Ecology and Evolution of Lanthipeptides in Marine Picocyanobacteria

By
Andrés Fernando Cubillos-Ruiz

BSc. Microbiology
Universidad de los Andes (2006)

MSc. Microbiology
Universidad de los Andes (2008)

Submitted to the Department of Biology in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Microbiology at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

Signature redacted

Signature of Author _____
Andrés F. Cubillos-Ruiz
Microbiology Graduate Program

Signature redacted

Certified by _____
Sallie W. Chisholm
Institute Professor
Thesis Supervisor

Signature redacted

Accepted by _____
Kristala Jones Prather
Associate Professor of Chemical Engineering
Chair, Microbiology Graduate Program

Ecology and Evolution of Lanthipeptides in Marine Picocyanobacteria

By

Andrés Fernando Cubillos-Ruiz

Submitted in partial fulfillment of the
requirement for the degree of Doctor of
Philosophy in Microbiology

ABSTRACT

Microbial secondary metabolites are among the most structurally and functionally complex molecules in nature. Lanthipeptides are ribosomally derived peptide secondary metabolites that undergo extensive post-translational modification. Most lanthipeptides are bactericidal but they are also known to act as signaling molecules or morphogenetic peptides, nevertheless the function of many lanthipeptides remains unknown. Prochlorosins are a diverse group of lanthipeptides produced by strains of the ubiquitous marine picocyanobacteria *Prochlorococcus* and *Synechococcus*. Unlike other lanthipeptide-producing bacteria, picocyanobacteria utilize an unprecedented mechanism of catalytic promiscuity for the production multiple structurally diverse lanthipeptides using a single biosynthetic enzyme. Also unprecedented is the production of lanthipeptides by single celled, planktonic gram-negative bacteria in a dilute nutrient-limited habitat, which suggests that they may have an unconventional biological function. The overarching goal of this thesis is to further our understanding of the ecology and evolution of the prochlorosins, and provide insights into their biological role in the marine environment. Here, we demonstrate that the prochlorosin genes are widespread in the ocean and that globally distributed populations of marine picocyanobacteria have the genetic potential of producing thousands of different lanthipeptide structures. The diversity of prochlorosin structures provides an interesting model to study the evolutionary forces that drive the creation of new lanthipeptide structures. We present evidence that there is a unique evolutionary interplay between the components of prochlorosin biosynthesis pathway; while the peptide substrates independently expand and diversify within the genome, the catalytically promiscuous biosynthetic enzyme evolves under a strong purifying selection that maintains its substrate tolerant state. This relationship indicates that the lanthipeptide production trait in marine picocyanobacteria might find its evolutionary advantage in the plasticity of the production of multiple cyclic peptides with diverse ring topologies. The remarkable diversity of prochlorosins poses many questions regarding their biological role in the marine environment. In laboratory experiments, we explore of some of the potential bioactivity of the prochlorosins, namely their potential as signaling molecules, antimicrobials and nutrient sources. The results from this exploration open new perspectives for the role of the lanthipeptides in the natural environment – more specifically the oligotrophic ocean.

Thesis Supervisor: Sallie W. Chisholm
Title: Institute Professor

To my parents, aunts and the memory of my grandmother

ACKNOWLEDGMENTS

I would like to express my deep gratitude to all the people that helped me throughout of my PhD studies. First of all, I thank my thesis advisor, Penny Chisholm, for her mentorship and guidance. It has been an honor to be her student and I am tremendously grateful for her support and encouragement both inside and outside of the lab. I would like to thank the members of my thesis committee: Michael Laub, Janelle Thompson and Jing-Ke Weng, for their advice and continued interest in my projects. I am also grateful to Wilfred van der Donk at the University of Illinois, who has been more than a collaborator of the prochlorosin project. I thank Wilfred for the invaluable advice and for his support on my career goals.

Many other people also helped me get to this point and have influenced my growth as a scientist. I want to thank Roberto Kolter and Michael Fischbach for their continued support and guidance throughout my scientific career. I also thank my former advisors Maria Mercedes Zambrano and Patricia del Portillo for their friendship and encouragement. I always look up to them as great examples of leadership and tenacity.

I have had the great fortune to work with excellent people. I thank all the past and current members of the Chisholm Lab for the help and advice needed to complete this work. I also want to thank Yanxiang (Nancy) Shi from the van der Donk group for her hard work and for providing me with precious recombinant prochlorosins. I would also like to thank the Microbiology Graduate Program, especially Bonnielee Whang for her assistance throughout the course of my PhD.

I want to thank all my friends in Boston who have made these past six years a fun journey. Most importantly, I want to thank my parents and family for their never-ending support. Everything I am, I owe to my family, and they are my greatest source of inspiration. Finally, I want to thank Alejandra for all her love and support. She is the greatest blessing God has given me and I would have not make it without her. I am incredibly fortunate to have such an amazing woman as my wife.

My research was supported by grants from the Gordon and Betty Moore Foundation, the National Science Foundation Center for Microbial Oceanography: Research and Education (C-MORE), the Simmons Foundation through the SCOPE initiative and the MIT-Masdar Institute. An MIT-BP Energy Initiative Fellowship and an International Student Research Fellowship from the Howard Hughes Medical Institute additionally supported my studies, for which I am deeply grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Microbial secondary metabolism produces a wealth of small molecules collectively known as natural products that are used for interspecies competition and communication (Riley and Wertz, 2002). These small molecules have been an important source of useful therapeutic agents such as antibiotics, antifungals, immunosuppressive agents and anticancer agents (Clardy and Walsh, 2004). Depending on the biogenesis mechanism, peptide-based bacterial natural products can be broadly classified into non-ribosomal peptides (NRP) and ribosomally produced and post-translationally modified peptides (RiPPs). For the synthesis of NRP the cell uses refined molecular machines known as non-ribosomal peptide synthetases (NRPS) that allow them the use of a wide array of amino acid substrates (Walsh and Nolan, 2008). In contrast, the synthesis of RiPPs involves the use of the cellular traditional translation machinery and the resulting peptide scaffolds cannot explore amino acids beyond the canonical 20 proteinogenic amino acids, limiting their structural diversity to some degree. However, RiPPs natural products can be astoundingly complex in structure and display an incredible functional diversity that results from the action of tailoring enzymes that provide a wide variety of post-translational modifications. Among the most important groups of RiPPs natural products with interesting structural features are the lanthipeptides (Willey and van der Donk, 2007), cyanobactins (Sivonen et al., 2010), microcins (Nolan and Walsh, 2009) and thiopeptides (Li and Kelly, 2010).

Compared to NRP antibiotics, ribosomally derived peptide antibiotics are less celebrated as potential small-molecule therapeutic agents, in part because of the erroneous perception of intrinsic limitations in their structural diversity and antimicrobial activity. Nevertheless, the generation of antibacterial peptides by the post-translational modification of ribosomal precursors is a well-known strategy that is used by prokaryotes to increase their relative fitness in ecologically relevant settings (Riley and Wertz, 2002). The advent of massive microbial genomic information in recent years has revealed that natural product biosynthesis using the ribosomal machinery is much more widespread than originally anticipated, and has led to fascinating discoveries outlining new biochemical transformations in secondary metabolism, that at the same time suggests that the true structural diversity of these compounds is just beginning to be appreciated.

Converting ribosomal peptides into bioactive molecules: leader peptide-directed biosynthesis.

The general scheme in the biosynthesis of RiPPs involves the synthesis of an N-terminal-extended precursor peptide that undergoes various types of post-translational modification followed by proteolytic cleavage to release the active peptide. These tailoring processes release the peptides from the structural and functional constraints imposed on natural ribosomal peptides, while at the same time restricting conformational flexibility to allow increased metabolic and chemical stability (Oman and van der Donk, 2010).

The biosynthesis of ribosomally synthesized bacterial natural products shares certain key features. First, all of the RiPPs follow the mechanics of leader peptide-directed biosynthesis (Fig. 1-1), in which the precursor genes encode a peptide that contains an N-terminal leader extension in addition to the C-terminal core peptide that is processed to the mature compound. For the vast majority of natural products of ribosomal origin, the initial precursor peptide is much larger than the final product. These precursors typically contain N-terminal leader peptides, and in some cases, C-terminal extensions that are removed in the last step of the maturation process. Second, the genes encoding precursor peptides are often found in biosynthetic gene clusters and are accompanied by genes encoding modifying enzymes, immunity, and secretion. Third, many of these precursor peptides contain specific motifs within the leader peptide that act to recruit modifying enzymes, identify the sites of proteolysis, or play other roles (Chatterjee et al., 2005b; McIntosh et al., 2009).

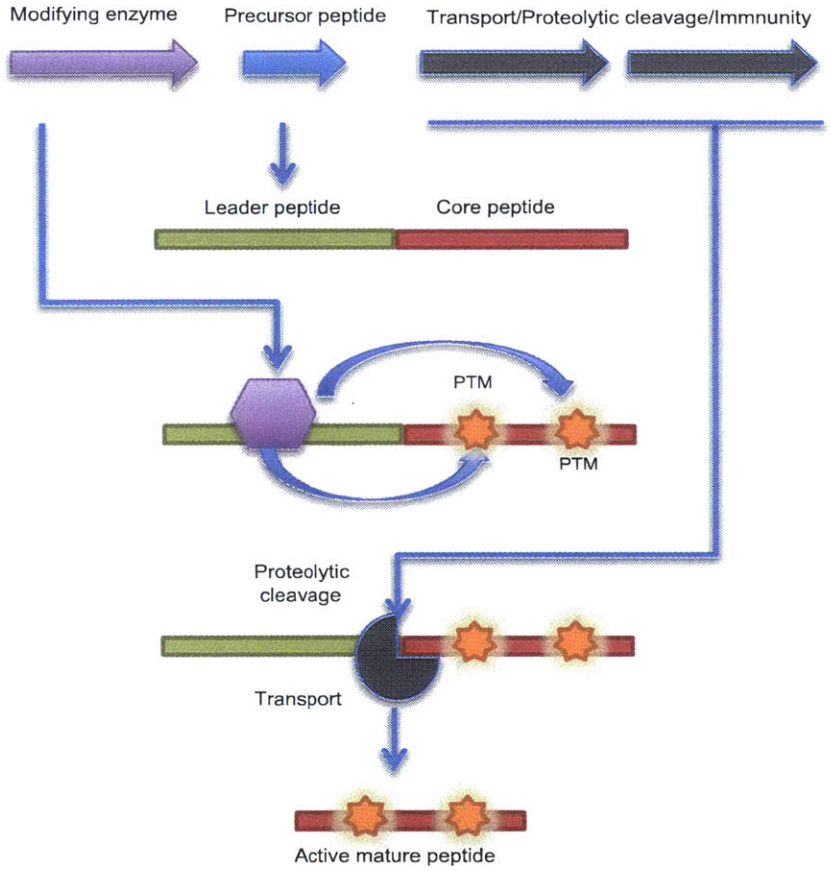

Figure 1-1: Leader peptide-directed biosynthesis.
The biosynthesis of RiPPs natural products involves the transcription and translation via ribosome of a pre-peptide. In general the pre-peptide contain a relatively conserved leader sequence directing enzyme modification (green) and a hypervariable core sequence that encodes the final natural product (red). The precursor peptide is post-translationally modified by enzymes (purple hexagon) that catalyze the formation of a large number of different chemical motifs (stars). In a final step, the precursor peptide can be subject of proteolytic cleavage and transported (black circle) outside the cells where it is found in its mature and active form. Sometimes the cell encodes an immunity protein that protects it from the antibiotic effect of the mature peptide. Post-translational Modification (PTM).

Many – but not all – of the post-translational modifications commonly thought of as unique to the NRP world are also found in RiPPs and allow them to explore a greater chemical space in a manner similar to NRP. RiPPs scaffolds, like those of NRP, can undergo a series of main-chain and side-chain modification (Table 1-1). Among the most important shared post-translational modifications are the occurrence of macrocyclization, the presence of D-amino acids, heterocycles derived from Cys, Ser, and Thr, isoprene-derived groups, and dehydrated residues (McIntosh et al., 2009). Notably, RiPPs scaffolds also display post-translational modifications that are unique to their realm. The most notable examples are well-known modifications involving the ligation of amino acid residues, these include the formation of disulfide linkages, lanthionine bridges such as in the lantibiotic nisin A, heteroaromatic rings such as in microcin B17, and macrolactam linkages such as in the lasso peptide microcin J25 (Nolan and Walsh, 2009).

Side-Chain Modifications	
	Present in NRP?
Dehydration of Ser and Thr	Yes
Lanthionine synthesis	No
Heterocyclization of Cys, Ser, and Thr	Yes
Prenylation Trp/Ser/Thr	Yes (Trp only)
Disulfide bonds	Infrequent
Amino acid a-modification	Yes
Epimerization	Yes

Main-Chain Modifications	
	Present in NRP?
Proteolysis	No
Macrocyclization	Yes
Formylation	Yes
Nucleotide base addition	Nucleoside
Lactone formation	Yes
Siderophore addition	No
Aminovinylcysteine bridge	Unknown

Table 1-1: Post-translational modification commonly found in RiPPs or NRP natural products
The information on this table was adapted from reference (McIntosh et al., 2009)

One of the highlights of the leader peptide-directed biosynthesis is the multifunctional nature of the leader peptide. One of the most commonly proposed roles for the leader peptides is that of a secretion signal, but intriguingly, the majority of leader peptides of RiPPs natural products have no homology with the peptides of the typical secretory translocation pathways that are used in bacteria (Oman and van der Donk, 2010). Another role attributed to the leader peptides is that it serves as a recognition motif for the post-translational modification enzymes. The overall peptide sequences are often variable, but they are able to remain as substrates for specific post-translational processes depending upon conserved precursor peptide sequences and enzyme selectivity (Oman and van der Donk, 2010). This property of the leader peptide-directed biosynthesis is the most attractive from a natural product engineering perspective as it may allow generation of novel products by attachment of core peptide variants to the leader peptides. Some other more traditional proposed tasks for leader peptides are to assist in folding of the precursor peptide, stabilizing the precursor against degradation, or keeping the precursor peptide inactive during biosynthesis inside the host until the appropriate time for secretion and proteolysis (Braun and Tommassen, 1998).

Expanding the repertoire of ribosomal natural products

The upsurge of complete microbial genome sequences has provided invaluable information to gain new insights into the genetic capacity of organisms to generate secondary metabolites. Recent genome-enabled studies have shown that many classes of peptide natural products that were initially suspected to be of non-ribosomal origin are in fact gene-encoded, and that structural motifs previously thought exclusive of non-ribosomal secondary metabolites are also found in ribosomally synthesized compounds. The discovery of the thiostrepton 2 biosynthetic gene cluster, for example, relied upon partial sequencing of the *Streptomyces laurentii* genome, a known thiopeptide producer (Kelly et al., 2009). Similarly, in the discovery of the biosynthetic machinery for thiocillin 3, two independent works deduced the ribosomal nature of this thiopeptide natural product, using as a base the mining of the genome of *Bacillus cereus* ATCC 14579 (Liao et al., 2009; Wieland Brown et al., 2009). Other notable examples of how genome mining has led to the disambiguation between ribosomal and non-ribosomal

biogenesis pathways are the discovery of widely distributed fungal and bacterial toxins (Hallen et al., 2007; Lee et al., 2008).

In the above-mentioned cases, the mining of the genome of a bacterial producer of a known natural product led to the identification of the biosynthetic pathway of the compound of interest. Remarkably, genome mining can also lead to the discovery of novel natural products in the opposite direction, i.e., a gene-to-molecule fashion. This approach has been widely exploited in the search of bioactive small molecules of the NRP type (Corre and Challis, 2009). Because large, highly conserved genes encode NRPS, their presence, distribution and abundance is easily assessed by means of standard homology searches like BLAST (Chatterjee et al., 2005a). Since identifying one gene means the others are close by, cloning gene clusters for complete biosynthetic pathways is now a straightforward strategy for NRP. In contrast, the genes for RiPPs natural products are harder to discover, especially in the absence of any chemical or bioactivity information. The precursor peptides are small and display low nucleotide identity to other precursor peptides, and so are often missed in automatic genome annotations (Corre and Challis, 2009). Similarly, new families of modifying enzymes are often not closely related enough to characterized relatives to be identified by BLAST searching (Haft et al., 2010).

In spite of these limitations, the information derived from biochemical and genetic investigations of known RiPPs systems enables more directed mining of ribosomal biosynthetic pathways of natural products. Recent work from Haft *et al.* (Haft et al., 2010) used a bioinformatics approach in which the genetic structure of the biosynthetic pathways of different ribosomally produced natural products families were converted into specific Hidden Markov Model-Based (HMM) protein family definitions. They applied these HMMs family definitions to more than 1,000 available bacterial genome sequences, specifically searching for cyclodehydratase protein sequences. From this, two new precursor peptide classes were discovered, both of which are related to larger proteins; one is related to a non-catalytic fragment of nitrile hydrolase (NHase) and the other to the Nif11 proteins involved in nitrogen fixation (Haft et al., 2010). This study provides evidence that the synthesis of secondary metabolites via the ribosome turns out to be

much more widespread than originally anticipated. The integration of genomic information and the knowledge of the biogenesis mechanisms of RiPPs have proven useful for the discovery of novel ribosomal small molecules in unexpected environments and microorganisms. Notably, this type of approach led to the discovery of lanthipeptides in strains of the ubiquitous planktonic marine cyanobacteria *Prochlorococcus* and *Synechococcus*, from which natural products had never been isolated (Li et al., 2010). The lanthipeptides are the central subject of this work and a general description of this family of RiPPs will be presented in the following sections. For an in-depth revision of lanthipeptides and other families of RiPPs please see (Arnison et al., 2013) and references therein.

Lanthipeptides

Lanthionine-containing peptides, or lanthipeptides, are small (usually < 40 amino acids) polycyclic peptides that go through different posttranslational modifications. They are characterized by the presence of the thioether-cross-linked amino acids lanthionine (Lan) and methyl-lanthionine (MeLan) (Arnison et al., 2013; Schnell et al., 1988). These posttranslational modifications originate from the enzymatic action of the lanthionine synthetase that catalyzes the dehydration of Ser and Thr residues to produce dehydroalanine (Dha) and dehydrobutyrine (Dhb) residues, respectively. Then, the thiol group of a Cys residue is added to either Dha or to Dhb to form Lan or MeLan bridges, respectively (Knerr and van der Donk, 2012; Willey and van der Donk, 2007). Lanthipeptides are classified into four different classes depending on the differences between the biosynthetic machinery involved in the processing and maturation of the precursor peptide. Based on their sequence homology to other enzymes of the primary metabolism, the mechanism of action and the structure of many lanthipeptide synthetases presumably have evolved from other posttranslational modification enzymes (Yu et al., 2013).

In class I lanthipeptides the formation of the Lan (or MeLan) is performed by two separate enzymes: the LanB catalyzes the dehydration of Ser and Thr residues and the cyclase LanC promotes the addition of a thiol group from Cys to either the Dha or Dhb resulting from the action of LanB (Arnison et al., 2013; Knerr and van der Donk, 2012).

While the Lan B enzyme does not show sequence homology with other dehydratases in other organisms, homologous domains of the LanC cyclase can be found in other organisms, including insects and plants (Johnston et al., 2007; Yu et al., 2013). The dehydration and cyclization processes in class II, III and IV lanthipeptides are carried out by bifunctional lanthionine synthases. Biosynthesis of class II lanthipeptides is catalyzed by the lanthionine synthetase LanM that show both dehydratase and cyclase activities. The N-terminal domain of LanM mediates the dehydration of Ser and Thr residues by first performing a phosphorylation reaction, followed by the elimination of phosphate groups (Chatterjee et al., 2005b). Similar to LanB, the dehydratase domain of LanM proteins does not display sequence homology when compared with different protein databases. In contrast, the C-terminal cyclase domain of LanM shows homology with the LanC cyclase from class I lanthipeptides (Yu et al., 2013).

Lanthionine synthetases from class III and IV are called LanKC and LanL, respectively (Yu et al., 2013). Similar to LanM synthetases from class II, these enzymes have shown kinase activity, at the central domain, by phosphorylating Ser/Thr residues to mediate dehydration to the substrate peptide (Goto et al., 2010; Muller et al., 2010). However, the dehydration does not involve elimination of the phosphate groups as it is observed in LanM enzymes, instead class III and IV enzymes contain a Lyase domain at the N-terminal that does not display homology with the LanM N-terminal. The N-terminal and central domain of class III and IV enzymes are very similar, however the cyclase domain in the C-terminal differ. Class IV enzymes show a canonical LanC cyclase domain and homology to the domains in proteins from class I and II including the metal binding resides, however these metal ligands seem to be absent in class III synthetases (Goto et al., 2010; Kodani et al., 2005). The cyclization reaction also seems to differ in class III proteins, in which the structure generated is labionin (commonly found in labyrinthopeptins), whereas class IV enzymes generate (methyl)-lanthionines (Goto et al., 2010; Meindl et al., 2010; Wang and van der Donk, 2012). The diversity in the biochemical mechanisms observed in these four pathways of biosynthesis of lanthionine-containing peptides highlights the biological relevance of installing Lan or MeLan bridges in peptides of ribosomal origin.

Bioactivities of Lanthipeptides

Most of the known lanthionine-containing peptides have antimicrobial activity and are referred to as lantibiotics. Lantibiotics were initially discovered in gram-positive bacteria, and have gained attention since then, because of their activity as bacteriocins: bacterial proteinaceous compounds that display antibiotic activity against strains of the same species or closely related species. The use of a combination of genetic and biochemical approaches in the last decade have contributed to establish a better understanding of the relationship between the structure of lanthipeptides and their biological activity, in particular for the description of the antimicrobial mechanism of lantibiotics. Nisin is the most extensively studied of the lantibiotics and belongs to the class I of lanthipeptides. Importantly, despite the fact that nisin has been used in the food industry for more than four decades no resistant bacterial isolates have been identified (Willey and van der Donk, 2007). There are three different structural variants of Nisin, all displaying antimicrobial activity due to their closely related structures: Nisin A (the best described and model template for the study of variants), Nisin Z (only differs by one amino acid when compared to Nisin A) and Nisin Q (differs by four amino acids when compared to Nisin A) (Delves-Broughton et al., 1996). Nisin binds to lipid II, an important precursor in the peptidoglycan biosynthesis pathway, resulting in pore formation and cell wall synthesis inhibition (Breukink et al., 1999; Brotz et al., 1998). The N-terminus of nisin is composed of two rings that bind to the pyrophosphate moiety of lipid II (Hasper et al., 2004; Hsu et al., 2002; Hsu et al., 2004). Then, the C-terminus of nisin is inserted into the bacteria bilayer surface vertically resulting in the formation of pores of about 2 to 2.5 mm diameter (Wiedemann et al., 2004).

The class II lantibiotics represent the largest of all the classes of lantibiotics, however, the mechanism of action of the different members in this group is not fully understood. Mersacidin is a highly relevant example of class II lanthipeptides because it displays a potent activity against methicillin-resistant *Staphylococcus aureus* (Kruszewska et al., 2004). Although mersacidin and other class II lantibiotics also bind lipid II, the mechanism of action of mersacidin is based on the inhibition of the transglycosylation step in peptidoglycan biosynthesis instead of membrane pore formation (Brotz et al.,

1997). However, the molecular details of the bioactivity of class II lantibiotics have not been fully described.

The class III lanthipeptides, SapB and SapT, are the only examples of lanthipeptide bioactivity outside antibiosis. These peptides, found in some strains of *Streptomyces*, function as morphogenetic peptides and are involved in the formation of aerial hyphae due to the amphiphilic and hydrophobic nature of their side chains. (Kodani et al., 2004; Kodani et al., 2005). Note that their function was discovered before it was known that they were lanthionine-containing peptides, highlighting that by screening only for antibiotic activity, novel bioactivities of lanthipeptides may be missed (Arnison et al., 2013).

Prochlorosins: a family of diverse lanthipeptides from marine picocyanobacteria

Marine picocyanobacteria, composed of *Prochlorococcus* and *Synechococcus* genera, are the most abundant photosynthetic organisms on Earth. Using a genome-enabled approach, Li et al discovered type II lanthipeptides in some strains of picocyanobacteria, from which natural products had never been isolated (Li et al., 2010). In most lanthipeptide-producing bacteria, the lanthionine synthetase modifies only a single precursor peptide. Remarkably, lanthipeptide-encoding strains of *Prochlorococcus* and *Synechococcus* are unique in that they show that a single organism can produce as many as 29 different lanthipeptide secondary metabolites – named prochlorosins – from distinct gene-derived precursors (ProcA) by using only one promiscuous biosynthetic enzyme (ProcM) (Fig. 1-2).

To characterize the prochlorosin biosynthesis pathway, Li *et al.* demonstrated *in vitro* that the ProcM enzyme from *Prochlorococcus* MIT9313 is able to catalyze the dehydration and cyclization of all of the eighteen different prochlorosin precursor peptide tested. Furthermore, to demonstrate that prochlorosins are also produced *in vivo,* the transcripts of the *procM* gene and several *procA* genes were detected, and in spite of low-yields of recovery, 3 out of the 29 prochlorosins of *Prochlorococcus* MIT9313 were detected in the spent medium (Li et al., 2010). Importantly, ESI-MSMS analysis of the three prochlorosins recovered from the spent medium demonstrated that the *in vitro* prepared

compounds have the same ring topologies as the ones naturally produced and that the leader peptide cleavage site is located at the anticipated Gly-Gly motif that is commonly found in other type II lanthipeptides (Willey and van der Donk, 2007).

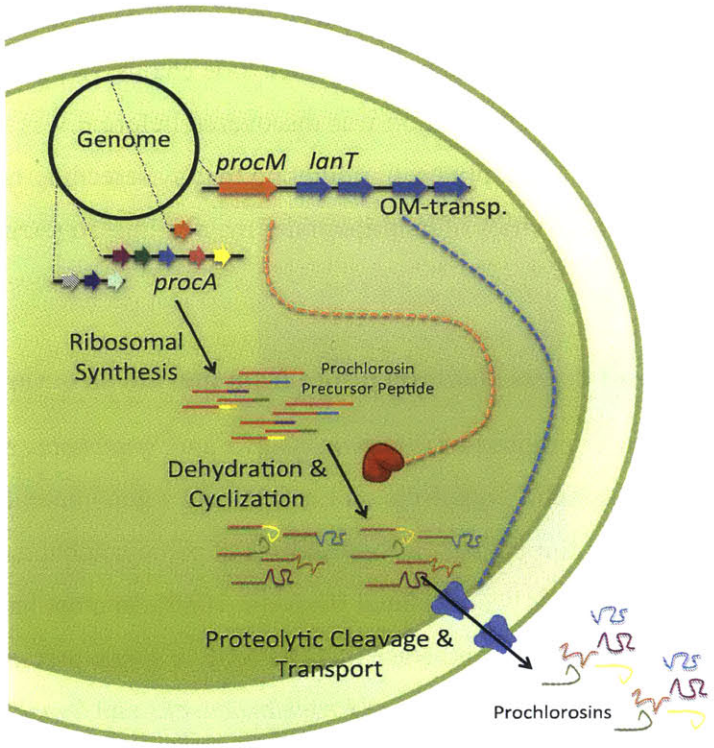

Figure 1-2. Prochlorosin biosynthesis pathway.
Prochlorosins are created from gene-encoded precursor peptides (*procA*) that are composed of a leader peptide and a core peptide region. These ribosomal peptides are tailored by a dedicated lanthionine synthetase (*procM*), which binds to the leader peptide region and catalyzes the dehydration and cyclization of distinct residues in the core peptide region. Following modification, the N-terminal leader peptide is removed from the precursor peptide by the action of a membrane associated ABC-transporter with a C39 proteolytic domain (*lanT*). In the prochlorosin pathway, it is presumed that the C-terminal core peptide is released to the environment as a mature active product by the combined action of the LanT and a TolC-like outer membrane transporter (OM-transp) that is usually encoded in the same locus.

The 'catalytic promiscuity' in the biosynthesis of prochlorosins is unprecedented both in nature and in the laboratory, and thus is of great interest to biologists and chemists alike. The ProcM lanthionine synthetase is an exceptionally substrate-tolerant enzyme that is able to introduce post-translational modification into vastly different core peptides. One property in which ProcM differ from other lanthionine synthetases enzymes is that its active site has three Cys ligands to the Zn^{2+} co-factor instead of the two Cys and one His, commonly found in other cyclization enzymes (Tang and van der Donk, 2012). The increased number of thiolates from Cys on the Zn^{2+} ions is known to increase the reactivity and might be related to the substrate tolerance (Mukherjee and van der Donk, 2014; Yu et al., 2015).

The prochlorosin system is also a remarkable example of structural diversity of lanthionine-containing peptides. Multiple sequence alignment of the 29 ProcA amino acid sequences reveals two striking features of the prochlorosins from *Prochlorococcus* MIT9313 (Fig. 1-3). First, none of the peptides have Ser/Thr and Cys residues in positions that would result in ring patterns similar to the currently known lanthipeptide ring topologies. Second, the sequence of the leader peptide is remarkably conserved among the peptides, which contrasts with the diversity of the core peptide region (Fig. 1-3). These unique characteristics of the prochlorosin pathway indicate that *Prochlorococcus* MIT9313 has the genetic capacity of producing as many different secondary metabolites as the model antibiotic-producing Actinomycetes but with a genome less than one-third in size (2.4MB compared with the ~9.0 MB). Notably, *Prochlorococcus* "lifestyle" is very different from that of microbes known to produce these types of compounds; they are single-celled and free-floating and live in a very dilute habitat where the function of secondary metabolites is not readily apparent.

