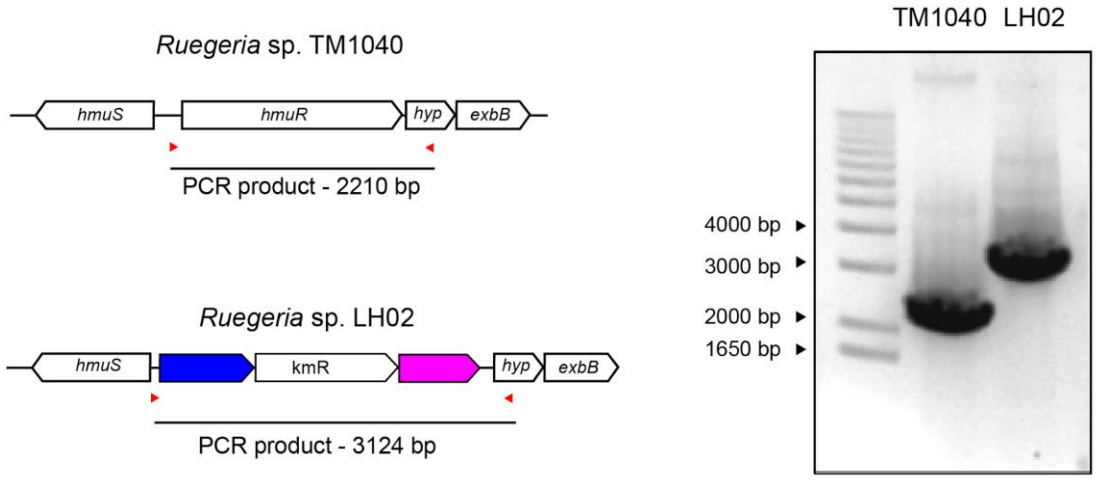


Figure 4.6 PCR verification of kanamycin insert in TM1040 strain LH02. Left: genomic organization of *hmuR* and surrounding genes in TM1040 WT (top) and TM1040 LH02 (bottom). Red arrows denote primers insert_cnfrm_fwd and insert_cnfrm_rev (Table 2) used to verify the double crossover event in LH02. The predicted PCR product lengths in each strain are listed below. In LH02 the upstream region from original *hmuR* is shown in blue while the downstream region is shown in pink. Right: gel showing PCR products using primers insert_cnfrm_fwd and insert_cnfrm_rev on TM1040 WT (left lane) and TM1040 LH02 (right lane). First lane contains 1-kb plus ladder (Invitrogen).

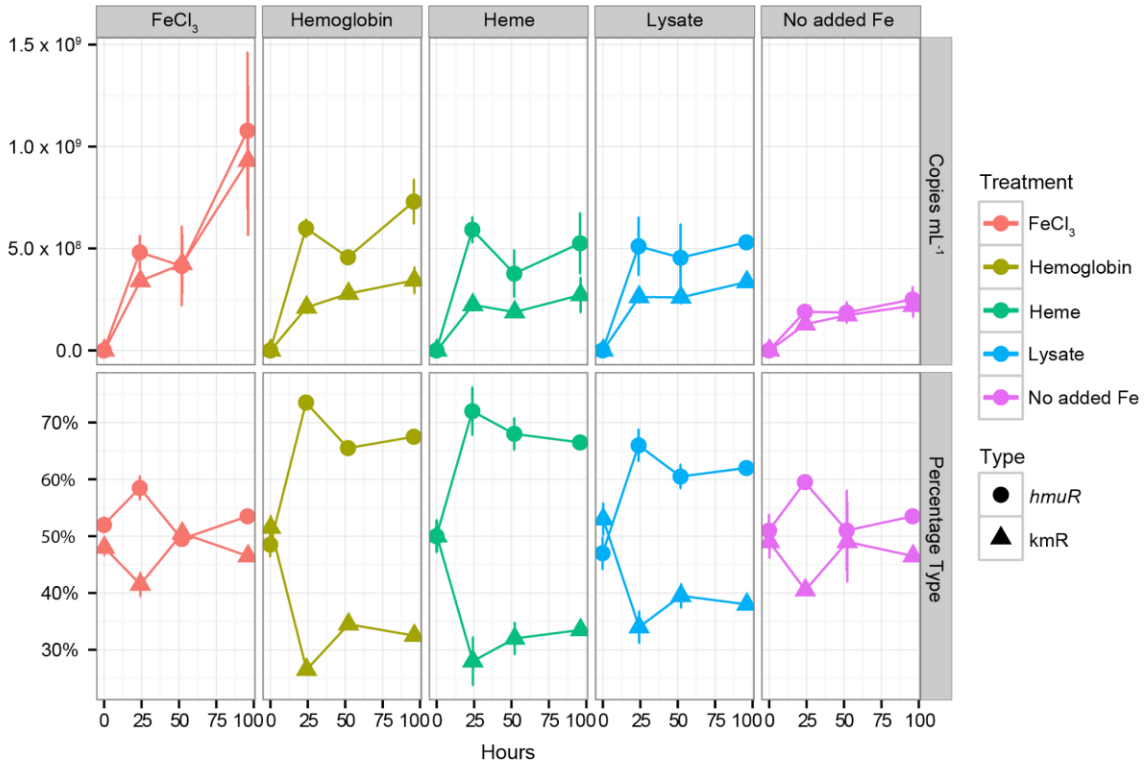


Figure 4.7 Competition between cocultured LH02 and wild type strains for different Fe sources. The TM1040 strain (Type: *hmuR*) is represented by circles while the LH02 mutant strain (Type: kmR) is represented by triangles. The proportion of each strain was determined by quantitative real-time PCR using primers targeting either the Kanamycin resistance cassette for LH02 or the region of the *hmuR* gene deleted in LH02 and retained in TM1040. The Y axis displays the absolute concentration of transcripts per mL of culture or the percentage of each strain with respect to time (X axis). $FeCl_3$ and heme were added at 500 nM, hemoglobin was added at ~17 µg/mL, and the algal lysate was added at a concentration providing 10.6 nM heme equivalents.

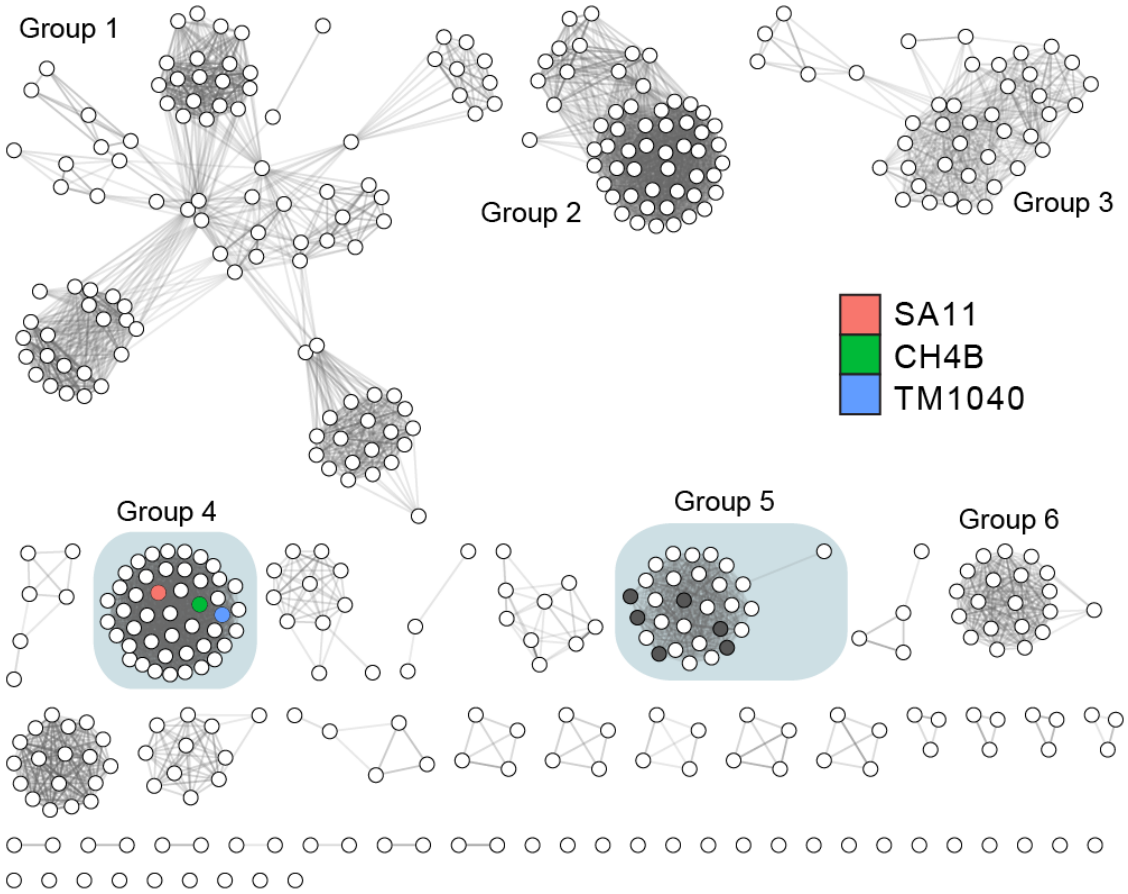


Figure 4.8 Sequence similarity network of TonB-dependent transporters from *Roseobacter* genomes. The network is displayed with an e-value threshold of 10^{-110} (corresponding to ~34% sequence identity). The 45 sequences in Group 4 all are located in genome neighborhoods with shared synteny to the TM1040 heme uptake gene clusters. 23 sequences in Group 5 are syntenous to the TM1040 heme uptake clusters (dark gray nodes indicate sequences that lack neighborhood synteny). *Sulfitobacter* sp. SA11, *Ruegeria* sp. TrichCH4B, and *Ruegeria* sp. TM1040 heme TonB-dependent transporters are highlighted for reference.

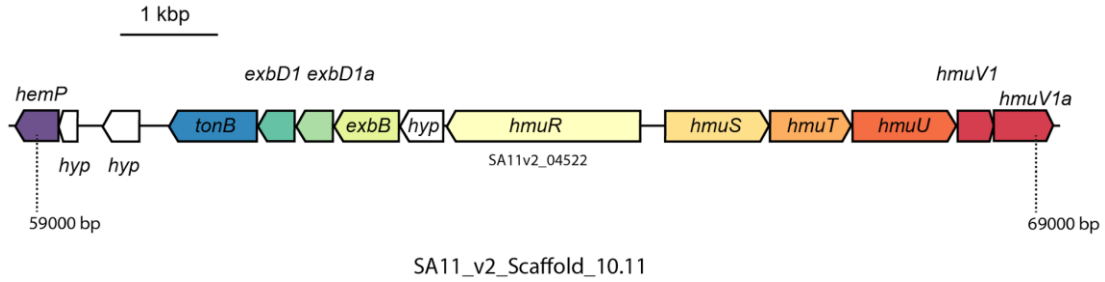


Figure 4.9 The putative heme uptake locus in *Sulfitobacter* sp. SA11. Genes homologous to the heme uptake locus in TM1040 are colored and labeled equivalently for both (see Figs. S1, S2). Genes in white are hypothetical proteins. For reference, the locus name of the putative heme outer membrane receptor is listed underneath *hmuR*, the scaffold name is listed below the diagram, and the base pair position on the scaffold is labeled with a dashed line at each end of the figure.

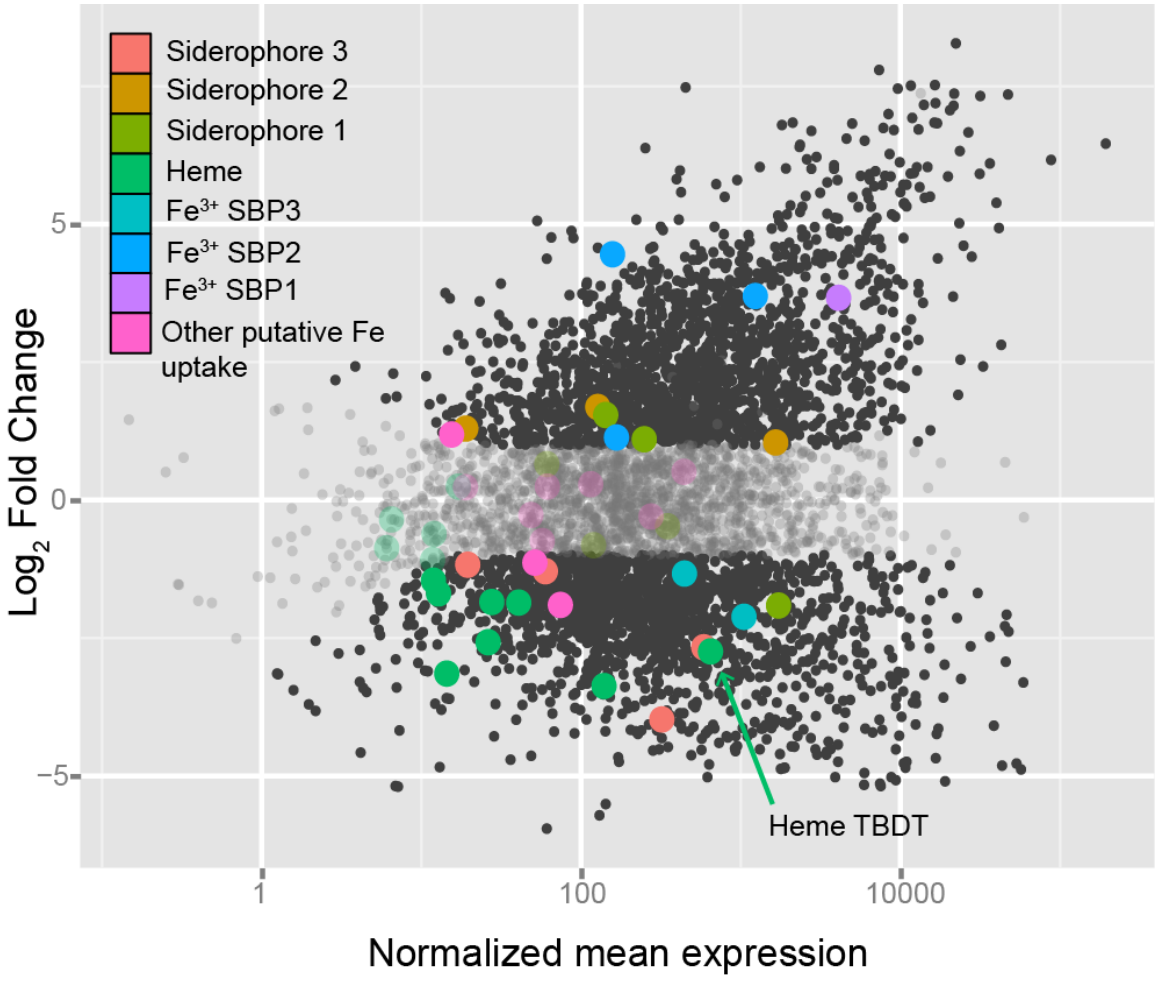


Figure 4.10 MA plot generated from a previously published RNA-seq experiment investigating interactions between a diatom, *Pseudo-nitzschia multiseries* PC9, and *Sulfitobacter* sp. SA11. Each SA11 gene detected in the RNA-seq experiment is represented as a dot. The X axis displays the variance stabilizing transformed expression of each transcript averaged over all samples (as implemented in the DESeq2 package). The Y axis displays the $\log_2$ fold change of SA11 transcripts. Genes with a positive Log_2 fold change were more abundant when SA11 was cocultured with *P. multiseries* (compared with axenic control), while those with a negative fold change were less abundant under coculture. Genes with an FDR corrected *P* value > 0.05 or a Log_2 fold change magnitude < 1 are shown with 50% opacity.

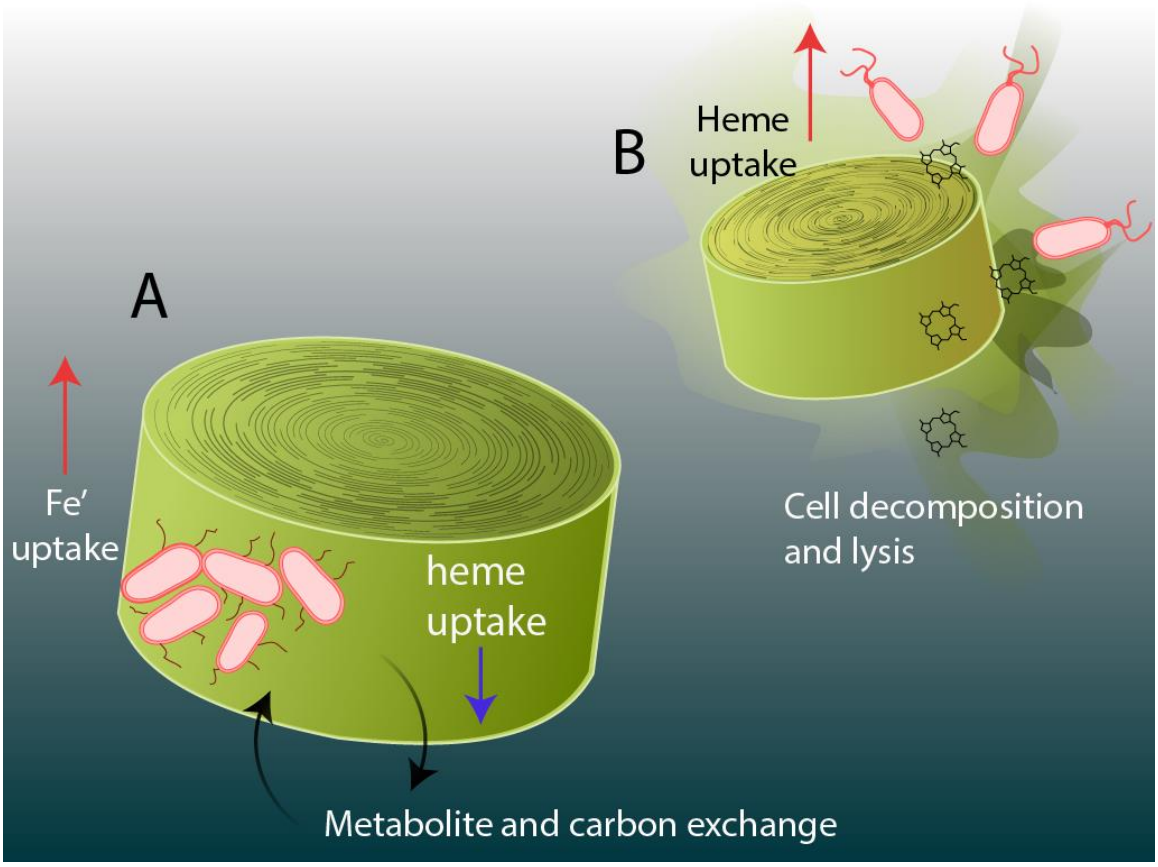


Figure 4.11 Model of *Roseobacter* heme utilization and algal host growth. During exponential growth of the host (A) attached bacteria and the algal cell exchange carbon and metabolites in a mutualistic relationship. Bacteria shift their Fe acquisition strategy away from heme (blue arrow, downregulation) uptake and towards other Fe' species (red arrow, upregulation). During algal cell decomposition (B) bacterial cells potentially transition to a motile growth phase in order to search out a new host while simultaneously targeting (red arrow, upregulation) newly abundant heme and hemoproteins as Fe resources.

4.8 References

1. de Baar HJW, Boyd PW, Coale KH, Landry MR, Tsuda A, Assmy P, Bakker DCE, Bozec Y, Barber RT, Brzezinski MA, Buesseler KO, Boyé M, Croot PL, Gervais F, Gorbunov MY, Harrison PJ, Hiscock WT, Laan P, Lancelot C, Law CS, Levasseur M, Marchetti A, Millero FJ, Nishioka J, Nojiri Y, van Oijen T, Riebesell U, Rijkenberg MJA, Saito H, Takeda S, Timmermans KR, Veldhuis MJW, Waite AM, Wong C-S (2005) Synthesis of iron fertilization experiments: From the Iron Age in the Age of Enlightenment. *J Geophys Res* 110(C9):C09S16.

2. Moore JK, Doney SC, Glover DM, Fung IY (2001) Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. *Deep Sea Res Part 2 Top Stud Oceanogr* 49:463–507.

3. Boyd PW, Ellwood MJ (2010) The biogeochemical cycle of iron in the ocean. *Nat Geosci* 3(10):675–682.

4. Moran MA (2015) The global ocean microbiome. *Science* 350(6266):aac8455.

5. Amin SA, Parker MS, Armbrust EV (2012) Interactions between diatoms and bacteria. *Microbiol Mol Biol Rev* 76(3):667–684.

6. Amin S a., Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, Morales RL, Berthiaume CT, Parker MS, Djunaedi B, Ingalls a. E, Parsek MR, Moran M a., Armbrust EV (2015) Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522(7554):98–101.

7. Durham BP, Sharma S, Luo H, Smith CB, Amin S a., Bender SJ, Dearth SP, Van Mooy B a. S, Campagna SR, Kujawinski EB, Armbrust EV, Moran M a. (2014) Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc Natl Acad Sci U S A* 112(2):453–457.

8. Mayali X, Azam F (2004) Algicidal Bacteria in the Sea and their Impact on Algal Blooms1. *J Eukaryot Microbiol* 51(2):139–144.

9. Seyedsayamdost MR, Carr G, Kolter R, Clardy J (2011) Roseobacticides: small molecule modulators of an algal-bacterial symbiosis. *J Am Chem Soc* 133(45):18343–18349.

10. Azam F (1998) Microbial Control of Oceanic Carbon Flux: The Plot Thickens. *Science* 280(5364):694–696.

11. Tortell PD, Maldonado MT, Price NM (1996) The role of heterotrophic bacteria in iron-limited ocean ecosystems. 383(6598):330–332.

12. Hogle SL, Cameron Thrash J, Dupont CL, Barbeau KA (2016) Trace metal acquisition by marine heterotrophic bacterioplankton with contrasting trophic

strategies. *Appl Environ Microbiol*. doi:10.1128/AEM.03128-15.

13. Boyd PW, Strzepek RF, Ellwood MJ, Hutchins DA, Nodder SD, Twining BS, Wilhelm SW (2015) Why are biotic iron pools uniform across high- and low-iron pelagic ecosystems? *Global Biogeochem Cycles* 29(7):2014GB005014.

14. Hogle SL, Barbeau KA, Gledhill M (2014) Heme in the marine environment: from cells to the iron cycle. *Metallomics* 6(6):1107–1120.

15. Hogle SL, Barbeau KA, Gledhill M (2014) Heme in the marine environment: from cells to the iron cycle. *Metallomics* 6(6):1107–1120.

16. Roe KL, Hogle SL, Barbeau K (2013) Utilization of heme as an iron source by marine alphaproteobacteria in the roseobacter clade. *Appl Environ Microbiol* 79(18):5753–5762.

17. The UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res* 43(D1):D204–D212.

18. Paz-Yepes J, Brahamsha B, Palenik B (2013) Role of a microcin-C-like biosynthetic gene cluster in allelopathic interactions in marine Synechococcus. *Proc Natl Acad Sci U S A* 110(29):12030–12035.

19. Tang D, Morel FMM (2006) Distinguishing between cellular and Fe-oxide-associated trace elements in phytoplankton. *Mar Chem* 98(1):18–30.

20. Gledhill M (2007) The determination of heme b in marine phyto- and bacterioplankton. *Mar Chem* 103(3-4):393–403.

21. Schauer K, Rodionov D, de Reuse H (2008) New substrates for TonB-dependent transport: do we only see the "tip of the iceberg"? *Trends Biochem Sci* 33(7):330–338.

22. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4(2):e4345.

23. Luo H, Moran MA (2014) Evolutionary Ecology of the Marine Roseobacter Clade. *Microbiol Mol Biol Rev* 78(4):573–587.

24. Howard EC, Henriksen JR, Buchan A, Reisch CR, Bürgmann H, Welsh R, Ye W, González JM, Mace K, Joye SB, Kiene RP, Whitman WB, Moran MA (2006) Bacterial Taxa that Limit Sulfur Flux from the Ocean. *Science* 314:649–652.

25. Alavi M, Miller T, Erlandson K, Schneider R, Belas R (2001) Bacterial community associated with Pfiesteria-like dinoflagellate cultures. *Environ Microbiol* 3(6):380–396.

26. Miller TR, Belas R (2003) Pfiesteria piscicida, P. shumwayae, and other Pfiesteria-like dinoflagellates. *Res Microbiol* 154(2):85–90.

27. Miller TR, Hnilicka K, Dziedzic A, Desplats P, Belas R (2004) Chemotaxis of Silicibacter sp. strain TM1040 toward dinoflagellate products. *Appl Environ Microbiol* 70(8):4692–4701.

28. Bell W, Mitchell R (1972) Chemotactic and growth responses of marine bacteria to algal extracellular products. *Biol Bull* 143(2):265–277.

29. Slightom RN, Buchan A (2009) Surface colonization by marine roseobacters: Integrating genotype and phenotype. *Appl Environ Microbiol* 75(19):6027–6037.

30. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R (2012) Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science* 336(2006):608–611.

31. Buchan A, LeCleir GR, Gulvik C a., González JM (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* 12(10):686–698.

32. Hopkinson BM, Roe KL, Barbeau K (2008) Heme uptake by Microscilla marina and evidence for heme uptake systems in the genomes of diverse marine bacteria. *Appl Environ Microbiol* 74(20):6263–6270.