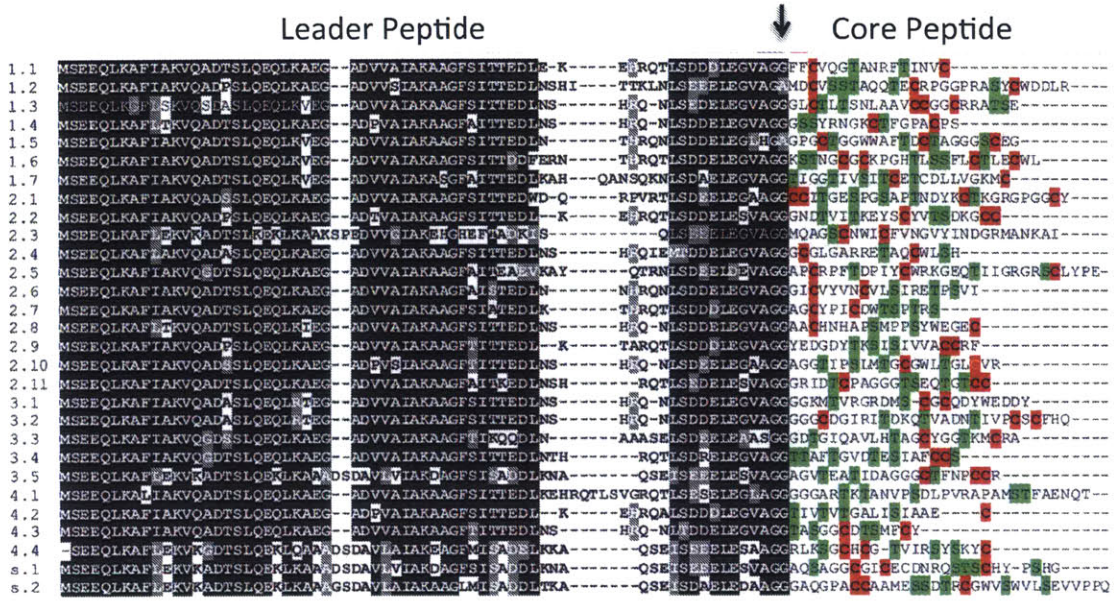

Figure 1-3. Prochlorosin precursor peptides of *Prochlorococcus* MIT9313 (Adapted from Li et al. 2010). Multiple sequence alignment of ProcA amino acid sequences reveals a remarkable conservation of the leader peptide. Outstanding hyper-variability in the core region begins right after the protease cleavage motif GG/GA (arrow)

Motivating Questions and Overview

Lanthipeptides have been studied in isolated microorganisms outside of any ecological context. Hence, our understanding of their diversity is derived primarily from studies on single microbial strains of interest because of their potential as producers of antimicrobials and, therefore, they are undoubtedly not representative of the full extent of lanthipeptide diversity and functionality. The discovery of these compounds in the globally abundant and genetically diverse picocyanobacteria *Prochlorococcus* and *Synechococcus* (Biller et al., 2015; Scanlan et al., 2009) opens opportunities to understand the forces shaping the evolution and diversity of this class of compounds. **Consequently, the overarching goal of this thesis is to further our understanding of the ecology and evolution of the diverse lanthipeptides of marine picocyanobacteria, and attempt to unravel their biological role in this extreme, but globally distributed, habitat.**

In Chapter 2, we investigate the ecology and evolution of prochlorosins. First, we make use of both previously and newly sequenced strains of *Prochlorococcus* and *Synechococcus* to ask: **What is the genetic diversity of the prochlorosin biosynthesis pathway in the marine picocyanobacterial phylogeny?** This phylogenomic analysis revealed genomic plasticity and structural diversity features of the prochlorosin pathway not present in any other biosynthesis pathway of secondary metabolites. Given that these features were observed in a small subset of picocyanobacterial genomes, we moved onwards to study wild populations of marine picocyanobacteria and posed the question: **How widespread and diverse is the prochlorosin biosynthesis pathway in globally distributed natural populations of marine picocyanobacteria?** To address this question, we employed a combination of biogeographic and metagenomic approaches that enabled us to uncover an unprecedented abundance and diversity of prochlorosin structures. The exceptional diversity of prochlorosins prompted us to ask the question: **What are the molecular evolutionary mechanisms that drive the diversification of prochlorosins?** For this, the information gathered from genomes and metagenomes was used to elucidate the unique evolutionary dynamics that govern the diversification of the prochlorosin biosynthesis pathway. We discuss how the mode of evolution of prochlorosin biosynthesis pathway contrasts with the canonical lantibiotic biosynthesis pathways. Moreover, we hypothesize a molecular mechanism by which hypervariable regions might arise in the core peptide region of the prochlorosins (Appendix A).

The prochlorosin system is the first experimentally characterized lanthipeptide pathway of gram-negative bacteria. The combinatorial mechanism in the biosynthesis of the prochlorosins and the potential diversity of its structures is unprecedented in any other family of natural products. Also unprecedented is the production of these types of compounds by single-celled, free-floating bacteria in an extremely dilute habitat, which suggests that they may have novel biological functions or mechanisms of action. In Chapter 3, we explore the potential biological activities of the prochlorosins. First, by using a set of four recombinant prochlorosins from *Prochlorococcus* MIT9313, we explored the effect of prochlorosins on their own producer strain and asked the questions: **Is the biosynthesis of prochlorosins autoregulatory? Could the prochlorosins be acting as autocrine signaling molecules in marine picocyanobacteria?** Subsequently,

we performed some exploratory experiments to address the possible effect of the recombinant prochlorosins on the growth other members of the marine microbial community. In particular, we asked: **Do prochlorosins display bacteriocin activity against other *Prochlorococcus* strains? What is the effect of prochlorosins on heterotrophic bacteria commonly co-isolated with *Prochlorococcus*?**

In addition, we set out find an ecologically relevant system to test the effect of prochlorosins and to think about their function from a perspective that is relevant to the nutrient-limited environment where they happen. **Could lanthipeptides serve as a nutrient source in the oligotrophic ocean?** To address this, we tested the effect of prochlorosins on a cultured representative of the SAR11 clade, the dominant sympatric heterotroph with *Prochlorococcus* in the oligotrophic ocean. We present proof-of-concept evidence that lanthipeptides can be used to supply some of the unique nutritional requirements of the SAR11 group.

In Chapter 4, we investigated the role of amino acids as a possible source of organic nitrogen for *Prochlorococcus*, which is not directly related to the function of prochlorosins, but to the idea of the importance of organic compounds as nutrient sources in the oligotrophic environment. In Chapter 4, we describe an unexpected phenomenon of amino acid toxicity and tolerance in strains of *Prochlorococcus* that is observed under culture conditions. Finally, in Chapter 5, we provide the concluding remarks of this work and discuss future directions for the study of the ecological role of prochlorosins, some of which have motivated a field experiment whose experimental design is presented in the Appendix B, and that is expected to bring insights into the biological role of prochlorosins in the near future.

This work illuminates important features of the lanthipeptide production trait in natural populations of picocyanobacteria that, in conjunction with further laboratory and field experiments, will help us unravel the role of these natural products in marine microbial ecology. In the long term, the understating of how nature creates, evolves and utilizes peptide-based natural products –and its biosynthetic enzymes– is necessary for the rational use of natural products in human health, biotechnology and further industrial applications.

REFERENCES

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., *et al.* (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep *30*, 108-160.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015). *Prochlorococcus*: the structure and function of collective diversity. Nat Rev Microbiol *13*, 13-27.

Braun, P., and Tommassen, J. (1998). Function of bacterial propeptides. Trends Microbiol *6*, 6-8.

Breukink, E., Wiedemann, I., van Kraaij, C., Kuipers, O.P., Sahl, H.G., and de Kruijff, B. (1999). Use of the cell wall precursor lipid II by a pore-forming peptide antibiotic. Science *286*, 2361-2364.

Brotz, H., Bierbaum, G., Reynolds, P.E., and Sahl, H.G. (1997). The lantibiotic mersacidin inhibits peptidoglycan biosynthesis at the level of transglycosylation. Eur J Biochem *246*, 193-199.

Brotz, H., Josten, M., Wiedemann, I., Schneider, U., Gotz, F., Bierbaum, G., and Sahl, H.G. (1998). Role of lipid-bound peptidoglycan precursors in the formation of pores by nisin, epidermin and other lantibiotics. Mol Microbiol *30*, 317-327.

Chatterjee, C., Miller, L.M., Leung, Y.L., Xie, L., Yi, M., Kelleher, N.L., and van der Donk, W.A. (2005a). Lacticin 481 synthetase phosphorylates its substrate during lantibiotic production. J Am Chem Soc *127*, 15332-15333.

Chatterjee, C., Paul, M., Xie, L., and van der Donk, W.A. (2005b). Biosynthesis and mode of action of lantibiotics. Chem Rev *105*, 633-684.

Clardy, J., and Walsh, C. (2004). Lessons from natural molecules. Nature *432*, 829-837.

Corre, C., and Challis, G.L. (2009). New natural product biosynthetic chemistry discovered by genome mining. Nat Prod Rep *26*, 977-986.

Delves-Broughton, J., Blackburn, P., Evans, R.J., and Hugenholtz, J. (1996). Applications of the bacteriocin, nisin. Antonie Van Leeuwenhoek *69*, 193-202.

Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M.J., and van der Donk, W.A. (2010). Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. PLoS Biol *8*, e1000339.

Haft, D.H., Basu, M.K., and Mitchell, D.A. (2010). Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. BMC Biol *8*, 70.

Hallen, H.E., Luo, H., Scott-Craig, J.S., and Walton, J.D. (2007). Gene family encoding the major toxins of lethal Amanita mushrooms. Proc Natl Acad Sci U S A *104*, 19097-19101.

Hasper, H.E., de Kruijff, B., and Breukink, E. (2004). Assembly and stability of nisin-lipid II pores. Biochemistry *43*, 11567-11575.

Hsu, S.T., Breukink, E., de Kruijff, B., Kaptein, R., Bonvin, A.M., and van Nuland, N.A. (2002). Mapping the targeted membrane pore formation mechanism by solution NMR: the nisin Z and lipid II interaction in SDS micelles. Biochemistry *41*, 7670-7676.

Hsu, S.T., Breukink, E., Tischenko, E., Lutters, M.A., de Kruijff, B., Kaptein, R., Bonvin, A.M., and van Nuland, N.A. (2004). The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. Nat Struct Mol Biol *11*, 963-967.

Johnston, C.A., Temple, B.R., Chen, J.G., Gao, Y., Moriyama, E.N., Jones, A.M., Siderovski, D.P., and Willard, F.S. (2007). Comment on "A G protein coupled receptor is a plasma membrane receptor for the plant hormone abscisic acid". Science *318*, 914; author reply 914.

Kelly, W.L., Pan, L., and Li, C. (2009). Thiostrepton biosynthesis: prototype for a new family of bacteriocins. J Am Chem Soc *131*, 4327-4334.

Knerr, P.J., and van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. Annu Rev Biochem *81*, 479-505.

Kodani, S., Hudson, M.E., Durrant, M.C., Buttner, M.J., Nodwell, J.R., and Willey, J.M. (2004). The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in Streptomyces coelicolor. Proc Natl Acad Sci U S A *101*, 11448-11453.

Kodani, S., Lodato, M.A., Durrant, M.C., Picart, F., and Willey, J.M. (2005). SapT, a lanthionine-containing peptide involved in aerial hyphae formation in the streptomycetes. Mol Microbiol *58*, 1368-1380.

Kruszewska, D., Sahl, H.G., Bierbaum, G., Pag, U., Hynes, S.O., and Ljungh, A. (2004). Mersacidin eradicates methicillin-resistant Staphylococcus aureus (MRSA) in a mouse rhinitis model. J Antimicrob Chemother *54*, 648-653.

Lee, S.W., Mitchell, D.A., Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V., and Dixon, J.E. (2008). Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci U S A *105*, 5879-5884.

Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P.J., Joewono, I., Rusch, D., Chisholm, S.W., and van der Donk, W.A. (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. Proc Natl Acad Sci U S A *107*, 10430-10435.

Li, C., and Kelly, W.L. (2010). Recent advances in thiopeptide antibiotic biosynthesis. Nat Prod Rep *27*, 153-164.

Liao, R., Duan, L., Lei, C., Pan, H., Ding, Y., Zhang, Q., Chen, D., Shen, B., Yu, Y., and Liu, W. (2009). Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. Chem Biol *16*, 141-147.

McIntosh, J.A., Donia, M.S., and Schmidt, E.W. (2009). Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. Nat Prod Rep *26*, 537-559.

Meindl, K., Schmiederer, T., Schneider, K., Reicke, A., Butz, D., Keller, S., Guhring, H., Vertesy, L., Wink, J., Hoffmann, H., *et al.* (2010). Labyrinthopeptins: a new class of carbacyclic lantibiotics. Angew Chem Int Ed Engl *49*, 1151-1154.

Mukherjee, S., and van der Donk, W.A. (2014). Mechanistic studies on the substrate-tolerant lanthipeptide synthetase ProcM. J Am Chem Soc *136*, 10450-10459.

Muller, W.M., Schmiederer, T., Ensle, P., and Sussmuth, R.D. (2010). In vitro biosynthesis of the prepeptide of type-III lantibiotic labyrinthopeptin A2 including formation of a C-C bond as a post-translational modification. Angew Chem Int Ed Engl *49*, 2436-2440.

Nolan, E.M., and Walsh, C.T. (2009). How nature morphs peptide scaffolds into antibiotics. Chembiochem *10*, 34-53.

Oman, T.J., and van der Donk, W.A. (2010). Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat Chem Biol *6*, 9-18.

Riley, M.A., and Wertz, J.E. (2002). Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol *56*, 117-137.

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., Post, A.F., Hagemann, M., Paulsen, I., and Partensky, F. (2009). Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev *73*, 249-299.

Schnell, N., Entian, K.D., Schneider, U., Gotz, F., Zahner, H., Kellner, R., and Jung, G. (1988). Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. Nature *333*, 276-278.

Sivonen, K., Leikoski, N., Fewer, D.P., and Jokela, J. (2010). Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. Appl Microbiol Biotechnol *86*, 1213-1225.

Tang, W., and van der Donk, W.A. (2012). Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria. Biochemistry *51*, 4271-4279.

Walsh, C.T., and Nolan, E.M. (2008). Morphing peptide backbones into heterocycles. Proc Natl Acad Sci U S A *105*, 5655-5656.

Wang, H., and van der Donk, W.A. (2012). Biosynthesis of the class III lantipeptide catenulipeptin. ACS Chem Biol *7*, 1529-1535.

Wiedemann, I., Benz, R., and Sahl, H.G. (2004). Lipid II-mediated pore formation by the peptide antibiotic nisin: a black lipid membrane study. J Bacteriol *186*, 3259-3261.

Wieland Brown, L.C., Acker, M.G., Clardy, J., Walsh, C.T., and Fischbach, M.A. (2009). Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. Proc Natl Acad Sci U S A *106*, 2549-2553.

Willey, J.M., and van der Donk, W.A. (2007). Lantibiotics: peptides of diverse structure and function. Annu Rev Microbiol *61*, 477-501.

Yu, Y., Mukherjee, S., and van der Donk, W.A. (2015). Product Formation by the Promiscuous Lanthipeptide Synthetase ProcM is under Kinetic Control. J Am Chem Soc *137*, 5140-5148.

Yu, Y., Zhang, Q., and van der Donk, W.A. (2013). Insights into the evolution of lanthipeptide biosynthesis. Protein Sci *22*, 1478-1489.

CHAPTER 2

Evolutionary Radiation of Lanthipeptides in Natural Populations of Marine Picocyanobacteria

Andres Cubillos-Ruiz[1,2]

Jessica W. Berta-Thompson[1,2]

Jamie Becker[2]

Sallie W. Chisholm[2]

[1]Microbiology Graduate Program, Massachusetts Institute of Technology.

[2]Departments of Biology and Civil and Environmental Engineering, Massachusetts Institute of Technology.

ABSTRACT

Lanthipeptides are ribosomally derived peptides that undergo extensive post-translational modifications. Prochlorosins are a recently discovered group of diverse lanthipeptides that are produced by certain strains of the globally distributed marine picocyanobacteria *Prochlorococcus* and *Synechococcus*. The discovery of these compounds in the globally abundant and genetically diverse genera *Prochlorococcus* and *Synechococcus* provides a model to investigate the forces shaping the evolution and diversity of lanthipeptides in a natural microbial population. In this work, analysis of the prochlorosin biosynthesis pathway genes in 9 picocyanobacterial genomes predicts the presence of 181 novel lanthipeptide products, out of which only a single pair of products displays similarity in their putative ring topologies. Our findings reveal genomic plasticity and structural diversity features not present in any other known family of lanthipeptides. Furthermore, using a biogeographic approach, we demonstrate that different wild marine picocyanobacterial populations from the Atlantic and Pacific Oceans harbor largely dissimilar collections of the prochlorosin precursor peptide genes, indicating that the selective pressures acting on these genes promote the diversification of lanthipeptide structures rather than the convergence of structures with similar ring topologies. Intrigued by how this unprecedented diversity could arise, we investigated the evolutionary forces driving the diversification of prochlorosins and determined that while the peptide substrates rapidly expand and diversify within a genome, the catalytically promiscuous lanthionine synthetase evolves under a strong purifying selection to maintain its substrate tolerant state. This unique evolutionary dynamic suggests that the lanthipeptide production trait in marine picocyanobacteria might find its evolutionary advantage in the plasticity of its biosynthesis pathway for the creation and rapid diversification of multiple cyclic peptides with diverse ring topologies, which contrasts with canonical lanthipeptide biosynthesis pathways that evolve towards creating a single molecule with a defined ring topology.

INTRODUCTION

Microbial secondary metabolism produces a wealth of small molecules collectively known as natural products, which are among the most structurally and functionally diverse molecules in nature. Cyanobacteria are a rich source of natural products with interesting biological activities, and the most widely recognized are the result of a non-ribosomal pathway (Nunnery et al., 2010). Nonetheless, ribosomally synthesized and post-translationally modified peptides (RiPPs), with potent biological activities, have been found in diverse cyanobacterial species (Nunnery et al., 2010; Sivonen et al., 2010) and the upsurge of complete microbial genome sequences and novel data-mining approaches have revealed that the synthesis of RiPPS is much more widespread than originally thought (Haft et al., 2010; Li et al., 2010; Schmidt, 2010; Wang et al., 2011). Notably, this genomic approach led to the discovery of lanthipeptides in strains of the ubiquitous planktonic marine picocyanobacteria *Prochlorococcus* and *Synechococcus*, referred to as prochlorosins (Li et al., 2010).

Lanthipeptides are small ribosomally derived peptides that undergo extensive posttranslational modifications. This results in complex polycyclic molecules formed by the action of a lanthionine synthetase that dehydrates select Ser and Thr residues and catalyzes the intramolecular addition of Cys thiols to the resulting unsaturated amino acids, forming lanthionine and methyl-lanthionine bridges, respectively. The general scheme in the biosynthesis of lanthipeptides (and in general of most RiPPs) involves the synthesis of an N-terminal-extended precursor peptide composed by a N-terminal leader peptide and a C-terminal core region. The latter undergoes various types of posttranslational modifications followed by proteolytic cleavage to release the active peptide (Willey and van der Donk, 2007). While the vast majority of known lanthionine-containing peptides are bactericidal (Chatterjee et al., 2005b; Willey and van der Donk, 2007), some can act as signaling molecules (Schmitz et al., 2006) or morphogenetic peptides (Willey et al., 2006). The function of prochlorosins is unknown.

In most lanthipeptide-producing bacteria, the enzyme lanthionine synthetase modifies only a single precursor peptide. Lanthipeptide-encoding strains of *Prochlorococcus* and *Synechococcus*, on the other hand, can produce multiple different lanthipeptides from

distinct gene-derived precursors (ProcA) by using only one highly substrate-tolerant biosynthetic enzyme (ProcM) (Li et al., 2010). Biochemical characterization of the ProcM enzyme from the strain *Prochlorococcus* MIT9313, which encodes 29 different *procA* precursor peptide genes, has revealed that this enzyme is able to catalyze the dehydration and cyclization of all the 18 prochlorosin precursor peptide substrates tested so far (Li et al., 2010), and it is likely that the remaining 11 are substrates as well. Sequence analysis of the prochlorosin precursor peptides from this strain revealed two striking features. First, the core regions of the 29 precursor peptides are highly dissimilar, and none have Ser/Thr and Cys residues in positions that would result in ring patterns similar to the currently known lanthipeptide ring topologies. Second, in striking contrast to the core regions, the sequence of the leader peptide is remarkably conserved (Li et al., 2010). These observations from a single strain of *Prochlorococcus* make the prochlorosins not only a remarkable example of combinatorial biosynthesis, but also an exceptional case of genetic variability within a single gene family.

Our understanding of lanthipeptides is derived primarily from studies on single microbial strains of interest largely because of their potential as producers of antimicrobials. The discovery of these compounds in the globally abundant and genetically diverse picocyanobacteria *Prochlorococcus* and *Synechococcus* (Martiny et al., 2009; Mazard et al., 2011; Rodrigue et al., 2009; Zwirglmaier et al., 2007) opens opportunities to understand the forces shaping the evolution and diversity of this class of compounds. These two genera have been extensively studied at the genomic, physiological, and ecological level and as such have become model organisms for integrative systems biology (Coleman and Chisholm, 2007, 2010; Scanlan et al., 2009) *Prochlorococcus* is restricted to mid-latitude oligotrophic waters, whereas *Synechococcus* is found from pole to pole, and also occupies coastal waters. Collectively, these two genera are everywhere in the illuminated surface waters of the global oceans. *Prochlorococcus* contains diverse lineages with specific adaptations to high-light (HL) and low-light (LL) habitats (Rocap et al., 2003), allowing it to colonize the entire euphotic zone of the oceans. These "ecotypes" are phylogenetically related but physiologically distinct populations that are differentially distributed along not only light gradients, but also temperature and nutrient gradients (Johnson et al., 2006). *Synechococcus* is a more genetically diverse group with

suites of adaptations that allow it to cope with horizontal gradients of nutrients and light quality (Mazard et al., 2011). Collectively, these marine picocyanobacteria are ubiquitous, extremely abundant, and readily sampled and enumerated from their natural environment, which is also more easily characterized relative to the more heterogeneous environments inhabited by terrestrial bacteria (Biller et al., 2015). Therefore, the presence of lanthipeptides in these populations represents a unique opportunity to study secondary metabolites in the context of a well-characterized natural microbial population.

Here we first analyze the genomes of previously and newly sequenced strains of *Prochlorococcus* and *Synechococcus* to explore the genetic diversity of the prochlorosin biosynthesis pathway. We then employ a biogeographic approach to determine the distribution and abundance of the prochlorosin trait in wild populations, and using metagenomic data, explore the diversity of lanthionine-containing peptide structures in these marine phototrophs. Finally, we analyze the sequence variation patterns in a large set of prochlorosin precursor peptide genes – from cultures and wild cells – to decipher novel evolutionary mechanisms that govern the diversification of lanthipeptides in these organisms.

RESULTS AND DISCUSSION

Phylogenomic analysis of the prochlorosin biosynthesis pathway

To investigate the phylogenetic distribution of the prochlorosin trait among our set of genomes, we first searched for homologs of the prochlorosin lanthionine synthetase (*procM*) and the prochlorosin precursor peptide genes (*procA*) in a total of 41 *Prochlorococcus* (Biller et al., 2015) and 17 *Synechococcus* publicly available genomes that are representative of the major marine picocyanobacterial clades (Scanlan et al., 2009). The prochlorosin biosynthesis genes were present in all of the five *Prochlorococcus* strains in clade IV of low-light adapted ecotypes (LL-IV), and in three *Synechococcus* strains classified within the clades I, II and IX of the sub-cluster 5.1 of marine *Synechococcus* (Fig. S2-1A).

Given the apparent association of the prochlorosin trait with the *Prochlorococcus* LL-IV clade, we expanded our repertoire of genomes from strains belonging to this clade through targeted isolation efforts (see methods). We obtained 9 new LL-IV *Prochlorococcus* strains (Table S2-1), out of which 2 where found to encode the complete prochlorosin biosynthesis pathway. We also screened unsequenced *Synechococcus* cultures in our collection for prochlorosin precursor peptide genes using PCR, and found three from the CRD1 clade (Moore et al., 2002), which were then sequenced. In total, 11 prochlorosin-encoding genomes were identified in the complete collection of *Prochlorococcus* and *Synechococcus* genomes. While our data set is small, the prochlorosin trait appears broadly distributed among 4 distantly related clades in *Synechococcus*, while within *Prochlorococcus* it has a narrow phylogenetic range and is found only in a subset of strains form the LL-IV clade, the most deeply branching clade of *Prochlorococcus* (Fig. S2-1B). To further refine our collection of genomes for analysis, we eliminated genomes that were identical in their Internal Transcribed Spacer (ITS) ribosomal DNA sequence, including only those that had more than 1% overall genome dissimilarity. In total, 4 *Prochlorococcus* strains (MIT9313, MIT9303, MIT0701 and MIT1327) and 5 *Synechococcus* strains (MITS9509, MITS9508, RS9916, WH8016 and KORDI-100) met these criteria and were used for subsequent analysis.

The prochlorosin pathway displays great flexibility in the genetic organization of its three principal components: the precursor peptide genes, the lanthionine synthetase gene and the genes that encode the inner and outer membrane transporters (Fig. S2-2). With the exception of *Synechococcus* WH8016 that only harbors one *procA* gene, all prochlorosin-encoding genomes contained multiple *procA* genes located in different regions of the genome – some organized in arrays of up to 13 genes. Interestingly, in some strains the *procM* and *lanT* genes are not located in the same locus, suggesting modularity in the components of the prochlorosin pathway. These observations contrast with the observed genetic organization of other lanthipeptide biosynthesis pathways and with canonical peptide-based secondary metabolite pathways, where all the necessary components for biosynthesis are found in a gene cluster that behaves as a single evolutionary unit (Fischbach et al., 2008). The flexibility in the configuration of the genetic components of the prochlorosin pathway implies an uncommon mode of evolution for this family of lanthipeptides.

Genetic Diversity of Prochlorosin Precursor Peptide Genes

As mentioned above, prochlorosin-encoding strains display great variation in the number of *procA* genes per genome. In the set of 4 closely related strains from the LL-IV clade of *Prochlorococcus*, for example, the number of *procA* genes can range from 9 up to 29. In the *Synechococcus* strains, which span a larger genetic distance, the gene number variation is more dramatic, from 1 *procA* gene in WH8016 to 80 *procA* genes in MITS9509. There are a total of 181 *procA* genes in this set of 9 genomes.

To assess the potential structural diversity of lanthipeptides encoded by these genomes, we created an amino acid sequence similarity matrix for the leader and core peptide regions of the prochlorosin precursor peptides (Fig. 2-1A). Strikingly, with the exception of one pair of precursor peptides from MIT9313 and MIT1327 that display 100% identity in their core region, all of the core regions in the dataset display less than 30% similarity. In contrast, the leader peptide region displays a high degree of inter- and intra-genome conservation, indicating that these precursor peptides are potential clients of their cognate lanthionine synthetases. Importantly, amidst the massive sequence diversity, the core peptides are highly enriched in Cys, Ser and Thr residues (Fig. 2-1B), and 91% of them

have the potential to result in a cyclic peptide. Remarkably, the variation in the precursor peptide is highly concentrated in the core peptide region; the amino acid diversity index is maximal in the positions of the precursor peptide following the Gly-Gly motif of proteolytic cleavage, meaning that all of the 20 amino acids are equally represented at different sites of the core peptide (Fig. 2-1C).

Interestingly, when searching for *procA* homologs we found high identity hits in the genome that did not correspond to coding regions as predicted by the genome annotation. These loci were found to be *procA* pseudogenes, which are otherwise structurally correct prochlorosin precursor peptide genes that contain a mutation that prevents the production of a functional polypeptide. The most common example is a non-sense mutation that introduces an early stop codon in the leader peptide region, but frame-shift mutations and substitutions that eliminate the start codon were also found (Fig. S2-3A). Putative core regions of pseudogenes have no sequence similarity with the core regions of other *procA* genes, and are also enriched in Cys, Ser and Thr, which indicates that before the inactivating mutations these genes could have encoded substrates for the creation of prochlorosins.