33. Strzepek RF, Harrison PJ (2004) Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* 431:689–692.

34. Nienaber A, Hennecke H, Fischer HM (2001) Discovery of a haem uptake system in the soil bacterium Bradyrhizobium japonicum. *Mol Microbiol* 41(4):787–800.

35. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, Luo H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* 110(28):11463–11468.

36. Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA (2014) Evolutionary analysis of a streamlined lineage of surface ocean Roseobacters. *ISME J* 8(7):1428–1439.

37. Hopkinson B, Barbeau K (2012) Iron transporters in marine prokaryotic genomes and metagenomes. *Environ Microbiol* 14(1):114–128.

38. Toulza E, Tagliabue A, Blain S, Piganeau G (2012) Analysis of the global ocean sampling (GOS) project for trends in iron uptake by surface ocean microbes. *PLoS One* 7(2):e30931.

39. Pedler BE, Aluwihare LI, Azam F (2014) Single Bacterial Strain Capable of Significant Contribution to Carbon Cycling in the Surface Ocean. *Proc Natl Acad Sci U S A* 111(20):7202–7207.

40. Dupont CL, McCrow JP, Valas R, Moustafa A, Walworth N, Goodenough U, Roth R, Hogle SL, Bai J, Johnson ZI, Mann E, Palenik B, Barbeau K a., Craig Venter J, Allen AE (2015) Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J* 9(5):1076–1092.

41. Sule P, Belas R (2013) A novel inducer of roseobacter motility is also a disruptor of algal symbiosis. *J Bacteriol* 195(4):637–646.

42. Belas R, Horikawa E, Aizawa S-I, Suvanasuthi R (2009) Genetic determinants of Silicibacter sp. TM1040 motility. *J Bacteriol* 191(14):4502–4512.

43. Sarthou G, Vincent D, Christaki U, Obernosterer I, Timmermans KR, Brussaard CPD (2008) The fate of biogenic iron during a phytoplankton bloom induced by natural fertilisation: Impact of copepod grazing. *Deep Sea Res Part 2 Top Stud Oceanogr* 55(5-7):734–751.

44. Boyd PW, Law CS, Hutchins D a., Abraham ER, Croot PL, Ellwood M, Frew RD, Hadfield M, Hall J, Handy S, Hare C, Higgins J, Hill P, Hunter K a., LeBlanc K, Maldonado MT, McKay RM, Mioni C, Oliver M, Pickmere S, Pinkerton M, Safi K, Sander S, Sanudo-Wilhelmy S a., Smith M, Strzepek R, Tovar-Sanchez A, Wilhelm SW (2005) FeCycle: Attempting an iron biogeochemical budget from a mesoscale SF6 tracer experiment in unperturbed low iron waters. *Global Biogeochem Cycles* 19(4):1–13.

45. Boyd PW, Ibisanmi E, Sander SG, Hunter K a., Jackson G a. (2010) Remineralization of upper ocean particles: Implications for iron biogeochemistry. *Limnol Oceanogr* 55(3):1271–1288.

46. Gledhill M, Achterberg EP, Honey DJ, Nielsdottir MC, Rijkenberg MJ a. (2013) Distributions of particulate Heme b in the Atlantic and Southern Oceans-Implications for electron transport in phytoplankton. *Global Biogeochem Cycles* 27(4):1072–1082.

47. Honey D, Gledhill M, Bibby T, Legiret F, Pratt N, Hickman A, Lawson T, Achterberg E (2013) Heme b in marine phytoplankton and particulate material from the North Atlantic Ocean. *Mar Ecol Prog Ser* 483:1–17.

48. González JM, Simó R, Massana R, Covert J, Casamayor E, Pedrós-Alió, Moran MA (2000) Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol* 66(10):4237–4246.

49. LeCleir GR, DeBruyn JM, Maas EW, Boyd PW, Wilhelm SW (2014) Temporal changes in particle-associated microbial communities after interception by nonlethal sediment traps. *FEMS Microbiol Ecol* 87(1):153–163.

50. Saito MA, Bertrand EM, Dutkiewicz S, Bulygin VV, Moran DM, Monteiro FM, Follows MJ, Valois FW, Waterbury JB (2011) Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph Crocosphaera watsonii. *Proc Natl Acad Sci U S A* 108(6):2184–2189.

51. Gledhill M, Gerringa L, Laan P, Timmermans KR (2015) Heme b quotas are low in Southern Ocean phytoplankton. *Mar Ecol Prog Ser* 532:29–40.

52. Rao AU, Carta LK, Lesuisse E, Hamza I (2005) Lack of heme synthesis in a free-living eukaryote. *Proc Natl Acad Sci U S A* 102(12):4270–4275.

53. Pishchany G, McCoy AL, Torres VJ, Krause JC, Crowe JE, Fabry ME, Skaar EP (2010) Specificity for human hemoglobin enhances Staphylococcus aureus infection. *Cell Host Microbe* 8(6):544–550.

54. Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* 30(9):e36–e36.

55. Gibson DG, Young L, Chuang RY (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):12–16.

56. Espinas N a., Kobayashi K, Takahashi S, Mochizuki N, Masuda T (2012) Evaluation of unbound free heme in plant cells by differential acetone extraction. *Plant Cell Physiol* 53(7):1344–1354.

57. Atamna H, Frey WH 2nd (2004) A role for heme in Alzheimer's disease: heme binds amyloid beta and has altered metabolism. *Proc Natl Acad Sci U S A* 101(30):11153–11158.

58. Parsons TR, Maita Y, Lalli CM (1984) 4.1 - Determination of Chlorophylls and Total Carotenoids: Spectrophotometric Method. *A Manual of Chemical & Biological Methods for Seawater Analysis*, ed Lalli TRPMM (Pergamon, Amsterdam), pp 101–104.

59. Markowitz VM, Chen I-M a., Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40(Database issue):D115–22.

60. Mashiyama ST, Malabanan MM, Akiva E, Bhosle R, Branch MC, Hillerich B, Jagessar K, Kim J, Patskovsky Y, Seidel RD, Stead M, Toro R, Vetting MW, Almo SC, Armstrong RN, Babbitt PC (2014) Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere. *PLoS Biol* 12(4):e1001843.

61. Gerlt JA, Allen KN, Almo SC, Armstrong RN, Babbitt PC, Cronan JE, Dunaway-Mariano D, Imker HJ, Jacobson MP, Minor W, Poulter CD, Raushel FM, Sali A, Shoichet BK, Sweedler JV (2011) The Enzyme Function Initiative. *Biochemistry* 50(46):9950–9962.

62. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 41(D1):D991–D995.

63. Love MI, Huber W, Anders S (2014) Fully formatted Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(550):1–21.

64. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita R a., Zhang D, Zheng C, Bryant SH (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(D1):D222–D226.

Chapter 5

Linking phytoplankton and bacterioplankton assemblage dynamics to iron-binding ligand production in a microcosm experiment

5.1 Abstract

Siderophore-producing marine bacteria were identified nearly two decades ago as potential sources for strong iron-binding ligands (L_1) detected in seawater, but specific linkages between ligands detected in natural water and microbial community dynamics remain unclear. We followed the production of different classes of Fe-binding ligands, Fe concentrations, and macronutrient concentrations in a series of iron amended incubations over a period of six days and examined the phytoplankton and bacterioplankton community composition at the termination of the incubation. High iron additions stimulated increased phytoplankton biomass relative to low iron samples, but both iron treatments resulted in generally similar diatom-dominated communities. Strikingly, L_1 ligands with iron-binding affinities akin to siderophores were only observed in high iron treatments. The abundance of transiently dominant copiotroph bacterial strains was correlated with the emergence of L_1 ligands and high iron conditions, but we observed no universal heterotrophic community association with increasing L_1 concentrations. Rather, incubations with similar L_1 concentrations and binding strengths each harbored generally distinct bacterial communities dominated by a handful of known copiotroph strains. We interpret these patterns as evidence for waves of heterotrophic bacterial succession during phytoplankton bloom progression and collapse with potential functional redundancy of

ligand production at fine scale taxonomic levels. Ultimately, the robust correlations between multiple copiotroph bacterial taxa and strong siderophore-like ligands reported here suggests a potential linkage between these taxa and ligand production in the marine environment.

5.2 Introduction

The concentrations, chemical forms, and spatial/temporal distributions of oceanic iron are important factors shaping the ecology of marine phytoplankton and the overall productivity of marine ecosystems (1). Iron that has been regenerated from sinking particulates - as opposed to “new” inputs from sources like atmospheric deposition is known to be important in sustaining surface ocean productivity in iron-limited regions (2, 3). The mechanisms by which particulate iron is recycled back to primary producers or into the wider microbial food web, a process herein referred to as iron remineralization, are largely uncharacterized and represent major unknowns in global iron budgets (4). The activity of microzooplankton and viruses are known to facilitate iron remineralization (5, 6), and heterotrophic bacteria also play a fundamental role as has been observed for other nutrients (7, 8). For example, heterotrophic bacterial activity has been shown to mobilize particulate Fe into smaller size fractions and shown to generate Fe-binding molecules from sinking algal particles (9). Fe-binding molecules drive Fe solubility in the oceans and are critical components of the marine Fe cycle (10).

Dissolved marine iron is predominantly bound by organic ligands of mostly uncharacterized structure. The concentrations and conditional stability constants, a measure of iron-binding strength, of natural marine ligands are typically measured using

an electrochemical technique known as competitive ligand exchange - adsorptive cathodic stripping voltammetry (CLE-ACSV). Many of the iron-binding ligands from seawater (11) or marine cultures (12) that can be structurally characterized are siderophores, bacterial secondary metabolites likely produced as an iron acquisition strategy. Marine siderophores are generally characterized by their small molecular weights, their high affinity for iron, and by some having amphiphilic side chains (13). Intriguingly, the binding strength of natural marine ligands of binding-strength class "L_1" as measured by CLE-ACSV is often very similar to or higher than those of known siderophores, leading to hypotheses that L_1 ligands in seawater are biologically derived (14) and may in fact be siderophores (15). Supporting these hypotheses, transporters for siderophores and siderophore-like complexes have been identified in the genomes of many marine bacterial strains (16). However, siderophore transporters and biosynthetic pathways are conspicuously absent in genome-streamlined organisms suggesting that siderophore production and transport is confined to specific microbial niches and is biased towards transiently dominant but metabolically diverse and highly active bacterial strains (40).

Siderophores are well-known examples of strong Fe-binding ligands, but other ligands (in this study classified as L_2 and L_3) that bind Fe comparatively weakly can also be abundant in marine systems (17). These weak ligands are hypothesized to be derived from a variety of sources, but it is unclear whether they are produced through defined microbial secondary metabolisms as are siderophores. Some weak ligands have been identified as phytoplankton-derived exopolymeric substances (18), but most are thought to be breakdown products from organic matter degradation (15, 19). The "ligand soup"

thought to exist throughout the water column is believed to be primarily comprised of weak ligand classes, while strong L_1 ligands are thought to dominate in surface waters where they outcompete weaker undefined complexes and are thought to play an important role in Fe bioavailability (10).

There are multiple lines of evidence in support of a biological source for strong iron-binding L_1 ligands. Numerous CLE-ACSV electrochemical studies have demonstrated a positive correlation between L_1 ligand concentrations and biological activity (20), but the precise biological sources of those ligands are difficult to ascertain through electrochemical methods alone. A recent study reported increased ligand concentrations coupled with increased rates of bacterial siderophore uptake and siderophore receptor expression after a mesoscale Fe fertilization experiment (21), but did not examine bacterial community composition or multiple ligand classes. A different study demonstrated that siderophore uptake and production strongly shape social interactions between particle-associated marine bacteria (22), but it is unknown how these microscale processes would relate to ligand dynamics at the spatial and temporal scales at which they are measured by current chemical techniques. Ultimately, it is clear that both siderophores and strong Fe-binding ligands (as determined by electrochemistry) are present in marine systems and are probably connected to the metabolic activities of marine bacteria and phytoplankton, but *direct* functional and mechanistic linkages between L_1, siderophores, and biological community structure at this point remain elusive.

Here we couple high-throughput 16S rRNA marker gene surveys with multiple analytical window CLE-ACSV chemical analysis in shipboard incubation experiments in

order to explore how phytoplankton and bacterioplankton community composition is connected to the production of different classes of iron-binding ligands. We conducted whole seawater microcosm incubations at two distinct NO_3^-:dissolved Fe ratios, an oceanographic parameter that has been demonstrated as an indicator of diatom Fe stress (23). Within this experimental framework we sought to (1) observe planktonic responses to differing levels of Fe stress in microcosm incubations, (2) characterize the chemical nature of iron-binding ligands during the course of the experiment, (3) compare the microbial community composition between high and low Fe treatments at experiment termination, and (4) examine associations between individual chemical parameters and microbial community composition.

5.3 Results

5.3.1 Nutrients

Incubation experiments were conducted using oligotrophic water collected from the offshore southern California Bight (33.879° N, -123.306° W) with low initial nitrate concentrations ([NO_3^-]; nitrate+nitrite) of 0.94 µmol L^{-1} (Table 5.1). All incubations were initiated in water with low *in situ* NO_3^- and dissolved Fe (dFe) concentrations (NO_3^- : dFe = 3.0), suggesting that the initial phytoplankton community was likely limited by NO_3^- and not dFe (23). The dFe concentration was 0.31 ± 0.04 nmol L^{-1} on day 0 in the initial offshore water. In order to stimulate a phytoplankton bloom, incubations were spiked with equal macronutrient concentrations as well as 1 and 5 nmol L-1 dFe for the Low Fe and High Fe treatments, respectively. These two Fe treatments were selected to simulate conditions approaching phytoplankton Fe-limitation (NO_3^- : dFe = 11) in Low Fe

treatments and Fe-replete conditions in High Fe treatments (NO_3^- : dFe = 2) at the onset of incubation. In the southern California Current, NO_3^-:dFe values in excess of 5 have been shown to limit diatom growth (24). At incubation termination on day 6, average ratios of NO_3^-:dFe were still elevated in the Low Fe treatment incubations (6.03) compared with High Fe treatments (0.20) (Table 5.1). Low Fe treatments had an average dFe concentration of 0.32 ± 0.10 nmol L^{-1} after six days, while High Fe treatments were on average 0.61 ± 0.01 nmol L^{-1} after day six (Table 5.1).

5.3.2 Phytoplankton community

The chlorophyll a concentration was 0.75 µg L^{-1} on day 0, and increased to 9.91±0.18 µg L^{-1} (n = 3) on average in Low Fe treatments (+1 nmol L-1) and 12.50±3.17 µg L^{-1} (n = 3) on average in High Fe (+5 nmol L-1) treatments (Figure 5.1). In both treatments, most of the biomass gain was due to an increase in the abundance of diatoms as indicated by the increase in fucoxanthin in Low Fe and High Fe incubations (Table 5.2, Figure 5.2) and direct observation in the cell counts (Table 5.3, Fig 5.3). Before Fe addition, the phytoplankton community was composed primarily of Pseudo-nitzschia species (67%) and other large diatoms (Table 5.3). *Pseudo-nitzschia* responded to Fe addition in both treatments, increasing to 78% of the community on average in Low Fe bottles and 82% in High Fe bottles. *Chaetoceros* spp. also increased in abundance in the treatments, increasing from 2% of the initial community to 9% and 8% by day 6 in Low Fe and High Fe treatments, respectively. Other large diatoms (≥ 10 µm) displayed a moderate increase in abundance by day 6. Zeaxanthin, chlorophyll b and divinyl chlorophyll a were all slightly higher on average in High Fe bottles potentially

representing increased abundance of *Synechococcus*, *Chlorophyceae*, and *Prochlorococcus* in High Fe treatments (25). Cell counts of specific taxonomic groups were not statistically different between High and Low Fe treatments when normalized to total phytoplankton cell counts, indicating that phytoplankton community composition was largely unchanged between Fe treatments even though total biomass increased in High Fe samples.

5.3.3 Dissolved Fe-binding ligands

Several classes of dFe-binding ligands were determined in each experimental treatment using CLE-ACSV. The dFe-binding ligand pools were drastically different in the Low Fe and High Fe treatments by day 6. Initially, L_2 (2.2 nmol L^{-1}) and L_3 (2.5 nmol L^{-1}) ligands were detected in situ in the initial water mass while the strongest L_1 ligands were conspicuously absent (Table 5.4, Figure 5.4), consistent with observations from others in the same oceanographic region (17, 26). After six days, Low Fe treatments displayed slightly higher L_2 and L_3 concentrations than in the initial CCE waters. L_2 concentrations also changed little between initial and final sampling points in High Fe treatments, but in contrast to Low Fe treatments, L_3 concentrations markedly increased at day six in the High Fe treatments (Student's t-test, $P < 0.05$) (Table 5.4, Figure 5.4). The most striking result was the presence of siderophore-like L_1 ligands exclusively in High Fe treatments at day 6, while no Fe-binding ligands of this strength were detected in Low Fe treatments. These results suggest that L_1 and L_3 production were dependent on Fe treatment and the resulting biological dynamics in these incubations.

5.3.4 Heterotrophic bacterial assemblage

DNA was collected at day six of the incubations in order to assess heterotrophic microbial diversity within and between Fe treatments. Sequencing of the V3-V4 region of the 16S ribosomal RNA gene resulted in an average of 1.2×10^5 reads per sample (n = 6) after post-processing (see methods, Table 5.5). Operational Taxonomic Units (OTUs) were binned at 97% sequence identity resulting in 826 unique OTUs of which 304 were retained for downstream analyses after filtering out eukaryotic reads, archaeal reads, and rare taxa. A non-metric multidimensional scaling (NMDS) analysis based on diversity and abundance of OTUs indicated that Low Fe treatments clustered, while High Fe treatments were highly variable between replicates (Figure 5.5). However, broad patterns in the composition of the heterotrophic microbial community appeared to align with the High / Low Fe treatment experimental structure (Fig 5.5, Figure 5.6). Additionally, many OTUs co-occurred almost exclusively with a single replicate from High Fe treatments. For example, the distinct microbial community in High Fe replicate C was dominated by a single OTU, OTU_13 which was binned to the *Pseudoalteromonas* genus of the order *Alteromonadales* (Figure 5.7). Continuous environmental variables (dFe, NO_3^-, PO_4^{3-}, $Si(OH)_4$, chl a, and ligand concentrations) were fit to the NMDS ordination and were tested for correlations at a significance level of $\alpha = 0.1$. L_1 and NO_3^- concentrations were significantly correlated with the ordination of bacterial OTUs with the direction of the NO_3^- gradient opposite that of L1 (Figure 5.5). OTUs associated with Low Fe treatments were positively correlated with higher NO_3 concentrations ($R^2 = 0.75$, $P < 0.06$), while L_1 concentrations were strongly positively correlated with OTUs more abundant in High Fe

treatments (L_1, $R^2 = 0.77$, $P < 0.05$). Chlorophyll *a* ($R^2 = 0.64$) and dFe ($R^2 = 0.62$) were also strongly correlated with the ordination, although not significantly.

There were 13 OTUs whose mean normalized abundance across all High Fe samples was at least 1.5 times greater than their abundance in Low Fe samples and whose mean abundance across all samples was greater than 250 (Figure 5.7). These OTUs were taxonomically binned at the order level to known copiotroph groups including *Altermonadales*, *Cytophagales*, *Flavobacterales*, *Rhodobacterales*, and *Sphingobacteriales*. In addition, there were ten different OTUs with a statistically significant difference in abundance across High and Low Fe treatments (see methods), two of which (OTU_4 and OTU_21; Figure 5.7) were in the previously mentioned 13 OTUs. Of the ten differentially abundant OTUs (Table 5.6), three were enriched in High Fe treatments and fell into the orders *Alteromonadales, Caulobacterales,* and *Verrumicrobiales*. Nearly 70% of the OTUs enriched in Low Fe treatments were from SAR11 and other *Alphaproteobacteria.* The nature of the dispersion modeling procedure used to test for differential abundance (27) is such that OTUs that are highly abundant in a single sample are discarded as outliers, for example OTU 13, and it serves to highlight OTUs that are consistently abundant or scarce between treatments. Because High Fe samples were largely different from each other in taxonomic composition, this procedure showed greatest sensitivity to OTUs that were consistently abundant in Low Fe treatments rather than High Fe treatments (Table 5.6). Alpha diversity, or the total species diversity in each incubation, was statistically lower for a number of different diversity metrics in High Fe treatments (Table 5.7, Figure 5.8B). High Fe replicate C generally displayed the lowest alpha diversity of all sample treatments which may reflect the

dominance of a *Pseudoalteromonas* OTU in that particular sample. The proportion of OTUs binned at class taxonomic level was largely similar between Fe conditions with *Alphaproteobacteri*a and *Gammaproteobacteria* composing the largest fractions in all conditions (Figure 5.8A).

5.4 Discussion

Here we report the effects of differing NO_3^- : dFe conditions on the development of bacterial and phytoplankton communities and Fe-binding ligand dynamics in microcosm incubations. After six days of grow-out, both NO_3^- : dFe treatments stimulated increases in phytoplankton biomass and produced communities dominated by diatoms. However, we report multiple lines of evidence suggesting that the Fe-binding ligand pool, macronutrient concentrations, and the heterotrophic microbial community differed substantially between High and Low Fe treatments. This work suggests the potential for unique relationships between phases of phytoplankton bloom development, Fe-binding ligand dynamics, and heterotrophic bacterial community.

Despite the similarities between the phytoplankton communities in each treatment, the Fe-binding ligand pool, particularly the strongest L_1 ligands, was distinct between Fe additions at day 6. The L_1 ligand classification, defined in this study as having a $\log K \geq 12$, is comparable with the binding affinities of siderophores found in cultures of marine bacteria (13). One explanation for the emergence of L_1 ligands in High Fe treatments is that they were produced actively by the phytoplankton community. It does appear that some eukaryotic phytoplankton can utilize Fe bound to siderophores, but genomic evidence suggests that in order to do so most eukaryotic phytoplankton rely on

multi-step reductive pathways instead of direct siderophore-complex uptake (28). Additionally, canonical siderophore biosynthesis pathways have not been detected in eukaryotic phytoplankton genomes, with the exception of a putative catecholate biosynthesis motif in the green alga *Ostreococcus lucimarinus* (29). *Pseudo-nitzschia* spp., the dominant diatom under both Fe conditions in our experiments, is known to produce significant amounts of domoic acid, a toxin with Fe chelating functionalities. However, this compound has a binding affinity approximately three orders of magnitude lower than that of the L1 ligand class reported in this study (30) making it unlikely that domoic acid production was responsible for the increase in L1 concentrations in High Fe conditions. Ultimately, the similarity of the phytoplankton communities between Fe treatments and genomic and physiological evidence from model organisms does not support the hypothesis that eukaryotic phytoplankton were producing the strong siderophore-like complexes detected in these incubations.

The composition of the microbial community in the High Fe incubations was broadly correlated with increasing L_1 concentrations and was mostly dominated by OTUs related to known copiotroph heterotrophic strains. These observations are consistent with other studies reporting enrichment of copiotroph bacterial taxa after dissolved organic matter (DOM) additions (31, 32). Although L_1 increases across High Fe treatments were generally uniform, there was no universal shift in microbial community in response to High Fe treatments. Low Fe replicate incubations harbored similar microbial communities (Figure 5.5, Figure 5.6), with general enrichments of OTUs related to SAR11 (Table 5.6). In contrast, the microbial community composition in High Fe replicates were largely different from Low Fe treatments and from one another, with

High Fe C being particularly distinct due to the apparent bloom of a unique *Pseudoalteromonas* OTU. However, High Fe treatments shared similarities in that they tended to be dominated by a small handful of unique OTUs classified as belonging to taxonomic groups of known copiotrophic bacteria (Figure 5.7, Table 5.6), for example *Alteromonadales*, *Rhodobacterales*, and *Cytophagales*. Although transiently abundant copiotrophs are frequently dismissed as "weed" species, a number of studies have demonstrated that the activities of these strains have disproportionately large impacts on biogeochemical cycling. For example, a copiotroph from coastal California waters, *Alteromonas* sp. ALT199, known to be abundant and transcriptionally active in particle size fractions (33), has been shown to effectively consume as much labile marine DOM as the rest of the microbial community combined (34). ALT199 has numerous putative siderophore-specific TonB receptors, a putative siderophore synthetase for petrobactin, and a putative siderophore-ferric complex reductase. Metatranscriptomic evidence suggests that other *Altermonadales* species synthesize siderophores and take up siderophore-ligand complexes in hydrothermal vent plumes (35). Another preliminary study on Fe availability to different microbial groups found that *Alteromonas* consumed disproportionately more Fe relative to its abundance in seawater incubations (36). Additionally, a particular *Verrucomicrobia* OTU has been found to dominate certain Baltic Sea microbial communities (37), and the fully assembled genome contains numerous carbohydrate degrading enzymes as well as the functional potential for siderophore uptake (38).