The genome of *Synechococcus* S9509 harbors 30 different *procA* pseudogenes that alongside the 80 *procA* genes suggest that this strain underwent a massive gene expansion and diversification that resulted in the capacity of producing more than one hundred different cyclic products. Importantly, *procA* pseudogenes are also found in some genomes of LL-IV *Prochlorococcus* that do not encode prochlorosin lanthionine synthetase (Table S2-1). In one of the few examples where we could identify a conserved *procA*-harboring genomic locus across multiple strains, we observe *procA* gene decay in the absence of the rest of the prochlorosin biosynthesis genes (*lanT* and *procM*), which confirms the evolutionary interdependence between the components of the prochlorosin biosynthesis pathway (Fig. S2-3B). The high frequency of *procA* pseudogenes in prochlorosin-encoding genomes is suggestive of a high turnover rate of paralogous alleles and denotes that the prochlorosin production is a highly dynamic

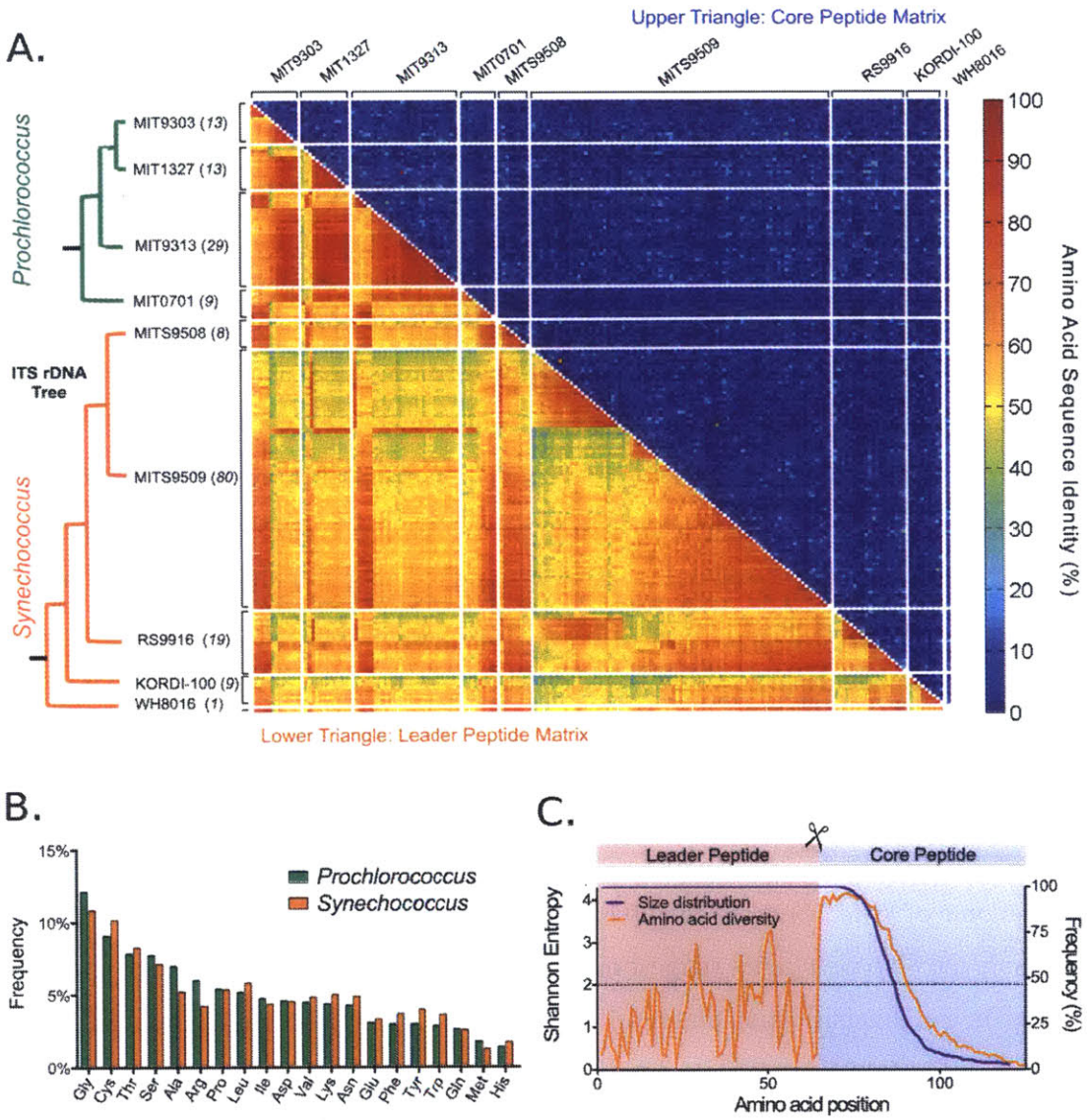
A.
Upper Triangle: Core Peptide Matrix
MIT9303
MIT1327
MIT9313
MIT0701
MITS9508
MITS9509
RS9916
KORDI-100
WH8016
Prochlorococcus
MIT9303 (13)
MIT1327 (13)
MIT9313 (29)
MIT0701 (9)
ITS rDNA Tree
Synechococcus
MITS9508 (8)
MITS9509 (80)
RS9916 (19)
KORDI-100 (9)
WH8016 (1)
Lower Triangle: Leader Peptide Matrix
Amino Acid Sequence Identity (%)
100
90
80
70
60
50
40
30
20
10
0
B.
Prochlorococcus
Synechococcus
Frequency
15%
10%
5%
0%
Gly
Cys
Thr
Ser
Ala
Arg
Pro
Leu
Ile
Asp
Val
Lys
Asn
Glu
Phe
Tyr
Trp
Gln
Met
His
C.
Leader Peptide
Core Peptide
Size distribution
Amino acid diversity
Shannon Entropy
Frequency (%)
Amino acid position
0
50
100
1
2
3
4
25
75

Figure 2-1 Genetic variation of the prochlorosin precursor peptide.
A. The matrix displays the percentage of amino acid identity of each pairwise comparison of the leader peptide region (lower triangle) and of the core peptide region (upper triangle) of the prochlorosin precursor peptides found in 9 different lanthipeptide-encoding strains of *Prochlorococcus* and *Synechococcus*. The precursor peptides in the matrix were grouped first by the phylogenetic relationship of the strain of origin (based on the Internal Transcribed Spacer (ITS) ribosomal DNA, left tree) and subsequently by the intra-genome amino acid identity of their leader peptide region. **B.** Amino acid composition of the prochlorosin core peptides found in *Prochlorococcus* and *Synechococcus* genomes reveal enrichment in amino acids necessary for the formation of ring structures (Cys, Ser and Thr). **C.** Amino acid diversity along the prochlorosin precursor peptide. Positions within the leader peptide region correspond to the consensus of the multiple sequence alignment of 181 leader peptides. Because the core region cannot be aligned, the positions for this region correspond to stacked sequences that use the Gly-Gly motif as a starting reference. The Shannon entropy values higher than 2 are considered variable. The maximum theoretical value of the Shannon entropy for 20 amino acids is 4.322 and signifies that all residues are equally represented in the position (Litwin and Jores, 1992).

genetic trait where expansion, diversification and elimination of *procA* genes is an evolutionary feature of lanthipeptide-encoding genomes.

Sequence analysis of the *procA* genes from these picocyanobacterial genomes alone does not provide a clear picture of how the expansion and diversification happens. The variation in the core peptide regions is so extreme that no meaningful sequence alignments can be obtained for the 3' end of the gene, hindering the reconstruction of a robust phylogeny for the whole precursor peptide that could inform the mechanisms of diversification of the genes. Using only the leader peptide region, however, one can generate a partial reconstruction of the phylogeny of these genes (Fig. S2-4), where the clade of *Prochlorococcus* leader peptides stem from a distinct branch of *Synechococcus* leader peptides, which is consistent with the general phylogeny of these two genera. The difference in the genetic variability of the leader and core peptide regions suggests an atypical molecular evolution mechanism for the *procA* gene that invites further exploration.

Lanthipeptides in Natural Populations of Marine Picocyanobacteria

The lanthipeptide production trait in marine picocyanobacterial cultures has genomic plasticity and structural diversity features not present in any other family of lanthipeptides. In only 9 genomes, we found 181 *procA* genes – an exceptional number by any measure – and only two of these would produce the same lanthipeptide structure. The suite of 9 was isolated from diverse oceanic locations (Table S2-1), indicating that the prochlorosin trait is broadly distributed in the oceans. One wonders then, how diverse are prochlorosins in the wild, and what is their distribution and abundance across oceanic environmental gradients? More importantly, what are the molecular evolutionary mechanisms that drive their diversification? The phylogenetic resolution afforded by cultivation-dependent discovery of *procA* genes does not enable a clear reconstruction of the molecular changes that give rise to new prochlorosins, therefore the investigation of the genetic variation patterns of precursor peptide genes from natural populations can provide insights into the evolutionary mechanisms of this family of genes.

To address these questions, we employed a biogeographic approach to identify distinct populations of marine picocyanobacteria that harbor the prochlorosin trait and characterize its distribution and abundance in different regions of the ocean. Also, we used high-throughput metagenomics to explore the diversity of prochlorosins in populations that display differences in their environmental distribution of prochlorosin biosynthesis genes. Using sequence information derived from this population level diversity survey, we infer the forces that shape the evolutionary diversification in the prochlorosins.

Distribution and Abundance of Prochlorosin Precursor Peptide Genes in Wild Picocyanobacterial Populations

We used quantitative PCR to measure the abundance of the prochlorosin precursor peptide and the prochlorosin lanthionine synthetase genes in samples collected at Station ALOHA in the North Pacific Subtropical Gyre and along 16 stations of a meridional transect in the Atlantic Ocean (AMT) that encompass multiple environmental gradients (Fig. 2-2A) (Johnson et al., 2006; Malmstrom et al., 2010). The primers targeting the prochlorosin precursor peptide genes were designed on regions of the leader peptide that

are conserved in *procA* alleles from both *Prochlorococcus* and *Synechococcus*. For the lanthionine synthetase gene (*procM)*, the set of primers targets a region conserved only in alleles from *Prochlorococcus* LL-IV strains; it is not able to detect *Synechococcus procM* alleles because of their high divergence at the nucleotide level. Thus while our PCR assay was able to identify prochlorosin precursor peptide genes from both *Prochlorococcus* and *Synechococcus* populations, we can only report data on prochlorosin lanthionine synthetase from *Prochlorococcus* LL-IV-like alleles.

The prochlorosin trait is widespread but patchily distributed in the ocean (Fig. 2-2B); the *procA* and *procM* genes display variable abundances throughout the euphotic zone and across multiple oceanic environments. Depending on the oceanic region, the prochlorosin precursor peptide genes (*procA)* and the lanthionine synthetase genes (*procM*) display distinctive distribution patterns that emerge, not surprisingly, as a result of the differences in the structuring of the marine picocyanobacterial populations (Fig. 2-2C). For instance, the distribution pattern at two subtropical locations such as Station ALOHA and AMT-S25 –known to be dominated by high-light adapted ecotypes of *Prochlorococcus* (which don't encode prochlorosins) and where *Synechococcus* is prevalent (Malmstrom et al., 2010; Zwirglmaier et al., 2007) – shows a low abundance of *procA* genes and the absence LL-IV-like *procM* genes throughout the mixed layer. Conversely, the abundance of *procA* and LL-IV-like *procM* genes increase in well-stratified waters below the mixed layer (dashed horizontal line (Fig. 2-2C) where LL-adapted ecotypes of *Prochlorococcus* (including LL-IV's, which harbor prochlorosin genes) are known to thrive (Malmstrom et al., 2010). The only instance in which LL-IV-like *procM* alleles were found in surface waters is when deeply mixed surface waters layers bring LL-adapted ecotypes of *Prochlorococcus* to the surface (i.e AMT-S72). In contrast to these subtropical waters, at low or high latitudes of the Atlantic Ocean (AMT-S78 and AMT-S3, respectively) where the picocyanobacterial population is dominated by *Synechococcus* and LL-adapted *Prochlorococcus* are relatively rare (Johnson et al., 2006), the *procA* alleles are abundant in the surface waters and the LL-IV-like *procM* alleles are less abundant throughout the water column. Recall that our *procM* primers will not detect this gene if it is in *Synechococccus,* so the disparity between *procA* and *procM* abundance at these *Synechococcus*-dominated sites is to be expected. In general, however, the widespread

distribution and abundance of the prochlorosin production trait in natural populations lends support to the functional relevance of lanthipeptides in the marine environment.

Despite that the fraction of prochlorosin-encoding cells is relatively low, the number of prochlorosin precursor peptide genes in the water column can reach high concentrations. For instance, the highest concentration of the LL-IV-like *procM* gene was 10^3 copies/ml. If we assume a single copy of the *procM* gene per cell, in this case the number of prochlorosin-producing cells only account for 1 to 5% of the total picocyanobacterial population. However, due to the multi-copy nature of the prochlorosin precursor peptide gene (*procA)*, its concentration may range from 10 to 10^4 copies/ml of seawater; an abundance that in some instances can match up or even exceed the abundance of the total marine picocyanobacterial population, i.e. between 50 and 100m depths of AMT-S25 (Fig. 2-2C). Moreover, this high resolution mapping of the prochlorosin precursor peptide genes enabled us to identify marine picocyanobacterial populations that display different distribution patterns of the prochlorosin trait so we can explore the diversity of prochlorosin structures in these distinct populations.

Global Comparative Analysis of Prochlorosin Populations

Given the startling variation found in the core region of the prochlorosins of different cultured strains of *Prochlorococcus* and *Synechococcus*, we next sought to investigate the diversity of prochlorosins in wild populations using a *procA*-targeted metagenomic approach. This is challenging given the hypervariability of the 3' end of the *procA* gene, making it impossible to anchor a primer that would enable the amplification of the entire coding sequence. Close inspection of the *procA* loci in our genomes, however, revealed a moderately conserved 3' intergenic region downstream of the *procA* gene that can serve as a primer-anchoring site and enable amplification of the coding region of the procA gene (Fig. S2-5). While this set of primers is not likely to amplify all the possible *procA* alleles, it can provide a lower bound estimate of the diversity of *procA* genes and facilitate the comparison of prochlorosin populations from different natural environments.

For this analysis, we selected 4-5 representative depths of the five oceanic regions that in our previous biogeographic analysis displayed distinct patterns in the distribution of the

procA and *procM* genes (Fig. 2-2C). In addition, to examine the temporal variation of prochlorosin populations from the same picocyanobacterial population, we used 9 monthly samples (June, 2005 to February, 2006) for the 125m and 175m depths of station ALOHA that were collected as part of the Hawaii Ocean Time Series (HOTS) program (Karl and Church, 2014; Malmstrom et al., 2010). We amplified the *procA* locus from a total of 39 different environmental DNA samples and the amplicons were sequenced using the MiSeq Illumina platform. The resulting *procA* reads were quality filtered and processed to remove possible chimeric products (see Methods). To determine the total number of different prochlorosin precursor peptides in the dataset, the *procA* open reading frames (ORF) were predicted from the cleaned reads from all samples and clustered at a genetic distance of 3% using the Uclust algorithm (Edgar, 2010) to account for possible amplification or sequencing errors. Each unique 97% identity cluster (excluding singletons) was considered to be an Operational Prochlorosin Unit (OPU) representing a unique prochlorosin precursor peptide from which the sequence information of the leader and core peptide regions can be obtained. After the clustering, a total of 1697 OPUs were identified and the cleaned *procA* ORF sequences from each site sampled were mapped back to the total OPU set to create a matrix of the OPU composition at each site. The number of OPUs found here is quite astounding, given that it represents a lower bound estimate of prochlorosin diversity because of the primer sites we used.

Rarefaction analysis of 1.6 million *procA* ORF sequence observations from all samples shows that the number of OPUs sampled does not tend to an asymptote (dashed line, Fig. 2-2D). Rarefaction curves for individual sites, however, show that the rate of discovery of new OPUs in most of the oceanic sites significantly decreases as the sampling effort reaches its maximum, which suggests that with the current sequence coverage we are sampling the majority of OPUs that can be detected with our primer set in each site (colored lines, Fig. 2-2D). The differences in the trends for total OPU's vs. site specific OPU's could reflect differences in the composition of OPU populations between samples. To explore this possibility we constructed a dissimilarity matrix based on the OPU composition at each site using the Jaccard distance and employed Principal Coordinate Analysis (PCoA) to represent the similarity between samples in ordination space. In

addition, Analysis of Similarity (ANOSIM) was used to assess the statistical significance of the observed Jaccard distances between samples (Fig. 2-2E).

Prochlorosin populations cluster mainly by geographic region (see color patterns Fig. 2-2E). All samples originating from the Pacific Ocean (ALOHA H175, HOTS-125m and HOTS175m) form a group distinct from all the samples from the Atlantic Ocean (ANOSIM, R=0.86-1; p-value <0.01). Within the samples from station ALOHA, there is a significant overlap between the OPU populations from the time series samples at 125m and 175m depths (ANOSIM, R=0.24; p-value<0.05) indicating temporal stability in prochlorosin populations from these depths. Despite the overall cohesion of prochlorosin populations from the Pacific Ocean samples, we observed a few distant samples from the main cluster that corresponded to the 25m, 85m and 100m depths and to samples of three months in the fall and one month in the winter of the 125m series, which suggests that environmental gradients along the water column and seasonal changes can also influence the prochlorosin population composition at station ALOHA. In the Atlantic Ocean, the samples from South Atlantic (AMT-S72 and AMT-S78), where *Synechococcus* represents the majority of the population, show some similarity in their OPU populations (ANOSIM, R=0.67; p-value <0.05) and form a close cluster in the PCoA. By contrast, the samples from stations of the North Atlantic (AMT-S3) and the subtropical Atlantic (AMT-S25) form two additional distinct clusters for which there is very low overlap between prochlorosin populations (ANOSIM, R=0.88-1; p-value <0.05). In summary, it appears that picocyanobacterial populations from different oceanic environments harbor significantly different sets of prochlorosin precursor peptide genes.

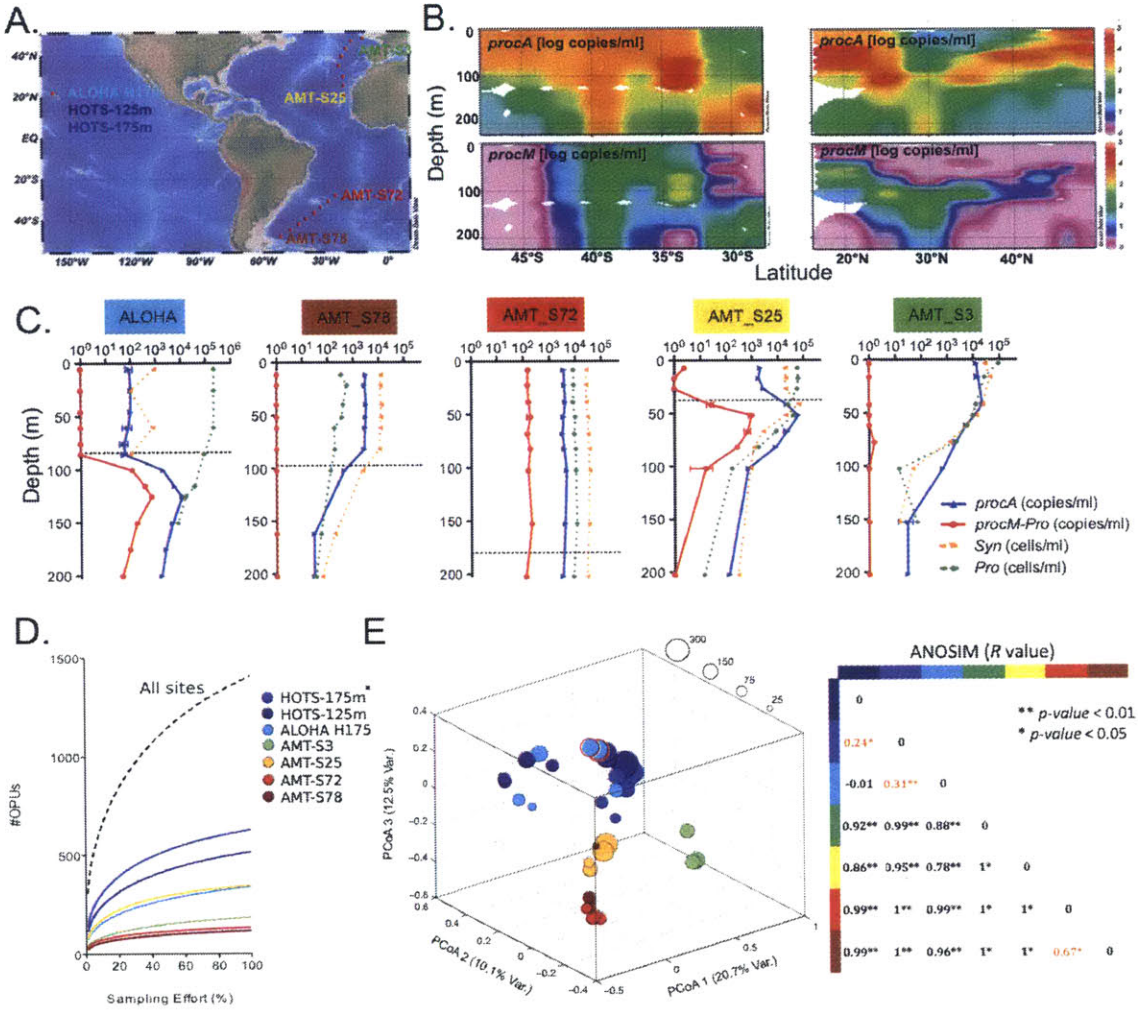

Figure 2-2. Biogeographic analysis of prochlorosin distribution and diversity in the ocean

A. Map of the oceanic sites used for this study. Red dot in the Pacific Ocean denotes the location of Station Aloha and the red dots in the Atlantic Ocean denote the stations of the AMT13 transect. Color labels correspond to stations that were analyzed in the global diversity analysis. **B.** Environmental distribution of *procA* and LL-IV-like *procM* genes along the first 200 m of the water column along Northern and Southern sections of the AMT13 transect. Color-scale indicates the logarithm of concentration in copies/ml. **C.** Distinct depth profiles of the abundance of *procA* (blue) and LL-IV-like *procM* (red) genes, and flow cytometry cell counts of *Prochlorococcus* (green) and *Synechococcus* (orange) along the water column in five oceanic sites. Dashed line indicates the depth of

the mixed layer, as determined by potential density differential from surface values by >0.125 kg m^{-3}. Mixed layer data not available for AMT-S3. **D.** OPU Rarefaction curves for each one of the 5 oceanic sites analyzed: AMT-S3 (9, 54, 79 and 104m) AMT-S25 (17, 53, 83 and 103m) AMT-S72 (9, 53, 74 and 154m) AMT-S78 (10, 50, 80 and 160m) and ALOHA (25, 85, 100, 125 and 175m), and two monthly time-series samples, HOTS-125m (June 2005-Feb 2006) and HOTS-175m (June 2005-Feb 2006). **E.** OPU clustering on the basis of a Jaccard dissimilarity matrix (visualized by Principal Coordinates Analysis). The size of the circle is proportional to the richness of OLUs in each sample. Three ALOHA H175- 125m replicate control samples are circled red. ANOSIM R=1 (complete dissimilarity), R=*0* (complete evenness). Values highlighted in orange correspond to samples that displayed significant overlap in their OPU populations.

Structural Diversity in Prochlorosins from Natural Populations

To explore in more detail differences among the 1697 OPUs we have identified, we calculated the frequency distribution of nucleotide identities among pair-wise comparisons. The vast majority of the OPUs fall in the range of 40 to 70% identity, indicating a great extent of diversity (Fig. 2-3A, inset). To characterize where in the prochlorosin precursor peptide the sequence diversity is concentrated, we translated the OPUs into protein and analyzed the identity of their leader peptide and core peptide regions. Similar to the sequence variation observed in the prochlorosin precursor peptides identified in the genomes, the vast majority of the core regions share less than 30% amino acid identity while the leader peptide region displays a greater degree of conservation (Fig. 2-3A).

The exceptional degree of variation in the core peptide region poses the question of whether the core regions of these OPUs possess structural properties that could result in a cyclic peptide. To investigate this, we determined the dipeptide composition of the core regions of the entire set of OPUs with the objective of identifying the most frequent amino acid associations that might represent underlying structural commonalities between peptides. The most frequent dipeptides are composed of a polar amino acid involved in cyclization (Cys, Ser and Thr) and a non-polar amino acid with a small side chain (Gly, Ala and Val) (Fig. 2-3B). Notably, the fact that core regions of the OPUs are enriched with residues involved in the formation of lanthionine bridges and that these are flanked by small amino acids has important functional implications for biosynthesis, as less bulky side chains pose a minimal energy barrier at the moment of cyclization. This amino acid sequence composition also represents a structural hallmark of other families of RiPPs (Ennahar et al., 2000; Haft et al., 2010; Jack et al., 1995). Further, the dipeptide signature is also present in the core peptide region of the prochlorosin precursor peptides from cultured genomes, and is different from the dipeptide frequency observed for the general proteome of prochlorosin-encoding strains (Fig. S2-6). Consequently, although there is no overall large sequence conservation between most of the core peptides of the OPUs, they display the general structural signatures of cyclic peptides.

Together with the genomic analysis of prochlorosin-encoding cultured strains, the results from the biogeographic approach demonstrate that the hypervariability of the prochlorosins is a feature that happens across scales of complexity. At the genome level, all of the *procA* genes found within a genome are different. At the population level, prochlorosin-encoding genomes from different picocyanobacterial clades harbor distinctive sets of *procA* genes, and in the global ocean, different natural marine picocyanobacterial populations contain largely dissimilar collections of *procA* genes. This observation is indicative that lanthipeptides in marine picocyanobacteria are undergoing a process of evolutionary diversification unprecedented in any other family of RiPPs.

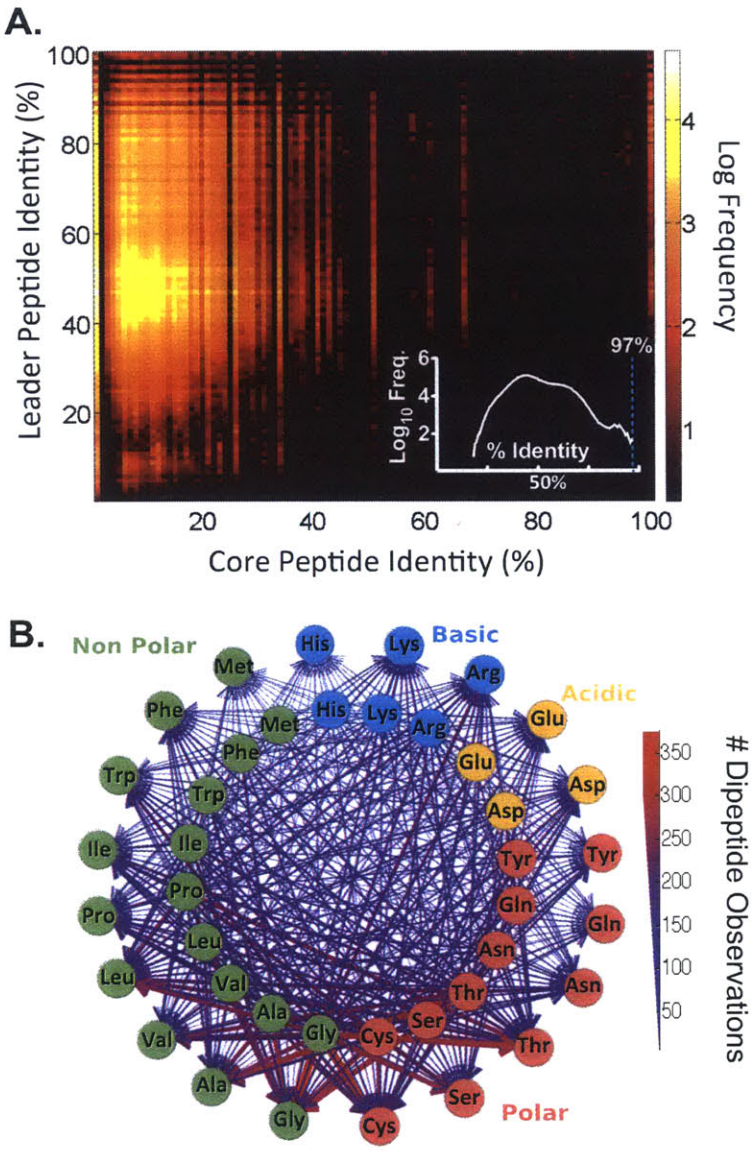

Figure 2-3. Characterization of the sequence diversity of OPUs
A. (Inset) Histogram of the nucleotide similarity of every pairwise comparison of 1697 OPUs. The blue dash line represents the 3% distance cutoff used in the *procA* ORF clustering. (Outset) Heat map of the amino acid identity between every pairwise OPU comparison for the core peptide and leader peptide region. Note the low frequency of core peptide regions with more than 30% amino acid identity between them **B.** Dipeptide composition of the prochlorosin core peptides identified in the entire set of OPUs. Inner and outer circles correspond to the N-terminal and C-terminal amino acid positions, respectively. Color and width of the scale is weighted for the number of dipeptide observations in the dataset.

Molecular Diversification of Prochlorosin Precursor Peptides

How do *procA* genes diversify to produce such amount of different lanthipeptide structures? As mentioned above, because *procA* genes from the genomes share little identity, we cannot establish the phylogenetic relationships necessary to infer the natural selection forces operating in this family of genes. However, having a large sample of *procA* genes in the form of OPUs from natural populations gives the opportunity to identify early steps in the diversification of the core peptide region and to characterize the types of molecular changes that give rise to new lanthipeptide structures. To this end we searched for OPUs that share 100% identity in their leader peptide region to find groups of genes that might have undergone recent changes in their core peptide region. We identified 156 of these groups, 8 of which contained more than 7 OPUs; the core peptide regions within groups were 0 to 88% identical – a significant range of variation considering they have identical leader peptides (Fig. 2-4A). To explore further the variation at the 3' end of the *procA* locus in these groups we recovered the sequence information of the 3' intergenic region downstream the *procA* ORFs that had been removed as part of the read processing and clustering procedure. Multiple sequence alignments of the full length of the *procA* locus revealed that the 3' intergenic region downstream of the ORF, like the leader peptide, is highly conserved, further suggesting that the changes in the core peptide region of these *procA* genes correspond to recent events.

Using the sequence information from the entire *procA* locus, it is possible to examine the molecular changes that gave origin to new *procA* genes in these closely related groups. Examining 9 *procA* loci of the G3 group of recently diverged OPUs, for example, we were able to deduce that the OPU116 is the ancestor of 8 other OPUs that arose as the result of individual deletion events in different parts of the 3' end of the locus (Fig. 2-4B). Notably, 5 out of the 8 deletions events comprised several nucleotides outside of the coding region of the original OPU116. Accordingly, compared to their ancestor OPU116, every single one of these deletions resulted in *procA* genes with core regions displaying large differences in their sequence composition, but that nonetheless still contain Cys, Ser and Thr residues in different positions that could serve as substrates for the creation of

lanthionine bridges (Fig. 2-4B). This represents a striking example of efficiency in sequence space exploration, where a single *procA* locus was rapidly diversified into 8 prochlorosin precursor peptide genes whose products would display completely different ring topologies.

In general, observed sequence variation patterns in groups of closely related OPUs are dominated by multiple insertion-deletion (indel) events of variable lengths, indicating that the majority of the polymorphisms in the core peptide region of recently diverged *procA* loci resulted immediately in large changes in the amino acid sequence of the final product. In contrast, when we looked for groups of OPUs that have 100% identity in their core peptide region, we found that the variation in the leader peptide region is dominated by single nucleotide substitutions, a large fraction of them, of synonymous nature (Fig. S2-7). Thus, the variation patterns observed in the OPUs indicate that the leader and core peptide regions of the prochlorosin precursor peptide gene are undergoing two different processes of molecular evolution. The leader peptide, whose function is to direct biosynthesis, follows the molecular evolution pattern of most proteins where large sequence polymorphisms are thought to be detrimental to the protein structure and therefore only small changes in the exploration of the sequence space are allowed. On the other hand, the core peptide region evolves mainly by the action of large sequence polymorphisms that greatly affect the overall composition of the core peptide region, disregarding any conservation of the ancestral structure. This pattern of variation implies that the observed diversity of the core peptide region of the prochlorosins is not likely to be the result of a slow step-wise exploration of the sequence space around a particular ring topology, but rather the result of a rapid diversification process that explores drastic changes in the sequence composition.