Transiently abundant but transcriptionally active copiotrophs have been mostly studied in the context of DOM utilization and degradation, but these strains may also

contribute heavily to the cycling of Fe. The chemical data from this study indicates that strong Fe-binding ligands (L1) with binding-strengths comparable to siderophores only appeared under High Fe treatments. High Fe treatments generated increased phytoplankton biomass and increased nutrient drawdown, but harbored relatively similar phytoplankton community compositions. In contrast, shifts in bacterial community composition from Low to High Fe conditions were strongly correlated with increases in L_1, L_3, and Chlorophyll *a* concentrations. The mostly even L_1 distributions in High Fe treatments relative to their individually distinctive microbial communities suggest that the potential for siderophore uptake and biosynthesis may have been functionally redundant across numerous copiotroph OTUs in these incubations. The increase in weaker L_3 ligands in High Fe treatments relative to Low Fe treatments (Figure 5.4) is also interesting and may be related to the release of polysaccharide-type material by either phytoplankton or bacteria, the release of domoic acid by *Pseudo-nitzschia*, and/or the release of intracellular material by lysing cells.

We suggest that the most likely scenario accounting for the emergence of L_1 in High Fe treatments is that heterotrophic bacteria were directly producing siderophores or other strong Fe-binding ligands in order to scavenge iron from lysing algal cells. Our data suggests that by day 6 the phytoplankton community in High Fe treatments had consumed all the available nitrate and had either begun to enter or had fully entered senescence, the initial phase of remineralization. The different High Fe replicates appeared to be in largely different states of maturity with respect to bacterial taxonomic composition (Figure 5.5). Additionally, Low Fe C appeared to be the most mature of the Low Fe treatments, with the lowest NO_3^-, SiO_3, and dFe concentrations and the highest

chlorophyll concentrations. Furthermore, the bacterial assemblage composition in Low Fe C was the closest of Low Fe samples to High Fe samples. We postulate that phytoplankton senescence was coupled with a release of phytoplankton-derived (DOM) and resulted in a stimulation of copiotroph heterotrophic bacterial strains. The higher chlorophyll concentrations in High Fe samples is consistent with the greater nitrate utilization in these treatments, and it is probable that copiotroph bacterial taxa were enriched due to the release of phytoplankton-derived dissolved organic matter (DOM) (39) and not necessarily by a direct stimulation of iron. We suspect that the difference in the heterotrophic community between Low and High Fe treatments was due to an initial stimulation by DOM release, while the differences in L_1 concentrations reflected siderophore production by DOM-responding strains in High Fe samples. Copiotroph bacterial strains tend to have large genomes with multiple Fe uptake pathways suggesting that they can access a large diversity of Fe species in their immediate environment (40). It may be that these organisms also have greater Fe demand or participate in "luxury uptake" and thus siderophore production may provide a competitive advantage for bacterial strains adapted to infrequent and ephemeral nutrient pulses.

Typically, siderophores are produced only under extreme Fe limitation so it is somewhat paradoxical that siderophore production would greatly increase under conditions with the greatest amount of added Fe. However, diatoms generally outcompete bacteria and other phytoplankton after Fe fertilization, and large-scale Fe-fertilization studies have demonstrated an increase in L_1 ligands after Fe addition (14, 41), during bloom decline (41), and heterotrophic bacteria may still retain a degree of Fe-limitation after Fe enrichment (21). Our experiments provide further insight to the results from

mesoscale studies and suggest that L_1 production may be related to phytoplankton bloom phase progression rather than direct stimulation by Fe itself. Thus the apparent paradox of L_1 concentration spikes post Fe fertilization may be explained by a mostly carbon-limited heterotrophic bacterial community that subsequently shifts to Fe limitation relative to the excess carbon derived from decaying phytoplankton. If this phenomenon is widespread it may serve as a significant source of L_1 type ligands in marine waters. Excess strong ligand production by heterotrophic bacteria during early bloom senescence may thus be important for overall iron recycling efficiency in microbial ecosystems and may serve to minimize Fe loss due to particle export in the upper ocean.

Although these proposed scenarios are hypothetical, the abundance of L_1 and L_3 in the High Fe incubations and their clear correlation with shifts in the heterotrophic bacterial community (Figure 5.5, Figure 5.6, Figure 5.7, Table 5.6) suggests a potential for a link between specific microbial taxa and biological ligand production, such as siderophores. Potentially, components of the ligand pool detected during the incubations can be attributed directly to copiotrophic members of the microbial community. More work is needed to clarify the nature and extent of microbial taxonomic succession after Fe enrichments. When paired with high resolution chemical measurements of Fe-binding ligand, surveys of microbial functional gene expression as well as the recovery of entire bacterial genomes should offer promise for elucidating further linkages between ligand production and specific taxonomic groups.

5.5 Materials and Methods

5.5.1 Oceanographic setting

Samples for the incubation experiment were collected aboard the R/V *Melville* in August 2012 in the Southern California Bight (33.879° N, -123.306° W). Water for the incubation experiment was collected in the surface waters of a warm oligotrophic anticyclonic eddy off Point Conception. Trace-metal clean seawater was collected from 35 meters using a rosette containing 5 L Teflon-coated external-spring Niskin bottles (Ocean Test Equipment) deployed on non-metallic hydroline. The Niskin bottles were immediately brought into a positive pressure clean van and the unfiltered water was dispensed into an acid-cleaned 50 L carboy. The water was mixed and then placed in six 2.7 L polycarbonate bottles.

5.5.2 Incubation set-up

Incubation bottles were placed in an on-deck flow-through incubator screened to 30% incident light levels. The incubation experiment contained three low Fe (1 nmol L^{-1} $FeCl_3$; Low Fe A, B and C) and three high Fe (5 nmol L^{-1} $FeCl_3$; High FeA, B and C) treatments. The two different dFe additions aimed to explore ligand production under two different NO_3^-:dFe conditions, which have been shown to be a proxy for Fe-limitation in diatoms (24). All six bottles in the experiment also had macronutrient additions of 10.0 µmol L^{-1} NO_3^-, 1.0 µmol L^{-1} PO_4^{3-}, and 9.3 µmol L^{-1} $Si(OH)_4$. The incubation was sampled on day 0 (initial conditions) and day 6 (final conditions) for dFe, dFe-binding ligands, chl a, macronutrients (NO_3^-, PO_4^{3-}, and $Si(OH)_4$), pigment concentrations, phytoplankton cell counts, and DNA. Nutrients and chl a were also sampled from all six bottles on days 3, 5 and 6 in order to track phytoplankton growth and nutrient consumption.

5.5.3 Nutrients, pigments, and phytoplankton

$Si(OH)_4$, PO_4^{3-}, and NO_3^- (nitrate + nitrite) samples were collected in 60 ml polypropylene centrifuge tubes and frozen at -20˚C until analysis on a Lachat QuickChem 8000. Samples for chl a and phytoplankton pigments were collected in amber bottles and filtered onto GF/F filters (Fisher Scientific). The chl a filters were extracted in acetone for 24 hours and analyzed on-board using a Turner Designs 10-AU Fluorometer. Pigment samples were stored in cryovials (Nalgene) in liquid nitrogen and were analyzed by high performance liquid chromatography (HPLC) as described previously (42). Phytoplankton cell counts were collected into 50 mL glass vials with 1% tetraborate buffered formalin. Each cell count sample was adjusted by volume to 60 mL, settled in a 50 mL Utermöhl settling chamber, and counted with a Zeiss phase-contrast inverted light microscope at 200x magnification. The counts were classified in categories of *Chaetoceros* spp., *Pseudo-nitzschia* spp., other diatoms (>10 µm), dinoflagellates, flagellates (< 10 µm), and ciliates. The sample volume enumerated was approximately 5.6 to 1.1 mL (1/9 of slide), with detectable cell abundances between 245 and 1,227 cells L^{-1}, depending on the settling volume.

5.5.4 DFe and dFe-binding organic ligands

DFe was analyzed using flow-injection analysis (FIA) with sulfite reduction after pre-concentration on a conditioned nitrilotriacetic acid (NTA) column (23, 24). The concentration and conditional binding strength of organic dFe-binding ligands were measured using established electrochemical methods which have been discussed in-depth elsewhere (15, 17, 26, 43). In brief, competitive ligand exchange-adsorptive cathodic

stripping voltammetry (CLE-ACSV) was employed using a hanging mercury drop electrode (BioAnalytical Systems Inc.), with salicylaldoxime (SA) as the added ligand. Three separate titrations were done on each sample with varying concentrations of added SA in order to detect a range of dFe-binding ligands, termed multiple analytical window (MAW) analyses. A lower concentration of SA has an effective competition strength that is lower than using higher concentrations of SA; thus targeting the weaker portion of the dFe-binding ligand pool. Titrations for each sample were done with 11 separate titration points, ranging in added dFe concentrations from 0-25 nmol L^{-1} depending on the [SA] used in the titration. The added dFe and 50 µl of 1.5 mol L^{-1} boric acid-ammonium buffer (pH 8.2, NBS scale) was equilibrated with the natural ligands for at least 2 hours, and then either 17.7, 25.0, or 32.3 µmol L^{-1} SA was added and equilibrated for at least 15 minutes before the first titration point was analyzed. Peak heights were analyzed using newly available software (ECDSOFT) (44), and all three titrations were analyzed concurrently as a unified dataset in ProMCC (44). Side-reaction coefficient parameters for $Fe(SA)_x$ in ProMCC were updated to the latest calibration (45). Initial guesses in ProMCC were derived by first fitting the three titrations individually using Ružić/van den Berg and Scatchard linearization (46, 47). The results from the analysis of each titration using linear techniques were input into ProMCC and then chemical speciation fitting mode was used in order to derive ligand concentrations and strengths at the 95% confidence interval. If ProMCC could not converge in 500 iterations on three ligand classes, then the ligand parameters were reduced to two and solved again in ProMCC. Ligand classes were defined as L_1 if the conditional stability constant or strength

($\log K^{cond}_{FeL1,Fe'}$) was ≥ 12.0, L_2 when $\log K^{cond}_{FeL2,Fe'}$= 11.0-12.0, and L_3 when $\log K^{cond}_{FeL3,Fe'}$= 10.0-11.0.

5.5.5 DNA extraction

At the final time point (day 6) 100 mL of seawater was collected from the incubation experiment by filtering onto 0.22 µm Sterivex-GV filters (Millipore) using a peristaltic pump. After collection, Sterivex filters were immersed in 500 µL of lysis buffer (50 mmol L^{-1} Tris-HCl, 40 mmol L^{-1} EDTA, 0.75 mol L^{-1} sucrose, pH = 8.4) and frozen at -80°C until analysis in the laboratory. DNA was extracted from the Sterivex filters by adding 100 µL lysozyme (125 mg mL^{-1}) and 1 µL RNAse A (500 µg mL^{-1}) to the filter and incubating at 37°C for one hour. Afterwards, 100 µL proteinase K and 100 µL 20% SDS were added and the cartridge was incubated at 55°C for two hours. The resulting lysate was removed from the filter cartridge, and nucleic acids were extracted twice with phenol:chloroform:IAA (25:24:1, Sigma) and once with chloroform:isoamyl alcohol (24:1, Sigma). Nucleic acids were concentrated using Centricon 100 filters (Millipore) to between 200-500 µL.

5.5.6 Sequencing

The V3-V4 region of the 16S ribosomal RNA gene was PCR amplified using an established primer pair, S-D- Bact-0341-b-S-17/S-D-Bact-0785-a-A-21, (48) modified with Illumina adapters (Illumina Inc., San Diego, CA, USA) and Q5 polymerase (New England Biolabs). PCR products were purified using AMPure XP Beads (Beckman Coulter), PCR tagged with dual-index barcodes, then purified again using the AMPure

system. DNA concentrations were measured using a Qubit fluorometer (Life Technologies) then samples were diluted and combined. Paired-end 300-bp sequencing of barcoded amplicons was performed on a MiSeq machine running v3 chemistry (Illumina Inc., San Diego, CA, USA). Runs were uploaded to Illumina BaseSpace, automatically trimmed of primer sequences and barcodes, and downloaded as fastq format to a local computer.

5.5.7 Bioinformatic Processing

Paired reads from each sample were merged using usearch (49), allowing for zero base differences in the overlap region. Reads were truncated at an overlapping read length of 250 and then discarded if they contained a base with a Q-score lower than 15. The resulting sequences from each sample were pooled, de-replicated, and singleton sequences were discarded. These sequences were grouped into operational taxonomic units (OTUs) at 97% minimum sequence identity based on default parameters of the UPARSE-OTU algorithm, and chimeric sequences were removed post-clustering by comparison to the UCHIME gold standard reference database. Taxonomies were assigned using the Naive Baseian Classifier of the Ribosomal Database Project (50). Unclassifiable OTUs at the kingdom level, OTUs matching mitochondrial or plastid-like sequences were excluded resulting in a total of 416 OTUs across all samples. Sample processing was performed using the PhyloSeq package in R (51).

5.5.8 Statistical Processing

Prior to ordination analysis and calculation of Simpson and Shannon indexes, OTUs that did not appear at least three times in 15% of the samples were discarded. Abundances of individual OTUs in each sample were standardized to the median sequencing depth of all samples using the formula $\frac{x}{\Sigma(x)} \times median$, where x is the abundance of a given OTU. The Simpson and Shannon indexes of alpha diversity were calculated on contingency tables obtained from OTU clustering and taxonomic classification procedures using the PhyloSeq package. The normalized OTU abundance table was transformed to the Bray-Curtis dissimilarity index using the R package Vegan (52) and subjected to non-metric multidimensional scaling (NMDS) analysis using the R package MASS (isoMDS). Continuous environmental variables measured during the incubation experiments (L_1, L_2, L_3, dFe, NO_3^-, $Si(OH)_4$, PO_4^{3-}, chl a) were fitted to the resulting NMDS ordination using the envfit function from the Vegan package. *P* values for the resulting correlations were computed using 999 random permutations. To test for differential abundance of OTUs between iron treatments we used negative binomial Wald Tests using standard maximum likelihood estimates for generalized linear models coefficients under a zero-mean normal prior distribution from the DESeq2 package (method: nbinomWaldTest) in R with RLE normalization (27) as implemented in the phyloseq package (53). All tests were corrected for multiple tests using the Benjamini-Hochberg method for controlling the false discovery rate.

5.6 Acknowledgements

Chapter 5 is currently being prepared for publication by Shane L. Hogle, Randelle A. Bundy, Jessica M. Blanton, Eric E. Allen and Katherine A. Barbeau. The dissertation

author was the primary investigator and first author of this paper. This work was funded by NSF GRFP grant DGE-144086 to S.L.H.

Table 5.1 Dissolved Fe, nutrient and chlorophyll *a* concentrations

Treatment	Day	NO_3^- (µM)	PO_4^{3-} (µM)	SiO_3 µM	Chl a (µg/L)	dFe (nM)	dFe SE	NO_3^- : dFe
Initial	0	0.94	0.17	2.48	0.75	0.312	0.038	3.03
Low Fe A	3	10.4	0.90	11.1	2.07	-	-	-
Low Fe B	3	10.3	1.06	11.1	2.66	-	-	-
Low Fe C	3	10.3	0.99	10.7	2.91	-	-	--
High Fe A	3	10.0	0.98	11.0	4.15	-	-	-
High Fe B	3	9.49	0.94	10.3	4.32	-	-	-
High Fe C	3	9.69	0.91	10.3	3.85	-	-	-
Low Fe A	5	7.59	0.83	8.65	4.90	-	-	-
Low Fe B	5	6.69	0.88	7.53	6.55	-	-	-
Low Fe C	5	6.70	0.87	6.59	6.24	-	-	-
High Fe A	5	4.29	0.79	6.05	11.98	-	-	-
High Fe B	5	3.99	0.76	6.07	10.56	-	-	-
High Fe C	5	4.31	0.70	6.32	10.49	-	-	-
Low Fe A	6	2.59	0.64	4.04	9.74	0.311	0.024	8.32
Low Fe B	6	1.62	0.52	3.07	9.90	0.422	0.032	3.85
Low Fe C	6	1.27	0.57	2.25	10.11	0.214	0.021	5.93
High Fe A	6	0.11	0.45	1.86	15.22	0.614	0.044	0.18
High Fe B	6	0.09	0.44	1.96	13.39	0.621	0.167	0.14
High Fe C	6	0.18	0.45	2.94	9.04	0.601	0.057	0.30

Concentrations of nutrients and Chlorophyll *a* in the six incubations after three, five, and six days and at the start of the experiment (Day 0). Biological replicates are displayed for each iron treatment.

Table 5.2 Phytoplankton pigment concentrations

Treatment	Day	19' but (µg/L)	Fuc (µg/L)	19' hex (µg/L)	Zea (µg/L)	Chl *b* (µg/L)	DV chla (µg/L)	Chl *a* (µg/L)
Initial	0	0.052	0.226	0.120	0.033	0.042	0.040	0.423
Low Fe A	6	0.101	4.263	0.240	0.019	0.148	0.023	1.757
Low Fe B	6	0.105	5.005	0.010	0.032	0.171	0.082	2.345
Low Fe C	6	0.110	5.741	0.229	0.025	0.203	0.181	3.364
High Fe A	6	0.119	5.701	0.239	0.024	0.262	0.560	5.604
High Fe B	6	0.122	6.019	0.227	0.051	0.279	0.361	6.824
High Fe C	6	0.169	4.222	0.186	0.072	0.230	0.668	5.611

Concentrations of phytoplankton pigments as measured by HPLC in the six incubations after six days and at the start of the experiment (Day 0). Biological replicates are displayed for each iron treatment. 19’ but (19 ′ -butanoyloxyfucoxanthin), Fuc (Fucoxanthin), 19’ hex (19′-hexanoyloxyfucoxanthin), Zea (Zeaxanthin), Chl b (Chlorophyll b), DV chla (Divinyl Chlorophyll a), Chl a (Chlorophyll a)

Table 5.3 Phytoplankton cell counts

Treatment	Day	Chaeto. (10^4 cells/L)	P-nitzschia. (10^4 cells/L)	Other diatoms (10^4 cells/L)	Dino. (10^4 cells/L)	Unk Dino. / Flag. (10^4 cells/L)	Meso. (10^4 cells/L)
Initial	0	0.368	12.5	3.31	1.47	1.23	0
Low Fe A	6	29.6	210	50.9	0.61	1.84	0
Low Fe B	6	26.1	304	47.1	1.84	2.09	1227
Low Fe C	6	34.7	291	29.9	1.60	1.60	1227
High Fe A	6	26.9	389	42.1	2.21	2.95	0
High Fe B	6	38.5	408	34.6	2.95	2.70	0
High Fe C	6	51.2	382	50.6	3.07	2.70	0

Concentrations of phytoplankton cells determined by microscopy in the six incubations after six days and at the start of the experiment (Day 0). Biological replicates are displayed for each iron treatment. Chaeto. (*Chaetoceros* species), P-nitzschia. (*Pseudo-nitzschia* species), Other diatoms (Rare or unknown diatom species ≥ 10 μm), Dino. (integrated Dinoflagellate species), Unk Dino / Flag. (Unknown Dinoflagellate and Flagellate species ≥ 10 μm), Meso. (*Mesodinium* species).

Table 5.4 Dissolved Fe-binding ligand concentrations

Treatment	Day	L_1 (nM)	logK1	L_2 (nM)	logK2	L_3 (nM)	logK3
Initial	0	nd	nd	2.19	11.81	2.51	10.76
Low Fe A	6	nd	nd	2.55	11.82	3.79	10.20
Low Fe B	6	nd	nd	3.67	11.20	2.58	10.10
Low Fe C	6	nd	nd	4.10	11.74	3.65	10.20
High Fe A	6	4.39	12.02	2.97	11.38	6.58	10.10
High Fe B	6	3.65	12.36	3.07	11.00	7.54	9.39
High Fe C	6	2.54	12.08	3.57	11.16	4.60	10.16

Concentrations of dissolved Fe (dFe) and Fe-binding ligands (L_x) in the six incubations after six days and at the start of the experiment (Day 0). dFe SE indicates the standard error of dissolved Fe concentrations from technical replicates measured by FIA. LogK(x) displays $\log_{10}$ of the conditional stability constant measured for each ligand class. The NO_3^- : dFe ratio is an indicator for Fe limitation. Values greater than or equal to six are potentially indicative of Fe-limitation. Biological replicates are displayed for each iron treatment.

Table 5.5 Sequencing Statistics

Sample	Raw read count	Filtered read count	Secondary filtered read count
High Fe A	178841	139971	139505
High Fe B	123833	98411	98293
High Fe C	137718	100105	100054
Low Fe A	147891	117249	117246
Low Fe B	149017	121659	121611
Low Fe C	157690	125740	125730

Raw read count indicates the number of reads after merging paired ends and initial read quality control (see methods). Filtered read count indicate removal of OTUs unclassifiable at the taxonomic domain, mitochondrial reads, archaeal reads, plastid/cyanobacterial reads and rare OTUs. Secondary filtered read count displays the number of reads after removing OTUs that did not appear at least three times in 20% of samples.

Table 5.6 Differentially abundant OTUs between Fe treatments

OTU	Base Mean	Log_2 Fold Change	FDR adjusted *P* value	Class	Family	Genus
Otu21	3340.45	-3.66	0.0006	*Gammaproteobacteria*	*Alteromonadaceae*	*Aestuariibacter*
Otu78	133.98	-2.74	0.0001	*Alphaproteobacteria*	*Hyphomonadaceae*	*Hyphomonas*
Otu4	5061.43	-2.17	0.0006	*Verrucomicrobiae*	*Rubritaleaceae*	*Rubritalea*
Otu2	7038.59	0.73	0.02	*Alphaproteobacteria*	*SAR11*	*Pelagibacter*
Otu245	109.48	1.06	0.03	*Alphaproteobacteria*	SAR11	*Pelagibacter*
Otu45	248.55	1.08	0.04	*Alphaproteobacteria*	*Rhodobacteraceae*	*Roseibacterium*
Otu32	1146.14	1.18	0.01	*Alphaproteobacteria*	*SAR11*	*Pelagibacter*
Otu57	119.82	1.21	0.002	*Alphaproteobacteria*	*Kordiimonadaceae*	*Kordiimonas*
Otu56	212.94	1.23	0.02	*Gammaproteobacteria*	*Oleiphilaceae*	*Oleiphilus*
Otu15	1776.33	1.52	0.01	*Flavobacteriia*	*Flavobacteriaceae*	*Corallibacter*

Significant differential abundances of OTUs between high and low Fe conditions after six days. Only OTUs with false discovery rate corrected *P* values less than 0.05 and with base means greater than 100 are displayed. Base mean indicates the mean OTU abundance across all samples. Log_2 Fold Change is the fold change from high Fe to low Fe samples (i.e. a negative value indicates enrichment in high Fe samples) and OTUs are ordered by increase fold change. *P* values are FDR corrected.

Table 5.7 OTU based alpha diversity metrics

Sample	*Observed	*Chao1	*ACE	Shannon	Simpson	Inv Simpson	*Fisher
High Fe A	435.00	438.67 ± 3.15	437.81 ± 9.50	3.92	0.95	22.16	55.44
High Fe B	447.00	470.28 ± 8.30	485.86 ± 10.90	3.66	0.94	17.83	60.40
High Fe C	428.00	442.49 ± 5.80	459.46 ± 10.49	3.16	0.92	12.92	55.62
Low Fe A	515.00	531.60 ± 6.41	544.77 ± 11.46	3.79	0.94	17.27	70.93
Low Fe B	471.00	482.64 ± 5.16	495.15 ± 10.99	3.58	0.92	12.06	62.43
Low Fe C	499.00	523.93 ± 8.70	537.41 ± 11.47	3.70	0.94	17.38	66.02

Common alpha diversity metrics of the heterotrophic microbial community across the six incubations after six days. Metrics with an asterisk were significantly different between High and Low Fe treatments (Student's t-test $P < 0.05$)

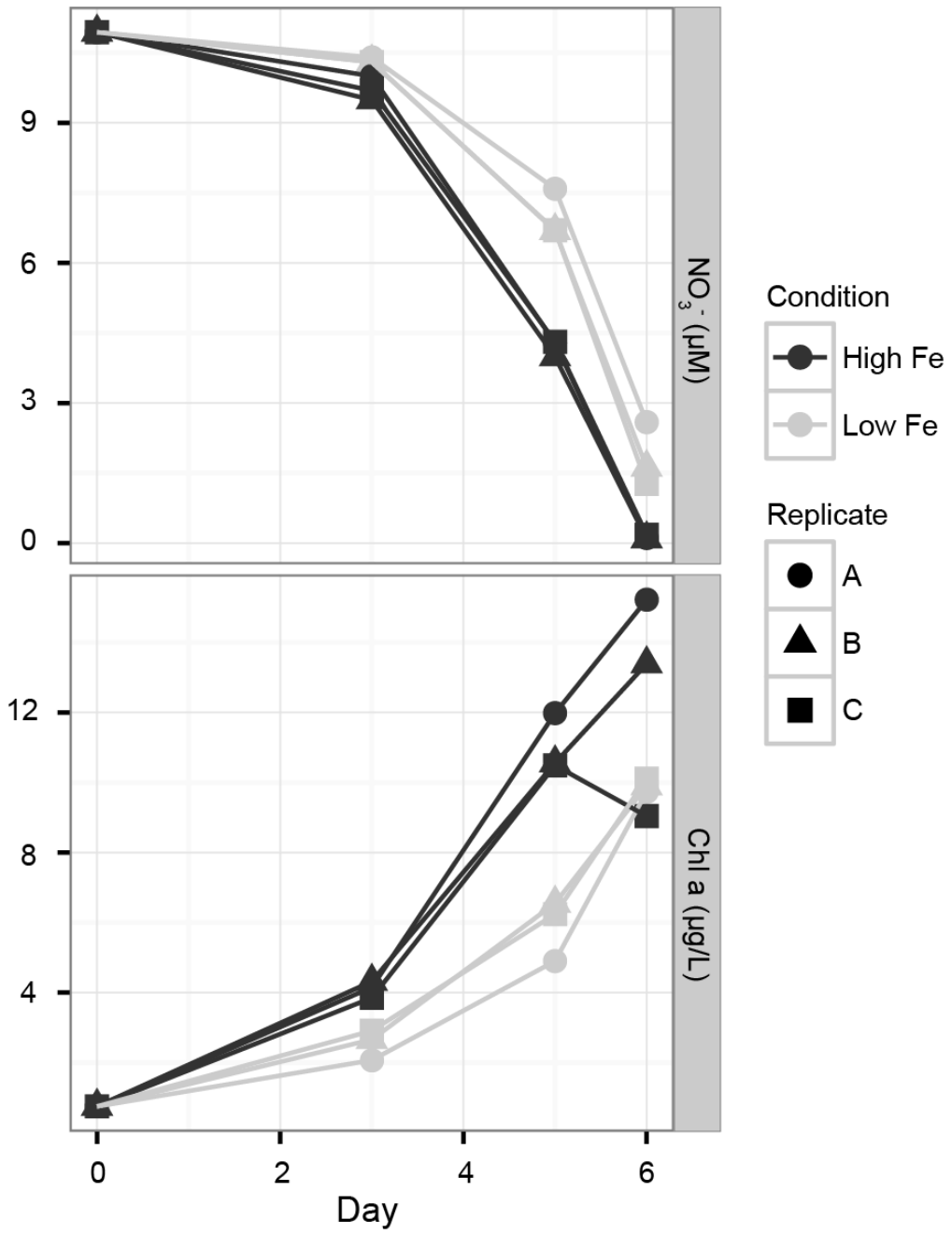


Figure 5.1 Changes in chlorophyll *a* and nitrate concentrations during the course of the high Fe and low Fe treatments. Nitrate concentrations at Day 0 are after the initial nutrient spike. High Fe incubations are in black, while Low Fe are in grey. Biological replicates are represented by shapes. Note different scales in each subplot.