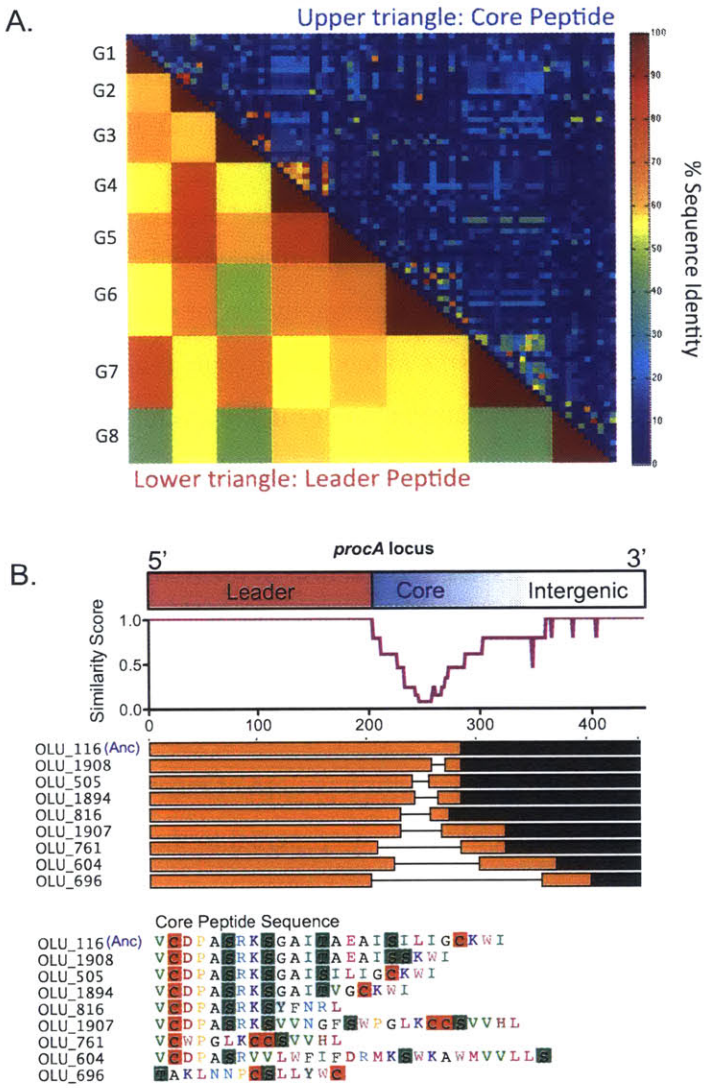

Figure 2-4. Rapid diversification of the *procA* locus.
A. Amino acid identity matrix of the core and peptide regions of 8 groups (G1-8) of OPUs that display complete conservation of their leader peptide sequence. **B.** Genetic variation in the *procA* locus of the G3 group of OPUs. Multiple sequence alignment of the *procA* locus indicates the presence of deletion events that occurred on the OPU116 ancestor (Anc). Orange and black regions in the alignment represent *procA* coding and intergenic regions, respectively. The predicted core peptide regions resulting of the diversification of this locus are presented below the alignment. Cys, Ser and Thr are highlighted to note that the given the positions of these residues, each peptide would display a different ring topology.

In the absence of a selective pressure, the diversification process of *procA* genes would create random peptides. This would result in a low proportion of peptides carrying all the residues required for cyclization, and we would expect to see random dipeptide signatures. However, 91% of the prochlorosin core peptides identified in the genomes and 81% of the core peptides in the OPUs contain the residues required for cyclization and the structural features that favor the formation of a cyclic product. Therefore, the observed variation in the core peptides is not likely to be the result of genetic drift.

Consequently, in our population-level analysis of precursor peptide genes, the predominance of mutations that dramatically affect the sequence composition of the core peptide region might be indicative of a diversifying selection process in which large sequence changes in the core regions (commonly observed in the population) are favored over small sequence changes (rarely observed). This might be an indication that the trait that is being selected in the prochlorosins is the capacity to rapidly evolve new lanthipeptide structures.

Evolutionary Pressure on the Substrate Tolerance of the Prochlorosin Lanthionine Synthetase

To study the evolutionary interplay between the lanthionine synthetases and their peptide substrates, we compared the evolutionary history of LanM enzymes from firmicutes, which usually modify a single precursor peptide substrate (Zhang et al., 2012), to the evolutionary history of the ProcM enzymes of marine picocyanobacteria for which there are multiple cognate precursor peptides (Fig. 2-5A). Does the rapid diversification of the prochlorosin core peptide substrates demand a fast evolution of the lanthionine synthetase? Surprisingly, the enzymes from the *procM* clade display shorter inter-clade evolutionary distances than the enzymes of the *lanM* clade of firmicutes, suggesting that for the ProcM clade there is no correlation between faster evolutionary rates and the number of cognate substrates. More likely, the differences in the evolutionary history between lanthionine synthetases from firmicutes and marine picocyanobacteria might be due to differences in the selection pressures that drive the evolution of new lanthipeptides in these two clades.

To detect patterns of selection in the molecular evolution in the lanthionine synthetase of the ProcM clade of marine picocyanobacteria and of the LanM clade of firmicutes, we determined the ratio of the rates of non-synonymous substitutions and synonymous substitutions (dN/dS) within enzymes of the same clade (Fig. 2-5B). Interestingly, the genes from ProcM clade display very low dN/dS ratios, which suggest the prochlorosin lanthionine synthetase is evolving under purifying selection. In contrast, the genes from the *lanM* clade display a higher number in non-synonymous substitutions suggesting that in this group of single-substrate lanthionine synthetases the evolutionary transition to a different substrate entails larger structural changes. Consequently, ProcM seems to be evolutionarily locked in a state that favors the maintenance of substrate promiscuity where the constant diversification of the *procA* substrates does not impose a commensurate change in the enzyme. In this scenario, the evolution of new prochlorosin structures is not constrained by the need of large changes in the lanthionine synthetase. An outstanding example of this are the *Prochlorococcus* strains MIT9313 and MIT9303, which encode vastly different sets of prochlorosin precursor peptides, however their *procM* genes are 96% identical, indicating that the ProcM lanthionine synthetases does not require large structural changes to be able to accommodate new precursor peptide substrates.

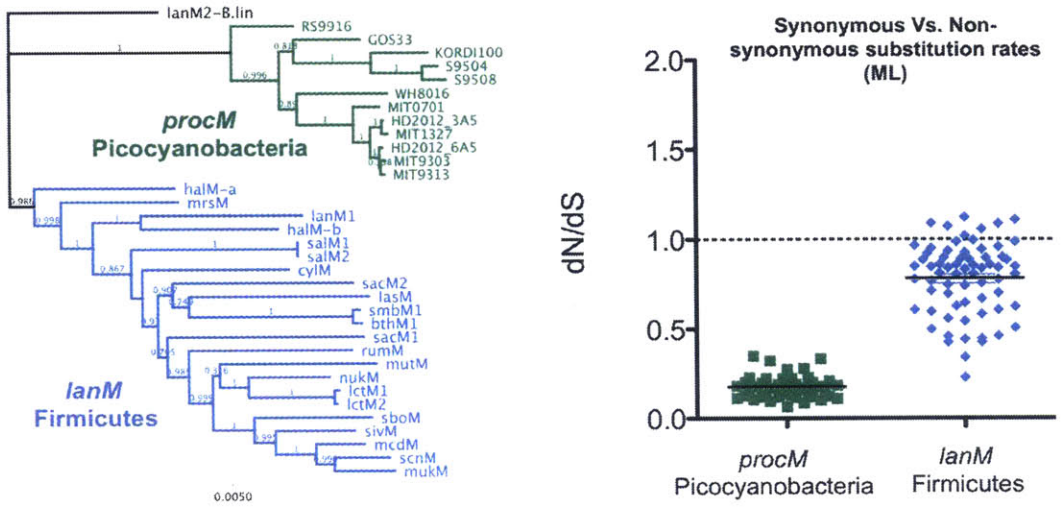

Figure 2-5. Natural Selection patterns in the prochlorosin lanthionine synthetase **A.** Maximum Likelihood Phylogenetic tree of the *procM* and *lanM* clades of lanthionine synthetases. **B.** Ratios of synonymous vs. non-synonymous substitution rates for every pairwise intra-lade comparison for *procM* and *lanM* genes.

Structural diversity is under selection in picocyanobacterial lanthipeptides

Most of our understanding of the evolution of lanthipeptide structures comes from studies of lantibiotics, lanthipeptides that display antimicrobial activity (Knerr and van der Donk, 2012). In these molecules, the ring topology resulting from the dehydration and cyclization of Cys, Ser and Thr residues located in specific positions of the core peptide is thought to be the major determinant of the bioactivity. For instance, in lantibiotics of the Nisin group, which target the lipid II of gram-positive bacteria, the variation found their core peptide sequences is mainly composed of small changes that revolve around the principal rings. In these peptides amino acid changes that do not have a major effect in the overall structure are tolerated without affecting the antimicrobial activity (Cotter et al., 2005; Zhang et al., 2012). Consequently, lantibiotics follow the evolutionary patterns of a protein with an established structure-function relationship, where structural similarity between variants is under selection. The case of prochlorosins seems to be different: the predominance in the population of prochlorosins that would display extensively different

ring topologies and the high frequency of large sequence polymorphisms that tend to change the structure of the core peptide, suggest that the potential bioactivity of the prochlorosin does not hold a relationship with one particular ring topology. Consequently, the evolutionary pattern of the prochlorosins indicates that the lanthipeptide production trait in marine picocyanobacteria might find its evolutionary advantage in the plasticity of the production of multiple cyclic peptides with diverse ring topologies.

CONCLUSION

Using a biogeographic approach we demonstrated that the genes involved in the biosynthesis of prochlorosins are found across different environmental gradients and throughout seasonal timescales, which supports the functional relevance of lanthipeptide biosynthesis in natural populations of picocyanobacteria. Our cross-scale analysis of the diversity of prochlorosin precursor peptide genes revealed that lanthipeptides in this group of marine microorganism are undergoing a process of evolutionary radiation. To achieve this, the prochlorosin biosynthesis pathway has evolved the capacity of generating of a suite of structurally diverse cyclic peptides by employing a combination of a substrate-tolerant lanthionine synthetase and a set of precursor peptide substrates that are encoded in a highly dynamic family of multi-copy genes that is poised for rapid expansion and diversification. The evolutionary strategy of the prochlorosins contrasts with the canonical lantibiotic biosynthesis pathways that evolve towards the efficient creation of one cyclic peptide with a defined ring topology.

MATERIALS AND METHODS

Prochlorococcus Isolations

Seawater was collected from the deep chlorophyll maximum (150 m) in the North Pacific Ocean at Station ALOHA (22.75°N, 158°W) in June 2013 as part of the Hawai'i Ocean Experiment - Phosphorous Rally HOE-PhoR research cruise (http://hahana.soest.hawaii.edu/hoephor/hoephor.html). *Prochlorococcus* isolations were performed as described in (Moore et al., 2007) with slight modifications. Briefly, seawater was gravity-filtered through a 1μm Nucleopore filter and the filtrate was amended with Pro2 medium nutrients (Moore et al., 2007), a related mix replacing the ammonia and urea N-sources with an equivalent concentration of nitrite, and in some cases supplementing with the hydrogen peroxide scavenging agent 1μM thiosulfate. The enrichments were kept under dim light conditions (1-3 $\mu E/m^2s$) until their arrival to the lab, where they were maintained at 24°C or 21°C under continuous light or diel conditions (1 $\mu E/m^2s$). The presence of *Prochlorococcus* was monitored with flow cytometry (BD/Cytopeia Influx), and samples containing *Prochlorococcus* were transferred periodically to fresh seawater-based media matching the original amendments. Dilution to extinction isolations for two enrichments containing consistently growing *Prochlorococcus* were performed in ProMM medium as described before (Berube et al.). Cell growth in the 96-well plates was monitored by eye and by fluorometry using a Synergy HT Microplate Reader (BioTek, Winooski, VT, USA). Wells that showed significant growth from the dilution experiment were scaled up in Pro2 and Pro99 media for subsequent identification, genome sequencing, cryopreservation and are maintained in batch culture in Pro99. Strain purity was confirmed via flow cytometry and a series of three purity test broths: ProMM, ProAC, and MPTB (Moore et al., 2007). These isolation experiments yielded 9 new *Prochlorococcus* strains with names beginning with MIT13XX (Table S1).

Genome Sequencing

DNA from 20 ml *Prochlorococcus* or *Synechococcus* cultures was isolated using the QIAamp DNA mini kit (Qiagen). Genomic DNA was sequenced using the MiSeq

Illumina Platform at the BioMicro Center of MIT using the protocol previously described in Biller *et al.* 2014 (Biller et al., 2014). Low quality regions of the sequencing reads were removed from the raw Illumina data using the quality_trim tool from CLC Assembly Cell package with the default settings in which at least 50% of the read must have a minimum quality score of 20 (V3.2, CLC bio). Draft genome sequences of MITS9504, MITS9508 and MITS9509 were assembled using the CLC Genomics Workbench (V3.2, from the CLC Assembly Cell package; CLC bio). Draft genome sequences of the strains MIT13XX were assembled using Spades v3.1.1 (Bankevich et al., 2012). Average coverage for the genome was higher than 50X. Cultures of strains MIT1313, MITS9504, MITS9508 and MITS9509 were non-axenic therefore contigs with a best match to a non-*Prochlorococcus* genome in BLASTN searches against the NCBI nt database were removed from the assembly. Contigs smaller than 500bp were also removed. The assembled contigs for each genome were annotated using the RAST and Prokka software (Aziz et al., 2008; Seemann, 2014).

Quantitative PCR Assay for *procA* and *procM* Genes

Samples used for the qPCR enumeration of the *procA* and *procM* genes were collected previously as part two biogeographic studies of *Prochlorococcus* in the Atlantic as part of the Atlantic Meridional Transect (AMT13) (Johnson et al., 2006) and in the Pacific Ocean as part of the Hawai'i Ocean Time Series program (HOTS) (Karl and Church, 2014; Malmstrom et al., 2010). Briefly, 100ml of seawater were filtered through a 25mm 0.2 μm polycarbonate filter using gentle vacuum (<10in. Hg), followed by 3ml of preservation solution (10 mM Tris (pH 8.0), 100 mM EDTA, 0.5 M NaCl). Cells from filters were resuspended in 650 μl of lysis buffer (10 μM Tris-HCl pH 8) in a beadbeater at 4800RPM for 2 minutes. Cells were lysed by incubating the tubes at 95°C for 15 minutes. Crude cell lysates were aliquoted and frozen at -80°C for later use. For the PCR quantification of the number *procA* and *procM* copies at each depth, cell lysates from duplicate filters of the same seawater sample were used as DNA template for duplicate PCR reactions. Standards for absolute quantification were processed and analyzed as described previously (Johnson et al., 2006; Malmstrom et al., 2010). The amplification for each sample was performed in a final volume of 15 μl in reactions consisting of 6 μl

of crude cell lysate, 1x QuantiTect SYBR Green PCR Mix (Qiagen, Germantown, MD, USA) and 1 μmol/L of each forward and reverse primer. Reactions were pre-incubated at 95°C for 15 min to activate the polymerase and then cycled 40 times at 95°C for 15 s, 55°C for 15 s, and 72°C for 30 s. Primer sequences for the *procA* genes were: procA-LP-Up 5'ATGTCAGAAGAACAACTCAA and procA-LP112Lo 5'CAGCAGCTTTGGCAAT. Primer sequences for the *Prochlorococcus* LL-IV-like *procM* genes were: procM-Nter-Up 5'ATTTGCTGATCCCAATT and procM-Nter-Lo 5'CTCCATATAGCCATAGCC. Melting curve analysis was used to assess specificity in each reaction as previously described (Malmstrom et al., 2010).

procA Locus Amplicon Libraries

The amplification for each one of the 39 samples selected from the Atlantic and the Pacific was performed in a final volume of 15 μl in reactions consisting of 6 μl of crude cell lysate, 1x QuantiTect SYBR Green PCR Mix (Qiagen, Germantown, MD, USA) and 1 μmol/L of each forward and reverse primer. Reactions were pre-incubated at 95°C for 15 min to activate the polymerase and then cycled 40 times at 95°C for 15 s, 55°C for 15 s, and 72°C for 30 s. Primer sequences for the creation of *procA* amplicon libraries were: procA-LP-Up 5'ATGTCAGAAGAACAACTCAA and procA3UTR 5'TTGAAGAAATAAAAAGCCC. The resulting amplicons were purified and MiSeq Illumina libraries were created and barcoded as previously described in Biller *et al.* 2014 (Biller et al., 2014).

Read Processing and Clustering of *procA* amplicons into OPUs

A total of 2.8 million *procA* raw reads were filtered and processed. Low quality regions of sequencing reads and low quality reads were removed from the raw Illumina data using quality_trim from CLC with the default settings in which at least 50% of the read must have a minimum quality of 20 (V3.2, from the CLC Assembly Cell package; CLC bio). We filtered out all the sequences that did not contain the procA-LP-Up and the procA3UTR primers (incomplete sequences, contamination or artifacts). Once selected, the primers were removed and UCHIME was used to remove possible chimeras resulting from the PCR amplification (Edgar et al., 2011). To obtain functional sequence

information about the precursor peptide from the full-length *procA* reads, we predicted the coding region (ORFs) excluding the N-terminal region of the leader peptide that correspond to the binding site of the procA-LP-Up primer. A total of 1.6 million of procA ORF sequences were predicted from the cleaned reads. These reads were then used to create 3% similarity clusters using the cluster_otu function of Usearch (Edgar, 2013). The resulting 3% similarity clusters were mapped back to the original *procA* ORF reads from each one of the sites sampled using Usearch with an identity threshold of 0.97. The resulting non-singleton clusters from this mapping are considered to be Operational Prochlorosin Units (OPUs).

Comparative analysis of OPU populations

Hypervariability in the core peptide region precludes the creation of a robust phylogenetic tree and therefore the use of methods for determining beta diversity that rely on measures of the branch length between taxa are not suitable for the study of diversity of different prochlorosin populations. For this reason we evaluated the differences in the composition of OPU populations by using cluster analysis and ordination (Buttigieg and Ramette, 2014). Calculation of the Jaccard dissimilarity matrix, Principal Coordinates Analysis (PcoA) and hypothesis testing using Analysis of Similarity (ANOSIM) were performed using the Fathom Toolbox for MATLAB. (Jones, D. L. 2014. Fathom Toolbox for Matlab: software for multivariate ecological and oceanographic data analysis. College of Marine Science, University of South Florida, St. Petersburg, FL, USA. Available from: http://www.marine.usf.edu/user/djones/)

Sequence Analyses

Multiple sequence alignments, genome sequence manipulations and phylogenetic trees were performed in the Genious 8 software (Kearse et al., 2012). Dipeptide Frequency Analysis was calculated using the sequence analysis toolbox from MATLAB (MATLAB R2014a, The MathWorks Inc., Natick, MA, 2014). Visualization of the dipeptide network was done using Biolayout Express 3D (Theocharidis et al., 2009). dN/dS Ratios were calculated using codeml from the PAML package (Yang, 2007).

REFERENCES

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., *et al.* (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics *9*, 75.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol *19*, 455-477.

Berube, P.M., Biller, S.J., Kent, A.G., Berta-Thompson, J.W., Roggensack, S.E., Roache-Johnson, K.H., Ackerman, M., Moore, L.R., Meisel, J.D., Sher, D., *et al.* (2015). Physiology and evolution of nitrate acquisition in *Prochlorococcus*. ISME J *9*, 1195-1207.

Biller, S.J., Berube, P.M., Berta-Thompson, J.W., Kelly, L., Roggensack, S.E., Awad, L., Roache-Johnson, K.H., Ding, H., Giovannoni, S.J., Rocap, G., *et al.* (2014). Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. Sci Data *1*, 140034.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015). *Prochlorococcus*: the structure and function of collective diversity. Nat Rev Microbiol *13*, 13-27.

Buttigieg, P.L., and Ramette, A. (2014). A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol Ecol *90*, 543-550.

Chatterjee, C., Paul, M., Xie, L., and van der Donk, W.A. (2005). Biosynthesis and mode of action of lantibiotics. Chem Rev *105*, 633-684.

Choi, D.H., Noh, J.H., and Lee, J.H. (2014). Application of pyrosequencing method for investigating the diversity of synechococcus subcluster 5.1 in open ocean. Microbes Environ *29*, 17-22.

Coleman, M.L., and Chisholm, S.W. (2007). Code and context: *Prochlorococcus* as a model for cross-scale biology. Trends Microbiol *15*, 398-407.

Coleman, M.L., and Chisholm, S.W. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. Proc Natl Acad Sci U S A *107*, 18634-18639.

Cotter, P.D., Hill, C., and Ross, R.P. (2005). Bacterial lantibiotics: strategies to improve therapeutic potential. Curr Protein Pept Sci *6*, 61-75.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods *10*, 996-998.

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

Ennahar, S., Sashihara, T., Sonomoto, K., and Ishizaki, A. (2000). Class IIa bacteriocins: biosynthesis, structure and activity. FEMS Microbiol Rev *24*, 85-106.

Fischbach, M.A., Walsh, C.T., and Clardy, J. (2008). The evolution of gene collectives: How natural selection drives chemical innovation. Proc Natl Acad Sci U S A *105*, 4601-4608.

Haft, D.H., Basu, M.K., and Mitchell, D.A. (2010). Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. BMC Biol *8*, 70.

Jack, R.W., Tagg, J.R., and Ray, B. (1995). Bacteriocins of gram-positive bacteria. Microbiol Rev *59*, 171-200.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M., and Chisholm, S.W. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. Science *311*, 1737-1740.

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., *et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., *et al.* (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet *3*, e231.

Knerr, P.J., and van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. Annu Rev Biochem *81*, 479-505.

Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P.J., Joewono, I., Rusch, D., Chisholm, S.W., and van der Donk, W.A. (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. Proc Natl Acad Sci U S A *107*, 10430-10435.

Litwin, S., and Jores, R. (1992). Shannon information as a measure of amino acid diversity. In Theoretical and Experimental Insights into Immunology, A. Perelson, and G. Weisbuch, eds. (Paris, Springer-Verlag), pp. 289-296.

Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., and Chisholm, S.W. (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. ISME J *4*, 1252-1264.

Martiny, A.C., Tai, A.P., Veneziano, D., Primeau, F., and Chisholm, S.W. (2009). Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. Environ Microbiol *11*, 823-832.

Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D.J. (2011). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. Environ Microbiol.

Moore, L., Coe, A., Zinser, E.R., Saito, M.A., Sullivan, M.B., Lindell, D., Frois-Moniz, K., Waterbury, J., and Chisholm, S.W. (2007). Culturing the marine cyanobacterium *Prochlorococcus*. Limnology and Oceanography: Methods *5*, 353-362.

Moore, L., Post, A.F., Rocap, G., and Chisholm, S.W. (2002). Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Limnology and Oceanography *47*, 989-996.

Nunnery, J.K., Mevers, E., and Gerwick, W.H. (2010). Biologically active secondary metabolites from marine cyanobacteria. Curr Opin Biotechnol *21*, 787-793.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., *et al.* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature *424*, 1042-1047.

Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R., and Chisholm, S.W. (2009). Whole genome amplification and de novo assembly of single bacterial cells. PLoS One *4*, e6864.

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., Post, A.F., Hagemann, M., Paulsen, I., and Partensky, F. (2009). Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev *73*, 249-299.

Schmidt, E.W. (2010). The hidden diversity of ribosomal peptide natural products. BMC Biol *8*, 83.

Schmitz, S., Hoffmann, A., Szekat, C., Rudd, B., and Bierbaum, G. (2006). The lantibiotic mersacidin is an autoinducing peptide. Appl Environ Microbiol *72*, 7270-7277.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068-2069.

Sivonen, K., Leikoski, N., Fewer, D.P., and Jokela, J. (2010). Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. Appl Microbiol Biotechnol *86*, 1213-1225.

Theocharidis, A., van Dongen, S., Enright, A.J., and Freeman, T.C. (2009). Network visualization and analysis of gene expression data using BioLayout Express(3D). Nat Protoc *4*, 1535-1550.

Wang, H., Fewer, D.P., and Sivonen, K. (2011). Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. PLoS One *6*, e22384.

Willey, J.M., and van der Donk, W.A. (2007). Lantibiotics: peptides of diverse structure and function. Annu Rev Microbiol *61*, 477-501.

Willey, J.M., Willems, A., Kodani, S., and Nodwell, J.R. (2006). Morphogenetic surfactants and their role in the formation of aerial hyphae in Streptomyces coelicolor. Mol Microbiol *59*, 731-742.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol *24*, 1586-1591.

Zhang, Q., Yu, Y., Velasquez, J.E., and van der Donk, W.A. (2012). Evolution of lanthipeptide synthetases. Proc Natl Acad Sci U S A *109*, 18361-18366.

Zwirglmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M., Zubkov, M.V., and Scanlan, D.J. (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. Environ Microbiol *9*, 1278-1290.

SUPPLEMENTAL MATERIAL

Strain		Number of contigs	Genome Size (Mb)	Place of Origin	*procA* genes	pseudo *procA* genes	Clonal Strain? (Representative Strain used)	Reference
Prochlorococcus	MIT9313	1	2.41	Gulf stream	29	0	No	Rocap et al., 2003
Prochlorococcus	MIT9303	1	2.68	Sargasso Sea	13	4	No	Ketter et al., 2007
Prochlorococcus	MIT0701	53	2.59	South Atlantic	9	1	Yes (MIT0701)	Biller at al., 2014
Prochlorococcus	MIT0702	61	2.58	South Atlantic	9	1	Yes (MIT0701)	Biller at al., 2014
Prochlorococcus	MIT0703	61	2.58	South Atlantic	9	1	Yes (MIT0701)	Biller at al., 2014
Prochlorococcus	MIT1306	12	2.50	ALOHA/North Pacific	0	4	No	This study
Prochlorococcus	MIT1313	28	2.59	ALOHA/North Pacific	0	1	No	This study
Prochlorococcus	MIT1318	27	2.58	ALOHA/North Pacific	0	1	No	This study
Prochlorococcus	MIT1320	26	2.50	ALOHA/North Pacific	2	3	No	This study
Prochlorococcus	MIT1323	26	2.44	ALOHA/North Pacific	0	1	No	This study
Prochlorococcus	MIT1342	27	2.55	ALOHA/North Pacific	0	2	No	This study
Prochlorococcus	MIT1327	34	2.59	ALOHA/North Pacific	14	5	Yes (MIT1327)	This study
Prochlorococcus	MIT1312	53	2.56	ALOHA/North Pacific	13	5	Yes (MIT1327)	This study
Prochlorococcus	MIT1303	47	2.51	ALOHA/North Pacific	0	1	No	This study
Synechococcus	RS9916	1	2.66	Gulf of Aqaba/Red Sea	19	1	No	N/A
Synechococcus	WH8016	1	2.69	Atlanic Ocean/Woods Hole	1	1	No	N/A
Synechococcus	KORDI-10	1	2.79	Tropical Pacific Ocean	9	2	No	Choi et al.,2014
Synechococcus	MIT9508	23	2.50	Equatorial Pacific	8	0	No	This study
Synechococcus	MIT9509	33	3.09	Equatorial Pacific	80	30	Yes (MIT9509)	This study
Synechococcus	MIT9504	34	3.09	Equatorial Pacific	78	29	Yes (MIT9509)	This study

Table S2-1. *Prochlorococcus* and *Synechococcus* genomes used in this work
Clonal strains refers to whether the strain has 100% identity in the rDNA ITS sequence and more than 99% whole-genome nucleotide similarity to other strain. Further information on previously sequenced strains can be found in references (Biller et al., 2014; Choi et al., 2014; Kettler et al., 2007; Rocap et al., 2003).

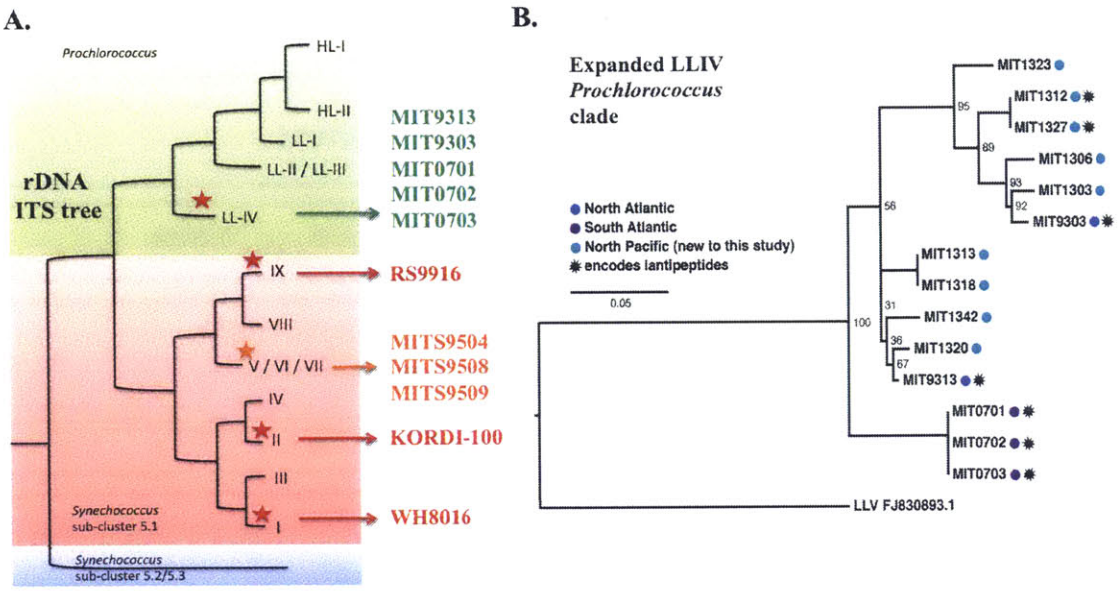

Figure S2-1. Phylogenetic Distribution of the prochlorosin biosynthesis pathway among *Prochlorococcus* and *Synechococcus*

(A) Cartoon representation of the major clades of marine picocyanobacteria. Clades marked with a star contain members that encode the prochlorosin genes. Strains names of the strains harboring the prochlorosin pathway are indicated to the right. *Synechococcus* strains presented in orange were sequenced for this study (B). Expanded rDNA ITS tree for the LL-IV clade of *Prochlorococcus*.