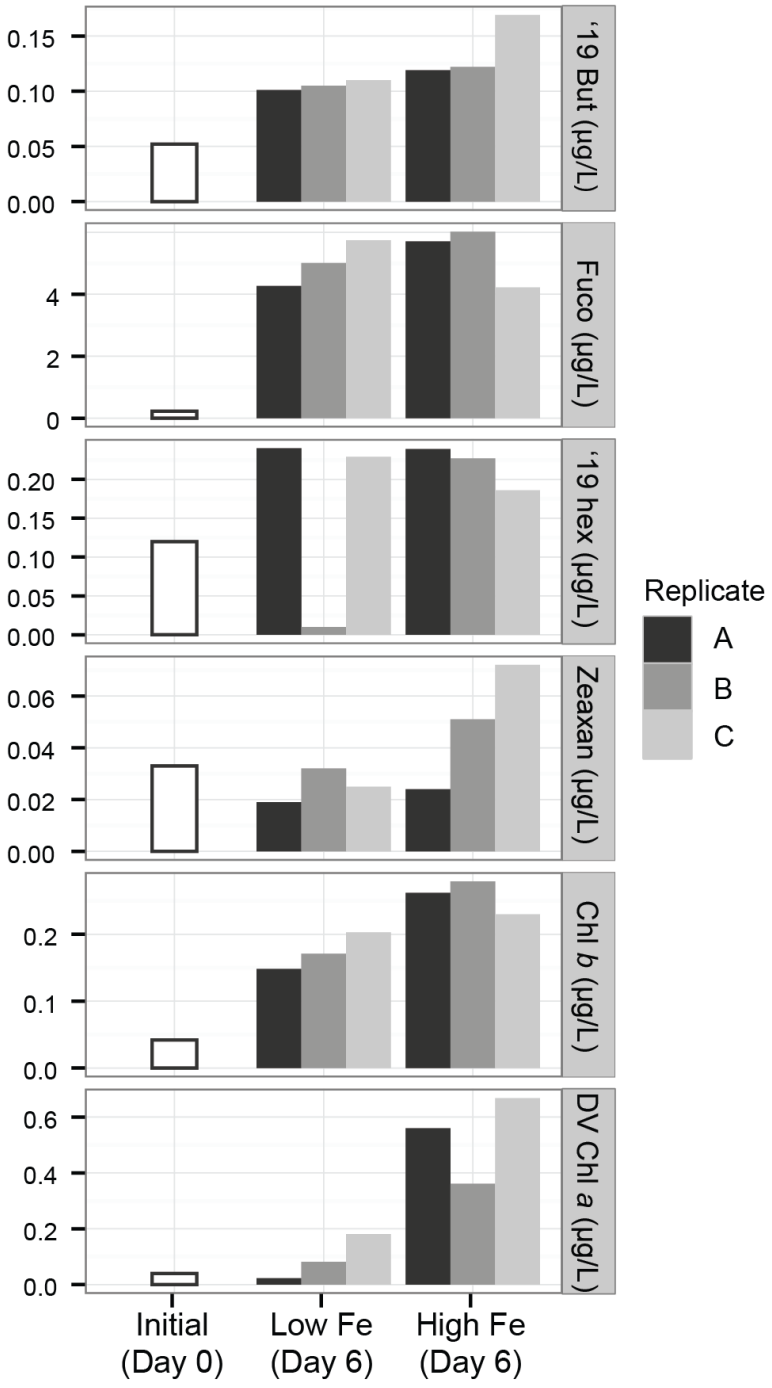


Figure 5.2 Phytoplankton pigment concentrations at initial and final sampling times in high Fe and low Fe treatments as assessed by HPLC. Initial values are from a single sample of seawater used to start the incubations and do not include biological replicates. Biological replicates are presented for Low Fe and High Fe conditions. Note different scales in each subplot.

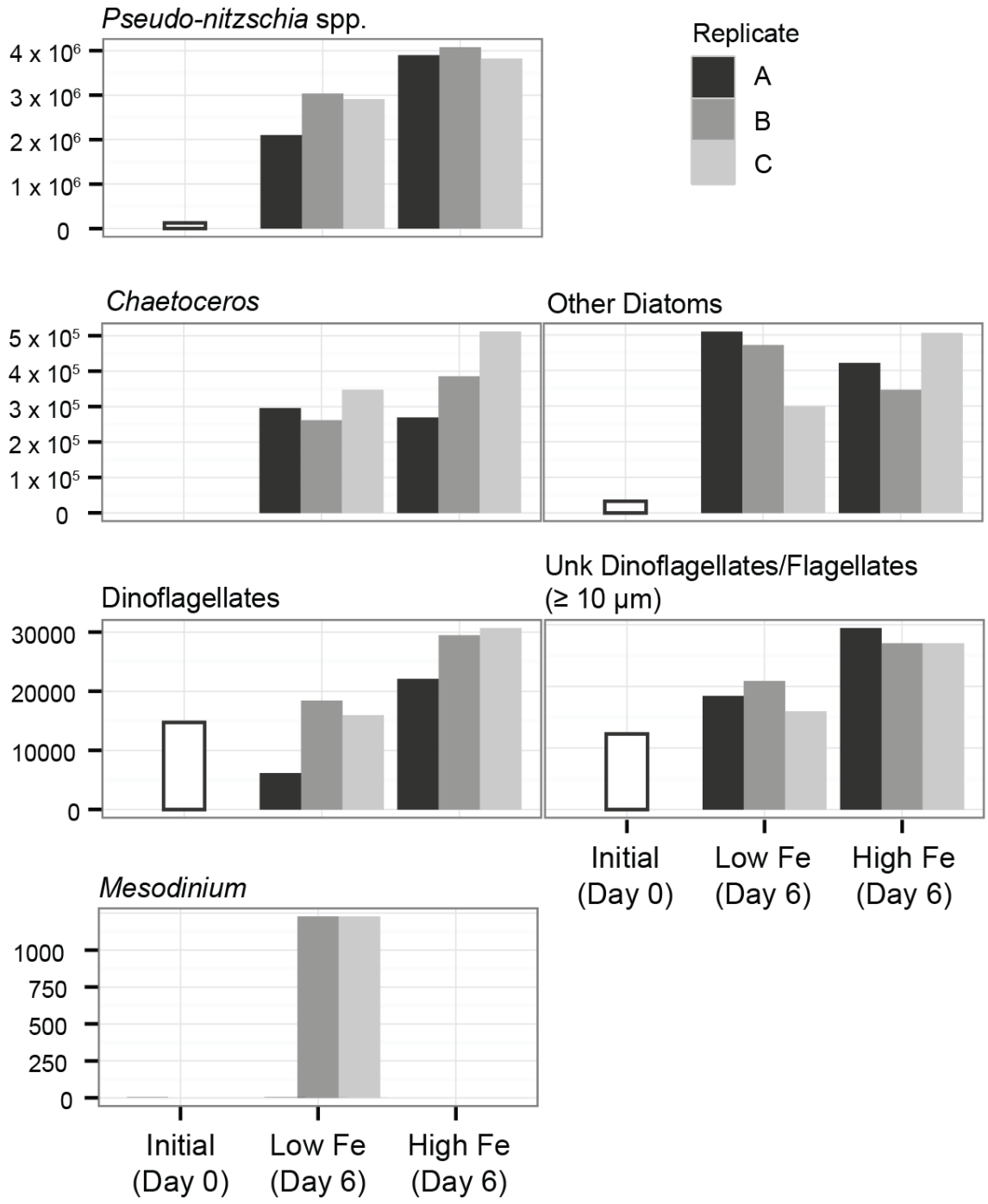


Figure 5.3 Phytoplankton cell counts at initial and final sampling times in high Fe and low Fe treatments as assessed by light microscopy. Initial values are from a single sample of seawater used to start the incubations and do not include biological replicates. Biological replicates are presented for Low Fe and High Fe conditions. Note different scales in each subplot.

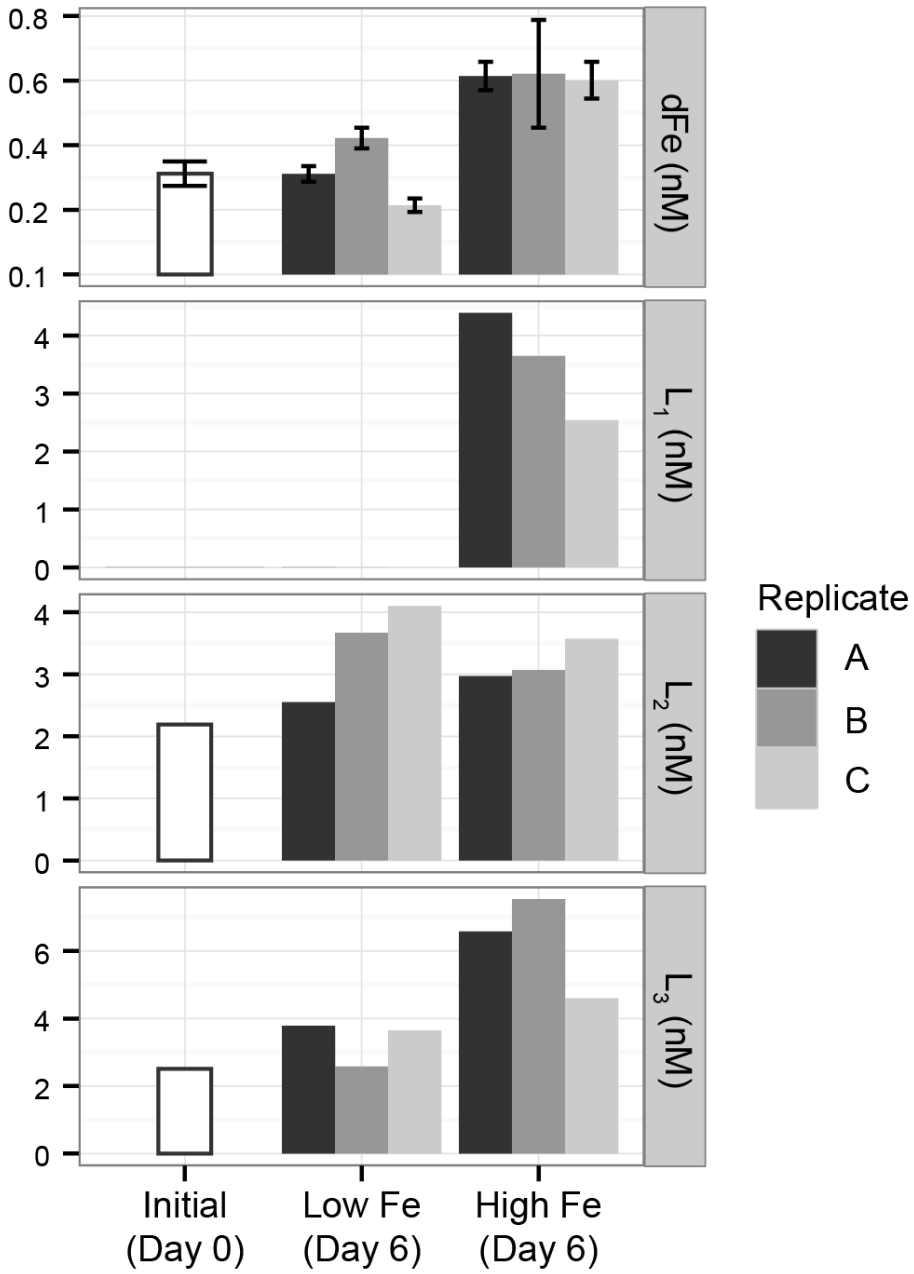


Figure 5.4 Dissolved Fe and Fe-binding ligand concentrations at initial and final sampling times in high Fe and low Fe treatments. Initial values are from a single sample of seawater used to start the incubations and do not include biological replicates. Biological replicates are presented for Low Fe and High Fe conditions. Error bars for dissolved Fe concentrations are technical replicates from Flow Injection Analysis (FIA). Note different scales in each subplot.

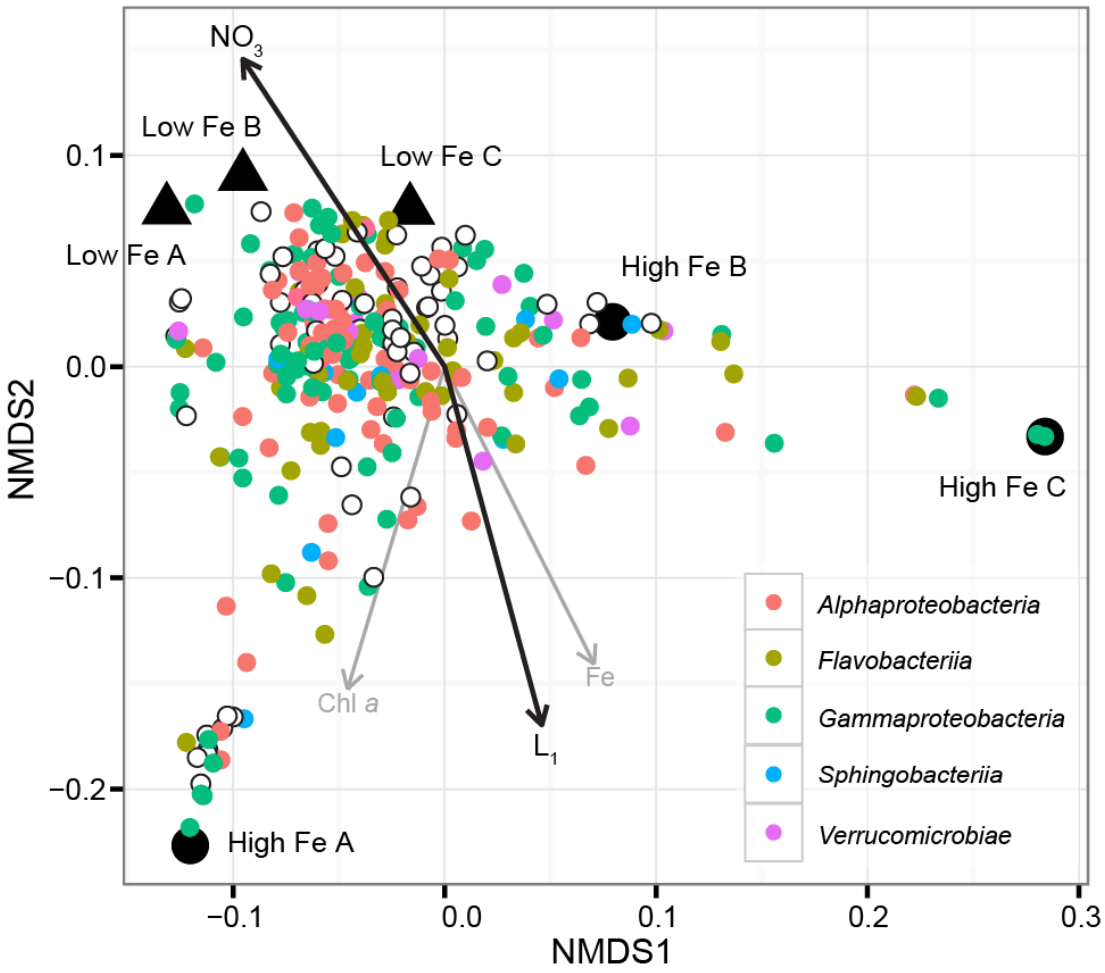


Figure 5.5 Nonmetric Multidimensional Scaling (NMDS) analysis based on the diversity and abundance of bacterial OTUs detected at Day 6 of the experiment. OTUs are colored according to the five most abundant taxonomic classes. The three High Fe samples (circles) and Low Fe samples (triangles) are plotted in addition to the ordination of individual OTUs (white circles). Arrows represent fitted vectors of continuous associated environmental variables and show the direction of the increasing gradient. Variables with a correlation *P* value < 0.1 are shown in black while those with *P* > 0.1 are shown in grey. Arrow length is proportional to the correlation between the variable and ordination. L_1, $R^2 = 0.77$; Fe, $R^2 = 0.62$; ChlA, $R^2 = 0.64$; NO_3, $R^2 = 0.75$.

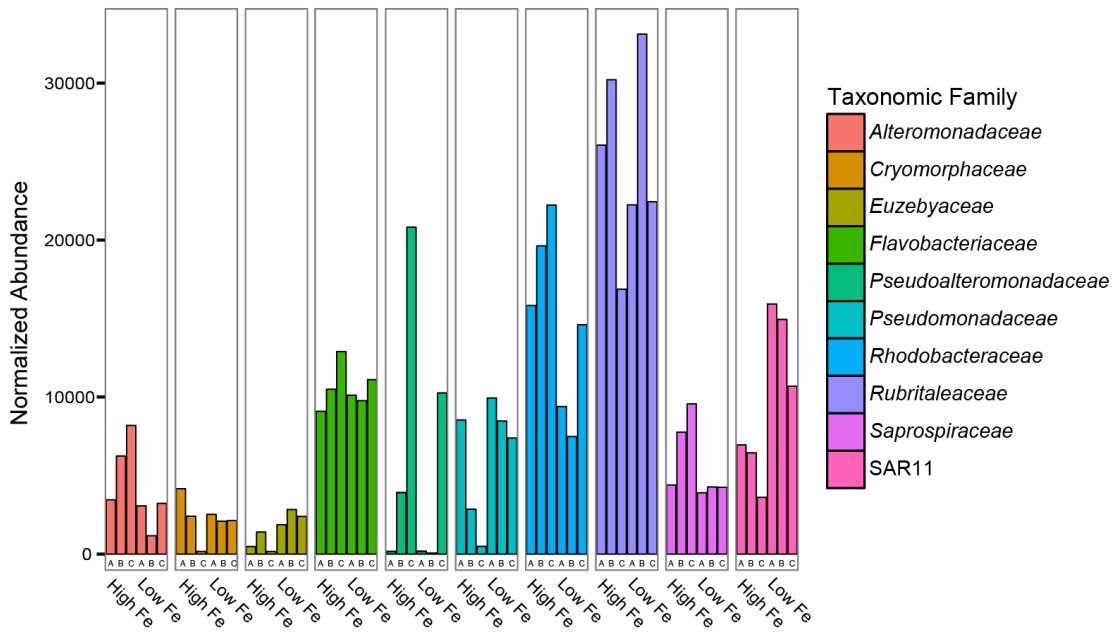


Figure 5.6 The 20 most abundant OTUs in the combined high and low Fe treatments binned by taxonomic family.

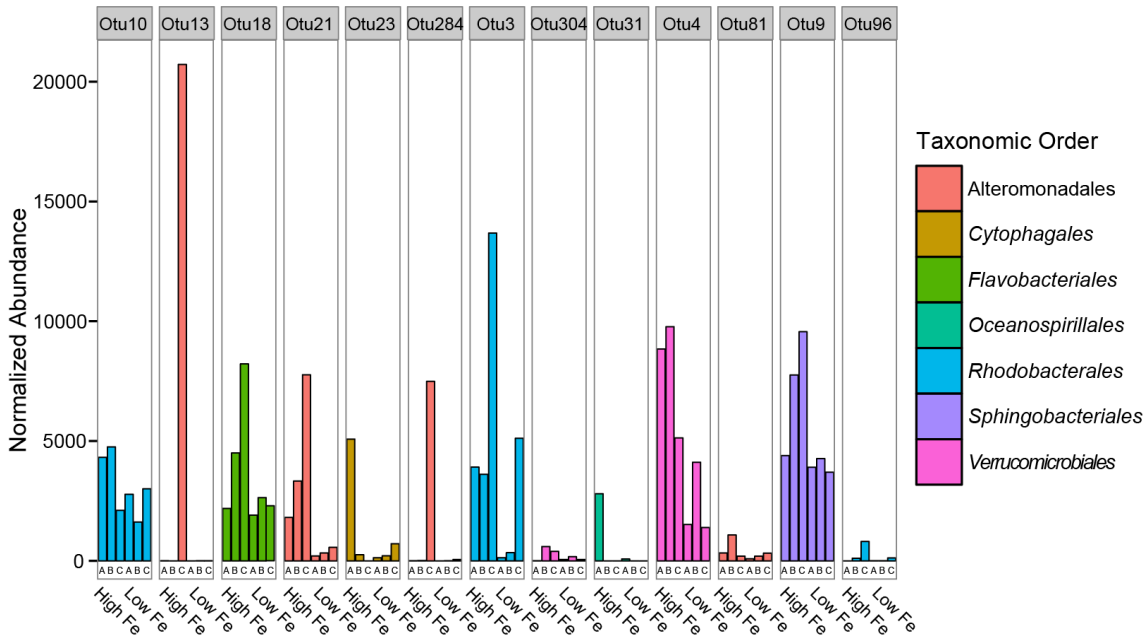


Figure 5.7 13 OTUs disproportionately represented in high Fe treatments. Bar plot showing 13 OTUs with a mean read count > 250 in High Fe samples and the High Fe mean is at least 1.5X greater than the Low Fe mean.

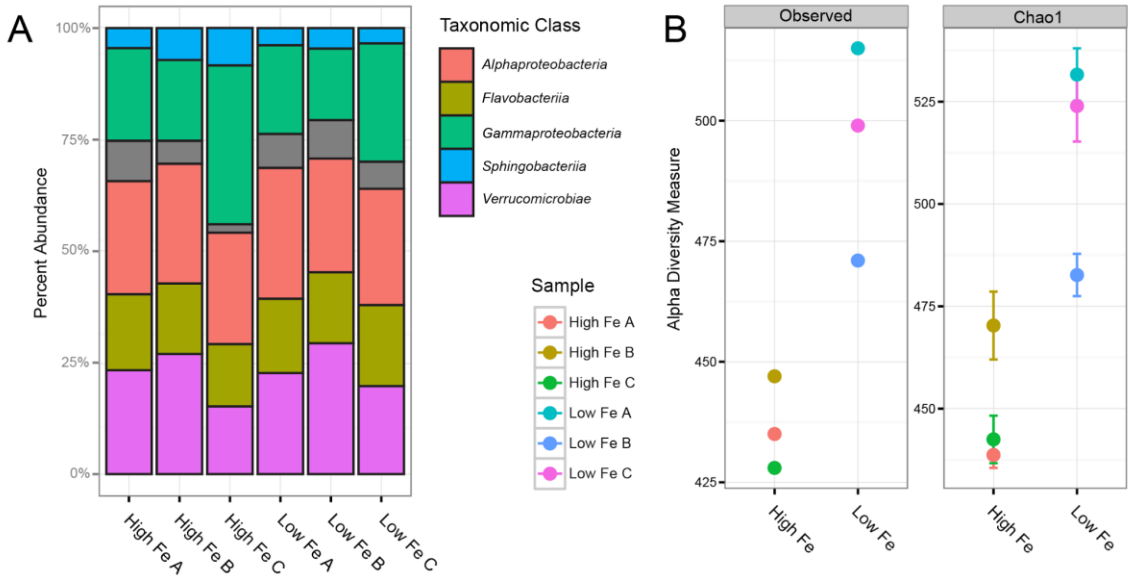


Figure 5.8 Taxonomic classes and alpha diversity in high and low Fe treatments. (A) Bar plot showing the percentage of taxonomic classes in each sample. (B) “Observed” and “Chao1” Alpha diversity metrics for High and Low Fe samples.