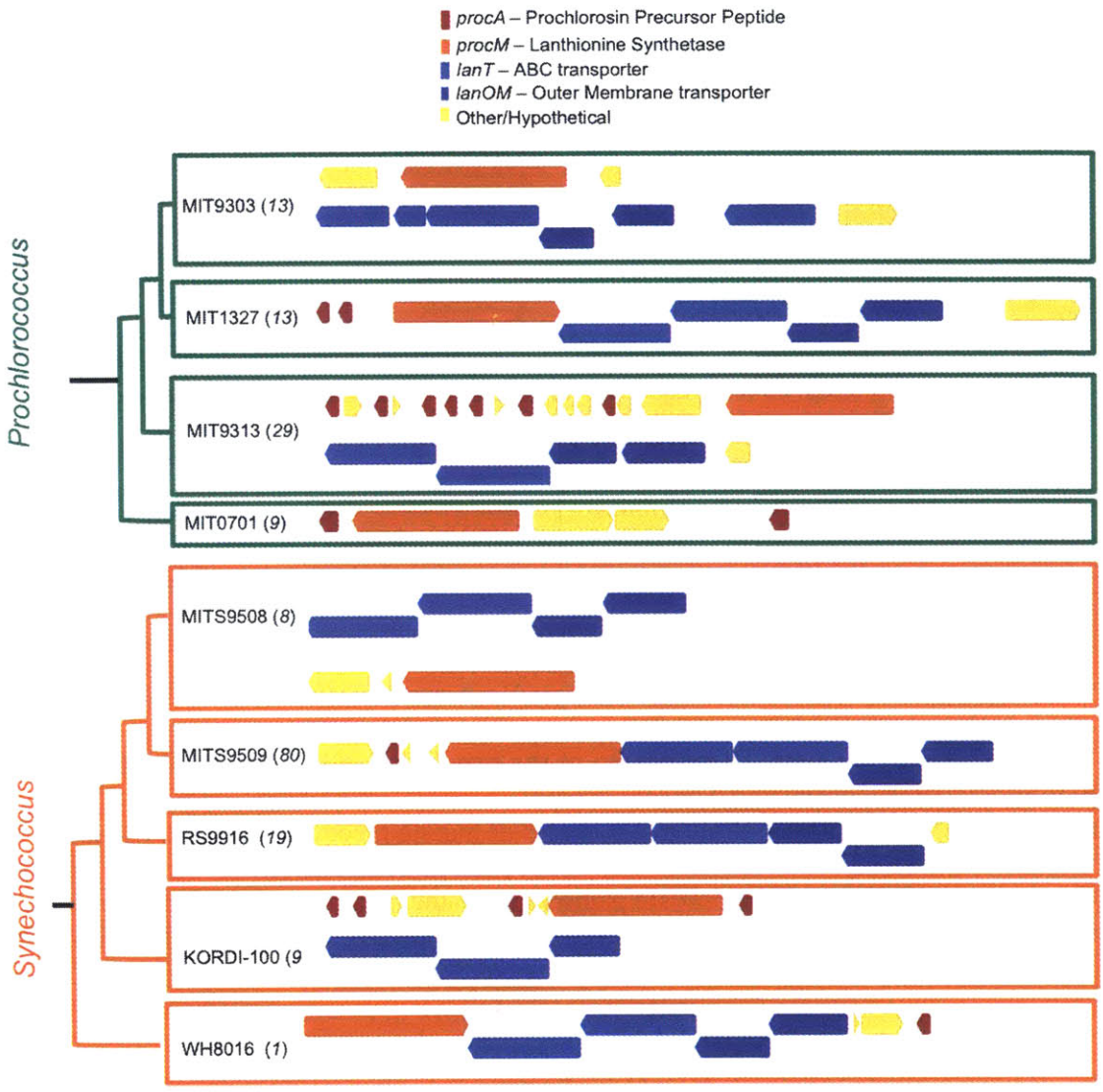

Figure S2-2. Genetic organization of the prochlorosin biosynthesis pathway
The dendrogram on the left indicates the phylogenetic relationship between strains according to the ITS rDNA marker. The total number of *procA* genes per strain is indicated in parenthesis. The color coding of the genes is as follows: prochlorosin precursor peptide gene (Red, only the *procA* genes in the vicinity of *procM* are shown), lanthionine synthetase (orange), lanthipeptide ABC transporters and Outer Membrane transporters (blue) and other/hypothetical genes unrelated to the prochlorosin pathway (yellow).

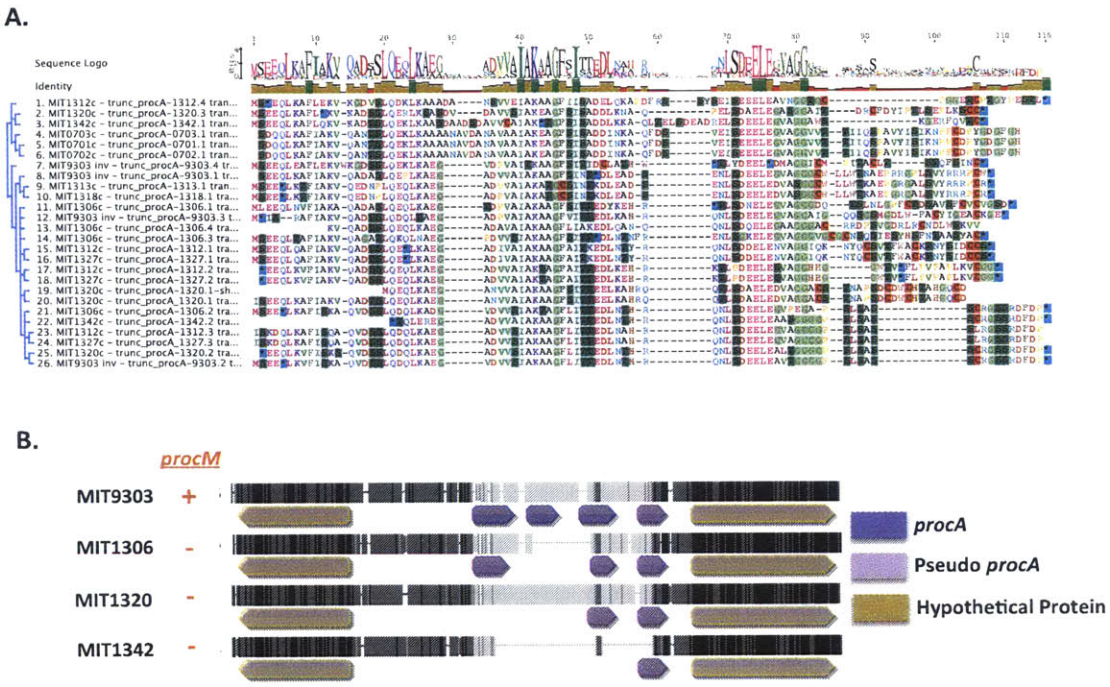

Figure S2-3: Prochlorosin precursor peptide pseudogenes and *procA* gene decay
A. Multiple sequence alignment of *procA* pseudogenes present in the LL-IV clade of *Prochlorococcus*. *procA* pseudo genes are characterized by the presence of a mutation that abolishes the production of a polypeptide (highlighted in blue). The most common example of this is a non-sense mutation that introduces an early stop codon (highlighted in blue) and therefore prevents the formation of a fully functional polypeptide. Frame-shifts and mutation of the start codon are other commonly found types of mutations that also create *procA* pseudogenes. Interestingly, core regions of pseudogenes are also enriched in Cysteine (red), Serine and Threonine residues (both highlighted in dark green), which indicates that these genes could have been functional before the inactivating mutations. **B.** Multiple sequence alignment of a conserved locus conserved between four *Prochlorococcus* genomes that contain *procA* genes (black color bars indicates the degree of sequence conservation). MIT9303 encodes a complete prochlorosin biosynthetic pathway and harbors three *procA* genes and one pseudo-*procA* gene in this locus. Strains MIT1306, MIT1320 and MIT1342 do not encode the complete biosynthesis pathway and the *procA* genes in this conserved locus have become pseudo-genes or lost entirely from the genome, indicating a process of gene decay.

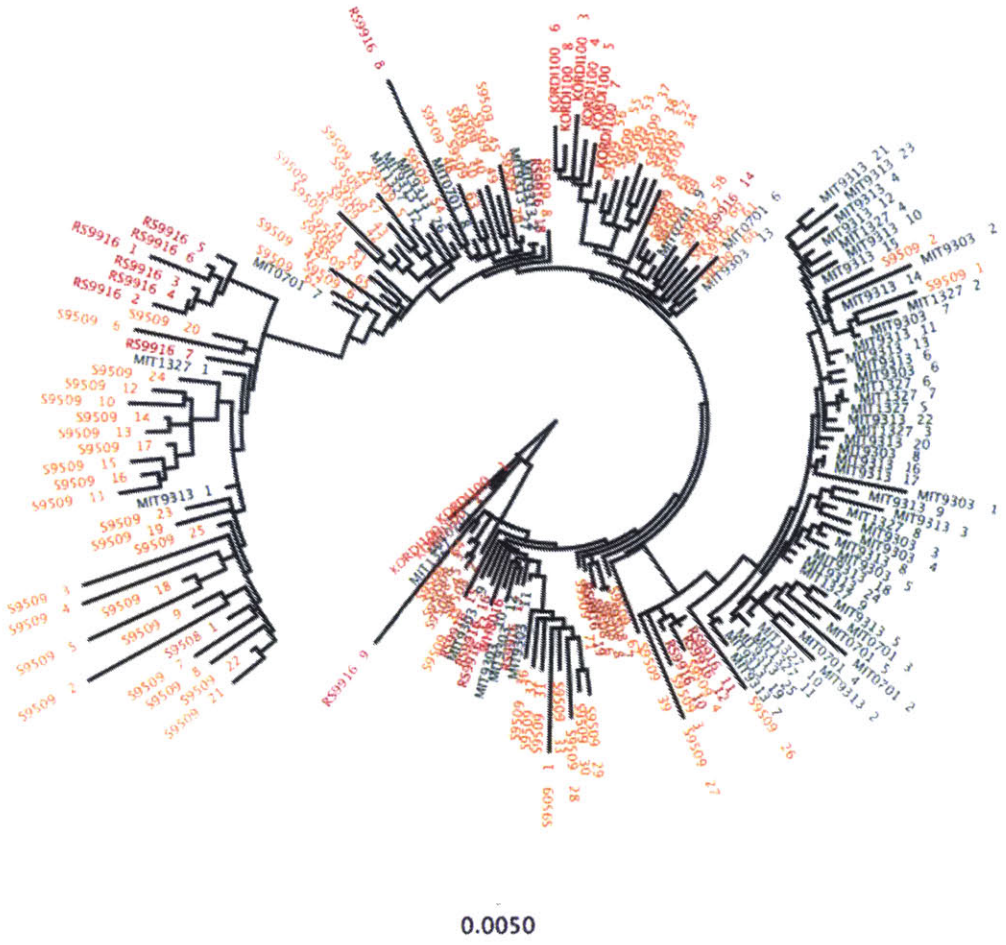

Figure S2-4: Phylogenetic tree of leader peptide region of the prochlorosin precursor peptide.
Maximum Likely hood tree for the amino acid sequence of the prochlorosin leader peptide region. The leader peptide region allows phylogenetic reconstruction of precursor peptides identifying two distinct groups. One group is dominated by *Prochlorococcus* (green) leader peptides and the second group is composed mostly of representatives of *Synechococcus* genomes (orange and red).

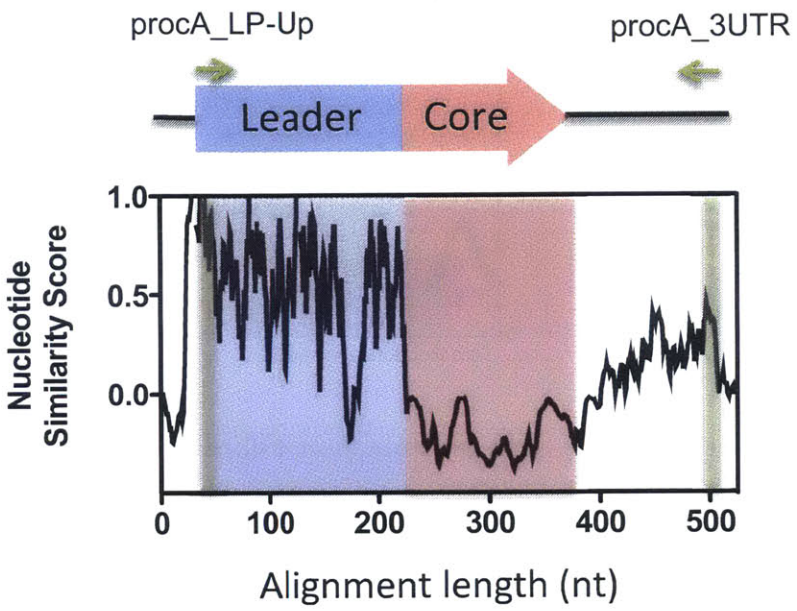

Figure S2-5: Sequence conservation of the procA locus.
Nucleotide similarity scores were calculated for every position in a multiple sequence alignment of 181 procA genes and their flanking regions. Regions highlighted in green were used to design primers for the amplification of the *procA* locus from environmental samples.

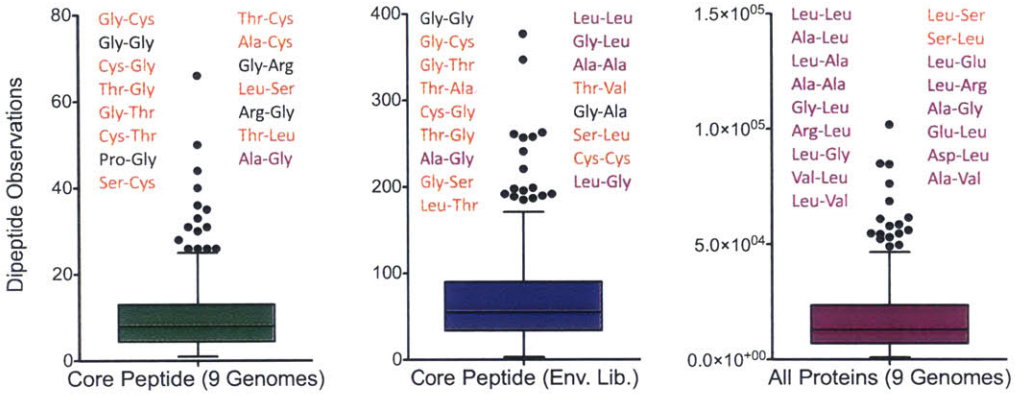

Figure S2-6: Dipeptide analysis in genomes, *procA* libraries and proteome. Boxplot analysis and Tukey test indicate the presence of outlier dipeptides indicating high frequency in the core peptides (black dots). Lists correspond to dipeptides that were considered outliers in their number of observations in the dataset. Dipeptides highlighted in red contain residues that participate in cyclization. Dipeptides highlighted in purple are frequently found in the proteome of *Prochlorococcus*. Green: Dipeptide observations in the 181 core peptides identified in 9 lanthipeptide-encoding picocyanobacterial genomes. Blue: Dipeptide observations in the 1354 OPUs with non-redundant core peptide regions from the environmental DNA libraries. Purple: Dipeptide observations in the predicted proteome of 9 lanthipeptide-encoding picocyanobacterial strains. Note that the most frequents dipeptides from the prochlorosin core peptides do not overlap with the ones from the predicted proteome.

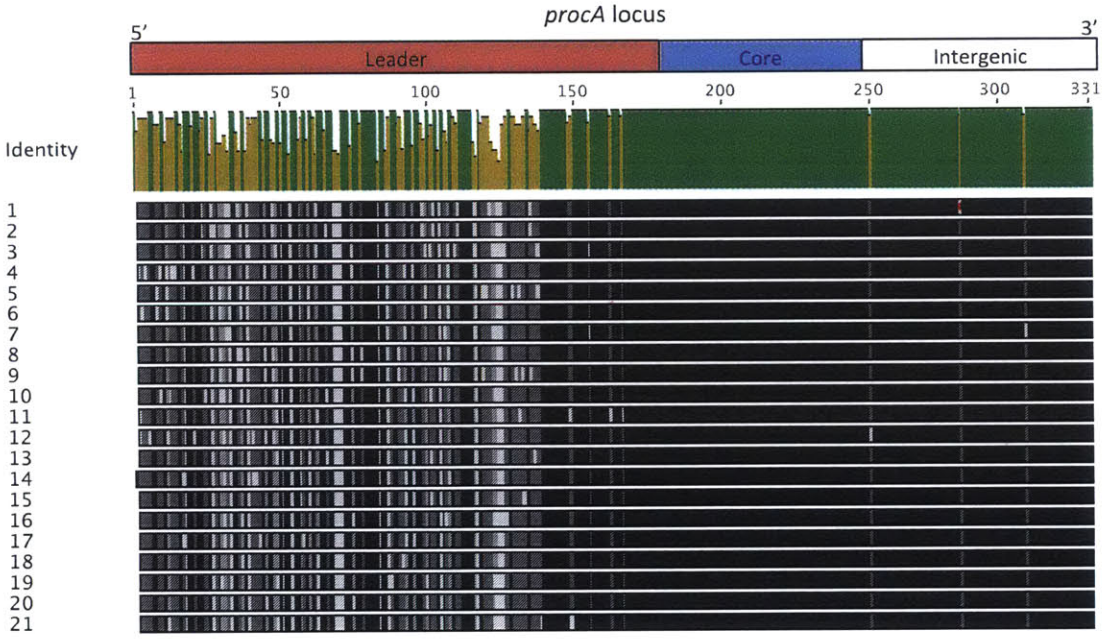

Figure S2-7: Sequence alignment of OPUs with complete identity in the core peptide region.

The variation found in the leader peptide region of closely related OPUs that display complete identity in the core region demonstrate that small sequence polymorphisms are the type of genetic changes that dominate the genetic variation among leader peptide sequences. Black color bars indicates the degree of sequence conservation.

CHAPTER 3

Exploring the Biological Role of Prochlorosins

Andres Cubillos-Ruiz [1,2]

Jamie Becker [2]

Yanxiang Shi [3]

Wilfred van der Donk [3]

Sallie W. Chisholm [2]

[1] Microbiology Graduate Program, Massachusetts Institute of Technology.

[2] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

[3] Department of Chemistry, University of Illinois.

ABSTRACT

Prochlorosins are a diverse group of lanthipeptides produced by strains of the ubiquitous marine picocyanobacteria *Prochlorococcus* and *Synechococcus*. Unlike other lanthipeptide-producing bacteria, some strains of picocyanobacteria are capable of producing multiple different prochlorosins from distinct gene-derived precursors, by using a single lanthionine synthetase. The 'catalytic promiscuity' in the biosynthesis of prochlorosins is unprecedented both in nature and in the laboratory. Also unprecedented is the production of lanthipeptides by single celled, free-floating gram-negative bacteria in an extremely dilute habitat, which suggests that they may have novel biological functions, or mechanisms of action. The experiments conducted here, using a set of four recombinant prochlorosins, are the first steps in the characterization of the biological function of the prochlorosins. We identified functional differences between the regulatory mechanism of the prochlorosin biosynthesis pathway from marine picocyanobacteria and the canonical lanthipeptide pathways from gram-positive bacteria, namely, the timing in the expression of the precursor peptide genes and the absence of autoregulation. The whole-transcriptome response of a producer strain to the addition of its own prochlorosins elicited a physiological response that resembles a self-immunity mechanism, in which the peptides are likely metabolized or sent back into the extracellular environment. In terms of the effect of prochlorosins to other marine bacteria, we did not observe any inhibition of growth of other *Prochlorococcus* strains or co-isolated heterotrophic bacteria. In contrast to an inhibitory role, we demonstrated that an isolate of SAR11, a ubiquitous clade of marine heterotrophic bacteria that is notoriously difficult to culture, is able to use lanthionine-containing peptides to meet some of its unique nutritional requirements for glycine and a reduced sulfur source. These findings open a new perspective in which the production of lanthipeptides in the oligotrophic ocean might mediate metabolic interactions between *Prochlorococcus* and groups of heterotrophic bacterioplankton; the nature of those interactions needs to be further studied.

INTRODUCTION

Lanthipeptides are small polycyclic ribosomally-derived secondary metabolites of bacterial origin. These natural products are created from gene-encoded precursor peptides that are tailored by a dedicated enzyme (lanthionine synthetase), which catalyzes the dehydration of select Ser and Thr residues and the intramolecular addition of Cys thiols to the resulting unsaturated amino acids, forming lanthionine and methyl-lanthionine bridges, respectively. Most of the known lanthionine-containing peptides have antimicrobial activity and are referred to as lantibiotics (Chatterjee et al., 2005b). These type of peptides are particularly potent against of Gram-positive bacteria, including many multiple antibiotic resistant pathogens (Willey and van der Donk, 2007). This fact combined with a low tendency to generate resistance, make lanthionine-containing peptides attractive candidates for clinical applications (Field et al., 2010).

The biosynthetic pathways for lanthipeptide production, and their biological function, have been extensively characterized in diverse groups of gram-positive bacteria, particularly Firmicutes and Actinobacteria (Willey and van der Donk, 2007). Although lanthipeptide biosynthetic gene clusters have been found recently in proteobacterial and cyanobacterial lineages (Haft et al., 2010; Zhang et al., 2012) – gram-negative bacteria – little is known about the biological function of the lanthionine-containing peptides from these groups. In fact, the only lanthipeptides that have been isolated and structurally characterized in gram-negative bacteria are Prochlorosins (Pcn); diverse lanthionine-containing peptides with unknown function produced by certain strains of the ubiquitous marine picocyanobacteria *Prochlorococcus* and *Synechococcus* (Li et al., 2010; Mukherjee and van der Donk, 2014; Tang and van der Donk, 2012).

Unlike other lanthipeptide-producing bacteria, some strains of picocyanobacteria are capable of using a single biosynthetic enzyme to produce multiple different prochlorosins from distinct gene-derived precursors (Li et al., 2010). One of the hallmarks of the prochlorosins is their startling structural diversity. Prochlorosin precursor peptide genes retrieved from the genome of 9 lanthipeptide-encoding strains of *Prochlorococcus* and

Synechococcus predict the presence of 181 novel lanthipeptide products, from which only a single pair of products displays similarity in their putative ring topologies. Moreover, metagenomic analysis of the prochlorosin precursor peptide genes from wild populations of marine picocyanobacteria revealed the presence of thousands of potentially different prochlorosin structures that are likely the result of an evolutionary radiation process (see Chapter 2). In a detailed analysis of the evolution of the prochlorosin biosynthesis pathway, we have shown that, while the peptide substrates rapidly expand and diversify within the genome, the catalytically promiscuous lanthionine synthetase evolves under a strong purifying selection that maintains its substrate tolerant state (Chapter 2). This unique evolutionary dynamic suggests that the selective advantage of prochlorosin production comes from the capacity of producing a suite of lanthipeptides with diverse ring topologies, which contrasts with canonical lanthipeptide systems that evolve towards creating a single molecule with a defined ring topology.

The production of prochlorosins by marine picocyanobacteria is intriguing, not only because of their unique combinatorial biosynthesis mechanism and exceptional diversity, but also because their dilute open ocean habitat seems an unconventional stage for natural products production. *Prochlorococcus* and *Synechococcus* are single-celled, free-floating microorganisms that live in a very dilute environment. On average, the distance between two neighboring microbes is approximately 200-cell lengths (Biller et al., 2015), thus it is difficult to imagine how prochlorosins could reach typical threshold concentrations at which other lanthipeptides display antimicrobial activity (Chatterjee et al., 2005b). These unique features of the prochlorosins suggest an unusual biological function/mechanism of action for these lanthipeptides.

There are two experimental barriers to study of the biological function of the prochlorosins. First, genetic tools do not exist for *Prochlorococcus*, which hampers the use of standard molecular genetics approaches to investigate their function. Second, prochlorosins are not readily isolated from cell cultures; only 3 of the 29 prochlorosins encoded by *Prochlorococcus* MIT9313 were detected in spent media from growing cultures, and they were recovered at a very low yield (<10 μg from 20L cell culture) (Li et al., 2010). These low yields make difficult the use of organic extracts to test their

biological function. To overcome this limitation, we used recombinant prochlorosins that were produced by the van der Donk group using an *E. coli*-based expression system (Shi et al., 2010). In this system, the prochlorosin precursor peptide has been engineered to contain an N-terminal His-tag and an artificial protease cleavage site at the end of the leader peptide. The ProcM lanthionine synthetase of *Prochlorococcus* MIT9313 is co-expressed with the prochlorosin precursor peptide, so the dehydration and cyclization reactions happen intracellularly. Following modification, the precursor peptide is isolated from the cytoplasmic content by affinity purification, the leader peptide is removed *in vitro* and the mature core peptide is purified by HPLC. Although, this approach offers the advantage of utilizing highly pure compounds, it is low-throughput so only a handful of prochlorosins can be tested (Shi et al., 2010).

Here, we use a set of four recombinant prochlorosins from *Prochlorococcus* MIT9313 to evaluate the role of these compounds as possible autocrine signaling molecules. We also use them to explore the possible effect of lanthipeptides on the growth other strains of *Prochlorococcus* and heterotrophic marine bacteria. In addition, we perform a proof-of-concept experiment to demonstrate the utility of lanthipeptides as a reduced-sulphur nutrient source for an abundant group of oligotrophic bacteria, SAR11, which cannot utilize oxidized forms of sulfur.

RESULTS AND DISCUSSION

Signaling Activity of Prochlorosins

The use of peptide products for bacterial signaling is well understood in gram-positive bacteria as part of their quorum sensing systems (Miller and Bassler, 2001). In fact, the biosynthesis of a number of lantibiotics is autoregulatory, with the transcription of biosynthetic and immunity operons under the control of quorum sensing systems (Kleerebezem, 2004). For instance, the autoregulation of the type I lantibiotics nisin and subtilin involves the sensing of subinhibitory concentrations of peptides via a two-component histidine kinase/response regulator signal transduction system, which in turn induces the expression of multiple transcriptional units within the operons that harbor genes for lantibiotic biosynthesis and immunity (Willey and van der Donk, 2007). In gram-negative bacteria, the vast majority of cell-to-cell signaling systems use N-acyl homoserine lactones and not peptide products (Miller and Bassler, 2001). However, it has been demonstrated that pathogenic bacteria, such as *Salmonella typhimurium* and *Pseudomonas aeruginosa*, are able to sense cationic antimicrobial peptides produced by the innate immune system through the action of two-component signal transduction systems and induce the expression of resistance mechanisms (Bader et al., 2003; McPhee et al., 2003). Could the prochlorosins be acting as autocrine signaling molecules in marine picocyanobacteria?

The prochlorosin system is the first experimentally characterized lanthipeptide production pathway of gram-negative bacteria. Consequently, it is unknown whether the expression of the lanthipeptides of this bacterial group follows a cell-density dependent response or if their biosynthesis is also autoregulatory.

To determine if the level of expression of prochlorosin biosynthesis genes is dependent on cell density, we quantified the expression of the transcripts for the lanthionine synthetase, the LanT ABC transporter and a group of precursor peptide genes, along the growth curve of *Prochlorococcus* MIT9313 (Fig. 3-1A). The highest levels of expression of the genes of the prochlorosin biosynthesis pathway are observed during exponential growth followed by decreased expression during stationary phase. The expression levels

of the *procA* genes are very similar during the mid-exponential and late-exponential phases of growth indicating that there is no dependency on cell-density in the expression of the precursor peptides (Fig. 3-1B). Also, the expression of the biosynthesis genes does not increase during stationary phase, which contrasts with the paradigmatic view of lanthipeptides as stationary phase-induced secondary metabolites. This experiment suggests that, under laboratory conditions at least, the expression of prochlorosin biosynthesis genes does not follow the typical behavior of a quorum sensing regulated system.

Whole-transcriptome Response to the Addition of Prochlorosins

To test whether the prochlorosins might have a signaling or autoregulatory role, we quantified the whole-transcriptome changes 12 and 36 hours after the addition of a mix of 4 recombinant prochlorosins to an exponentially growing culture of *Prochlorococcus* MIT9313 (Fig. 3-2). By twelve hours after the addition of the prochlorosins, 60 genes were differentially expressed relative to a control, consisting only of the addition of the buffer in which the prochlorosins had been resuspended. Forty-four of the 60 transcripts were elevated in abundance, and 16 were reduced relative to the control (Table S3-1). By 36h there were only 9 differentially abundant transcripts – all of them differentially less abundant– indicating that the effect of the prochlorosin addition on the transcriptome is short-lived. Importantly, in this experiment there were no differences in the bulk fluorescence or the cell density along the growth curve between the prochlorosin and the control treatments, which suggest that the observed changes in the gene expression are not likely due to changes in the growth rate of the cultures. Accordingly, this result demonstrates that *Prochlorococcus* MIT9313 is able to sense the presence of its own lanthipeptides and respond to them by modulating the expression of a small subset of genes (<2% of its genome).

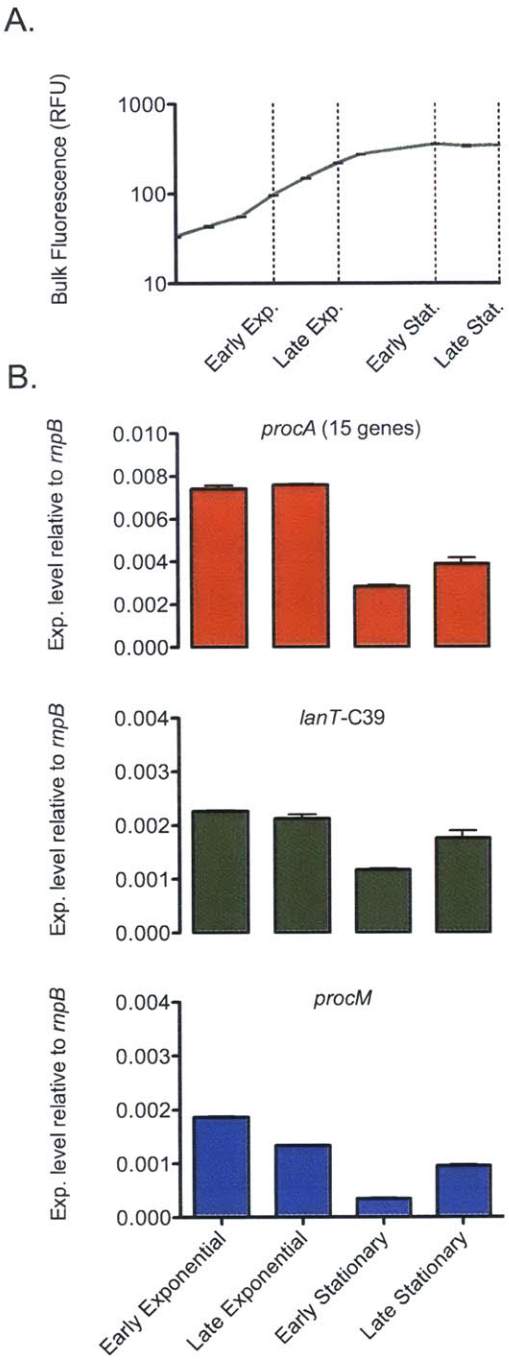

Figure 3-1. Expression of prochlorosin biosynthesis genes along the growth curve of *Prochlorococcus* MIT9313

A. Growth curve of triplicate cultures of *Prochlorococcus*. Dashed lines denote the time points at which the expression of genes was evaluated. Error bars (SD) are smaller than the size of the data points. **B.** qPCR quantitation of the expression of prochlorosin biosynthesis genes at four stages of the growth curve. Red: a group of 15 *procA* genes that shared identical priming sites in the leader peptide (see materials and methods). Green: Prochlorosin ABC transporter *lanT*, C39 protease domain-containing subunit. Blue: Prochlorosin lanthionine synthetase. Error bars reflect the standard deviation (SD).