5.7 References

1. Boyd PW, Jickells T, Law CS, Blain S, Boyle EA, Buesseler KO, Coale KH, Cullen JJ, de Baar HJ, Follows M, Harvey M, Lancelot C, Levasseur M, Owens NP, Pollard R, Rivkin RB, Sarmiento J, Schoemann V, Smetacek V, Takeda S, Tsuda A, Turner S, Watson AJ (2007) Mesoscale Iron Enrichment Experiments 1993-2005: Synthesis and Future Directions. Science 315(5812):612–617.

2. Bowie R, Maldonado MT, Frew RD, Croot PL, Achterberg EP, Mantoura RF, Worsfold PJ, Law CS, Boyd PW. (2001) The fate of added iron during a mesoscale fertilisation experiment in the Southern Ocean. *Deep Sea Res Part 2 Top Stud Oceanogr* 48(11-12):2703–2743.

3. Boyd PW, Law CS, Hutchins D, Abraham ER, Croot PL, Ellwood M, Frew RD, Hadfield M, Hall J, Handy S, Hare C, Higgins J, Hill P, Hunter K, LeBlanc K, Maldonado MT, McKay RM, Mioni C, Oliver M, Pickmere S, Pinkerton M, Safi K, Sander S, Sanudo-Wilhelmy S, Smith M, Strzepek R, Tovar-Sanchez A, Wilhelm SW. (2005) FeCycle: Attempting an iron biogeochemical budget from a mesoscale SF6 tracer experiment in unperturbed low iron waters. *Global Biogeochem Cycles* 19(4):1–13.

4. Völker C, Tagliabue A (2015) Modeling organic iron-binding ligands in a three-dimensional biogeochemical ocean model. *Mar Chem* 173:67–77.

5. Strzepek RF, Maldonado MT, Higgins JL, Hall J, Safi K, Wilhelm SW, Boyd PW. (2005) Spinning the "Ferrous Wheel": The importance of the microbial community in an iron budget during the FeCycle experiment. *Global Biogeochem Cycles* 19(4):1–14.

6. Barbeau K, Kujawinski E, Moffett J (2001) Remineralization and recycling of iron, thorium and organic carbon by heterotrophic marine protists in culture. *Aquat Microb Ecol* 24:69–81.

7. Smith DC, Simon M, Alldredge AL, Azam F (1992) Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. *Nature* 359(10):139–142.

8. Bidle KD, Azam F (1999) Accelerated dissolution of diatom silica by marine bacterial assemblages. *Nature* 397:508–512.

9. Boyd PW, Ibisanmi E, Sander SG, Hunter K a., Jackson G a. (2010) Remineralization of upper ocean particles: Implications for iron biogeochemistry. *Limnol Oceanogr* 55(3):1271–1288.

10. Boyd PW, Ellwood MJ (2010) The biogeochemical cycle of iron in the ocean. *Nat Geosci* 3(10):675–682.

11. Mawji E, Gledhill M, Milton J, Tarran G, Ussher S, Thompson A, Wolff G, Worsfold PJ, Achterberg EP. (2008) Hydroxamate siderophores: occurrence and importance in the Atlantic Ocean. *Environ Sci Technol* 42(23):8675–8680.

12. Boiteau RM, Fitzsimmons JN, Repeta DJ, Boyle E a. (2013) Detection of iron ligands in seawater and marine cyanobacteria cultures by high-performance liquid chromatography-inductively coupled plasma-mass spectrometry. *Anal Chem* 85(9):4357–4362.

13. Vraspir JM, Butler A (2009) Chemistry of Marine Ligands and Siderophores. *Ann Rev Mar Sci* 1(1):43–63.

14. Rue EL, Bruland KW (1997) The role of organic complexation on ambient iron chemistry in the equatorial Pacific Ocean and the response of a mesoscale iron addition experiment. *Limnol Oceanogr* (Iii):901–910.

15. Rue EL, Bruland KW (1995) Complexation of iron(III) by natural organic ligands in the Central North Pacific as determined by a new competitive ligand equilibration/adsorptive cathodic stripping voltammetric method. *Mar Chem* 50:117–138.

16. Hopkinson B, Barbeau K (2012) Iron transporters in marine prokaryotic genomes and metagenomes. *Environ Microbiol* 14(1):114–128.

17. Bundy RM, Biller DV, Buck KN, Bruland KW, Barbeau KA (2014) Distinct pools of dissolved iron-binding ligands in the surface and benthic boundary layer of the California Current. *Limnol Oceanogr* 59(3):769–787.

18. Norman L, Worms IA, Angles E, Bowie AR, Nichols CM, Ninh Pham A, Slaveykova VI, Townsend AT, David Waite T, Hassler CS. (2015) The role of bacterial and algal exopolymeric substances in iron chemistry. *Mar Chem* 173:148–161.

19. Wu J, Boyle E, Sunda W, Wen LS (2001) Soluble and colloidal iron in the oligotrophic North Atlantic and North Pacific. *Science* 293(5531):847–849.

20. Gledhill M, Buck KN (2012) The organic complexation of iron in the marine environment: a review. *Front Microbiol* 3(February):69.

21. Adly CL, Tremblay JE, Powell RT, Armstrong E, Peers G, Price NM. (2015) Response of heterotrophic bacteria in a mesoscale iron enrichment in the northeast subarctic Pacific Ocean. *Limnol Oceanogr* 60(1):136–148.

22. Cordero OX, Ventouras L-A, DeLong EF, Polz MF (2012) Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc Natl Acad Sci U S A* 109(49):20059–20064.

23. King AL, Barbeau KA (2011) Dissolved iron and macronutrient distributions in the

southern California Current System. *J Geophys Res* 116(C3):C03018.

24. King AL, Barbeau K (2007) Evidence for phytoplankton iron limitation in the southern California Current System. *Mar Ecol Prog Ser* 342:91–103.

25. Mackey MD, Mackey DJ, Higgins HW, Wright SW (1996) CHEMTAX - a program for estimating class abundances from chemical markers:application to HPLC measurements of phytoplankton. *Mar Ecol Prog Ser* 144:265–283.

26. Bundy RM, Abdulla HA, Hatcher PG, Biller DV, Buck KN, Barbeau KA. (2015) Iron-binding ligands and humic substances in the San Francisco Bay estuary and estuarine-influenced shelf regions of coastal California. *Mar Chem* 173:183–194.

27. Love MI, Huber W, Anders S (2014) Fully formatted Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(550):1–21.

28. Morrissey J, Bowler C (2012) Iron utilization in marine cyanobacteria and eukaryotic algae. *Front Microbiol* 3(March):43.

29. Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, Zhou K, Otillar R, Merchant SS, Podell S, Gaasterland T, Napoli C, Gendler K, Manuell A, Tai V, Vallon O, Piganeau G, Jancek S, Heijde M, Jabbari K, Bowler C, Lohr M, Robbens S, Werner G, Dubchak I, Pazour GJ, Ren Q, Paulsen I, Delwiche C, Schmutz J, Rokhsar D, Van de Peer Y, Moreau H, Grigoriev IV. (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A* 104(18):7705–7710.

30. Rue E, Bruland K (2001) Domoic acid binds iron and copper: a possible role for the toxin produced by the marine diatom Pseudo-nitzschia. *Mar Chem* 76(1–2):127–134.

31. Nelson CE, Carlson CA (2012) Tracking differential incorporation of dissolved organic carbon types among diverse lineages of Sargasso Sea bacterioplankton. *Environ Microbiol* 14(6):1500–1516.

32. McCarren J, Becker JW, Repeta DJ, Shi Y, Young CR, Malmstrom RR, Chisholm SW, DeLong EF. (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci U S A* 107(38):16420–16427.

33. Dupont CL, McCrow JP, Valas R, Moustafa A, Walworth N, Goodenough U, Roth R, Hogle SL, Bai J, Johnson ZI, Mann E, Palenik B, Barbeau K, Craig Venter J, Allen AE. (2015) Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J* 9(5):1076–1092.

34. Pedler BE, Aluwihare LI, Azam F (2014) Single Bacterial Strain Capable of Significant Contribution to Carbon Cycling in the Surface Ocean. *Proc Natl Acad Sci*

U S A 111(20):7202–7207.

35. Li M, Toner BM, Baker BJ, Breier JA, Sheik CS, Dick GJ. (2014) Microbial iron uptake as a mechanism for dispersing iron from deep-sea hydrothermal vents. *Nat Commun* 5:3192.

36. Fourquez M, Obernosterer I, Blain S (2012) A method for the use of the radiotracer 55Fe for microautoradiography and CARD-FISH of natural bacterial communities. *FEMS Microbiol Lett* 337(2):132–139.

37. Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* 5(10):1571–1579.

38. Herlemann DP, Lundin D, Labrenz M, Jürgens K, Zheng Z, Aspeborg H, Andersson AF. (2013) Metagenomic De Novo assembly of an aquatic representative of the verrucomicrobial class Spartobacteria. *mBio* 4(3):1–9.

39. Buchan A, LeCleir GR, Gulvik C a., González JM (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* 12(10):686–698.

40. Hogle SL, Cameron Thrash J, Dupont CL, Barbeau KA (2016) Trace metal acquisition by marine heterotrophic bacterioplankton with contrasting trophic strategies. *Appl Environ Microbiol*. doi:10.1128/AEM.03128-15.

41. Kondo Y, Takeda S, Nishioka J, Obata H, Furuya K, Johnson WK, Wong CS. (2008) Organic iron (III) complexing ligands during an iron enrichment experiment in the western subarctic North Pacific. *Geophys Res Lett* 35(12):L12601.

42. Zapata M, Rodríguez F, Garrido JL (2000) Separation of chlorophylls and carotenoids from marine phytoplankton:a new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases. *Mar Ecol Prog Ser* 195:29–45.

43. Buck KN, Bruland KW (2007) The physicochemical speciation of dissolved iron in the Bering Sea, Alaska. *Limnol Oceanogr* 52(5):1800–1808.

44. Omanović D, Garnier C, Pižeta I (2015) ProMCC: An all-in-one tool for trace metal complexation studies. *Mar Chem* 173:25–39.

45. Abualhaija MM, van den Berg CMG (2014) Chemical speciation of iron in seawater using catalytic cathodic stripping voltammetry with ligand competition against salicylaldoxime. *Mar Chem* 164:60–74.

46. Mantoura RFC, Riley JP (1975) The analytical concentration of humic substances from natural waters. *Anal Chim Acta* 76(1):97–106.

47. Scatchard G (1949) The attractions of proteins for small molecules and ions. *Ann N Y Acad Sci* 51(4):660–672.

48. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41(1):e1.

49. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10(10):996–998.

50. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16):5261–5267.

51. McMurdie PJ, Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8(4). doi:10.1371/journal.pone.0061217.

52. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MH, Wagner H. (2013) vegan: Community Ecology Package. Available at: http://cran.r-project.org/package=vegan.

53. McMurdie PJ, Holmes S (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol* 10(4). doi:10.1371/journal.pcbi.1003531.

Chapter 6

Molecular and geochemical evidence for iron limitation in subsurface chlorophyll maxima of the eastern subtropical Pacific Ocean.

6.1 Abstract

While iron limitation is well-documented to be a significant control on phytoplankton productivity and species composition in large areas of the surface ocean mixed layer, much less is known about the potential for iron limitation in phytoplankton communities at the subsurface chlorophyll maximum (SCM) observed ubiquitously in permanently and seasonally stratified areas of the global ocean. Here we present studies of a range of SCM communities off the southern California coast, extending from inshore waters to the offshore oligotrophic gyre. Metatranscriptomes of *in situ* phytoplankton assemblages from the SCM harbored strikingly abundant concentrations of transcripts matching known iron stress responsive genes and these iron stress markers were enriched in the SCM relative to surface waters. SCM stations were often characterized by strong nitricline and ferricline decoupling, leading to elevated nitrate to dFe (N:dFe) ratios, and indicating *in situ* Fe limitation. Additionally, silicic acid:nitrate ratios (Si:N) in time-series data sets from the region indicate that the effects of iron limitation on subsurface phytoplankton communities may be spatially widespread and temporally frequent, particularly in areas within 450 km of the coast where SCM communities represent maxima in both phytoplankton biomass and productivity. These results suggest that iron limitation may be widespread in SCM layers off the California coast and have

implications for biogeochemical processes like carbon export in this important eastern boundary current ecosystem.

6.2 Introduction

Iron limitation is currently believed to be a controlling factor for phytoplankton primary productivity in up to 40% of the world ocean (1). For the most part, these are oceanic areas characterized by continuous surface accumulations of nitrate, at times in excess of 20 μM (2), or seasonal coastal upwelling of high nitrate, low-Fe subsurface waters. In these areas, numerous bottle-scale and mesoscale iron addition experiments have demonstrated that large phytoplankton, like diatoms, are particularly limited by iron supply, with associated implications for atmospheric CO_2 uptake and export to the ocean interior (3). In the rest of the ocean where nitrate levels are generally undetectable at the surface, iron supply is not believed to limit phytoplankton growth, aside from the potential limitation of nitrogen-fixing marine cyanobacteria such as *Trichodesmium* (4). In these permanently or seasonally stratified low-nutrient regions of the ocean, nitrate is considered to be the proximate limiting nutrient in the mixed layer, which is dominated by a community of small phytoplankton such as cyanobacteria, chlorophytes, and haptophytes which grow quickly and can exist on recycled forms of nitrogen.

Below the surface mixed layer, at the base of the euphotic zone, stratified seas often harbor a distinct subsurface chlorophyll maximum (SCM) (5, 6) where the taxonomic composition of the phytoplankton assemblage can vary significantly from surface waters (7–10). Although this chlorophyll maximum can be decoupled from the water column biomass maximum due to cellular adaptations to low light (5), the SCM

generally marks a depth of high primary productivity and carbon export relative to the surface (11). The mechanisms that form and sustain SCMs are still debated to some extent, but most studies converge towards a generally unified explanation where the SCM results from two competing gradients of light and nutrient concentrations (6). For example, new nutrients are injected into the ocean euphotic zone by physical processes like diffusion or through mesoscale processes like upwelling and eddies. However, water column stratification, particularly in oligotrophic oceans, can prevent the injection of new nutrients from deeper water into the upper reaches of the euphotic zone where light levels are replete (12). Furthermore, biological uptake in these highly stratified layers with already minimal nutrient input can reduce euphotic zone nutrient concentrations to nearly undetectable levels. During upwelling vertically transported nutrients are first utilized by phytoplankton residing at the base of the euphotic zone, and uptake and remineralization processes within this layer have important consequences for the subsequent phytoplankton layers above. Cumulatively, quasi-stable upper ocean stratification generally works to sustain two different biogeochemical strata in the vertical dimension of the euphotic zone. In the upper layer, phytoplankton are strongly nutrient limited but light replete while the lower layer is always light limited. Interactions between irradiance levels and nutrients transported from below makes this lower stratum susceptible to co-limitation of both light and nutrients, and it is within this depth range that SCMs are situated.

SCMs can be found throughout the global ocean in both coastal and pelagic regions, and many of the regions where SCMs are formed are typically classified as macronutrient limited. The influences of micronutrients have received comparatively

little attention in these regimes even though micronutrients are known to have significant impact on phytoplankton growth (13). Although nitrate availability is certainly a limiting factor in upper layers of SCMs, phytoplankton communities at the base of SCMs where the nitricline begins may be predominantly limited by the availability of light and the efficiency of the photosynthetic apparatus. Indeed, the seminal culture-studies of Sunda and Huntsman showed a synergistic effect of iron and light co-limitation whereby light limitation increased the amount of cellular Fe required to support growth in four marine eukaryotic algal species (14). This work laid the theoretical foundation for a physiological link between iron availability, light intensity, and cell size and also implied the potential for Fe-limitation of phytoplankton communities situated in the lower reaches of the euphotic zone where nitrogen stress is relieved but light intensity remains low. A more recent study provided evidence for Fe-light co-limitation of phytoplankton communities in SCMs sampled from mesotrophic and oligotrophic sites in the California bight and the eastern tropical north Pacific (15). These coastal California regions have been historically considered as solely nitrogen limited, but contemporary studies now highlight an important role for Fe in structuring phytoplankton communities and in modulating carbon export in this highly productive eastern boundary current system (16–19).

Given that phytoplankton iron quotas can vary as a function of light intensity and that iron is increasingly being recognized as a factor influencing ecosystem productivity in regions previously thought to be macronutrient limited, we investigated the role of Fe as a limiting factor in SCMs along a productivity gradient in the eastern subtropical Pacific Ocean. Unlike in prior studies we explored the role of Fe from combined

geochemical and “omic” perspectives that allowed us to assess the role of iron from two independent, but complementary points of view. We collected hydrographic parameters, macronutrient and dissolved Fe concentrations, measurements of phytoplankton community composition, biomass, primary production rates, and nitrogen uptake rates from SCM and surface waters along CalCOFI line 93, the southernmost line in the CalCOFI program. Additionally, community RNA and DNA were sequenced from four SCM and surface stations in order to compare the taxonomic composition and function of expressed genes of phytoplankton communities at these stations as has already been described (20). Finally, we determined silicic acid:nitrate (Si:N) ratios using publically available data from the CalCOFI program, which provided a broader spatial and temporal perspective on the biogeochemical status of the southern California Current System (CalCS). The intersections of these data ultimately provided multiple independent lines of evidence for the occurrence of Fe stress and limitation in SCMs of the eastern subtropical Pacific Ocean.

6.3 Results and Discussion

6.3.1 Overview of nutrient and biomass distributions on CalCOFI line 93

Sampling conditions along the 700 km transect (Figure 6.1) ranged from mesotrophic near shore conditions to oligotrophic conditions in the offshore north Pacific oligotrophic gyre. There was a general shoaling of the nutricline (NO_3, $Si(OH)_4$) and isopycnals closer to the coast (Figure 6.2) which co-occurred with increased chlorophyll *a* (chl *a*) concentrations at both the SCM and surface in coastal stations (Figure 6.3A). However, nitrate and silicic acid concentrations were always low at the surface (nitrate

was always lower than 1 μM and silicic acid was never greater than 2 μM), and chl *a* concentrations were always at maximum at subsurface depths. The SCM occurred between depths of 5 to 30 m in the roughly first 300 km of the transect (stations 93.35 to 93.70) with the bulk of maxima residing at approximately 30 m depth. At approximately 400 km (station 93.80) the SCM began to track the nutricline to deeper waters (approximately 60 m) and SCMs at the oligotrophic offshore stations (93.110, 93.120) in final 100 km moved deeper in the water column to depths of approximately 110 m (Figure 6.3A).

6.3.2 Nitrate to dissolved Fe ratios on CalCOFI line 93: indicators of Fe limitation

Previous work in the Southern CalCS has demonstrated that the ratio of nitrate to dissolved Fe (N:dFe) is an index of overall community Fe stress (18, 19, 21). Phytoplankton communities from source waters with N:dFe (μM NO_3:nM dFe) ratios of approximately 6 to 12 have been observed to respond to Fe additions with increased Chl a concentrations, cell counts, and nitrate drawdown, while phytoplankton communities from water masses with N:Fe ratios of ≤ 5 have little to no response to added Fe (18, 19, 21). The physiological origin for this phenomenon is presumably due to the obligate Fe requirement of the nitrate reductase enzyme in diatoms (22, 23), and phytoplankton are observed to display highest growth rates when total cell Nitrogen:Fe quotas range from approximately 1 to 5 μM N:nM Fe (16, 24). Based on these observations, phytoplankton communities growing in upwelled source waters with an N:dFe ratio greater than 12 (the most conservative estimate) are likely to be primarily or co-limited by Fe. Waters at the base of the SCM have previously been observed with N:dFe ratios within the range

indicative of Fe-limitation and growth rates of endemic SCM phytoplankton communities have been shown to respond to both added Fe and light (15).

Total dFe concentrations in the line 93 transect were low (average of 0.22 nM) with a general trend towards increased concentration at depth (Figure 6.4). This trend was more pronounced in near-shore stations where the ferricline (as defined by the presence of ~0.2 nM dissolved iron) was shallower. Stations in the first 400 km (93.40, 93.60, and 93.80) displayed a decoupling of the nitricline (as defined by the presence of 1 μM nitrate) and the ferricline. This decoupling resulted in elevated nitrate to dissolved Fe (N:Fe) ratios at depths near or below the SCM (Figure 6.4). N:dFe ratios in the line 93 transect were frequently greater than or equal to 12 at the depth of the SCM thus indicating a strong potential for Fe-limitation. Elevated N:dFe ratios were particularly striking at station 93.40 where they approached 100 at the SCM (Figure 6.4). Station 93.40 resides in the coastal transition zone, a region 50-150 km offshore characterized by intermediate temperature, salinity, and isopycnal shoaling due to wind stress curl and eddy associated upwelling (25). A prior study reported elevated N:dFe ratios in the mixed layer of the transition zone during the spring and summer months of 2003 and 2004 (26) which is consistent with our observations of elevated N:dFe ratios at station 93.40 here. Stations 93.60 and 93.80 located approximately 230-400 km offshore both had elevated but smaller N:dFe ratios ranging from 6 to 12 at the depth of the SCM suggesting that Fe limitation may have been less dramatic than at 93.40. These stations occupied the beginning of the offshore regime characterized by low surface nitrate and dFe concentrations. Pelagic stations 93.110 and 93.120 were located more than 600 km offshore and were characterized by deep SCMs at around 100 meters depth and with very

low dFe and nitrate concentrations. However, N:dFe ratios were highly elevated just below the chl *a* maximum (Figure 6.4). The elevated N:dFe ratios at or near all the SCMs sampled on line 93 strongly indicate that phytoplankton communities were potentially experiencing Fe limitation at the time of sampling.

6.3.3 Excess silicic acid on CalCOFI line 93 and from the CalCOFI archive: Indicators of diatom Fe stress

In addition to the elevated N:dFe ratios at SCMs on line 93, distributions of excess silicic acid (Si_{ex}) provide complementary evidence for Fe limitation in SCMs from the eastern subtropical Pacific Ocean. The Si_{ex} index is sensitive to Fe dynamics and is defined as the excess concentration of silicic acid in the water column with respect to nitrate. Negative Siex values indicate an enrichment of nitrate relative to silicic acid in the water column and are thought to be related to diatom nutrient demand relative to Fe availability (27, 28). Diatoms acquire nitrate and silicic acid at a roughly equimolar ratio (Si:N = 1) when grown in nutrient replete conditions (29), but diatoms have been observed to utilize excess silicic acid relative to nitrate when stressed for iron (27, 28). If upwelled waters supply silicic acid and nitrate at an equimolar ratio (26) then an observed deviation from that ratio in the water column should reflect differences in biological uptake. Positive Siex values would indicate preferential uptake of nitrate relative to silicic acid which in turn would imply that non-silicifying organisms are consuming the bulk of nitrate. Alternatively, negative Siex values would indicate a preferential utilization of silicic acid relative to nitrate by the phytoplankton community, which in turn should reflect the dominance of diatoms. Importantly, diatom growth under nutrient replete

conditions or growth by non-silicifying organisms should serve to drive Si_{ex} to positive values. Thus the large deficit of silicic acid relative to nitrate in the water column is strongly indicative of a diatom-dominated phytoplankton community facing Fe-limiting conditions.

On line 93 during late July 2007, Si_{ex} values were zero or positive in surface waters of the transect with the exception of station 93.40 where there were slightly negative Si_{ex} values in surface waters. Deeper in the water column there was a strong silicic acid deficit roughly tracking the σ_t = 25.5 isopycnal (Figure 6.3B). Si_{ex} was most negative at depths just below the SCM layer at all stations, and negative Si_{ex} mirrored elevated N:dFe ratios at stations 93.40 and 93.60 consistent with Fe limitation at this depth range (Figure 6.4). We also leveraged the CalCOFI data archives from the summer of 2007 (06/28/2007 to 07/13/2007) in order examine patterns of nutrient distributions at a broader spatial scale. The CalCOFI program does not measure dFe as a part of their surveys due to the many analytical challenges associated with making the measurement. However, nitrate, silicic acid, and chl *a* concentrations (among many other parameters) are measured and archived in the CalCOFI surveys which allowed us here to calculate Si_{ex} at a broader spatial scale. An archived CalCOFI cruise section at line 93 from 06/28/2007 to 07/01/2007 matches many of the same trends we observed in the late July 2007 cruise transect such as a band of negative Si_{ex} values roughly following the lower bounds of the SCM (Figure 6.5B) and a general increase in chl a levels closer to the coast (Fig 5.5A).

Si_{ex} values calculated from archival CalCOFI data (Summer 2007: 06/28/2007 to 07/13/2007) displayed a similar trend in the greater southern California coastal region

(Figure 6.6, Figure 6.7). In summer of 2007, Si_{ex} values calculated at the depth of maximum chl *a* concentrations from 30-35 °N and 118-124 °W were strongly negative in a large portion of the CalCOFI sampling grid (Figure 6.6). Approximately 400 km offshore Si_{ex} values transitioned to mostly positive values and were relatively stable which is consistent with a more stratified and generally macronutrient limited system. chl *a* concentrations were generally high near the coast and declined to values less than 1 μg/L offshore (Figure 6.7). The most negative Si_{ex} values were often superimposed with the highest total chl *a* concentrations suggesting a major contribution of diatoms in a roughly 200 km band residing nearly 100 km offshore (Figure 6.7), which is consistent with prior surveys of the phytoplankton community in this region (30).