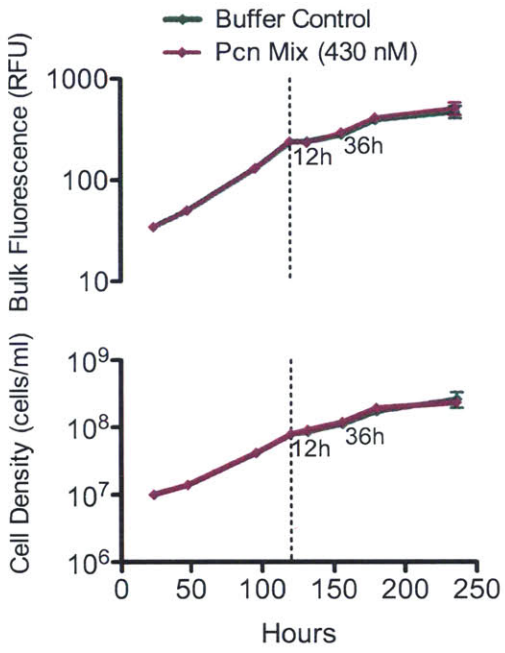

Figure 3-2: Growth response of *Prochlorococcus* MIT9313 to the addition of recombinant prochlorosins
Both cell numbers and bulk chlorophyll fluorescence are shown. Duplicate cultures of exponentially growing cells of *Prochlorococcus* MIT9313 were treated with an equimolar mix of four recombinant prochlorosins (Pcn1.7, Pcn 2.11, Pcn 3.2 and Pcn 3.3) at a final concentration of 430 nM (100ng/ml). The control treatment consisted in an equal volume of the buffer used to resuspend the prochlorosins (50 µl of 50mM HEPES buffer pH 8). Dashed line denotes the time points at which prochlorosins were added. Libraries for Illumina RNA-seq were created for the samples collected at 12h and 36 h after the addition of the peptides. Error bars are smaller than the data points and correspond to SD.

One question is whether the biosynthesis of prochlorosins is autoregulatory, as it has been observed in other families of lanthipeptides (Kleerebezem, 2004). None of the 29 prochlorosins precursor peptide genes, the lanthionine synthetase or the *lanT* transporters were differentially expressed upon the addition of the prochlorosins in this experiment (Table S3-1), suggesting that at least the subset of 4 prochlorosins tested do not have a direct autoinducing effect in the biosynthesis pathway. Prochlorosin addition did,

however, trigger the differential expression of 13 different genes located in 3 out of the 4 loci encoding the prochlorosin pathway in *Prochlorococcus* MIT9313 (Fig. 3-3). This suggests that the signaling cascade resulting from sensing of extracellular lanthipeptides is in fact transcriptionally linked to prochlorosin-associated loci, but does not exert any regulatory function on their biosynthesis. Rather, it suggests that some other genes in the prochlorosin loci might accomplish a role associated in the post-production of lanthipeptides. For instance, two differentially more abundant genes, PMT_2138 and PMT_0929, located in the *procA* Clusters 2 and 3, respectively, encode putative peptidases with signal peptides that suggest they might be trafficked beyond the inner membrane to either the periplasm or the outer membrane where the prochlorosins might be residing. In addition to these two *procA*-associated peptidases, the expression of 9 other proteins containing signal peptides was induced elsewhere in the genome, including another putative peptidase (PMT_1940) and an outer membrane protein with a peptidoglycan-binding domain (PMT_0915). Notably, a putative Signal Peptidase I (PMT_0382), the protein in charge of cleaving signal peptides, was also differentially induced suggesting that upon the addition of the prochlorosins there is a coordinated response to traffic peptidases and other proteins to the outer compartments of the cell.

Among the genes whose expression was differentially less abundant by both 12h and 36h after the addition of the prochlorosins are four short genes, with identical amino acid sequences, located near the *procA* Cluster 4; they are annotated as kinesin motor domain proteins. Kinesins are motor proteins that move along microtubules of eukaryotic cells, however their function in cyanobacteria is unknown but it might suggest that prochlorosins could trigger cell-wall related responses. Similarly, within the *procA* Cluster 1, we found three genes that had been differentially induced by 12h after the prochlorosin addition (Fig. 3-3). One of these genes is annotated as a possible profilin, an actin-binding protein that has been postulated to be the result of a horizontal gene transfer event between a eukaryote and cyanobacteria (Guljamow et al., 2012). It is not clear how these proteins might act in response to the prochlorosins, but given that they are located in a prochlorosin-harboring locus, they might have a functional relationship.

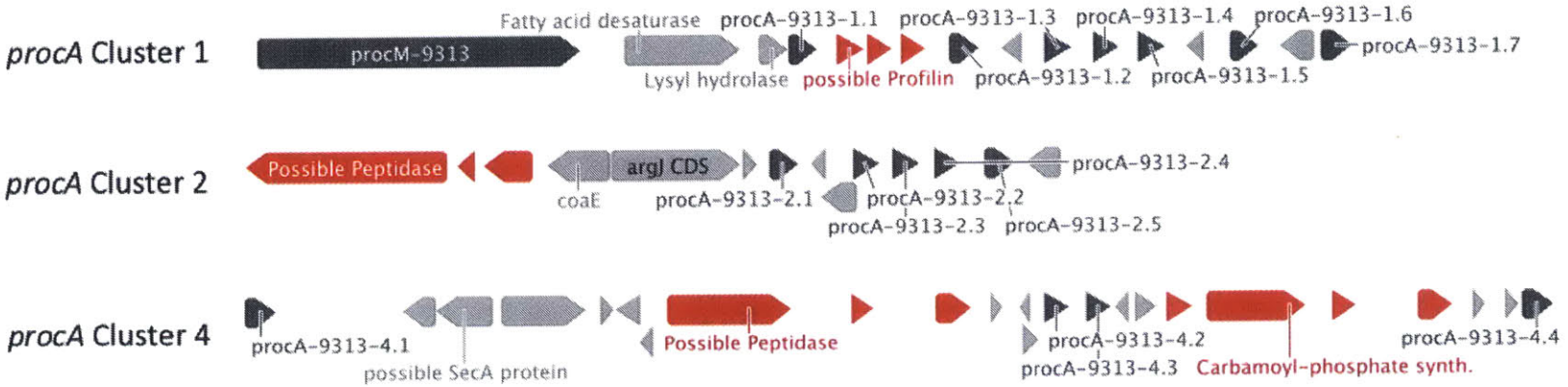

Figure 3-3: Prochlorosin-responsive genes located in *procA* loci.
Representation of three *procA* loci some of which were differentially expressed in response to the addition of prochlorosins. Red gene arrows indicate that the expression of the gene was differentially more abundant. Grey gene arrows indicate the genes were not differentially expressed. Dark grey gene arrows denote prochlorosin precursor peptide genes and the *procM* lanthionine synthetase.

Induction of a possible peptide efflux pump

Within the genes whose expression is differentially more abundant there is a locus composed of 8 genes, of which 4 encode a possible ABC-type antimicrobial peptide transport system (Fig. 3-4A). The most similar genes to this transporter are related to the DevCBA-family of glycolipid transporters known only to be present in cyanobacteria (Staron et al., 2011). The transporter genes, however, display some differences from the DevCBA-family that indicate that its function might be related to the transport of peptides rather than glycolipids. Besides having a different, non-operon gene structure organization, the main difference from the DevCBA-family resides in the composition of the inner membrane factors: this transporter has two, not one, putative DevC-like permease components. These permease components, as well as the ATPase component (DevA-like), are more related to the MacAB family of macrolide-specific ABC-type efflux carrier, which is not common for inner membrane factors of other members of the DevCBA family (Piddock, 2006). It is important to note that an outer membrane factor for this putative efflux pump is missing in this locus, however it is known that TolC-like outer membrane factors, such as the one found in the secretion pathway of the prochlorosins (PMT_0979/0980), can associate with ABC transporters of the DevCBA family (Staron et al., 2014), allowing the export of the peptides back to the extracellular environment. The strong transcript induction of this ABC transporter (Fig. 3-4B) might be suggestive of a possible efflux-pump-like immunity mechanism resembling the ones present in lantibiotic producing gram-positive bacteria (Willey and van der Donk, 2007). However, *Prochlorococcus* is a gram-negative bacterium and therefore functional inferences from structural homology to the well-known immunity LanEFG ABC transporters are not straightforward. If *Prochlorococcus* were a genetically tractable organism, directed mutagenesis experiments would help to elucidate the whether this ABC-transporter acts as a possible self-protection mechanism.

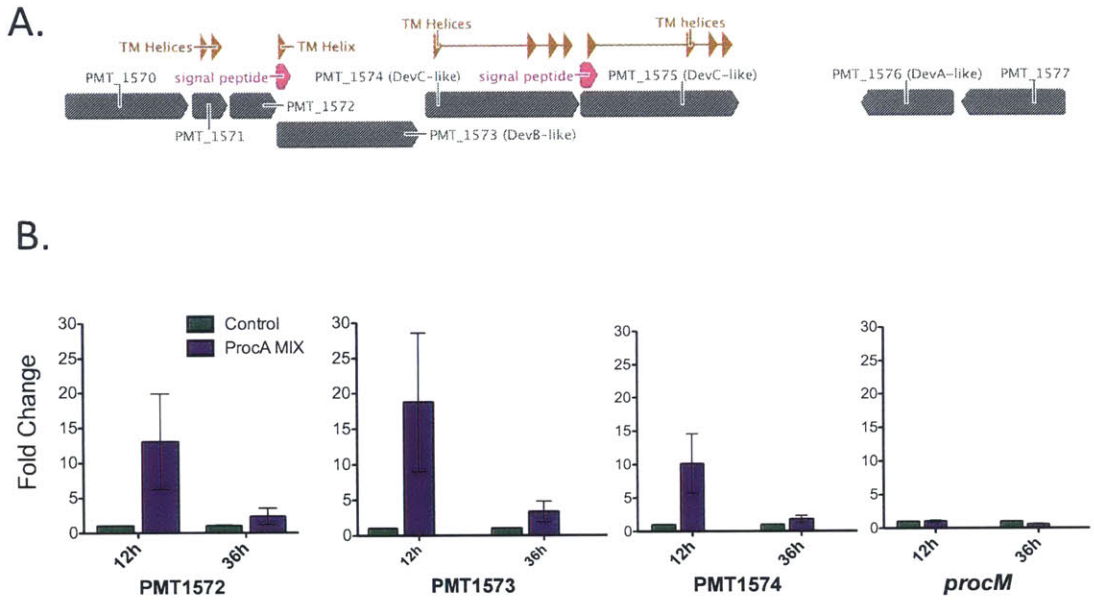

Figure 3-4: Prochlorosins trigger the expression of a putative efflux-pump. **A**. Genetic organization of the locus containing an ABC-type antimicrobial peptide transporter. PMT1572 is a putative sterol binding protein. PMT1573 and PMT1576 are the DevB-like and DevA-like inner membrane components, respectively. PMT1574 and PMT1575 are the two putative periplasmic membrane fusion protein. The other genes in the locus are conserved hypothetical proteins. Pink and yellow arrows denote export signal peptides and transmembrane helices, respectively. **B.** qPCR results for the differences in the expression levels of three genes in the ABC transporter locus and also for *procM*, whose expression levels was not changed in the RNA-seq data or in the qPCR data. Importantly, the induction of expression of those genes is greater than what was identified from RNA-seq data. Differences in fold change values between qPCR and RNA-seq might be observed because the rRNA normalization performed for the library construction reduced the dynamic range of RNA-seq, especially for highly expressed genes. The results from the qPCR experiment are more reliable and thus the response of this operon to the lanthipeptides is stronger than what we appreciate with the values of RNA-seq.

Thus it appears that genes involved in peptide transport and metabolism dominate the general transcriptomic response to the prochlorosins. The induction of putative exopeptidases, outer membrane proteins and a potential peptide efflux-pump is indicative of a possible concerted mechanism that prevents the entry of fully modified peptides back into the cytosol and that might degrade prochlorosins residing in the outer compartments of the cell. This degradative response could explain why 36h after the addition of prochlorosins, all the initially induced genes had gone back to levels of expression similar to the untreated control.

It is important to note that this experiment cannot resolve whether the set of genes that were differentially expressed reflect a specific response to lanthipeptides or a response to peptides in general. The differential expression of a number of genes encoded in the same loci of the *procA* genes indicates that the former possibility is more likely; however further testing will be required.

Effect of Prochlorosins on Closely Related Strains of *Prochlorococcus* and Heterotrophic Marine Bacteria

The results thus far demonstrate that prochlorosins are likely bioactive molecules that can trigger specific physiological responses upon contact with the producer cells. But lantibiotics were initially discovered, and have gained much attention since then, because of their activity as bacteriocins: bacterial proteinaceous compounds that display antibiotic activity against strains of the same species or closely related species. So the next question is, what is the effect, if any, of prochlorosins on other members of the marine microbial community? This is a challenging question to address in laboratory experiments, as the tremendous diversity of prochlorosins in the ocean and the complex nature of marine microbial communities makes difficult to identify experimental conditions that would represent ecologically relevant prochlorosin-target relationships. To begin to address this, we conducted exploratory experiments to test the possibility that the four recombinant prochlorosins from *Prochlorococcus* MIT9313, by some general mechanism, could inhibit the growth of closely related strains, as well as selected strains of co-occurring heterotrophic bacteria.

For the experiments with different *Prochlorococcus* strains we selected three strains from different clades, only one of which is a prochlorosin producer: MIT9303, a deeply branching low-light adapted strain belonging to the LL-IV clade, encodes 13 prochlorosins; NATL2A is also a low light adapted strain (more recently diverged, LL-I clade) and does not encode lanthipeptides; MED4 is a high-light adapted strain (HL-I clade), does not encode lanthipeptides, and is quite distantly related. The addition of the prochlorosin mix to a final concentration of 1μg/ml did not have a significant effect on the overall growth of the tested strains when compared to the untreated cultures, or to the cultures that were treated with the same concentration of the unmodified "linear" version of the Pcn 3.2 (Fig. 3-5). A treatment with kanamycin was included as a positive control of inhibition. Thus under the conditions tested here, the combination of four recombinant prochlorosins from *Prochlorococcus* MIT9313 does not display bacteriocin activity against other strains.

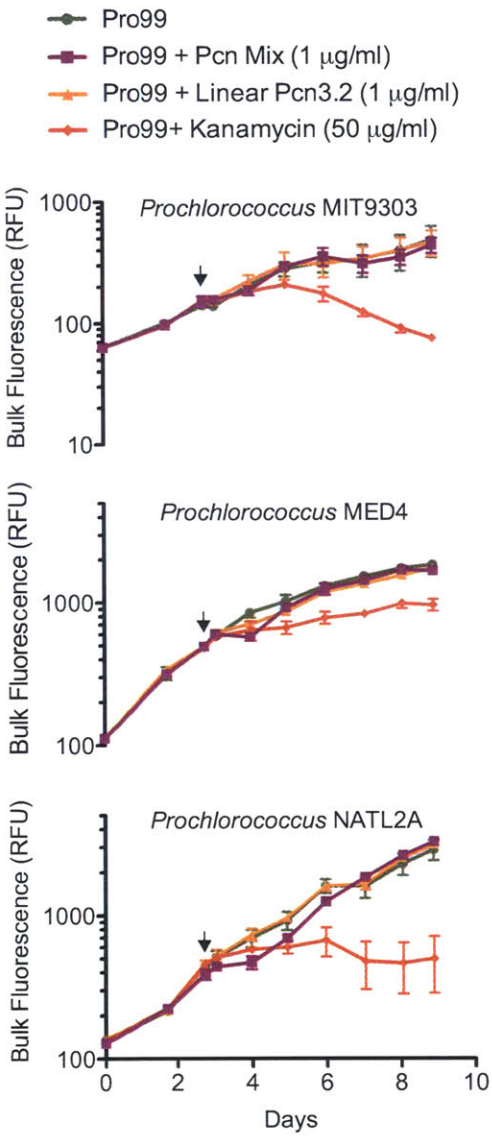

Figure 3-5: Effect of four recombinant prochlorosins from *Prochlorococcus* MIT9313 in the growth of different strains of *Prochlorococcus*.
Each *Prochlorococcus* culture was grown to early exponential phase and transferred to 96-well plates and on day 3 (arrow) the following treatments (triplicate) were made: Purple: 1 μg/ml Pcn Mix (Pcn1.7, Pcn 2.11, Pcn 3.2 and Pcn 3.3). Orange: 1 μg/ml Linear Pcn3.2 Linear (unmodified core peptide). Red: 50 μg/ml Kanamycin (antibiotic control) and Green: 50 mM HEPES (Buffer Control). Green: Pro99 Medium (No treatment control).

To further explore the possible role of prochlorosins as antibacterial agents, we tested the effect of the prochlorosin mix on four different gram-negative heterotrophic bacteria that were co-isolated with different *Prochlorococcus* strains: *Alteromonas alvinellae* Med4c2:1 and an undefined Alpha-Proteobacterium Med4c1:1 were co-isolated with *Prochlorococcus* MED4, while *A. macleodii* 9313c2 and *Halomonas* 9313c3 were co-isolated with *Prochlorococcus* MIT9313. The prochlorosin mix did affect the growth of these heterotrophic bacterial strains (Fig. 3-6). However, the presence of the linear peptide in the cultures produced a slightly more robust growth compared with the prochlorosin treated cultures, which indicates that there might be subtle differences in the growth of strains when they are exposed to linear versus cyclic peptides. For instance, in the *Alpha-proteobacterium* Med4c1:1 the treatment with the linear peptide yielded a higher final density compared to the no peptide control and the prochlorosin treatment.

While these exploratory experiments did not yield any results that point us in a direction that might solve the mystery of function, we certainly cannot rule out the possibility that the prochlorosins are acting as antimicrobial agents in the marine environment. On the contrary, these exploratory experiments only scratched the surface of the possible interactions among the thousands of microbes that co-occur with prochlorosin-producing *Prochlorococcus* strains. Furthermore, there are general caveats to consider regarding the possible bioactivity of the selected recombinant prochlorosins: first, we assume that no other posttranslational modifications happen naturally in *Prochlorococcus* MIT9313 and that the modifications conferred on the recombinant prochlorosin by the action of the ProcM are solely responsible of the bioactivity. Second, the potential mechanism of action does not require any other specialized cellular structure (i.e. tethering to the outer membrane or vesicle encapsulation) such that freely diffusing molecules would be sufficient for bioactivity. In the exploration of the biological function of prochlorosins, this was the logical first set of experiments to try.

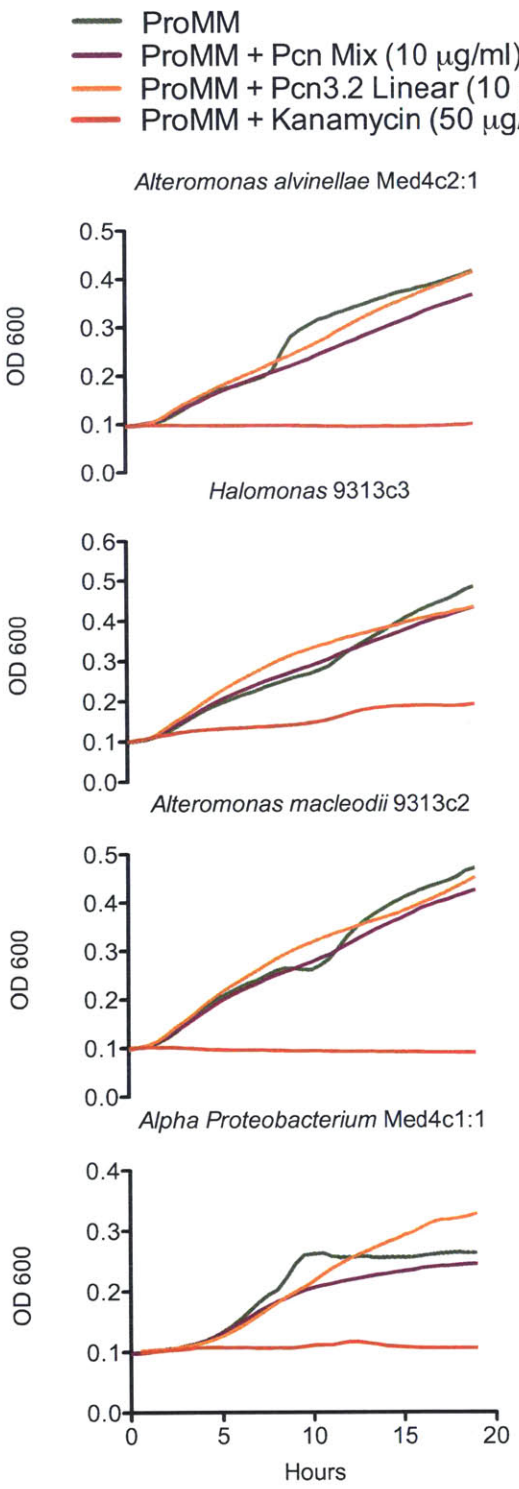

Figure 3-6: Effect of four recombinant prochlorosins from *Prochlorococcus* MIT9313 in the growth of different strains of marine heterotrophic bacteria. Each heterotrophic bacterial culture was grown overnight in ProMM medium, cells were diluted to an OD600 (Optical Density at 600 nm) of 0.1 and transferred into 96-well plates where the following treatments (triplicate) were made: Purple: 10 μg/ml Pcn Mix (Pcn1.7, Pcn 2.11, Pcn 3.2 and Pcn 3.3). Orange: 10 μg/ml Linear Pcn3.2 Linear (unmodified core peptide). Red: 50 μg/ml Kanamycin (antibiotic control) and Green: 50 mM HEPES (Buffer Control). Lines correspond to kinetic reads of every 30 mins. Data points and error bars are omitted for the ease of visualization. The presence of the error bars does not change the interpretation.

Could Lanthipeptides serve as a Source of Reduced Sulphur for the dominant sympatric heterotroph with *Prochlorococcus* in the Oligotrophic Ocean?

The results described above motivated us to find a more ecologically relevant system to test the effect of prochlorosins and to think about the function from a different perspective; these complex molecules are being produced and potentially released into a nutrient-limited environment where they are potentially metabolized by other microbes as part of the dissolved organic matter pool. Although produced in relatively low amounts, it is possible that they could satisfy the requirements of specific microbial species. Natural microbial communities form a dynamic network of interactions that promotes the evolution of an intricate web of interconnected metabolisms (Faust and Raes, 2012; Seth and Taga, 2014). Bacteria have developed specific adaptations that often involve interspecies interactions to efficiently exploit the resources available in the environment (Little et al., 2008; Morris et al., 2012). *Prochlorococcus* thrives in oligotrophic environments but little is known about how it interacts with other abundant bacterial groups that co-exist in this extremely nutrient-limited environment.

Marine members of the SAR11 clade of α-Proteobacteria can comprise up to 50% of all bacteria inhabiting the euphotic zone of the oligotrophic ocean, where *Prochlorococcus* is the main primary producer. They are the most abundant free-living organisms on planet Earth (Logares et al., 2010; Morris et al., 2002). These two extremely abundant microbes – one an autotroph and one a heterotroph, and both with streamlined genomes – have likely co-evolved a complex set of metabolic interactions to cope with the challenges of surviving in an oligotrophic environment. Contained within these interactions is important information regarding the cycling of many of Earth's most biologically important elements, including carbon, nitrogen and sulfur, however, to date, these two organisms have not been studied together in the laboratory – largely because SAR11 is legendarily difficult to culture. Given the widespread co-occurrence of *Prochlorococcus* and SAR11 groups in the global oceans, we investigated the effect of prochlorosins on a cultured representative of the SAR11 group (strain HTCC7211[1]).

[1] This strain was obtained from Steve Giovannoni, OSU, through collaborative agreement.

We first simply investigated whether the presence of one of the 29 prochlorosins made by *Prochlorococcus* MIT9313 (Pcn 2.11) could have an effect on the growth of SAR11 cells (Fig. 3-7). Growth experiments were conducted in AMS1, a defined artificial seawater medium recently developed for the cultivation of marine SAR11 isolates (Carini et al., 2013). Cells treated with Pcn 2.11 did not suffer any defect in their growth (Fig. 3-7). On the contrary, they displayed a 2-fold enhancement in maximum yield compared to the cells treated with the linear version of the peptide (which also had a Cys to Ser substitutions to prevent spontaneous disulfide bond formation) and the untreated control. The boost in yield displayed by the Pcn 2.11-treated cells suggests that prochlorosins might be metabolized by SAR11 and used as a nutrient source. Also, this result is consistent with our previous experiments, in which we found that prochlorosins do not inhibit the growth of other marine bacteria, and further weakens the hypothesis that prochlorosins might be acting as lantibiotics in the marine environment

Why might cyclized prochlorosins promote the growth of SAR11 while the linear form would not? Cultivation of representatives of the SAR11 clade revealed unusual nutritional requirements for reduced sulfur compounds and organic nitrogen, which have been proposed to be the result of selection for efficient resource utilization in nutrient-limited oceanic regions (Morris et al., 2002). Unlike many other known free-living bacteria in aerobic marine habitats, members of the SAR11 clade lack a complete pathway for assimilatory sulfate reduction and it has been shown that they rely exclusively on reduced sulfur compounds that originate from other plankton (Tripp et al., 2008). This relatively unique requirement for an exogenous source of reduced sulfur for growth seems particularly surprising, as reduced sulfur concentrations in the oligotrophic ocean are 10^6 to 10^7 times lower than that of sulfate (Tripp, 2013). Yet recent studies have shown that SAR11 is not alone in this requirement; marine group II archaea and SAR86 also appear to have reduced sulfur-dependent metabolism (Dupont et al., 2012; Iverson et al., 2012). It is noteworthy also that the hyper-streamlined genome of SAR11 also lacks the canonical genes necessary to synthesize glycine from glyceraldehyde-3-phosphate, and it thus appears to at least have conditional auxotrophy for glycine; glycine has been shown to be the preferred source of organic nitrogen for SAR11 under laboratory conditions (Tripp et al., 2009).

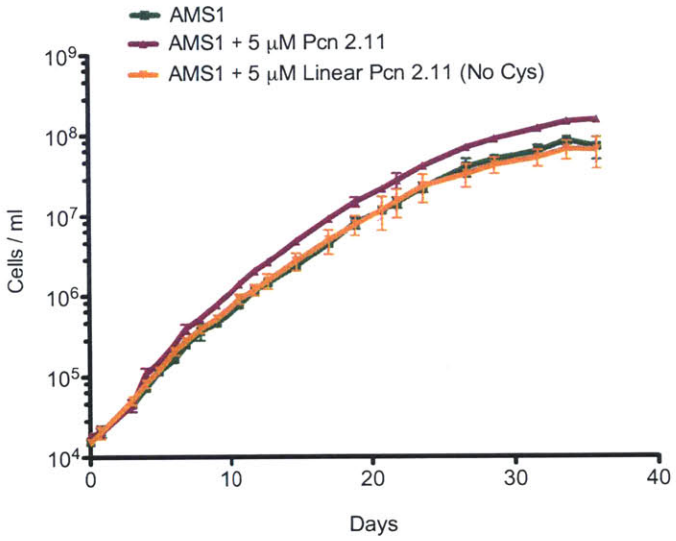

Figure 3-7: Influence of prochlorosins on the growth of SAR11.
Growth of SAR11 strain HTCC7211 in an artificial seawater medium (AMS1) amended with methionine (10uM), glycine (50uM) and pyruvate (50uM) to meet the unique organic sulfur, nitrogen, and carbon requirements of SAR11. Green: cells amended with a small quantity of sterile water to maintain consistent total volume. Orange: cells amended with ca. 5uM linear prochlorosin 2.11, which does not contain Cys residues. Purple: cells amended with ca. 5uM prochlorosin 2.11. Points are the average density of either triplicate or duplicate cultures and error bars indicate ± 1.0 s.d. When error bars are not visible is because they are smaller than the size of the symbols.

So where do prochlorosins fit in this picture? Lanthipeptides can be simplistically viewed as amino acid polymers that are tightly packed by lanthionine bridges. These modifications transform them into soluble, hydrolysis-resistant compounds, which make them stable nutrient sources that would have high nutritional value in the oligotrophic environments where SAR11 dominates. Interestingly the two most frequent amino acids in the prochlorosin core peptides from different genomes and metagenomes are glycine and cysteine (Fig. 3-8A). Consequently, we hypothesize that prochlorosins might fulfill an important part of the unique organic nutritional requirements of SAR11 in two ways:

(1) Prochlorosins can serve as a source of reduced sulfur. Thiol groups present in cysteine residues are converted into lanthionine (or methyl-lanthionine) bridges by the action of the lanthionine synthetase. This modification of the thiol groups within core peptides creates a stable peptide that can harbor between 1 and up to 10 sulfur atoms, depending on the prochlorosin. The stable thiol groups in prochlorosins might therefore serve as an important sulphur-containing substrate for these cells.

(2) Prochlorosins may act as a source of organic nitrogen for SAR11. While almost any peptide contains a significant amount of organic carbon and nitrogen, prochlorosins are particularly enriched in glycine, the preferred organic nitrogen source for SAR11 to supply its auxotrophy. Importantly, the nutritional value of prochlorosins cannot be matched by any random peptide derived from the *Prochlorococcus* proteome; Cys residues are ~5-fold more frequent and Gly residues are ~2-fold more frequent in prochlorosins compared with the average amino acid composition of proteins in *Prochlorococcus* (Fig. 3-8B).

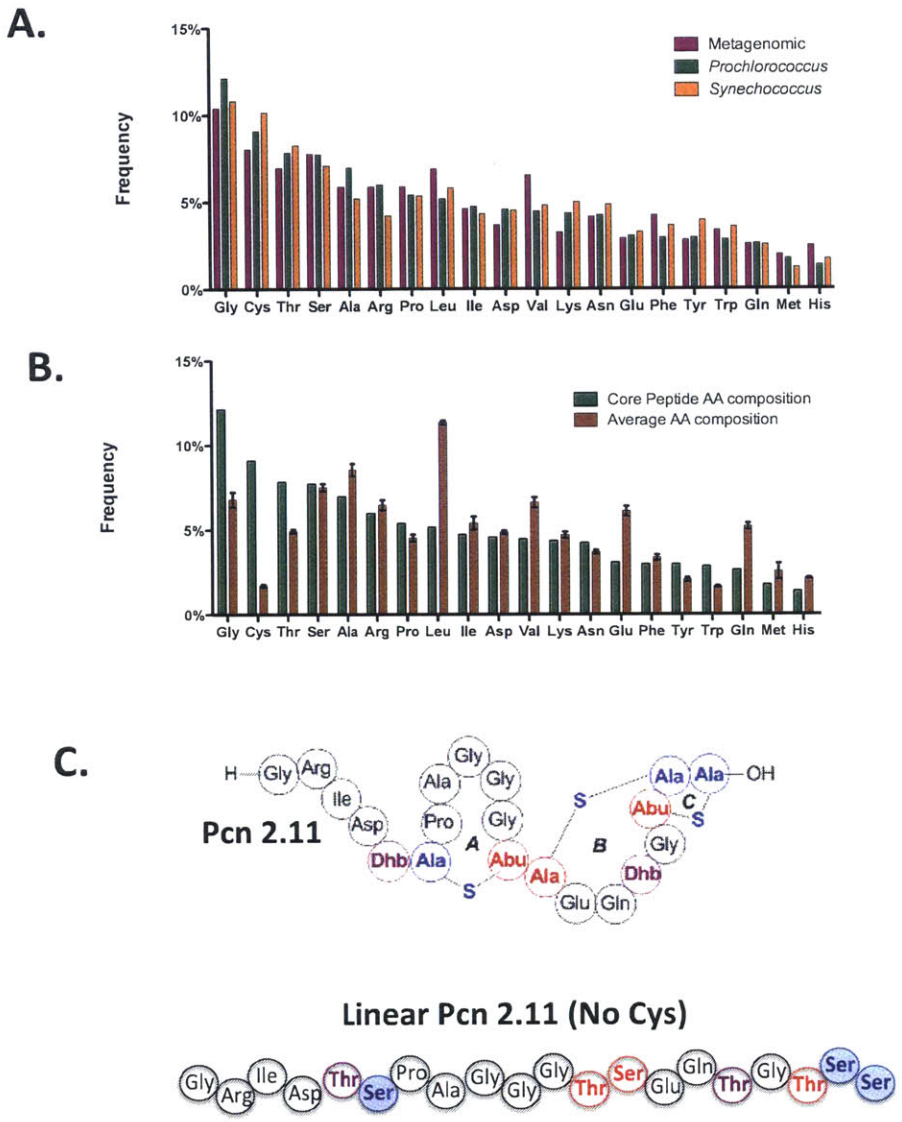

Figure 3-8: Amino acid content of prochlorosins.
A. Amino acid composition of the prochlorosin core peptides found in *Prochlorococcus* and *Synechococcus* genomes, and in marine picocyanobacterial shotgun sequences from metagenomic databases. **B**. Comparison of the amino acid composition of prochlorosins from *Prochlorococcus* genomes and the average amino acid composition of lanthipeptide-encoding genomes of *Prochlorococcus*. **C**. Structural conformation of the peptides used in the assays reported here. Lanthionines are represented as Ala-S-Ala and methyl-lanthionines as Abu-S-Ala. Dehydroalanine (Dha), 2-aminobutyric acid (Abu) In the linear version of Pcn 2.11, the cysteine residues (precursor of the final Ala residues) have been replaced by a serine. Also, the serine and threonine residues are not dehydrated in the linear version.

To test the hypothesis that prochlorosins might serve as a particularly attractive source of reduced sulfur and glycine for SAR11, we omitted all vitamins, glycine and methionine from the defined medium to create reduced sulfur and organic nitrogen limited conditions, and then provided prochlorosin Pcn 2.11 as a potential source of these nutrients. As a control, we also tested the linear version of Pcn 2.11 that lacks cysteine residues, and therefore does not provide any sulfur to the system (Fig. 3-8C). Cultures that were amended with Pcn 2.11 grew to a higher cell density than the control treatments, demonstrating that SAR11 can take up and metabolize the reduced sulfur present in prochlorosin 2.11 to support growth (Fig. 3-9). While additional experiments are needed to verify these initial findings (see Chapter 5), these proof-of-principle results suggest that lanthipeptides might play an unexpected role in the marine environment; they may be a stable source of reduced sulfur and organic nitrogen for groups of heterotrophic bacterioplankton that are highly abundant in oligotrophic regions.

What are the implications of prochlorosins as a "public good" in the ocean? The synthesis and secretion of each prochlorosin is a considerable energetic investment; it costs one ATP for every Ser/Thr and Cys to be converted into a lanthionine (or methyl-lanthionine), in addition, the transport machinery also uses ATP to export the modified peptides to the extracellular environment. Consequently, the benefit that would bring a biological function for prochlorosins that is independent of their condition as a nutrient source, must offset not only the cost of the production but also has to factor in the possibility that these molecules are degraded along the way by prevalent co-existing groups of SAR11 bacteria. On the other hand, their biological function could be related to this previously unknown dimension of lanthipeptides as compounds of high nutritional value that could mediate specific metabolic interactions between prochlorosin-producing picocyanobacteria and abundant groups of oligotrophic heterotrophic bacteria. In such case, it will be necessary to determine how specific these interactions might be and what could be the bilateral effects on the metabolism.

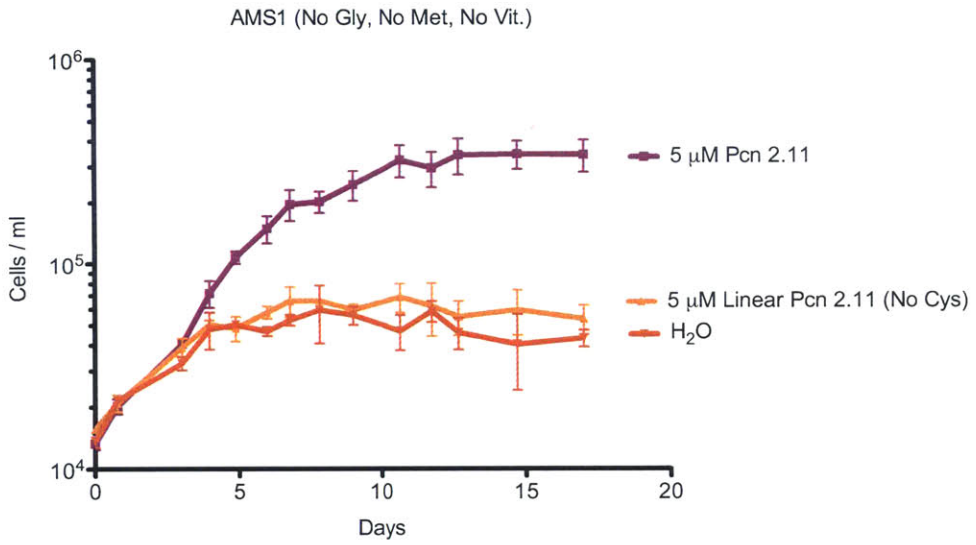

Figure 3-9: Lanthipeptides as a source of reduced sulphur for SAR11
Growth of SAR11 strain HTCC7211 in an artificial seawater medium (AMS1 minus vitamins) amended with pyruvate (50um) to meet the organic carbon requirements of SAR11. Vitamins were withheld to remove all potential sources of reduced sulfur and organic nitrogen. Red: cells amended with a small quantity of sterile water to maintain consistent total volume. Orange: cells amended with ca. 5uM linear prochlorosin 2.11, containing no reduced sulfur source. Purple: cells amended with ca. 5uM prochlorosin 2.11. Points are the average density of either triplicate or duplicate cultures and error bars indicate ± SD.

CONCLUSION

Investigation of the regulation of the prochlorosin biosynthesis pathway indicated that, under laboratory conditions, the expression of the prochlorosin precursor peptide genes is not dependent on cell density and the maximal levels of expression occur during exponential phase of growth. Also, the biosynthesis of prochlorosins does not seem to be autoregulatory; the addition of 4 recombinant prochlorosins did not modify the expression of the genes involved in the production of the prochlorosins, but did modulate the expression of genes in *procA*-associated loci. Thus, the differences between the regulatory mechanisms of the prochlorosin biosynthesis pathway and the pathways of well-characterized lanthipeptides may underlie distinct biological roles. In an initial exploration of the biological activity of these peptides, we did not observe any significant inhibitory effect of a subset of prochlorosins when tested against a small panel of *Prochlorococcus* and marine heterotrophic bacteria. This result does not necessarily rule out antimicrobial activity, but does challenge us to think about alternative roles that are more relevant to the dilute, nutrient-limited environment where these peptides are being produced. For instance, we demonstrate that the unique amino acid composition of the prochlorosins make them a valuable nutrient source that can supply the unique requirements for reduced sulfur and organic nitrogen that a cultured strain of the SAR11 clade needs. This finding opens the possibility that lanthipeptides, traditionally considered as secondary metabolites, might play a key role in the primary metabolism of important groups of bacterioplankton that co-exist with *Prochlorococcus* in the oligotrophic ocean.

MATERIALS AND METHODS

qPCR Determination of expression of prochlorosin biosynthesis genes

Triplicate cultures of *Prochlorococcus* MIT9313 were grown in Pro99 medium (Moore et al., 2007) at an irradiance of 20μE $m^{-2}s^{-1}$ and a temperature of 24°C. Sampling of 2 ml of the cultures was done during early exponential (day 4), mid exponential (day 6), early stationary (day 9) and later stationary (day 11) phases. RNA was extracted using the Zymo Quick RNA (Zymo) kit and residual DNA was removed using Turbo DNase (Ambion). 100 ng of DNA-free RNA were reverse transcribed using iScript (Bio-rad). Real-time PCR reactions were performed in duplicate with the resulting cDNA diluted 5-fold in 10mM Tris, pH 8, using the Qiagen SYBR green kit following the manufacturer conditions. Primers used are listed below. The LPg2 pair of primers allows the detection of the following *procA* genes: *procA* 1.1, *procA* 1.2, *procA* 1.5, *procA* 1.7, *procA* 2.1, *procA* 2.4, *procA* 2.5, *procA* 2.6, *procA* 2.9, *procA* 2.11, *procA* 3.3 and *procA* 4.3. The expression levels were calculated relative to the normalizer gene *rnpB*.

Name	Sequence	Annealing Temp. (°C)
ProcMup	5'CGACGCTTCCTCTACGAC	50
ProcMlo	5'AATTGCCAGTTTGGATTCTC	
LPg2g4Up	5'ATGTCAGAAGAACAACT	52
LPg2Lo	5'GCAATAGCAACAACA	
ABCTup	5'ATGTCACTTGGCACCTTAGT	50
ABCTlo	5'TAACGGGTCATCTTCACTTT	

Quantification of whole-transcriptome changes upon addition of prochlorosins

Quadruplicate cultures of *Prochlorococcus* MIT9313 were grown in 50 ml of Pro99 medium at an irradiance of 20μE $m^{-2}s^{-1}$ and a temperature of 24°C. When cultures reached mid-exponential phase of growth, two of them were amended with 50 μl of a cocktail of four prochlorosins for a resulting final concentration of 430 nM. The two other cultures were amended with 50 μl of 50mM HEPES buffer pH 8 as control treatment. Sampling of 12 ml of each culture was performed after 12 and 36 hours after

the addition. Cells were collected by centrifugation (8500 rpm) for 8-10 minutes, then were re-suspended in 1ml of RNA storage buffer (200mM sucrose, 10mM NaOAc pH 5.2, 5mM EDTA pH5.2), flash frozen in liquid nitrogen, and stored at -80°C. The mirVana miRNA kit (Ambion) was used to extract RNA following the manufacture's conditions. RNA-Seq Illumina libraries were constructed using a method adapted from (Giannoukos et al., 2012). Briefly, strand-specific RNA-Seq libraries were prepared using the dUTP second-strand approach, the double-stranded cDNA was prepared and enrichment by amplifying fragments ligated to Y-adapters, subsequently rRNA was removed using a duplex-specific normalization approach and then rRNA-depleted libraries were barcoded and sequenced using the Illumina GA-II platform. Low quality sequence reads were removed from the raw data using quality_trim (from the CLC Assembly Cell package, CLC bio) with default settings. Alignment to the reference genome of *Prochlorococcus* MIT9313 was done with the Burrows- Wheeler Aligner (Li and Durbin, 2010), and resulting alignment files parsed with aid of the SAMtools package (Li et al., 2009). Differentially expressed genes were calculated using DEseq (Anders and Huber, 2010).

Bioactivity assays on *Prochlorococcus* Strains

All *Prochlorococcus* cultures were grown in Pro99 medium under continuous light and a temperature of 24°C. *Prochlorococcus* MIT9313, MIT9303 and NATL2A were grown at an irradiance of 20μE $m^{-2}s^{-1}$. *Prochlorococcus* MED4 was grown at an irradiance of 35μE $m^{-2}s^{-1}$. Each *Prochlorococcus* culture was grown in 50ml glass tubes to early exponential phase, at this point were transferred to 96-well plates were they were allowed to acclimate to the new vessel. Each one of the triplicate treatments was added as a small volume of 5 μl so the salinity conditions in the wells is not drastically changed. Pcn Mix consisted in 1 μg/ml of Pcn1.7, Pcn 2.11, Pcn 3.2 and Pcn 3.3. Control treatments were the linear version of Pcn3.2 at 1 μg/ml, a buffer control consisting of 5 μl of 50 mM HEPES and antibiotic control of Kanamycin at 50 μg/ml. Fluorescence was used as a proxy for growth and was quantified every 24 hours.

Bioactivity assays on heterotrophic strains

Each heterotrophic bacterial culture was grown overnight at room temperature in ProMM medium (Moore et al., 2007), cells were diluted to an OD600 of 0.1 and transferred into 96-well plates. Each one of the triplicate treatments was added as a small volume of 5 µl so the salinity conditions in the wells are not drastically changed. Pcn Mix consisted in 10 µg/ml of Pcn1.7, Pcn 2.11, Pcn 3.2 and Pcn 3.3. Control treatments were the linear version of Pcn3.2 at 10 µg/ml, a buffer control consisting of 5 µl of 50 mM HEPES and antibiotic control of Kanamycin at 50 µg/ml. Plates were incubated for 18 hours in the SynergyOne plate reader (Biotek) were OD600 measurements were taken every 30 minutes.

SAR11 culture conditions

SAR11 strain HTCC7211 cultures were grown artificial seawater medium AMS1 in acid-washed and autoclaved polycarbonate flasks at 22C with shaking at 60 r.p.m under a 12-h/12-h light/dark cycle as indicated in (Carini et al., 2013). Cells were stained with SYBR Green I (Molecular Probes, Inc., Eugene, OR, USA) and cell density was determined by flow cytometry using a Guava Technologies instrument (Millipore, Billerica, MA, USA). Organic nutrient replacements from the AMS1 media were performed as described in the text.

REFERENCES

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

Bader, M.W., Navarre, W.W., Shiau, W., Nikaido, H., Frye, J.G., McClelland, M., Fang, F.C., and Miller, S.I. (2003). Regulation of Salmonella typhimurium virulence gene expression by cationic antimicrobial peptides. Mol Microbiol *50*, 219-230.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015). *Prochlorococcus*: the structure and function of collective diversity. Nat Rev Microbiol *13*, 13-27.

Carini, P., Steindler, L., Beszteri, S., and Giovannoni, S.J. (2013). Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. ISME J *7*, 592-602.

Chatterjee, C., Paul, M., Xie, L., and van der Donk, W.A. (2005). Biosynthesis and mode of action of lantibiotics. Chem Rev *105*, 633-684.

Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.J., Richter, R.A., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J.D., Haft, D.H., *et al.* (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J *6*, 1186-1199.

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. Nat Rev Microbiol *10*, 538-550.

Field, D., Hill, C., Cotter, P.D., and Ross, R.P. (2010). The dawning of a 'Golden era' in lantibiotic bioengineering. Mol Microbiol *78*, 1077-1087.

Giannoukos, G., Ciulla, D.M., Huang, K., Haas, B.J., Izard, J., Levin, J.Z., Livny, J., Earl, A.M., Gevers, D., Ward, D.V., *et al.* (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. Genome Biol *13*, R23.

Guljamow, A., Delissen, F., Baumann, O., Thunemann, A.F., and Dittmann, E. (2012). Unique properties of eukaryote-type actin and profilin horizontally transferred to cyanobacteria. PLoS One *7*, e29926.

Haft, D.H., Basu, M.K., and Mitchell, D.A. (2010). Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. BMC Biol *8*, 70.

Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., and Armbrust, E.V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science *335*, 587-590.

Kleerebezem, M. (2004). Quorum sensing control of lantibiotic production; nisin and subtilin autoregulate their own biosynthesis. Peptides *25*, 1405-1414.

Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P.J., Joewono, I., Rusch, D., Chisholm, S.W., and van der Donk, W.A. (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. Proc Natl Acad Sci U S A *107*, 10430-10435.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Little, A.E., Robinson, C.J., Peterson, S.B., Raffa, K.F., and Handelsman, J. (2008). Rules of engagement: interspecies interactions that regulate microbial communities. Annu Rev Microbiol *62*, 375-401.

Logares, R., Brate, J., Heinrich, F., Shalchian-Tabrizi, K., and Bertilsson, S. (2010). Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). Mol Biol Evol *27*, 347-357.

McPhee, J.B., Lewenza, S., and Hancock, R.E. (2003). Cationic antimicrobial peptides activate a two-component regulatory system, PmrA-PmrB, that regulates resistance to polymyxin B and cationic antimicrobial peptides in Pseudomonas aeruginosa. Mol Microbiol *50*, 205-217.

Miller, M.B., and Bassler, B.L. (2001). Quorum sensing in bacteria. Annu Rev Microbiol *55*, 165-199.

Moore, L., Coe, A., Zinser, E.R., Saito, M.A., Sullivan, M.B., Lindell, D., Frois-Moniz, K., Waterbury, J., and Chisholm, S.W. (2007). Culturing the marine cyanobacterium *Prochlorococcus*. Limnology and Oceanography: Methods *5*, 353-362.

Morris, J.J., Lenski, R.E., and Zinser, E.R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio *3*.

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. Nature *420*, 806-810.

Mukherjee, S., and van der Donk, W.A. (2014). Mechanistic studies on the substrate-tolerant lanthipeptide synthetase ProcM. J Am Chem Soc *136*, 10450-10459.

Piddock, L.J. (2006). Multidrug-resistance efflux pumps - not just for resistance. Nat Rev Microbiol *4*, 629-636.

Seth, E.C., and Taga, M.E. (2014). Nutrient cross-feeding in the microbial world. Front Microbiol *5*, 350.

Shi, Y., Yang, X., Garg, N., and van der Donk, W.A. (2010). Production of Lantipeptides in Escherichia coli. J Am Chem Soc.

Staron, P., Forchhammer, K., and Maldener, I. (2011). Novel ATP-driven pathway of glycolipid export involving TolC protein. J Biol Chem *286*, 38202-38210.

Staron, P., Forchhammer, K., and Maldener, I. (2014). Structure-function analysis of the ATP-driven glycolipid efflux pump DevBCA reveals complex organization with TolC/HgdD. FEBS Lett *588*, 395-400.

Tang, W., and van der Donk, W.A. (2012). Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria. Biochemistry *51*, 4271-4279.

Tripp, H.J. (2013). The unique metabolism of SAR11 aquatic bacteria. J Microbiol *51*, 147-153.

Tripp, H.J., Kitner, J.B., Schwalbach, M.S., Dacey, J.W., Wilhelm, L.J., and Giovannoni, S.J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature *452*, 741-744.

Tripp, H.J., Schwalbach, M.S., Meyer, M.M., Kitner, J.B., Breaker, R.R., and Giovannoni, S.J. (2009). Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. Environ Microbiol *11*, 230-238.

Willey, J.M., and van der Donk, W.A. (2007). Lantibiotics: peptides of diverse structure and function. Annu Rev Microbiol *61*, 477-501.

Zhang, Q., Yu, Y., Velasquez, J.E., and van der Donk, W.A. (2012). Evolution of lanthipeptide synthetases. Proc Natl Acad Sci U S A *109*, 18361-18366.

SUPPLEMENTARY MATERIAL

Locus_tag	Name	Differential Expression Ratio	*p*-value
	Differentially expressed after 12h		
PMT_0168	possible Negative factor, (F-Protein)	3.4	1.20E-82
PMT_0169	possible Chitin synthase CDS	4.4	0
PMT_2324	Hypothetical protein CDS	1.9	0
PMT_2325	Hypothetical protein CDS	2	0
PMT_0246	possible Profilin CDS	1.7	2.40E-07
PMT_0265	two component sensor histidine kinase CDS	2.2	1.30E-04
PMT_0327	FAD linked oxidase, N-terminal CDS	2	3.50E-03
PMT_0328	hypothetical protein CDS	2.7	0
PMT_0382	possible Signal peptidase I CDS	1.9	4.30E-09
PMT_0483	possible Malic enzyme CDS	2.3	0
PMT_0631	hypothetical CDS	2.2	0
PMT_0915	Bacterial outer membrane protein CDS	2.3	3.60E-09
PMT_0916	hypothetical CDS	2.5	2.10E-19
PMT_0923	hypothetical CDS	1.8	2.00E-07
PMT_2496	Hypothetical protein CDS	1.9	2.00E-13
PMT_0924	possible Carbamoyl-phosphate synthase	3.2	0
PMT_0925	hypothetical CDS	2	2.10E-31
PMT_0928	hypothetical CDS	2	8.30E-09
PMT_2502	Hypothetical protein CDS	2	2.60E-14
PMT_0929	Hemolysin-type RTX. Peptidase	3.7	3.40E-33
PMT_0943	hypothetical CDS	2	1.40E-56
PMT_2608	Hypothetical protein CDS	2.8	3.90E-19
PMT_2609	similar to P9303 10091 CDS	2	0
PMT_2620	Hypothetical protein CDS	2.1	4.70E-05
PMT_1223	hypothetical CDS	2.5	7.40E-34
PMT_2648	Hypothetical protein CDS	3	0
PMT_1224	Biotin/lipoate A/B protein ligase family CDS	2.9	2.50E-19
PMT_1389	hypothetical protein CDS	3.4	3.90E-05
PMT_2712	Hypothetical protein CDS	2.5	4.00E-27
PMT_2713	Hypothetical protein CDS	2.1	0
PMT_2714	Hypothetical protein CDS	2.5	9.00E-26
PMT_1570	conserved hypothetical protein CDS	2.8	1.00E-48
PMT_1571	hypothetical CDS	3.1	2.00E-64
PMT_1572	conserved hypothetical protein CDS	4.1	1.40E-57
PMT_1573	ABC-transporter component CDS	5.1	4.40E-35
PMT_1574	ABC transporter component CDS	5.5	8.50E-21

PMT_1576	ABC transporter, ATP binding protein CDS	2.7	0
PMT_1577	hypothetical protein CDS	3.2	0
PMT_2823	Conserved hypothetical protein CDS	2.1	6.40E-33
PMT_1940	possible Peptidase family C9 CDS	2.3	8.50E-15
PMT_1941	hypothetical protein CDS	2.3	6.60E-03
PMT_2137	hypothetical CDS	2.1	3.70E-19
PMT_2904	Hypothetical protein CDS	2.2	8.10E-08
PMT_2138	conserved hypothetical protein CDS peptidase	3.5	0
PMT_2283	Hypothetical protein CDS	-2.5	2.00E-07
PMT_2287	Conserved hypothetical protein CDS	-3.7	4.10E-03
PMT_0956	possible Kinesin motor domain CDS	-2.3	1.80E-15
PMT_0959	possible Kinesin motor domain CDS	-2.5	2.00E-17
PMT_0962	possible Kinesin motor domain CDS	-2.5	4.80E-18
PMT_0971	possible bZIP transcription factor CDS	-2.1	1.40E-08
PMT_2536	Conserved hypothetical protein CDS	-2.1	3.10E-04
PMT_0983	possible Kinesin motor domain CDS	-2.5	3.20E-17
PMT_0990	hli5 CDS	-2.4	1.30E-144
PMT_0991	hypothetical transmembrane protein CDS	-2.2	4.40E-10
PMT_2555	Hypothetical protein CDS	-2.1	5.90E-14
PMT_1564	possible Phosphate-binding protein CDS	-2.2	8.80E-10
PMT_1582	conserved hypothetical CDS	-2.4	5.10E-03
PMT_2809	Hypothetical protein CDS	-2.4	1.20E-03
PMT_2876	Hypothetical protein CDS	-2.1	5.40E-05
PMT_2116	hypothetical CDS	-2.1	8.70E-18
	Differentially expressed after 36h		
PMT_0956	possible Kinesin motor domain CDS	-2.1	1.80E-13
PMT_0959	possible Kinesin motor domain CDS	-2.5	2.30E-15
PMT_0962	possible Kinesin motor domain CDS	-2.2	4.60E-15
PMT_0971	possible bZIP transcription factor CDS	-2	1.20E-11
PMT0983	possible Kinesin motor domain	-2.1	4.40E-10
PMT2240	Formate and nitrite transporters	-2.5	5.90E-14
PMT2241	possible Homeobox domain	-2.2	8.80E-10
PMT1152	hli9 possible high light inducible protein	-2	2.00E-17
PMT1153	hli6 possible high light inducible protein	-2.1	3.40E-14

Table S3-1. Differentially expressed genes upon the addition of recombinant prochlorosins.

CHAPTER 4

Amino Acid Toxicity and Tolerance in *Prochlorococcus*

Andres Cubillos-Ruiz [1, 2]

Audrey Olshefsky [2, 3]

Sallie W. Chisholm [2]

[1] Microbiology Graduate Program, Massachusetts Institute of Technology.

[2] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

[3] UROP program, Massachusetts Institute of Technology.

SUMMARY

While studying the potential utilization of amino acids as a nitrogen source for *Prochlorococcus,* we discovered an unexpected inhibition of growth when amino acids were present in the culture media. We decided to further characterize this phenomenon and observed that after an initial period of inhibition some of the cultures were able to resume growth. The resulting strains were no longer inhibited by the presence of the amino acid. In addition, the newly emerged strains were also capable of tolerating the inhibitory effect of other amino acids to which they had not been exposed. The amino acid tolerance could be stably inherited across generations even in the absence of the amino acids in the media. Strikingly, whole-genome re-sequencing of the strains used in this study indicated that there were no genetic differences between amino acid-tolerant and parental strains, which suggest that the observed amino acid-tolerant strains are phenotypic variants that arise in response to the inhibitory effect of the amino acids. We discuss how this could represent a case of epigenetic inheritance in *Prochlorococcus.*

BACKGROUND AND MOTIVATION

Prochlorococcus utilize a diverse set of pathways for nitrogen assimilation. Several routes of inorganic nitrogen assimilation have been extensively studied and are thought to be the main mechanisms that provide the cells with nitrogen (Garcia-Fernandez et al., 2004). Nevertheless, field studies have shown that *Prochlorococcus* is able to take up multiple organic compounds, including amino acids (Zubkov et al., 2003). This agrees with genomic analyses that predict the presence of various metabolic pathways that would enable the assimilation of nitrogen from free amino acids and oligopeptides (Garcia-Fernandez et al., 2004). These findings highlight the importance of studying the potential use of organic compounds as an alternative source of nutrients for primary metabolism. It has been demonstrated that *Prochlorococcus* is able to take up amino acids and that this process is light enhanced (Church et al., 2004). Nevertheless, the metabolic processes that underlie the use of exogenous amino acids and the consequences that it might have for nitrogen nutrition and fitness advantage remain unknown. Consequently, the characterization of the pathways for the use of amino acids in *Prochlorococcus* is critical to shed the light on the importance of organic nitrogen nutrition in the ecological success of this bacterium in oligotrophic environments.

Although our initial interest was to identify conditions under which amino acids could provide a growth advantage for *Prochlorococcus*, we discovered that some amino acids had an inhibitory effect on *Prochlorococcus* NATL2A. Moreover, we observed the emergence of strains that could resume growth after an initial period of inhibition. Intrigued by this phenomenon, we decided to further characterize it with the hope that it could bring insights into unique physiological characteristics of *Prochlorococcus*. Here we describe the phenomenon of amino acid toxicity and tolerance on *Prochlorococcus* strain NATL2A.

RESULTS AND DISCUSSION

Effect of amino acids in the growth of *Prochlorococcus* NATL2A

To determine whether *Prochlorococcus* NATL2A can use amino acids as the sole nitrogen source to support growth, we subjected the cells to a 48h period of nitrogen starvation and then provided different amino acids to determine if the cells could assimilate them and rescue growth (Fig. 4-1). Some of the amino acids provided a mild rescue for the first 1-2 days, but, compared to the rescue by the inorganic nitrogen source (NH_4Cl), none of the 20 amino acids rescued the cells from the nitrogen starvation to provide a continued robust growth (Fig. 4-1). This experiment indicates that *Prochlorococcus* NATL2A cannot directly assimilate the nitrogen contained in the amino acids to supply its nitrogen requirements for growth. However, the differences in the dynamics of the rescue curves between the no addition control and the amino acid treatment suggests that the amino acids might have another effect in the growth. For instance, the addition of cysteine rapidly increased the bulk fluorescence of the culture that was then followed by a sharp decrease. In other instances, the amino acid treatment yielded a lower fluorescence than the no rescue control (Fig. 4-1).

To investigate the effect of amino acids on the growth of *Prochlorococcus* NATL2A, we added the amino acids to cells growing under nutrient-replete conditions and monitored their growth. Surprisingly, we discovered that 15 out of the 20 amino acids had a negative effect on the growth of *Prochlorococcus* NATL2A (Fig. 4-2). The addition of arginine, leucine, glutamate, aspartate or phenylalanine did not affect the cultures, but the other 15 amino acids led to a steady decrease in bulk fluorescence, suggesting inhibition of growth[2] (Fig. 4-2). To further characterize the apparent inhibitory effect of the amino acids, we selected a subset of 5 amino acids to test whether there is a concentration-dependent effect. They not only did not display the same

[2] There can be differences between cell number and culture fluorescence under non-steady state conditions. Cell density determination will be performed to confirm that the decrease in bulk fluorescence is directly related to inhibition of growth.