6.3.4 Metatranscriptomes from CalCOFI line 93:

Metatranscriptome samples were collected on the SCM cruise transect of line 93 at the surface and SCM at four stations (Figure 6.1, Figure 6.3A). Shotgun sequencing was performed on cDNA amplified from *in vivo* polyadenylated mRNA (herein called poly mRNA) and on cDNA amplified from total rRNA-depleted RNA (herein called total mRNA) which was *in vitro* polyadenylated for subsequent sequencing. The total mRNA fraction was intended to target both eukaryotes and prokaryotes, while the polyA fraction was presumably targeted to eukaryotes because prokaryotes polyadenylate a much smaller and more variable proportion of their total mRNA pool compared with eukaryotes (31). Both total mRNA and poly mRNA fractions were derived from the same filter which captured cells between 0.1 - 20 μm. After rRNA removal, there was a total of 4.67 million reads in the polyA mRNA libraries and 1.08 million reads in the total

mRNA library (Table 6.1). Between 22-33% and 27-36% of the polyA mRNA libraries and 12-44% and 15-60% of the mRNA libraries could be annotated taxonomically or to a PFAM protein domain, respectively. Both polyA and total mRNA libraries were combined to generate a hybrid assembly and reads from both library types at each sampling site were mapped to contigs to generate a single hybrid read count dataset. In the hybrid dataset there were 1451 assembled contigs that were differentially abundant between surface and SCM samples (Figure 6.8) which was approximately 2% of the total number of contigs. Of those 1451 contigs, 665 (46%) had a functional annotation while 1133 (78%) had a taxonomic annotation (Figure 6.9). The 12 most abundant taxonomic groups that were differentially expressing transcripts between the surface and SCM included a variety of dinoflagellates, chlorophytes, haptophytes, pelagophytes, and a single diatom genus (Table 6.2, Figure 6.9). An overview of the sequencing libraries is described in full by Dupont *et al.* (20).

6.3.5 Molecular indicators of Fe stress: ISIPs and Flavodoxins

Compared with prokaryotes relatively few systems for iron acquisition and homeostasis have been experimentally characterized in eukaryotic algal genomes, and the bulk of this characterization has been done in the model freshwater chlorophyte *Chlamydomonas reinhartii (32)*. However, recent studies are revealing new molecular systems related to Fe uptake in marine eukaryotic algae (33, 34). Past studies examining the iron nutritional status of *C. reinhartii* identified genes (Fea1 and Fea2) that are consistently upregulated under conditions of iron deficiency (35, 36). In marine plankton, orthologs to Fea1 exist in pelagophytes, chlorophytes, dinoflagellates, and diatoms.

Recently a new Fe acquisition system (ISIP2a) with little to no functional precedent in other organisms has been identified in the diatom *Phaeodactylum tricornutum* (34, 37). ISIP2a is distantly related to Fea1 but functions in a copper-independent and thermodynamically controlled manner to concentrate Fe at the cell surface and mediate uptake (34). Other iron-starvation induced proteins (ISIP1 and ISIP3) are highly expressed under iron deficient conditions in *P. tricornutum* (37), but their exact functions are currently unclear. Nonetheless, studies are increasingly showing that ISIP and Fea1 proteins are robust and consistent indicators of iron stress in marine phytoplankton (38, 39).

The substitution of flavodoxin, which does not require Fe as a cofactor, for the Fe-containing redox protein ferredoxin is an additional effect of Fe limitation in phytoplankton (40–42). Diatoms are particularly well-known examples displaying this substitution effect whereby the reduced redox efficiency of flavodoxin is compensated for its reduced Fe requirement. Cultured isolates have been shown to increase flavodoxin protein quotas by a factor of nearly 25 under Fe limitation (43) and in the field flavodoxins have been observed to accumulate in Fe limiting conditions while being subsequently reduced after Fe fertilization (42, 44). However, *increases* in flavodoxin gene expression have also been detected after alleviation of Fe stress in natural diatom assemblages suggesting a more complicated relationship between Fe and flavodoxin/ferredoxin at the community level (38). A recent species specific approach demonstrated consistent upregulation of flavodoxin in a cultured *Thalassiosira oceanica* isolate as well as in *T. oceanica* from phytoplankton communities in the northeastern Pacific Ocean (45). Regardless of the variability of field results it remains clear that

flavodoxin proteins are consistently responsive and valuable tracers of phytoplankton metabolic responses to altered Fe availability.

Because we were interested in SCMs as potential regions of Fe limitation, we searched for known iron-responsive transcripts that were differentially abundant between surface waters and the chlorophyll maximum at depth. We searched the combined 16 different metatranscriptomes from surface and SCM waters at each station for molecular biomarkers of iron nutrition in phytoplankton, and focused searches on the Fea1 and ISIP protein domains as well as the classical Fe-limitation induced flavodoxin protein domain. In general, ISIP domains were one of the most abundant of all domains in the dataset (Figure 6.8) and were mostly overrepresented in the SCM (Figure 6.10). Flavodoxin-like domains were generally less abundant than ISIP but also skewed towards the SCM. Both flavodoxin and ISIPs were strongly and significantly differentially expressed (in terms of Log_2 fold change) between SCM and surface waters (Figure 6.8) with one ISIP3 contig presumably belonging to the diatom genus *Chaetoceros* being the second most differentially expressed contig in the SCM. Three out of the five ISIP proteins were in the top 25 of the most abundant contigs detected as differentially abundant in both surface and SCM waters. In the surface waters of station 93.80 an ISIP3 contig assigned to *Prasinoderma singularis* RCC927 was significantly more abundant than in the SCM, but recruited only 11% of the total reads recruited to other ISIP contigs differentially abundant in the SCM. Because ISIP proteins were relatively recently discovered and are structurally novel we examined sequences of putative ISIP contigs and aligned them with trusted alignments (Jeff McQuaid, personal communication, January 2016). In four out of the five cases, the metatranscriptome contigs could be reliably aligned to conserved

residues in the full ISIP1, ISIP2a, or ISIP3 alignments suggesting that our annotations are authentic (Figure 6.8, Figure 6.10).

Prior metatranscriptomic surveys have focused mainly on the expression of ISIP transcripts in diatoms (38, 46). The elevated expression of ISIP-like genes has been consistently found to be associated with Fe-limiting conditions, and ISIP expression in diatoms is repressed upon relief of Fe stress in many but not all cases. However, ISIP sequences are also common in numerous non-diatom phytoplankton strains (34), and in one Fe amended incubation study numerous ISIP sequences were taxonomically binned to chlorophytes (38). Here one of the most highly expressed ISIP sequences in the SCM was taxonomically binned to *Bacillariophyceae* (diatoms), but the remaining differentially expressed ISIP contigs were binned predominantly to *Prasinophyceae* (chlorophytes) and one pelagophyte. Differentially expressed flavodoxins were also mostly taxonomically binned to *Prasinophyceae* and one dinoflagellate (Figure 6.10). The SCM metagenomes and metatranscriptomes from this dataset were previously reported to be highly enriched in *Ostreococcus*, *Micromonas*, and *Pelagomonas*-like sequences suggesting that they were particularly dominant players in the SCM relative to the surface waters (20). Phytoplankton accessory pigment data from that report also qualitatively supported an overrepresentation of *Pelagophyceae* in the SCM and haptophytes in surface waters. The taxonomic assignment of differentially expressed ISIP and flavodoxin sequences to abundant *Pelagophyceae* and *Prasinophyceae* groups mirrors this previously reported trend and also suggests that the SCM phytoplankton communities were relatively Fe stressed compared with surface waters.

6.4 Conclusions

6.4.1 Intersections between N:dFe ratios, Si_{ex}, and gene expression

Here we have presented geochemical tracer and molecular biomarker data which provides multiple independent lines of evidence for the importance of Fe availability at SCMs in the eastern subtropical Pacific Ocean. Elevated N:dFe ratios (> 6) corresponding to depths at or just below the SCM are consistent with the presence of an Fe-limited phytoplankton community unable to completely utilize available nitrate. The presence of a persistent negative band of Si_{ex} just below the depth of the SCM during the late July 2007 sampling transect on line 93 and in nearly half of the total CalCOFI survey area from the month of July 2007 is strongly indicative of the metabolism of Fe-stressed diatoms. Finally, the differential abundance of transcripts coding for known phytoplankton Fe-stress responses in the SCM metatranscriptomes highlights the SCM as a region with comparatively greater Fe stress than overlying surface waters.

On an initial inspection it might appear that the dominance of *Pelagophyceae* and *Prasinophyceae* transcripts and the relative paucity of diatom transcripts in the SCM is at odds with the negative Si_{ex} values at or near many of the coastal SCMs. Negative Si_{ex} values should be indicative of Fe-stress-induced silicification by diatoms, and it would follow that diatoms should be expressing the majority of transcripts assignable to Fe-stress responsive genes. However, the negative Siex band at line 93 may have a spatio-temporal component not captured by this study. For example, recent modeling studies have indicated a prominent role for mesoscale physical oceanographic features in structuring biological production in the CalCS (47), and a significant fraction of nutrients and organic carbon in the CalCS are likely laterally advected into offshore oligotrophic

waters by mesoscale processes like eddies and fronts (48). Nagai *et al.* recently demonstrated through a coupled physical-biogeochemical model that recently upwelled nearshore waters can be pushed offshore initially by wind-driven Ekman transport and then subducted to depths of ~100-200 m at sharp lateral density gradients of frontal features and filaments (48). After subduction these waters are propagated westward up to 800 km by cyclonic and anticyclonic eddies carrying with them significant nutrient and organic carbon concentrations. One explanation for the persistently negative Si_{ex} signature at the base of the SCMs is that some SCMs formed from water masses that originally upwelled closer to shore producing nearshore Fe-limited diatom growth. These waters could have subsequently been advected offshore and then subducted carrying with them the negative Si_{ex} signature of past diatom blooms. Another explanation would involve *in situ* diatom growth. If diatoms in the phytoplankton assemblage were truly growth rate limited by Fe, ephemeral Fe pulses and internal Fe recycling may have supported highly marginal, but persistent diatom growth that slowly stripped silicic acid from the water column and generated a negative Si_{ex} signature over a prolonged period of time (19, 27). This last scenario is consistent with diatom sequences comprising a relatively small portion of the phytoplankton community in the metatranscriptomes, but also with the well-documented capability of diatoms to rapidly respond to and quantitatively consume episodic Fe inputs (3, 38). It should be noted that these two explanations are not mutually exclusive and may have operated synergistically to produce the observed geochemical signature. Regardless of potential taxonomic inconsistencies, the metatranscriptomic data and geochemical proxies in this study clearly point to the SCM as a zone sensitive to the biological dynamics of Fe and indicate that Fe-limitation

could potentially be a widespread and consistent phenomenon in SCMs of the eastern subtropical Pacific Ocean.

6.4.2 Fe limitation in the California Current System and implications for carbon export

Although macronutrient limitation is certainly important in the upper euphotic zone of the eastern subtropical Pacific Ocean, we suggest that Fe-limitation at the base of the SCM may be an important factor structuring phytoplankton community composition and potentially overall carbon export to the deep ocean. At the base of the euphotic zone interactions between light limitation, low Fe concentrations, and relatively high nitrate levels may serve to generate a system ultimately limited by the efficiency of the photosynthetic apparatus (14). Indeed, earlier incubation-based field studies have demonstrated Fe and light co-limitation in many of the SCM layers in the CalCS (15). Some of the better-known examples of Fe-limitation in coastal California waters have been demonstrated in waters off the central coast (~ 36.8° N) where narrow continental shelves do not facilitate interactions between shelf sediments and upwelled waters and thus remain strongly depleted in Fe relative to nitrate (16, 17, 27, 49). Fewer studies have examined Fe limitation in the southern CalCS which is characterized by a more limited areal extent of strong coastal upwelling. Southern CalCS waters upwelled offshore due to wind stress curl and eddy dynamics interact to a lesser extent with shelf sediments and are depleted in Fe relative to nitrate implicating Fe limitation as a structuring force on phytoplankton communities in the surface mixed layer (18, 26). Here we contribute to the knowledge of Fe dynamics in the region by providing evidence for pronounced Fe

availability effects on phytoplankton communities at SCMs extending from coastal southern California waters to the edge of the offshore oligotrophic gyre.

Temporally dynamic sub-mesoscale to mesoscale events are increasingly being recognized as important structuring forces in the CalCS region (48, 50, 51), and these forces have been shown to impact Fe dynamics as well (19). By modulating phytoplankton ecosystem dynamics, Fe availability at the SCM may be an important component determining the forms of carbon exported to depth and the efficiency of nutrient export relative to internal recycling. Ephemeral Fe pulses to the SCM may serve to stimulate diatom production and in turn enhance export flux of ballasted diatom cells (15). Alternatively, chronic Fe-light co-limitation of lower SCM phytoplankton communities may cause enhanced silicification of endemic diatom cells relative to light/nutrient replete cells, thus increasing cellular grazing defense and sinking rates while sustaining relatively large export efficiencies in more marginally productive systems (19). Ultimately, SCMs throughout the global ocean may be unexpected regions of enhanced carbon export efficiency if Fe limitation and its downstream effects are pervasive features in these layers.

6.5 Materials and Methods

6.5.1 Study region and Sampling

An overview of the study region and cruise transect has already been published by Dupont *et al.* (20) in which the features of the metagenomes and metatranscriptomes were highlighted. Briefly, we sampled 15 oceanographic stations on the approximately 700 km CalCOFI transect line 93 in July of 2007 (Figure 6.1). We sampled all 15 stations

for hydrographic parameters (conductivity, temperature, pressure), oxygen concentrations, nitrate concentrations, and chl *a* fluorescence with a CTD rosette. We collected discrete water samples from each of the 15 stations in order to generate macronutrient depth profiles (silicic acid, nitrate, phosphate, ammonium), and we additionally sampled for total particulate carbon, particulate nitrogen, phytoplankton pigments, primary production, and nitrogen source uptake rates at six depths from 12 of the stations as described earlier (20). We also collected samples for total dissolved Fe concentrations from five to seven depths at five different stations (93.40, 93.60, 93.80, 93.110, and 93.120) using trace-metal clean techniques as described by King and Barbeau (18, 26). Metagenomic and metatranscriptomic samples were collected at four stations (93.40, 93.80, 93.110, 93.120) at the surface (3-5 m) and at the SCM (30-110 m).

6.5.2 Measurements

Silicic acid, phosphate, nitrate + nitrite, and ammonium concentrations were measured by autoanalyzer analysis on shore. Chl *a* was extracted from GF/F filters with 90% acetone at -20C for 24 hours then analyzed onboard using a Turner Designs Fluorometer. Taxon specific phytoplankton pigments were analyzed by high performance liquid chromatography on shore. Macronutrients, chl *a*, and phytoplankton pigments were analyzed using standard protocols as recommended by the CCE LTER program (http://cce.lternet.edu/data/methods-manual). Particulate organic carbon, particulate organic nitrogen, primary production rates, and nitrogen uptake rates were measured as described by Dupont *et al.* (20). Dissolved Fe concentrations were measured using flow-injection analysis (FIA) with sulfite reduction (18, 26).

The Si_{ex} value is calculated from a modified formula (26, 52) formally expressed by the equation: Si_{ex} = [silicic acid] - ([nitrate] × $R_{Si:NO3}$) (1) where $R_{Si:NO3}$ = ~1.0 and brackets denote nutrient concentration. The $R_{Si:NO3}$ value represents the initial ratio of silicic acid to nitrate (Si:N) in upwelled source water prior phytoplankton utilization, and we used an $R_{Si:NO3}$ value of 1.0 as it has been shown to be an appropriate estimate for the silicic acid:nitrate ratio in upwelled waters in the CalCS (26). Although this value can vary by ±0.2 in upwelled waters from this region (26), modifying $R_{Si:NO3}$ resulted in no change of the general trend displayed in Figs. 5.3, 5.5, 5.6, and 5.7.

6.5.3 Accession and processing of archival CalCOFI data

We obtained data for the CalCOFI cruise transects from the CalCOFI website and database (http://calcofi.org/data.html). We selected cruise data for the entire CalCOFI grid from the summer of 2007 (06/28/2007 to 07/13/2007) and used this data to generate isosurface plots of Si_{ex} and chl *a* in a region of approximately 35 °N to 30 °N and 124 °W to 118 °W. CalCOFI line 93 was sampled from 06/28/2007 to 07/01/2007.

6.5.4 Metagenomics and metatranscriptomics

The processes for collecting, storing, sequencing, quality control and annotation of community DNA and RNA have already been described (20). Briefly, we extracted total community RNA (0.2 - 20 μm) and utilized subtractive hybridization of rRNAs (53) to generate approximately 30 ng of rRNA-depleted total RNA from each sample. After subtractive hybridization we polyadenylated an aliquot of the total rRNA-depleted RNA from each sample, reverse transcribed the polyadenylated subsample using oligo(dT)

primers and sequenced the resulting cDNA using 454 pyrosequencing (herein called total mRNA). A separate aliquot of rRNA-depleted total RNA was reverse transcribed with oligo(dT) primers without prior polyadenylation in order to selectively enrich the eukaryotic component of the community. The resulting cDNA from polyA mRNA libraries was sequenced using 454 pyrosequencing. After sequencing, each resulting metatranscriptome or metagenome was assembled *de novo* using CLC Genomics workbench and gene calls were made using FragGeneScan (54). Afterwards reads were processed to remove artificial duplicates from 454 sequencing errors, were compared against SILVA (55) and an inhouse rRNA database in order to filter out rRNA, and were annotated using IMG (56), KEGG (57), Genbank (58), Ensembl (59), the PyloDB database (https://github.com/allenlab/PhyloDB) and other inhouse assemblies of algal transcriptomes. The processing, assembly, and annotation pipeline is described in full by Dupont *et al.* (20).

6.5.5 Metatranscriptome data analysis:

We tested for differential abundance of transcripts between SCM and surface samples using the DESeq2 analysis procedure (60). DESeq2 is a statistical framework for the comparison of diverse forms of count data, and it particularly excels in situations with few sample replicates and sparse count data. DESeq2 has been used to detect differential abundances of genes across sampling conditions in datasets as diverse as human RNA-seq data (61) and also marine 16S rRNA marker gene surveys (62). The DESeq2 framework models read counts as a negative binomial distribution in a generalized linear model (GLM) while attempting to correct for within-group variability (i.e. between

replicates) by controlling the dispersion parameter for different genes by estimating dispersions using maximum likelihood (ML) and pooling dispersion estimates among genes with similar counts. It then uses Bayesian priors for each gene to shrink ML estimated dispersions towards the Bayesian prior mean of the model which results in a final maximum *a posteriori* estimation of dispersion. Finally, estimates of log fold change (LFC) that account for within-group dispersion between sampling groups are calculated and corrected for dependence on the mean by shrinking LFC towards zero more strongly for genes with low mean counts. Differences between LFCs of sampling groups are tested for significance using a negative binomial Wald test and corrected for multiple testing by controlling the false discovery rate. Finally, count outliers are removed based on a Cook's distance threshold before significance testing, thus removing the potential for large but inconsistent counts among replicates to influence the significance of LFC estimates.

Within the DESeq2 framework, we treated metatranscriptomes from SCM and surface samples as condition replicates. Although within-group environmental variability of SCM samples was significant, the GLM-based normalization procedure and Wald significance tests are appropriate for the dataset as it is comprised of a large amount of count data from individual assembled contigs but the count data is sparse (i.e. has many zeros) and there is a relatively small number of samples (four) comprising each sampling group. By filtering out genes with high intra-sample group variability, the DESeq2 procedure should highlight transcripts that are *consistently* differentially abundant between the surface and SCM and should therefore provide a conservative estimate of

genes and their functions that are enriched or depleted in SCMs despite other co-occurring ecological gradients.

For Wald tests we used the null hypothesis of zero LFC between genes expressed in the SCM and surface, and selected a false discovery rate corrected *P* threshold of < 0.1 - i.e. we are willing to accept that 10% of the total number of genes flagged as differentially expressed have an actual LFC of zero. To visualize differences in transcript abundances between conditions and samples, counts were normalized by sequencing depth with an added pseudocount of 0.5 to allow for log scale presentation.

6.6 Acknowledgements

Chapter 6 is currently being prepared for publication by Shane L. Hogle, Chris L. Dupont, and Katherine A. Barbeau. The dissertation author was the primary investigator and first author of this paper. This work was funded by NSF GRFP grant DGE-144086 to S.L.H.

Table 6.1 Metatranscriptome sequencing statistics

Station	Depth	polyA mRNA reads	polyA %PFAM	polyA Tax %	total mRNA reads	total %PFAM	total %Tax
93.40 (GS272)	Surface	499770	33	29	95364	12	15
93.40 (GS271)	SCM	752922	27	36	163326	33	49
93.80 (GS265)	Surface	412298	20	27	96293	20	25
93.80 (GS266)	SCM	874427	22	29	209920	27	35
93.110 (GS267)	Surface	450932	22	29	182244	18	21
93.110 (GS268)	SCM	914645	24	31	128030	31	39
93.120 (GS269)	Surface	206280	25	32	133069	26	32
93.120 (GS270)	SCM	758859	25	35	77791	44	60

The numbers listed in the polyA mRNA reads and total mRNA reads columns are the number of reads in each library after the rRNA removal procedure. The Station column denotes stations by their CalCOFI designation as well as by their GOS identification. %PFAM indicates the percentage of reads in each library that could be annotated to PFAM family and %Tax indicates the percentage of reads that could be taxonomically annotated.

Table 6.2 Taxonomic identity of differentially expressed transcripts

Genus	Order	Class	Phylum
Bathycoccus	*Mamiellales*	*Prasinophyceae*	*Chlorophyta*
Ostreococcus	*Mamiellales*	*Mamiellophyceae* (prasinophyte)	*Chlorophyta*
Alexandrium	*Gonyaulacales*	*Dinophyceae*	*Dinoflagellata*
Oxyrrhis	*Oxyrrhinales*	*Dinophyceae*	*Dinoflagellata*
Karenia	*Gymnodiniales*	*Dinophyceae*	*Dinoflagellata*
Karlodinium	*Gymnodiniales*	*Dinophyceae*	*Dinoflagellata*
Noctiluca	*Noctilucales*	*Dinophyceae*	*Dinoflagellata*
Haptolina	*Prymnesiales*	*Prymnesiophyceae*	*Haptophyta*
Chrysochromulina	*Prymnesiales*	*Prymnesiophyceae*	*Haptophyta*
Phaeocystis	*Phaeocystales*	haptophyte	*Haptophyta*
Pelagomonas	*Pelagomonadales*	*Pelagophyceae*	*Heterokonta*
Pseudo-nitzschia	*Bacillariales*	*Bacillariophyceae* (diatom)	*Heterokonta*

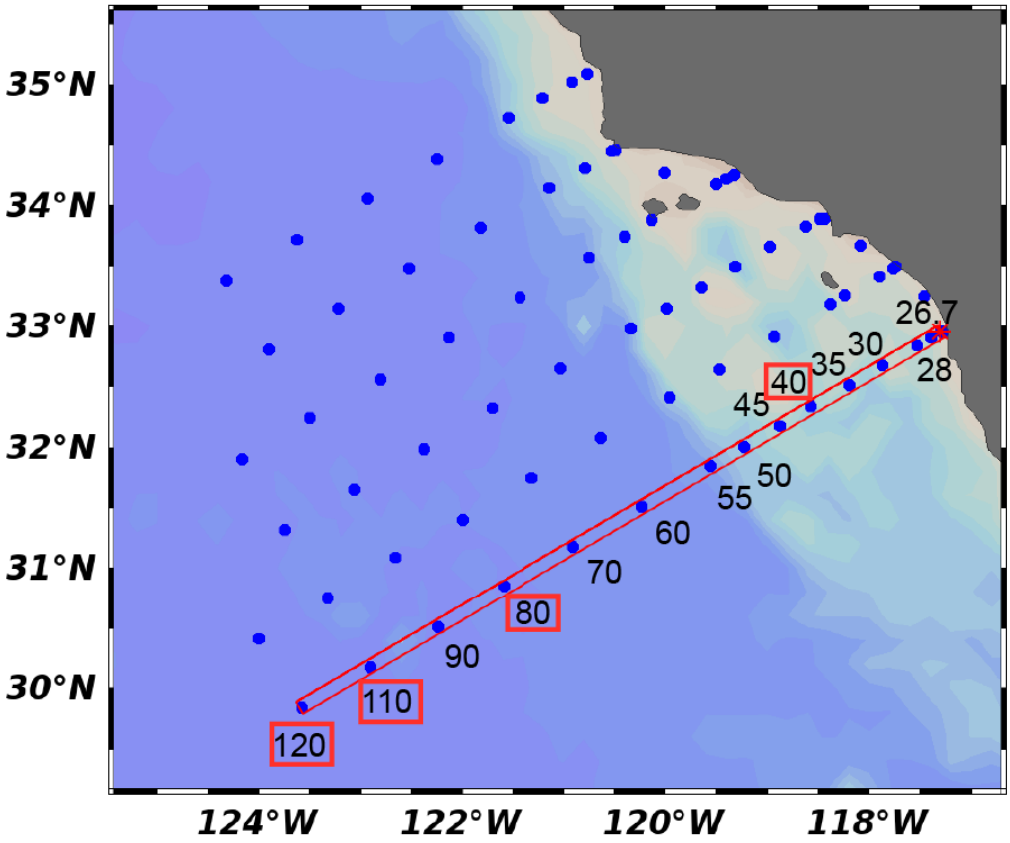


Figure 6.1 Map of the southern California Current System (CalCS). Sampling stations of the CalCOFI program are represented as blue dots and station numbers on line 93 are displayed. CalCOFI line 93 is highlighted in by an overlaid red rectangle. Stations where dFe, hydrographic, and metatranscriptomic samples were collected during the late July 2007 cruise are highlighted in red.

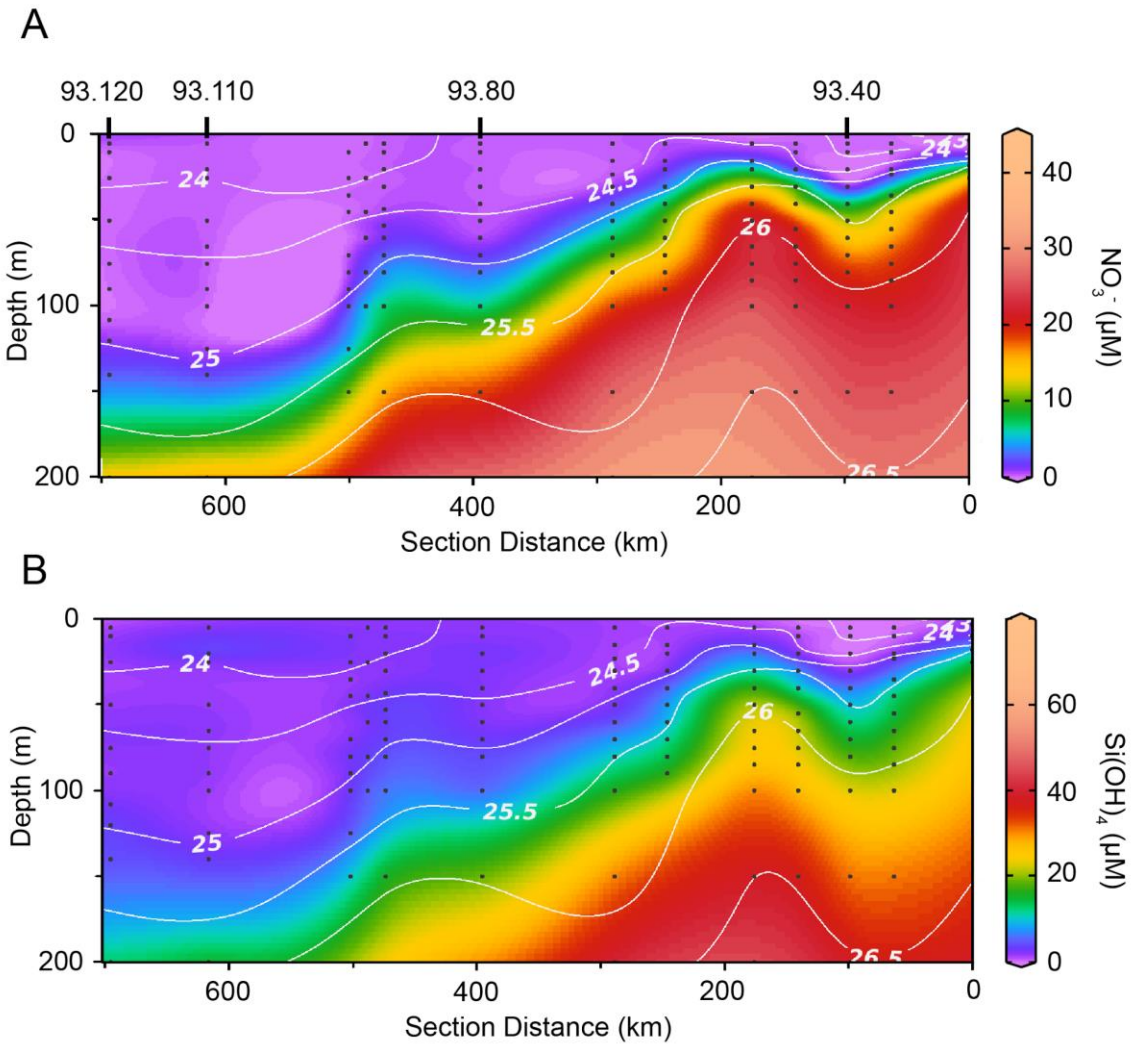


Figure 6.2 Nitrate and silicic acid section plots on CalCOFI line 93. Sampled stations and depths are indicated by black dots in each panel, and stations where dFe, hydrographic, and metatranscriptomic samples were collected during the late July 2007 cruise are labeled. **(A)** Nitrate concentrations (µmol L^{-1}) are shown by color gradients while white isopycnals indicate potential density anomaly. **(B)** Same layout as in panel A, but for silicic acid concentrations (µmol L^{-1}).

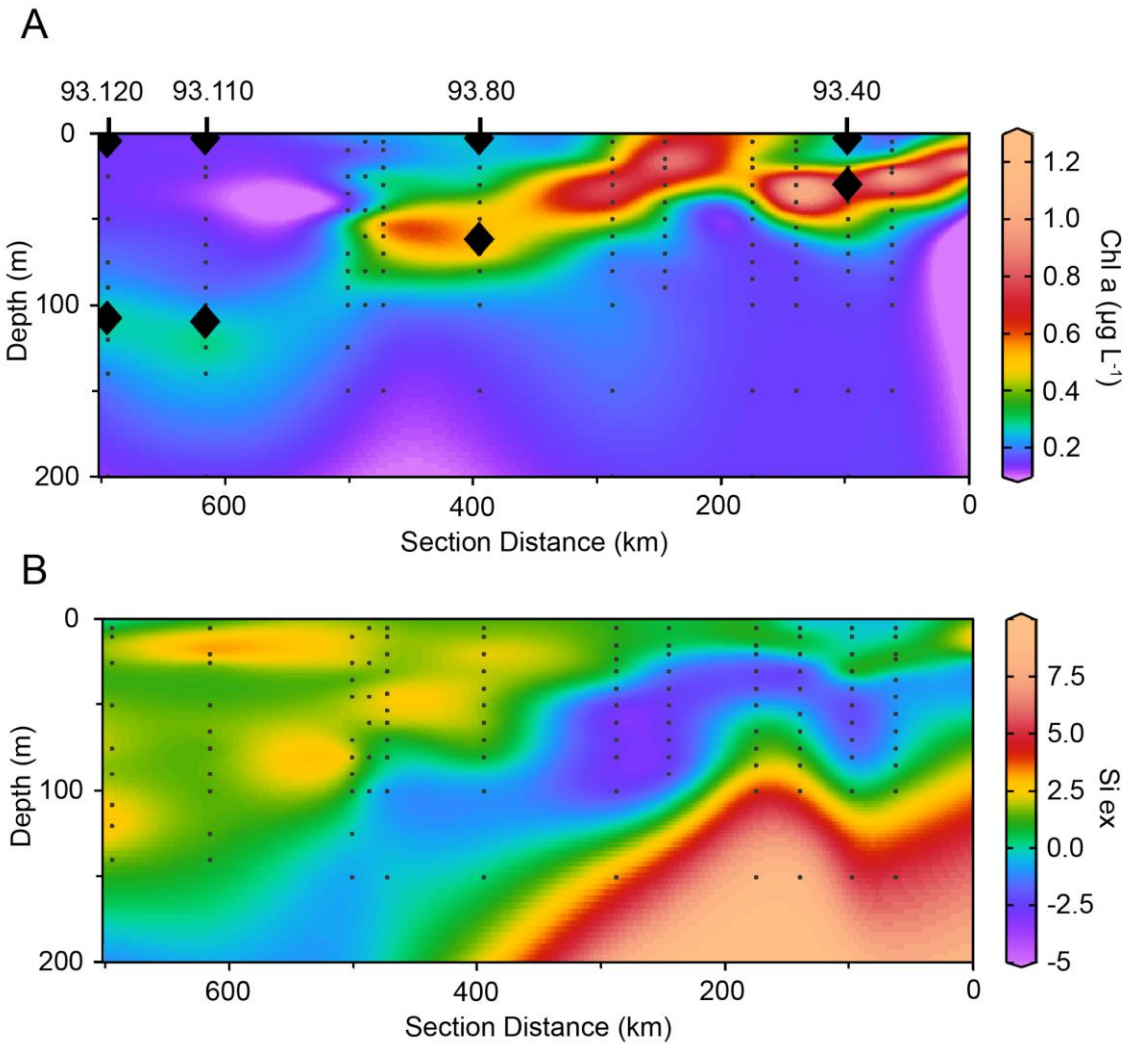


Figure 6.3 Chlorophyll *a* and Si_{ex} section plots on CalCOFI line 93. Sampled stations and depths are indicated by black dots in each panel, and stations where dFe, hydrographic, and metatranscriptomic samples were collected during the late July 2007 cruise are labeled. Layout is the same as in Figure 6.2. Chl *a* concentrations (µg L^{-1}) are displayed in **(A)** and values of the Si_{ex} tracer are displayed in **(B)**. In panel A, diamonds indicate depths where metatranscriptomes were collected.

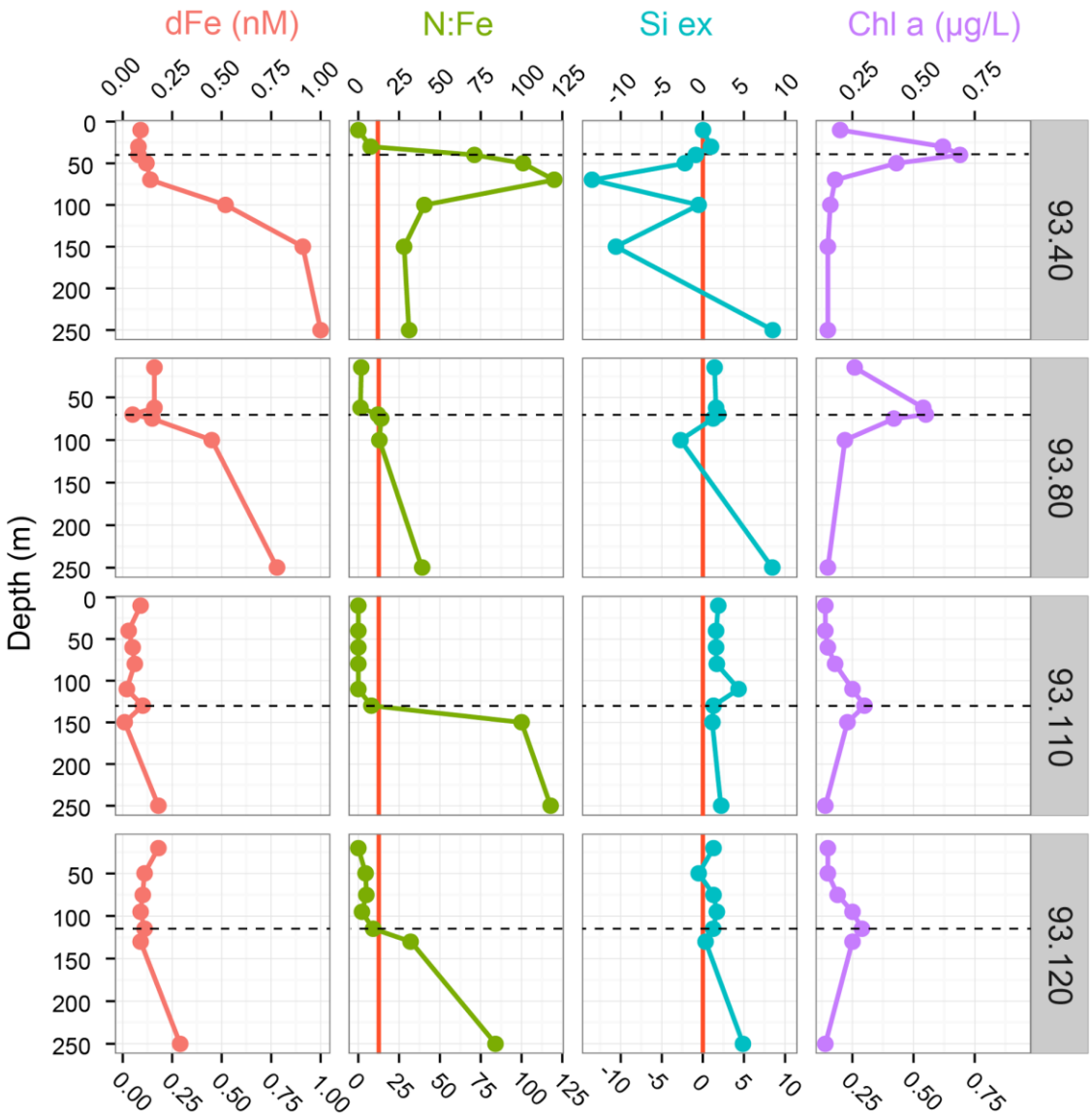


Figure 6.4 Dissolved Fe, N:dFe ratios, Si_{ex}, and chlorophyll *a* profiles at CalCOFI stations 93.40, 93.80, 93.110, and 93.120. Profile plots display dissolved Fe concentrations (red), N:Fe ratio (green), Si_{ex} (blue), and chl *a* concentrations (purple) at the four CalCOFI stations 93.40, 93.80, 93.110, and 93.120 sampled in late July, 2007. Depth is displayed on the y axis while measured values are displayed on the x axis. The vertical red bar for N:Fe indicates the threshold value for phytoplankton Fe limitation (i.e. > 12 indicates limitation), while the vertical red bar indicates the negative Siex threshold (i.e. < 0 indicates silicate is depleted relative to nitrate). Horizontal dashed lines indicate the depth of the Chlorophyll maximum.

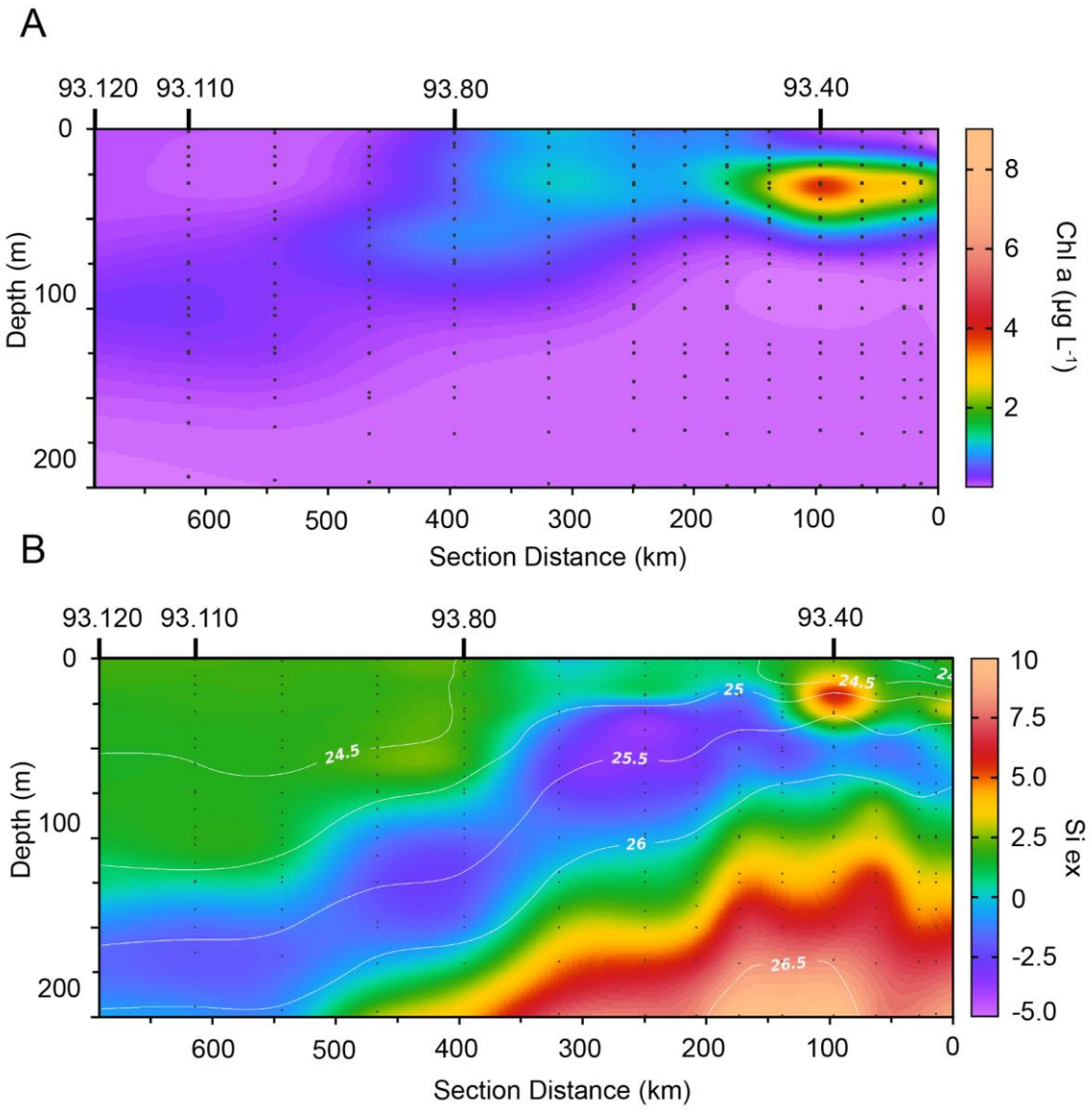


Figure 6.5 Chlorophyll *a* and Si_{ex} section plots on CalCOFI line 93 from archival CalCOFI cruises performed from 06/28/2007 to 07/01/2007. Layout is the same as in Figure 6.3. Chl *a* concentrations (µg L^{-1}) are displayed in **(A)** and values of the Si_{ex} tracer are displayed in **(B)**.

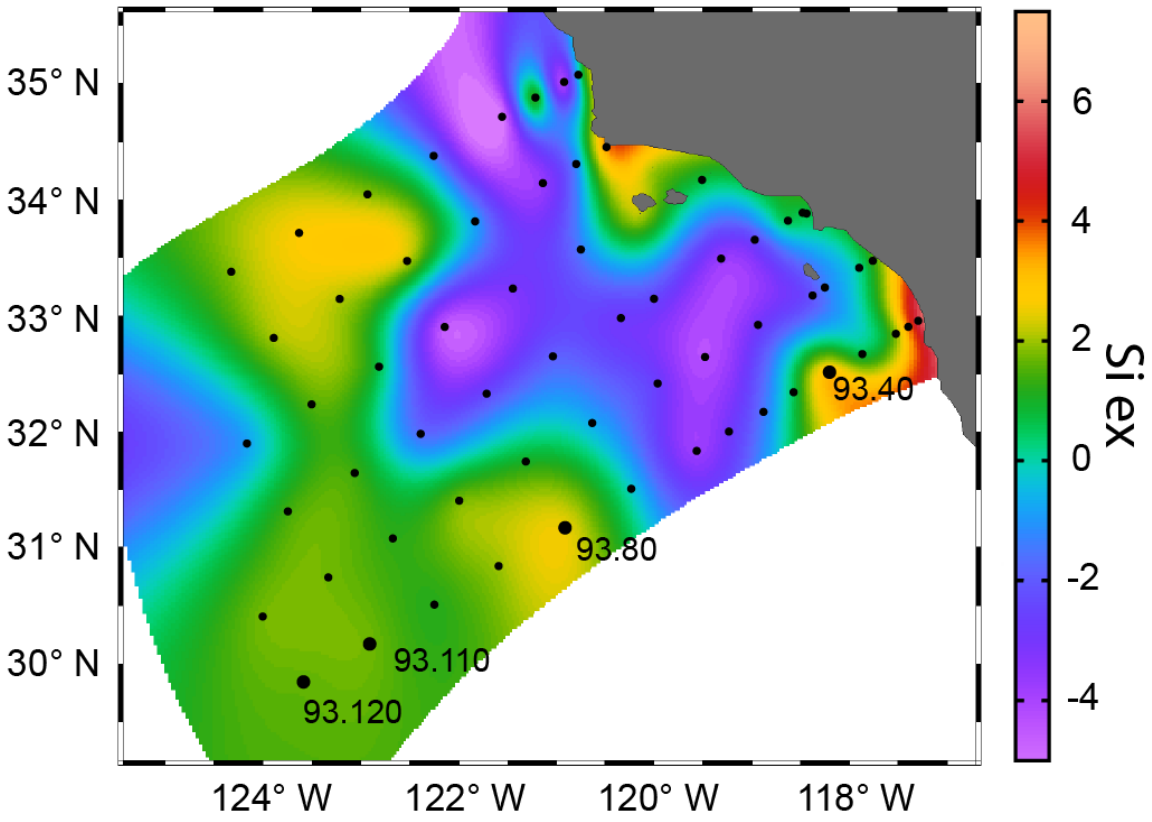


Figure 6.6 Si_{ex} calculated at the depth of the SCM for the entire CalCOFI grid from archival CalCOFI cruises performed during the summer of 2007. Stations where dFe, hydrographic, and metatranscriptomic samples were collected during the late July 2007 cruise are represented with enlarged circles. Negative Si_{ex} values indicate depleted water column silicic acid relative to nitrate concentrations.

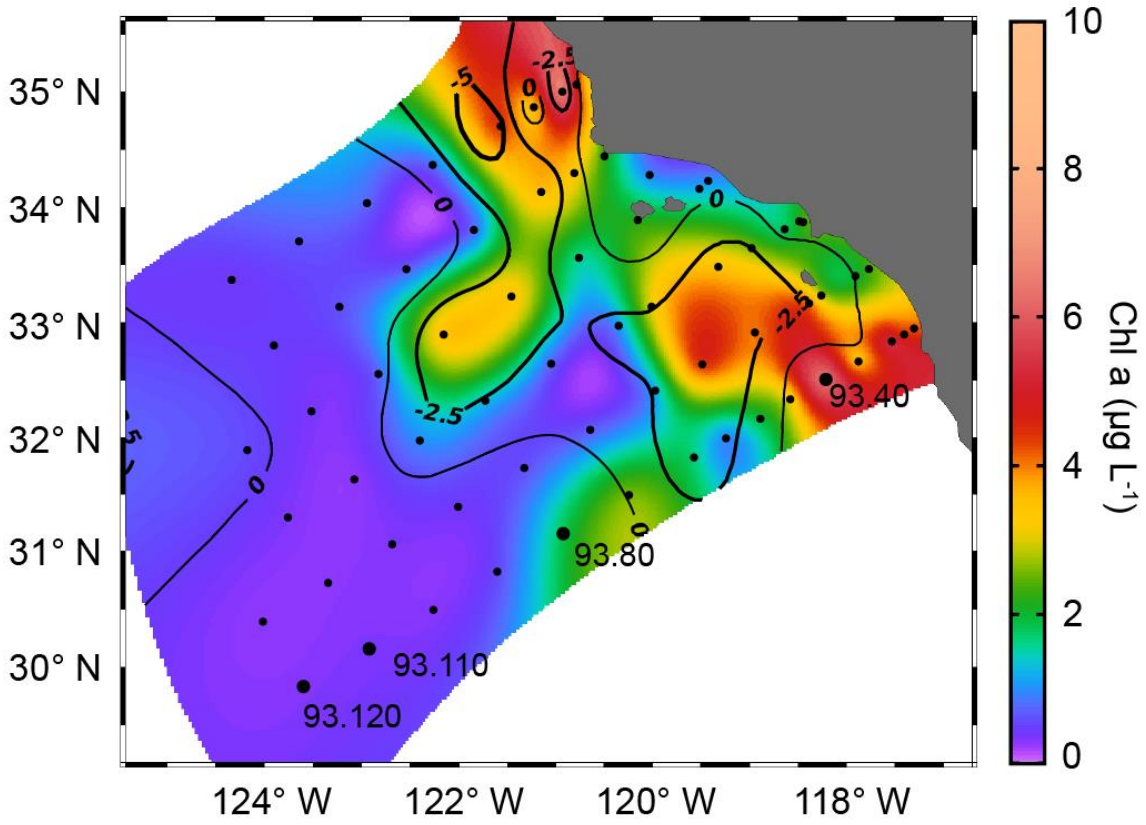


Figure 6.7 Chlorophyll *a* at the depth of the SCM for the entire CalCOFI grid from archival CalCOFI cruises performed during the summer of 2007. Negative Si_{ex} isobaths from Figure 6.5 are overlaid on the chl *a* concentrations. Stations where dFe, hydrographic, and metatranscriptomic samples were collected during the late July 2007 cruise are represented with enlarged circles. Negative Si_{ex} values indicate depleted water column silicic acid relative to nitrate concentrations.

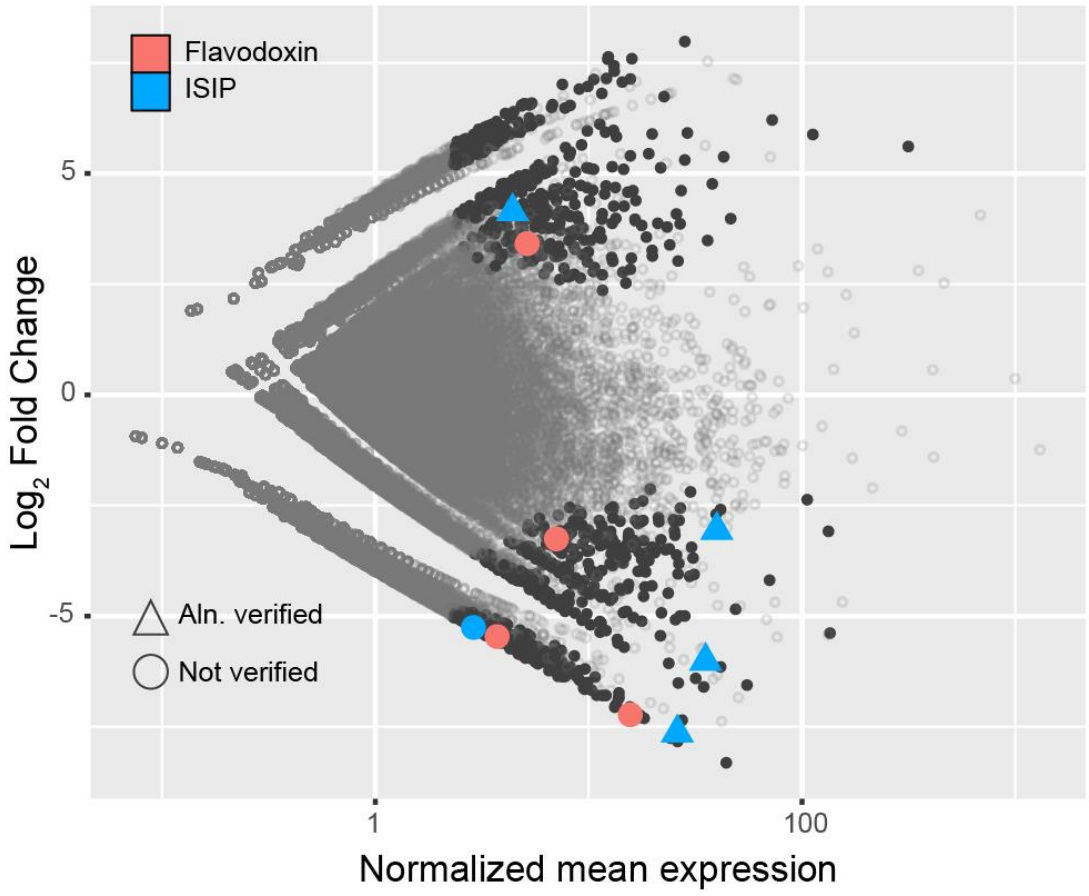


Figure 6.8 MA plot of differentially abundant metatranscriptomic contigs between surface and SCM metatranscriptomes at the four sampling sites. Contigs with an FDR corrected *P* value > 0.05 are displayed as empty light grey circles, while the 1451 contigs that are significantly differentially abundant are displayed as filled black circles. The X axis displays the variance stabilizing transformed expression of each contig averaged over all samples, and the Y axis displays the $\log_2$ fold change of contigs from surface to SCM (i.e. negative values indicate overabundance in SCM). Flavodoxin and ISIP contigs are highlighted by color. ISIP contigs with conserved residues aligning to ISIP multiple sequence alignments from sequenced phytoplankton genomes are denoted by triangles.

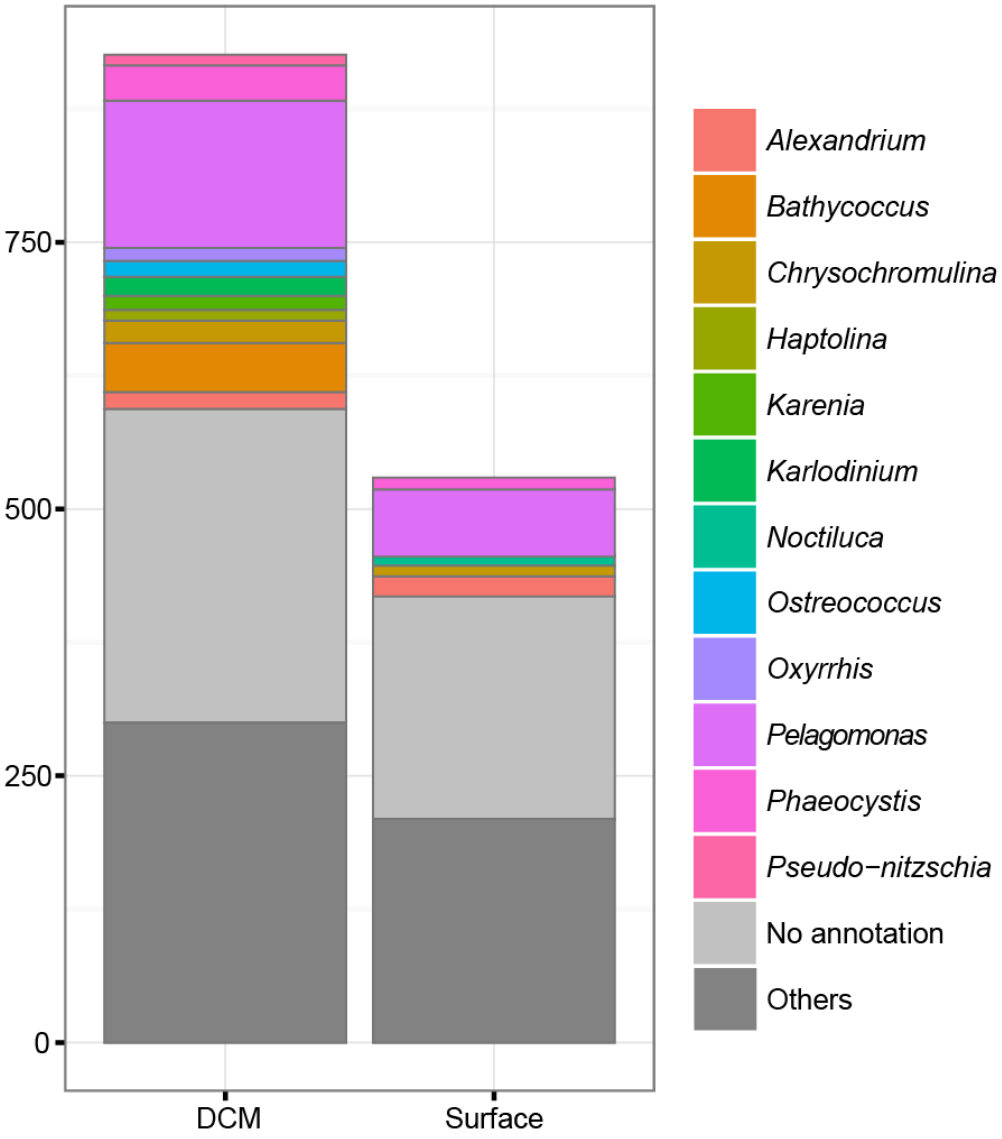


Figure 6.9 Taxonomic composition of the 1451 differentially expressed transcripts between surface and SCM samples. Transcripts (see Figure 6.8) are binned to the 12 most abundant taxonomic groups (Table 6.2). Transcripts with no taxonomic annotation are in light grey while those binned to less abundant taxonomic groups are grouped together and are represented in dark grey.

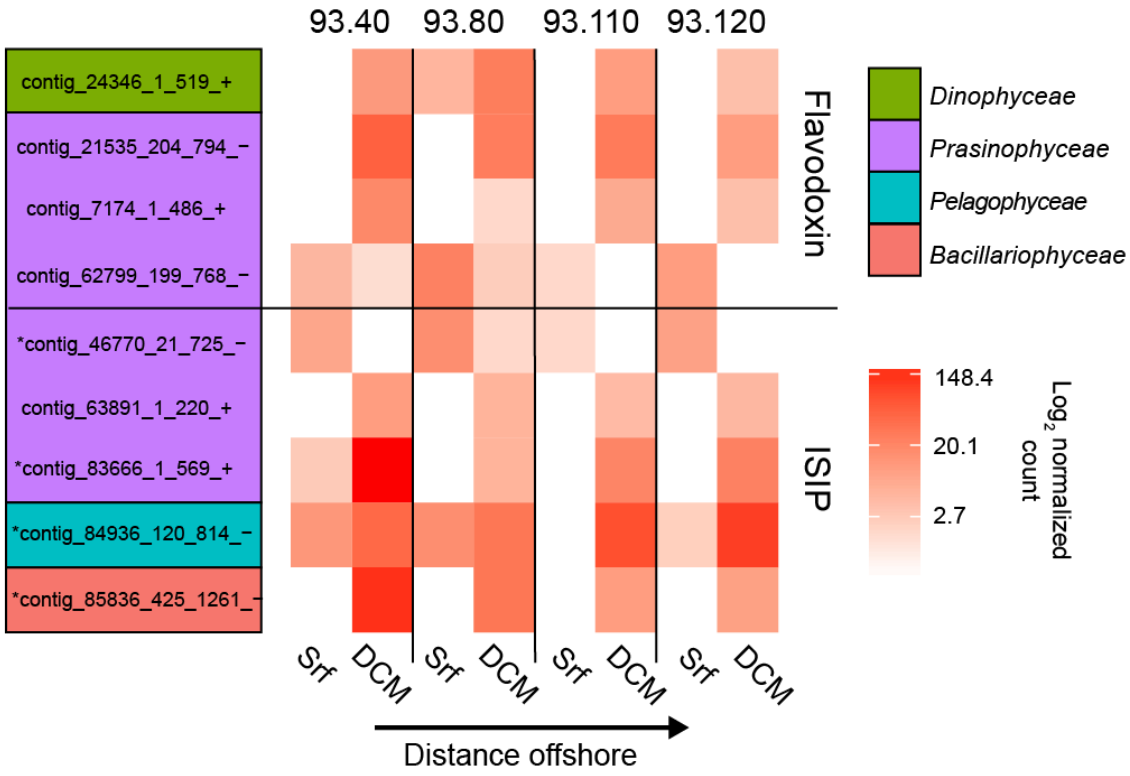


Figure 6.10 Heatmap of Flavodoxin and ISIP contig abundances. Transcripts are normalized to sequencing depth with an added pseudocount of 0.5 in order to plot on a Log_2 scale. Also note the Log_2 scale for the color bar. Contigs are colored by taxonomic Class and asterisks denote contigs matching conserved residues in an ISIP full sequence alignment. Sampling stations and depths are respectively labeled on the top and bottom of the heatmap.

6.7 References

1. de Baar HJ, Boyd PW, Coale KH, Landry MR, Tsuda A, Assmy P, Bakker DC, Bozec Y, Barber RT, Brzezinski MA, Buesseler KO, Boyé M, Croot PL, Gervais F, Gorbunov MY, Harrison PJ, Hiscock WT, Laan P, Lancelot C, Law CS, Levasseur M, Marchetti A, Millero FJ, Nishioka J, Nojiri Y, van Oijen T, Riebesell U, Rijkenberg MJ, Saito H, Takeda S, Timmermans KR, Veldhuis MJ, Waite AM, Wong CS. (2005) Synthesis of iron fertilization experiments: From the Iron Age in the Age of Enlightenment. *J Geophys Res* 110(C9):C09S16.

2. Boyd PW, Watson AJ, Law CS, Abraham ER, Trull T, Murdoch R, Bakker DC, Bowie AR, Buesseler KO, Chang H, Charette M, Croot P, Downing K, Frew R, Gall M, Hadfield M, Hall J, Harvey M, Jameson G, LaRoche J, Liddicoat M, Ling R, Maldonado MT, McKay RM, Nodder S, Pickmere S, Pridmore R, Rintoul S, Safi K, Sutton P, Strzepek R, Tanneberger K, Turner S, Waite A, Zeldis J. (2000) A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* 407(6805):695–702.

3. Boyd PW, Jickells T, Law CS, Blain S, Boyle EA, Buesseler KO, Coale KH, Cullen JJ, de Baar HJ, Follows M, Harvey M, Lancelot C, Levasseur M, Owens NP, Pollard R, Rivkin RB, Sarmiento J, Schoemann V, Smetacek V, Takeda S, Tsuda A, Turner S, Watson AJ (2007) Mesoscale Iron Enrichment Experiments 1993-2005: Synthesis and Future Directions. Science 315(5812):612–617.

4. Chappell PD, Moffett JW, Hynes AM, Webb EA (2012) Molecular evidence of iron limitation and availability in the global diazotroph Trichodesmium. *ISME J* 6(9):1728–1739.

5. Cullen JJ (1982) The deep chlorophyll maximum: comparing vertical profiles of chlorophyll a. *Can J Fish Aquat Sci* 39:791–803.

6. Cullen JJ (2015) Subsurface chlorophyll maximum layers: enduring enigma or mystery solved? *Ann Rev Mar Sci* 7:207–239.

7. Gomi Y, Fukuchi M, Taniguchi A (2010) Diatom assemblages at subsurface chlorophyll maximum layer in the eastern Indian sector of the Southern Ocean in summer. *J Plankton Res* 32(7):1039–1050.

8. Lyngsgaard MM, Richardson K, Markager S, Nielsen MH, Olesen M, Christensen JPA. (2014) Deep primary production in coastal pelagic systems: importance for ecosystem functioning. *Mar Ecol Prog Ser* 517:15–33.

9. Venrick EL (1988) The vertical distributions of chlorophyll and phytoplankton species in the North Pacific central environment. *J Plankton Res* 10(5):987–998.

10. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, Martinez A,

Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311(2002):496–503.

11. Coale KH, Bruland KW (1987) Oceanic stratified euphotic zone as elucidated by 234Th: 238U disequilibria1. *Limnol Oceanogr* 32(1):189–200.

12. Karl DM (1999) A Sea of Change: Biogeochemical Variability in the North Pacific Subtropical Gyre. *Ecosystems* 2(3):181–214.

13. Morel FMM, Price NM (2003) The biogeochemical cycles of trace metals in the oceans. *Science* 300:944–947.

14. Sunda WG, Huntsman SA (1997) Interrelated influence of iron, light and cell size on marine phytoplankton growth. *Nature* 390(6658):389–392.

15. Hopkinson BM, Barbeau KA (2008) Interactive influences of iron and light limitation on phytoplankton at subsurface chlorophyll maxima in the eastern North Pacific. *Limnol Oceanogr* 53(4):1303–1318.

16. Bruland KW, Rue EL, Smith GJ (2001) Iron and macronutrients in California coastal upwelling regimes: Implications for diatom blooms. *Limnol Oceanogr* 46(7):1661–1674.

17. Hutchins DA, DiTullio GR, Zhang Y, Bruland KW (1998) An iron limitation mosaic in the California upwelling regime. *Limnol Oceanogr* 43(6):1037–1054.

18. King AL, Barbeau K (2007) Evidence for phytoplankton iron limitation in the southern California Current System. *Mar Ecol Prog Ser* 342:91–103.

19. Brzezinski MA, Krause JW, Bundy RM, Barbeau KA, Franks P, Goericke R, Landry MR, Stukel MR. (2015) Enhanced silica ballasting from iron stress sustains carbon export in a frontal zone within the California Current. *J Geophys Res C: Oceans* 120(7):4654–4669.

20. Dupont CL, McCrow JP, Valas R, Moustafa A, Walworth N, Goodenough U, Roth R, Hogle SL, Bai J, Johnson ZI, Mann E, Palenik B, Barbeau K, Craig Venter J, Allen AE. (2015) Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J* 9(5):1076–1092.

21. King AL, Buck KN, Barbeau KA (2012) Quasi-Lagrangian drifter studies of iron speciation and cycling off Point Conception, California. *Mar Chem* 128–129:1–12.

22. Price NM, Ahner BA, Morel FMM (1994) The equatorial Pacific Ocean: Grazer-controlled phytoplankton populations in an iron-limited ecosystem1. *Limnol Oceanogr* 39(3):520–534.

23. Timmermans KR, Stolte W, de Baar HJW Iron-mediated effects on nitrate reductase in marine phytoplankton. *Mar Biol* 121(2):389–396.

24. Sunda WG, Huntsman SA (1995) Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar Chem* 50(1–4):189–206.

25. Di Lorenzo E (2003) Seasonal dynamics of the surface circulation in the Southern California Current System. *Deep Sea Res Part 2 Top Stud Oceanogr* 50(14–16):2371–2388.

26. King AL, Barbeau KA (2011) Dissolved iron and macronutrient distributions in the southern California Current System. *J Geophys Res* 116(C3):C03018.

27. Hutchins DA, Bruland KW (1998) Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* 393(6685):561–564.

28. Takeda S (1998) Influence of iron availability on nutrient consumption ratio of diatoms in oceanic waters. *Nature* 393(6687):774–777.

29. Brzezinski MA (1985) The Si:C:N ratio of marine diatoms: Interspecific variability and the effect of some environmental variables. *J Phycol* 21(3):347–357.

30. Venrick EL (2002) Floral patterns in the California Current System off southern California: 1990-1996. *J Mar Res* 60(1):171–189.

31. Sarkar N (1997) Polyadenylation of mRNA in prokaryotes. *Annu Rev Biochem* 66(1):173–197.

32. Blaby-Haas CE, Merchant SS (2012) The ins and outs of algal metal transport. *Biochim Biophys Acta* 1823(9):1531–1552.

33. Groussman RD, Parker MS, Armbrust EV (2015) Diversity and Evolutionary History of Iron Metabolism Genes in Diatoms. *PLoS One* 10(6):e0129081.

34. Morrissey J, Sutak R, Paz-Yepes J, Tanaka A, Moustafa A, Veluchamy A, Thomas Y, Botebol H, Bouget FY, McQuaid JB, Tirichine L, Allen AE, Lesuisse E, Bowler C. (2015) A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Curr Biol* 25(3):364–371.

35. Rubinelli P, Siripornadulsil S, Gao-Rubinelli F, Sayre RT (2002) Cadmium- and iron-stress-inducible gene expression in the green alga Chlamydomonas reinhardtii: evidence for H43 protein function in iron assimilation. *Planta* 215(1):1–13.

36. Allen MD, del Campo J a., Kropat J, Merchant SS (2007) FEA1, FEA2, and FRE1, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with FOX1 and FTR1 in iron-deficient Chlamydomonas reinhardtii. *Eukaryot Cell* 6(10):1841–1852.

37. Allen AE, Laroche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C. (2008) Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc Natl Acad Sci U S A* 105(30):10438–10443.

38. Marchetti A, Schruth DM, Durkin C, Parker MS, Kodner RB, Berthiaume CT, Morales R, Allen AE, Armbrust EV. (2012) Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc Natl Acad Sci U S A* 109(6):E317–25.

39. Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska M, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, Laroche J. (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol* 13(7):R66.

40. Erdner DL, Price NM, Doucette GJ, Peleato ML, Anderson DM (1999) Characterization of ferredoxin and flavodoxin as markers of iron limitation in marine phytoplankton. *Mar Ecol Prog Ser* 184:43–53.

41. McKay RML, Geider RJ, LaRoche J (1997) Physiological and Biochemical Response of the Photosynthetic Apparatus of Two Marine Diatoms to Fe Stress. *Plant Physiol* 114(2):615–622.

42. La Roche J, Boyd PW, McKay RML, Geider RJ (1996) Flavodoxin as an in situ marker for iron stress in phytoplankton. *Nature* 382(6594):802–805.

43. La Roche J, Murray H, Orellana M, Newton J (1995) Flavodoxin expression as an indicator of iron limitation in marine diatoms. *J Phycol* 31(4):520–530.

44. Maldonado MT, Boyd PW, LaRoche J, Strzepek R, Waite A, Bowie AR, Croot PL, Frew RD, Price NM. (2001) Iron uptake and physiological response of phytoplankton during a mesoscale Southern Ocean iron enrichment. *Limnol Oceanogr* 46(7):1802–1808.

45. Chappell PD, Whitney LP, Wallace JR, Darer AI, Jean-Charles S, Jenkins BD. (2014) Genetic indicators of iron limitation in wild populations of Thalassiosira oceanica from the northeast Pacific Ocean. *ISME J* 9(3):592–602.

46. Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, Post AF, Sipler RE, Spackeen JL, Xu K, Bronk DA, Hutchins DA, Allen AE. (2015) Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proceedings of the National Academy of Sciences* 112(32):9938–9943.

47. Gruber N, Lachkar Z, Frenzel H, Marchesiello P, Münnich M, McWilliams JC, Nagai T, Plattner GK. (2011) Eddy-induced reduction of biological production in eastern boundary upwelling systems. *Nat Geosci* 4(11):787–792.

48. Nagai T, Gruber N, Frenzel H, Lachkar Z, McWilliams JC, Plattner GK. (2015) Dominant role of eddies and filaments in the offshore transport of carbon and nutrients in the California Current System. *J Geophys Res C: Oceans* 120(8):5318–5341.

49. Biller DV, Bruland KW (2014) The central California Current transition zone: A broad region exhibiting evidence for iron limitation. *Prog Oceanogr* 120:370–382.

50. Landry MR, Ohman MD, Goericke R, Stukel MR, Barbeau KA, Bundy R, Kahru M. (2012) Pelagic community responses to a deep-water front in the California Current Ecosystem: overview of the A-Front Study. *J Plankton Res* 34(9):739–748.

51. Krause JW, Brzezinski MA, Goericke R, Landry MR, Ohman MD, Stukel MR, Taylor AG. (2015) Variability in diatom contributions to biomass, organic matter production and export across a frontal gradient in the California Current Ecosystem. *J Geophys Res C: Oceans* 120(2):1032–1047.

52. Gruber N, Sarmiento JL (1997) Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochem Cycles* 11(2):235–266.

53. Stewart FJ, Ottesen EA, DeLong EF (2010) Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* 4(7):896–907.

54. Rho M, Tang H, Ye Y (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 38(20):e191.

55. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(Database issue):D590–6.

56. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40(Database issue):D115–22.

57. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42(1):D199–205.

58. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. (2012) GenBank. *Nucleic Acids Res* 40(Database issue):D48–53.

59. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, Humphrey J, Kerhornou A, Khobova J, Aranganathan NK, Langridge N, Lowy E, McDowall MD, Maheswari U, Nuhn M, Ong CK, Overduin B, Paulini M, Pedro H, Perry E, Spudich G, Tapanari E, Walts B,

Williams G, Tello-Ruiz M, Stein J, Wei S, Ware D, Bolser DM, Howe KL, Kulesha E, Lawson D, Maslen G, Staines DM. (2016) Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res* 44(D1):D574–80.

60. Love MI, Huber W, Anders S (2014) Fully formatted Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(550):1–21.

61. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segrè AV, Djebali S, Niarchou A, GTEx Consortium, Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigó R. (2015) Human genomics. The human transcriptome across tissues and individuals. *Science* 348(6235):660–665.

62. McMurdie PJ, Holmes S (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol* 10(4). doi:10.1371/journal.pcbi.1003531.

Chapter 7

Conclusion

The marine iron cycle is spatially and temporally dynamic and influenced by intersecting biological, geochemical, and physical forces. However, iron has been understudied in the marine environment when compared with macronutrients such as carbon and nitrogen. A deficit in knowledge currently exists with regard to how interactions between microorganisms and their environment operate to influence the marine iron cycle. The trace metal biogeochemistry community is making rapid progress identifying geochemical sources and sinks of Fe to the ocean (1), and biogeochemical model inputs and outputs are becoming better constrained with field observations (2). However, current models of the marine iron cycle still lag far behind the ability of existing models to capture the dynamics of the marine carbon, nitrogen, and phosphorous cycles (3). Newer sophisticated models must describe more than geochemical inputs and outputs in order to capture the observed dynamics of the marine iron cycle (3). An important first step in producing more representative models is the parameterization of biological processes that determine rates of regeneration and uptake from different marine iron pools. Ultimately, a greater understanding of the marine iron cycle and a better ability to model it will provide better predictions of the global carbon cycle and future global climate states in a rapidly changing world.

The impetus behind much of the research presented throughout this dissertation was born from a desire to contribute to the understanding of how marine microorganisms

interact with iron. Much of my dissertation research examines specifically how heterotrophic bacteria use and acquire iron under diverse conditions. From this perspective, my dissertation work contributes to the greater understanding of the role of heterotrophic bacteria in remineralizing and recycling biogenic iron in the marine environment by highlighting specific compounds that they utilize. However, the final chapter addresses the role of iron in limiting phytoplankton growth in a geographical region mostly thought to be limited by nitrogen. This chapter underscores the importance of understanding how phytoplankton physiology intersects with iron, light, and macronutrient concentrations at different strata in the water column. Ultimately, the central theme of this chapter remains similar to that of the prior chapters; namely, that an understanding of microbial physiologies, metabolisms, and genomics adds what I would consider to be an essential perspective to the marine iron cycle. It is my hope that work from this dissertation will be informative and useful for marine biogeochemists attempting to understand and describe iron uptake and remineralization and for modelers trying to more accurately represent the marine iron cycle.

7.1 References

1. Resing JA, Sedwick PN, German CR, Jenkins WJ, Moffett JW, Sohst BM, Tagliabue A. (2015) Basin-scale transport of hydrothermal dissolved metals across the South Pacific Ocean. *Nature* 523(7559):200–203.

2. Tagliabue A, Williams RG, Rogan N, Achterberg EP, Boyd PW (2014) A ventilation-based framework to explain the regeneration-scavenging balance of iron in the ocean. *Geophys Res Lett* 41(20):2014GL061066.

3. Tagliabue A, Aumont O, DeAth R, Dunne JP, Dutkiewicz S, Galbraith E, Misumi K, Moore JK, Ridgwell A, Sherman E, Stock C, Vichi M, Völker C, Yool A. (2016) How well do global ocean biogeochemistry models simulate dissolved iron distributions? *Global Biogeochem Cycles*:2015GB005289.