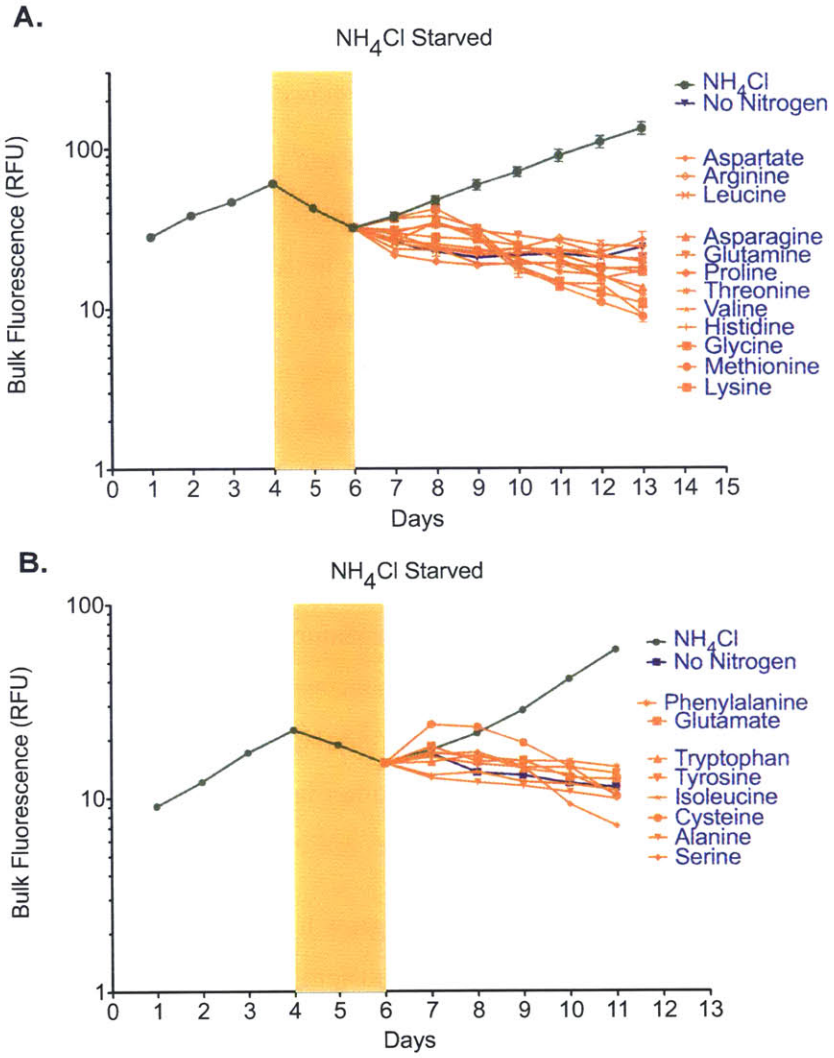

Figure 4-1: Effect of amino acid addition on cells of *Prochlorococcus* NATL2A under nitrogen-deplete conditions.
The evaluation of the nitrogen starvation rescue by the total set of 20 amino acids was performed in two independent experiments: **A.** A set of 12 amino acids. **B.** The remaining 8 amino acids. In both instances (**A** and **B**), following a period of 48h of nitrogen starvation (yellow area) *Prochlorococcus* NATL2A cultures were challenged to use amino acids as the only nitrogen source. All the amino acids (orange) and the inorganic nitrogen control (green) were added at a final concentration of 0.8 mM that is equivalent to the standard nitrogen content of the Pro99 medium. The no nitrogen rescue control is presented in red. Bulk culture chlorophyll fluorescence was used a proxy for culture biomass. Data points correspond to the average triplicate cultures and error bars represent the SD.

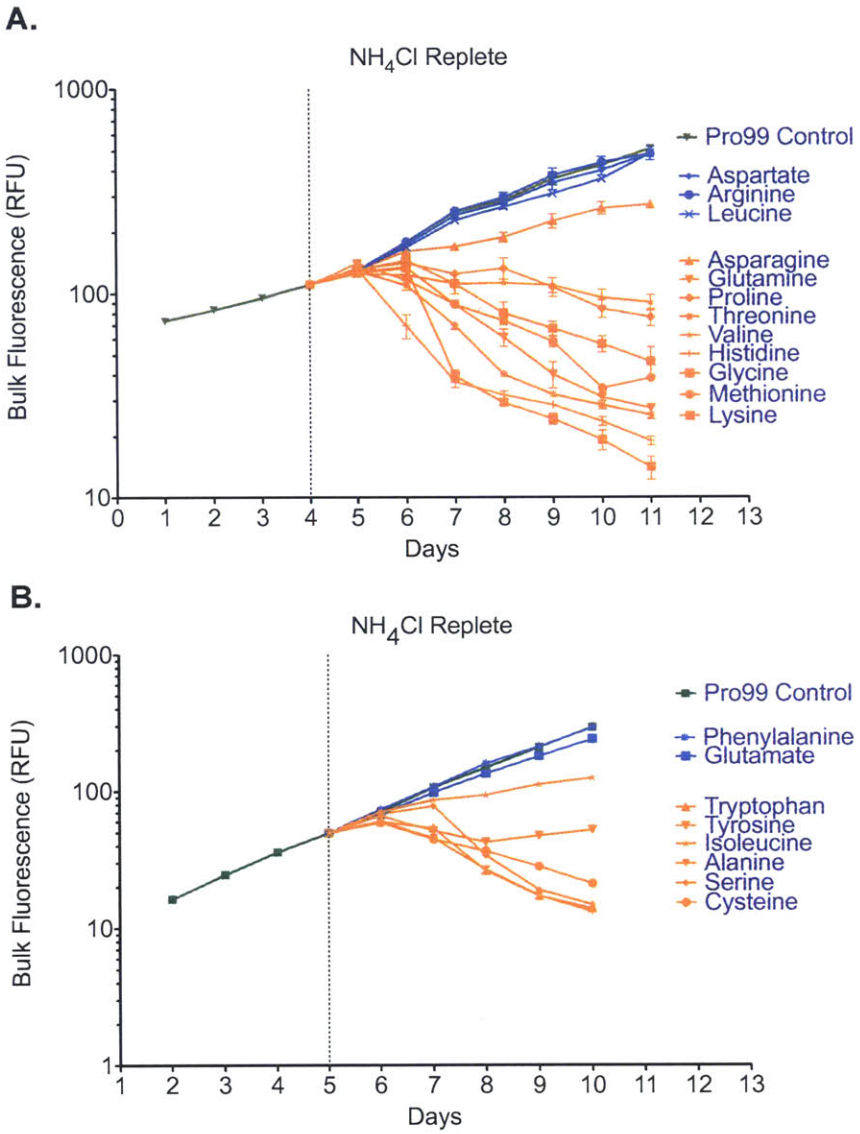

Figure 4-2: Effect of amino acid addition to cells of *Prochlorococcus* NATL2A under nitrogen-replete conditions.

The evaluation of the total set of 20 amino acids was performed in two independent experiments: **A.** First set of 12 amino acids. **B.** Second set of the remaining 8 amino acids. In both instances (**A** and **B**), amino acids were added at a final concentration of 0.8 mM during exponential growth (dashed line). Orange: Amino acid treatments that caused a defect in growth. Blue: Amino acids that did not interfere with growth. Green: No amino acid addition control. Bulk fluorescence was used a proxy for culture biomass. Data points correspond to the average triplicate cultures and error bars represent the SD.

minimum inhibitory concentration (MIC), defined here as the lowest amino acid concentration that causes a defect in growth compared with the untreated control, but they also had different dose-response patterns: lysine and proline display a typical concentration-dependent effect, whereas for isoleucine and valine there was no dose dependency at the concentrations used (Fig. 4-3).

One wonders why some amino acids are toxic to *Prochlorococcus* while others are not. We can see no clear relationship between toxicity and the properties of the amino acid side chain. For instance, although isoleucine has an MIC of 1 μM, leucine is not toxic even at 800 μM. Likewise, tyrosine only differs from phenylalanine in one hydroxyl group, however the former is toxic and the later is not. It is difficult to imagine what physiological process could turn these 15 amino acids into compounds that lead to inhibition of growth.

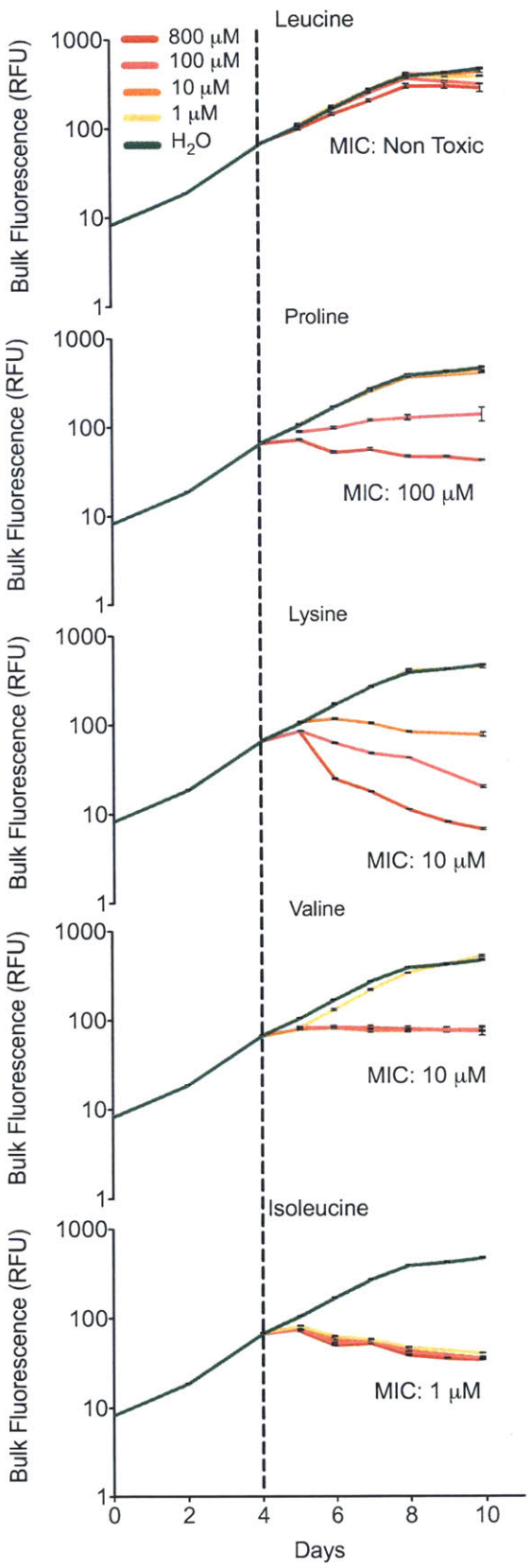

Figure 4-3: Dose-response effect of amino acid additions to cultures of *Prochlorococcus* NATL2A. Four different concentrations of one non-toxic (Leucine) and four toxic (Proline, lysine, Isoleucine and Valine) amino acids were added to exponentially growing cultures of *Prochlorococcus* NATL2A at the time indicated by the dashed line. Minimum Inhibitory Concentration (MIC) was determined as the lowest concentration at which the amino acid caused a clear decline in bulk culture fluorescence, Data points correspond to the average triplicate cultures and error bars represent the SD, some of them are smaller than the data points.

Emergence of Strains Tolerant to the Amino Acid Toxicity

In the course of these experiments we noticed that if we left the "inhibited" cultures to sit for some time, some of the strains were able to "recover" after the initial inhibition period. The first example of this occurred in the experiment with the addition of 0.8 mM lysine to the medium. Following a 7-day decline in bulk fluorescence, the cells recovered and grew to the same extent as the untreated cells grown in Pro99 medium alone (Fig. 4-4A). More significantly, those cells were able to grow as fast as the control when transferred into fresh Pro99 medium supplemented with lysine (Fig. 4-4B). This strain *Prochlorococcus* NATL2A that is no longer sensitive to the presence of lysine was named LysR. Further testing showed that although not inhibited by lysine, LysR still is unable to use of lysine as the only nitrogen source (Fig. 4-4C), thus these traits appear unrelated.

To discriminate whether this tolerance is a transient (i.e., acclimation) or stable (i.e., mutation/adaptation) trait, we removed the selective pressure of the amino acid from the LysR culture with the aim of relaxing the conditions that might have promoted acclimation; in the absence of this selection cells are expected to return to their default state (sensitive to lysine). On the other hand, if the tolerance is a stably inherited trait the removal of the selective pressure should not affect the tolerance to lysine when it is presented again. Interestingly, cells that grew in the absence of lysine for several transfers into new medium were still able to tolerate the toxicity when lysine was presented again, indicating that newly acquired the tolerance to lysine is a stably maintained trait (Fig. 4-5).

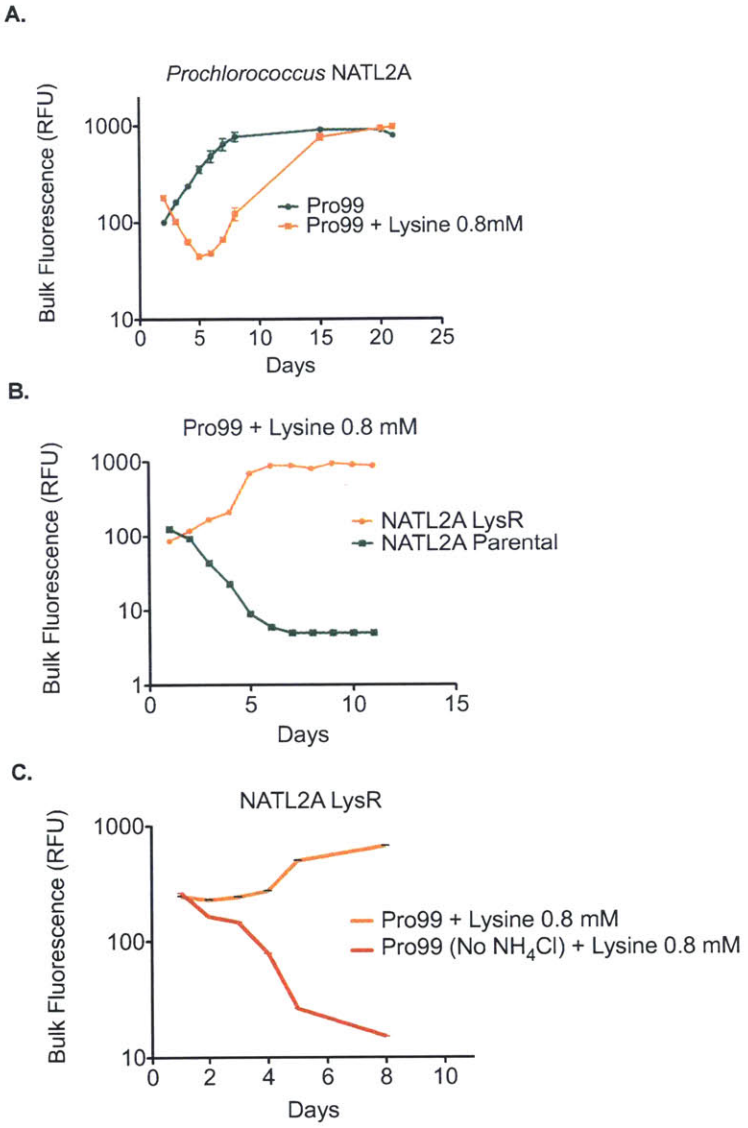

Figure 4-4: Emergence of amino acid-tolerant strains of *Prochlorococcus* NATL2A. **A.** Upon the addition of lysine to an exponentially growing culture of *Prochlorococcus* NATL2A a sharp decrease in the bulk culture fluorescence was observed, after 7 days the culture had decreased its fluorescence by 10-fold. After this inhibition period it was able to resume growth at rates comparable to the control. This culture was named LysR (Orange). **B.** The LysR cells that were transferred into fresh medium supplemented with 0.8 mM lysine were able to grow normally, while NATL2A Parental (unexposed to lysine) did not. **C.** The LysR transferred into Pro99 medium without NH_4Cl and supplemented with 0.8 mM lysine is not able to grow. This suggests that the incapacity of the assimilation of the nitrogen in the lysine might not be directly related to its toxic effect.

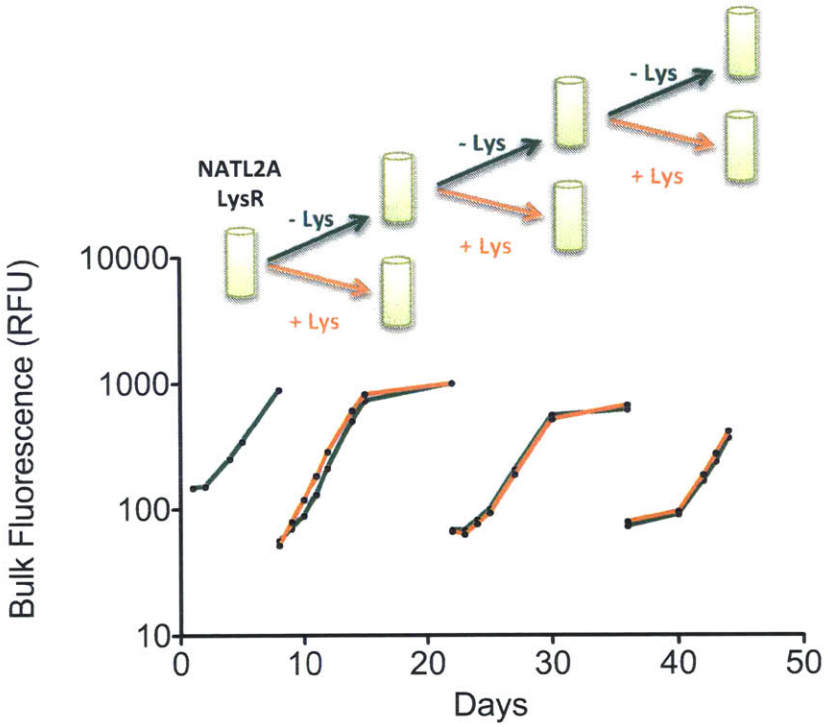

Figure 4-5: Lysine tolerance of LysR is maintained in the absence of the amino acid. LysR cells were removed from the media containing lysine and grown in Pro99 without lysine for four cycles (-Lys, green lines). At each cycle an aliquot of the cells was transferred into media containing 0.8 mM lysine (+Lys, orange line) to verify whether they still display the lysine tolerance. In every cycle, the cells grown in media with lysine behaved identically to cells grown in medium without lysine indicating that the tolerance is a stable trait that can be maintained in the absence of the selective pressure of the amino acid.

Extended-tolerance to other Amino Acids

To investigate whether this tolerance trait was exclusive to the inhibitory effect of lysine, cells from the LysR culture were tested for their ability to grow on media supplemented with other amino acids (Fig. 4-6). Interestingly, LysR cells also displayed tolerance to amino acids to which they had not been exposed. In this experiment, none of the amino acid treatments yielded the same inhibitory dynamics observed in the unexposed/parental strain, however the treatments with cysteine and histidine resulted in a delayed inhibition (Fig. 4-6). This extended-tolerance state is also stable over time and does not disappear when the cells are grown in the absence of amino acids; after six months of growing the LysR cells in the absence of lysine we re-tested growth on a subset of amino acids and found the same results (Fig. 4-7), suggesting that the underlying mechanism of tolerance to amino acid toxicity might be common for most of the amino acids and is not specific for singular amino acid structures.

The observation that the amino acid-tolerant state is stable over many generations in the absence of the selective pressure suggests that an inheritable trait might have occurred in the parental *Prochlorococcus* NATL2A strain, leading to tolerance not only to the original amino acid the cells were exposed to, but also to other amino acids. The lack of a genetic system for *Prochlorococcus* hinders the use of reverse genetics to screen for mutants that would display the same behavior as the LysR strain. We could use forward genetics, however, and compare the genome sequence the LysR strain with the parent strain.

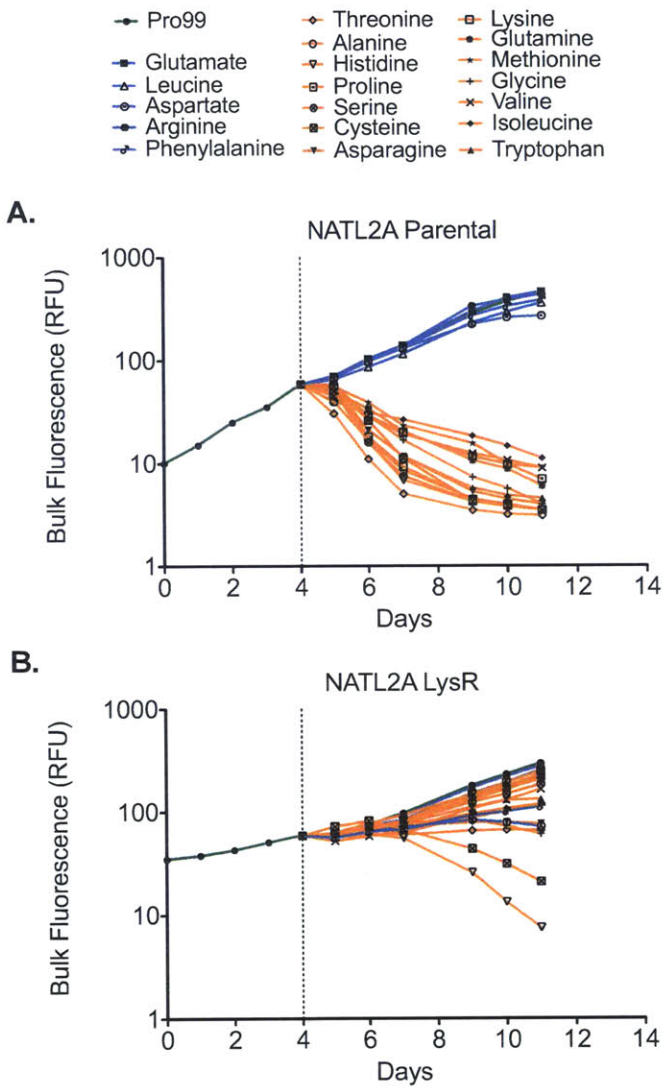

Figure 4-6: Extended-tolerance to other amino acids in the LysR strain.
The complete set of 20 amino acids were added to growing cells of the NATL2A parental strain and the LysR strain to a final concentration of 0.8 mM. The Parental strain (**A**) displayed the previously observed patterns of toxicity, while the LysR strain (**B**) is no longer sensitive to the presence of the amino acids in the media, with the exception of cysteine and histidine that had a late inhibitory effect. The curves for each treatment were colored according to the effect of the amino acids on the NATL2A parental strain: Non-toxic (blue) and toxic (orange). Non-treated control (green).

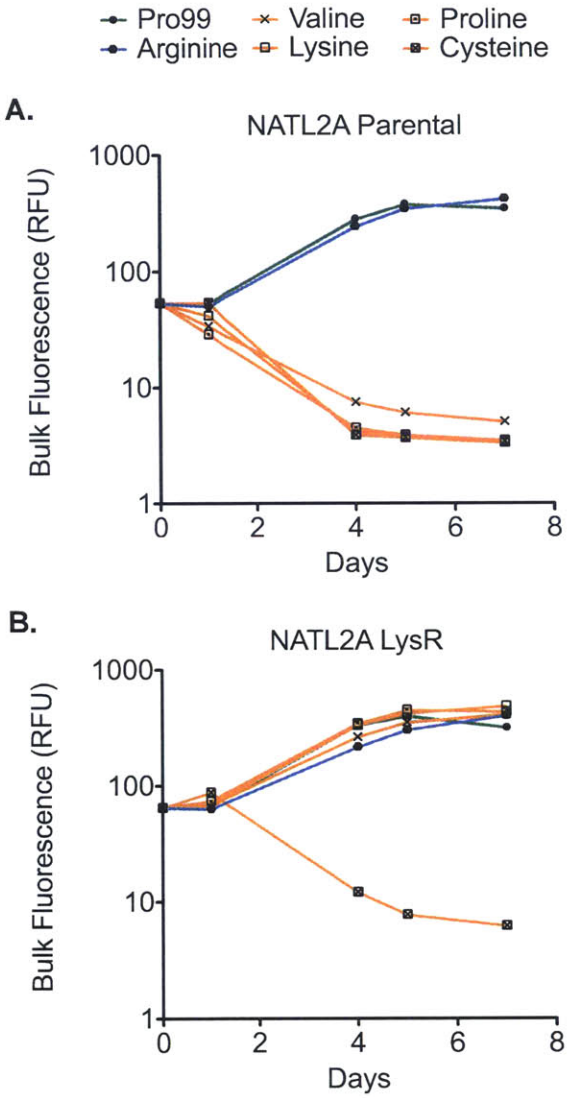

Figure 4-7: Cross-tolerance is stably maintained in the LysR strain over time. After being removed from amino acid selection for 6 months, the NATL2A parental strain (**A**) and the LysR (**B**) were exposed to a subset of 5 amino acids at a final concentration of 0.8 mM to evaluate the maintenance of the tolerant state. As observed previously, the LysR strain also displays tolerance to the original amino acid –lysine, and to others as well (in this case valine and proline). Likewise, as shown before the extended-tolerance does not protect against Cysteine. The curves for each treatment were colored according to the effect of the amino acids on the NATL2A parental strain: Non-toxic (blue) and toxic (orange). Non-treated control (green).

Investigation of the Inheritable Nature of the Amino Acid Tolerance

We investigated whether the inheritable nature of the amino acid tolerance has a genetic basis by examining potential genetic changes between parental and amino acid-tolerant strains. For this, we started a parental culture of *Prochlorococcus* NATL2A that was subjected to the treatment of seven different amino acids and was monitored to observe the emergence of amino acid-tolerant strains (Fig. 4-8). For this experiment we chose arginine and leucine as examples of non-toxic amino acids, and lysine, isoleucine, proline and valine as examples of toxic amino acids for which tolerance might emerge. In accordance with previous results, the addition of leucine and arginine did not inhibit the growth of the cells and the cultures displayed a typical growth curve with no significant differences compared to the Pro99 control (Fig. 4-8). The cultures that were amended with the toxic amino acids showed different dynamics of inhibiton and emergence of tolerance. For instance, the cultures treated with lysine and alanine were strongly inhibited and after a month of monitoring; tolerant cells did not emerge. Compared to lysine and alanine, the treatments with valine, proline and isoleucine underwent a milder initial period of inhibition, after which the cultures were able to resume growth. Interestingly, there were differences in the timing of the emergence of the amino acid-tolerant strains; while all the replicate cultures exposed to valine and proline resumed growth after 13 days, two of the replicates of the isoleucine treatment resumed growth earlier at day 5 after the amino acid addition and the third replicate only resumed growth after day 18 (Fig. 4-8). The fact that this time we did not observe the emergence of tolerance to lysine and the differences in the timing of the response to the amino acids suggests an intrinsic heterogeneity in the emergence of the amino acid-tolerant state.

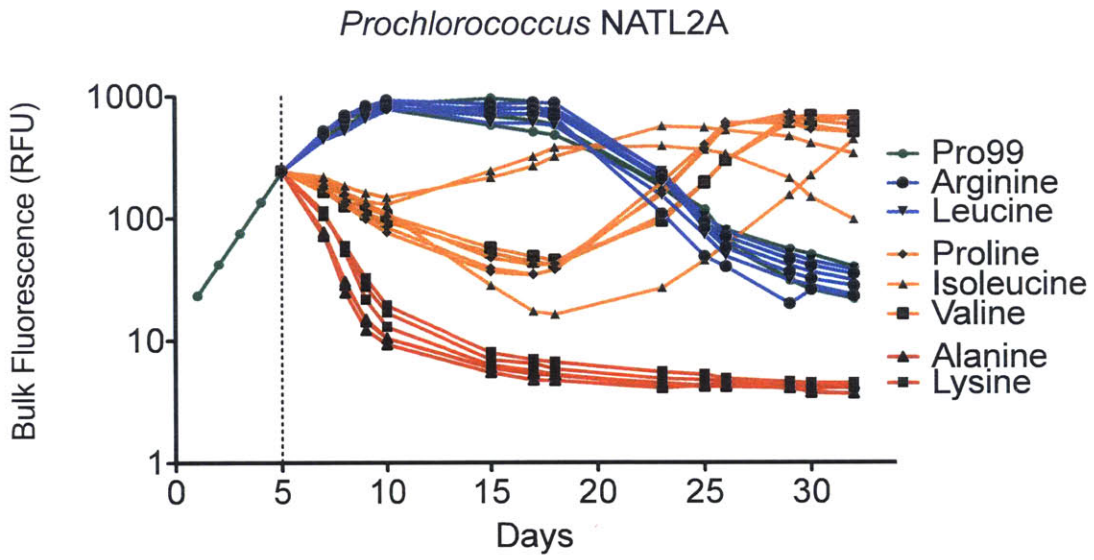

Figure 4-8: Dynamics of the emergence of tolerance to different amino acids in *Prochlorococcus* NATL2A.
An exponentially growing culture of *Prochlorococcus* NATL2A was treated at day 5 (dashed line) with 0.8 mM of each one of the amino acids listed. The emergence of amino acid tolerance was observed for cells treated with isoleucine, proline and valine (Orange). After 30 days of monitoring emergence of tolerance was not observed for the lysine and alanine treatments (Red). Arginine and leucine do not have a toxic effect (Blue). Each treatment was done in triplicate and the lines represents one replicate.

These strains gave us an opportunity to use a forward genetics approach to study the possible genetic basis of the tolerance to amino acid toxicity. In this population-level approach, if the amino acid tolerance is the result of a genetic trait, then we would expect to find a genetic variant that is present at a high frequency in the "tolerant" populations and that is found at a low frequency (or absent) in the parental population. To this end, we performed whole genome re-sequencing of the NATL2A parental (named Parental 1) and the LysR strain from Figure 4-4. We also re-sequenced the parental strain (named Parental 2), the triplicates of the proline and valine tolerant strains (named ProR1-3 and ValR1-3, respectively) and the duplicates of the isoleucine tolerant strains that emerged at day 5 (named IsoR1-2) in Figure 4-8. Each one of these genomes was re-sequenced at a coverage of ~150X. The parental 1 and parental 2 strains are technically the same, as neither has been exposed to amino acids.

To identify possible genetic variants in the population, we determined the frequency of single nucleotide polymorphisms (SNPs) for each re-sequenced strain with respect to the reference genome of *Prochlorococcus* NATL2A (sequenced circa 2006) (Kettler et al., 2007). We found a total of 21 SNPs that were at least 25% frequent in the sequence population of each genome (Figure 4-9 and Table 4-1). Out of the 21 SNPs, 11 were present at a high frequency (>96%) in the read population, however they were present in both the parental and the amino acid-tolerant strains (Figure 4-9). This suggests that these SNPs correspond to mutations that have happened since the genome was determined in 2006 and that have been fixed in the population, but they are not related with the emergence of the amino acid tolerance trait. The other 10 SNPs found were present at low frequency in the population and none of them was consistently found only in the amino acid-tolerant strains (Figure 4-9). Consequently, the result from this forward genetics approach indicates that there are not particular mutations in the population that can be directly related with the emergence of the amino acid-tolerant condition.

The absence of DNA changes between the parental and the amino acid-tolerant populations suggests that *Prochlorococcus* can generate phenotypic variants in response to the inhibitory effect of amino acids. Also, the inheritable nature of the amino acid tolerance trait might be indicative that this process could be related to epigenetic inheritance. Epigenetic inheritance is the transfer of physiological states from one generation to the next without the involvement of DNA alterations (Veening et al., 2008). The classic example of epigenetic inheritance in bacteria is DNA methylation. Methylation of promoter regions can lead to transcriptional regulation; switching between alternative DNA methylation patterns can in turn result in different phenotypic lineages originating from the same clonal population (Casadesus and Low, 2006). In model bacteria, it has been demonstrated that epigenetic mechanisms can regulate gene expression to generate different cell types within a single population in response to external stimuli such as nutritional or environmental stresses (Bierne et al., 2012; Veening et al., 2008). Inheritance of phenotypic variation can also occur through the transfer of transcriptional regulators during cell division that maintain the regulatory program (positive feedback) in the daughter cells (Veening et al., 2008). These mechanisms are also DNA independent and changes in the phenotype can be transmitted for multiple generations (Kaufmann et al., 2007; Veening et al., 2008).

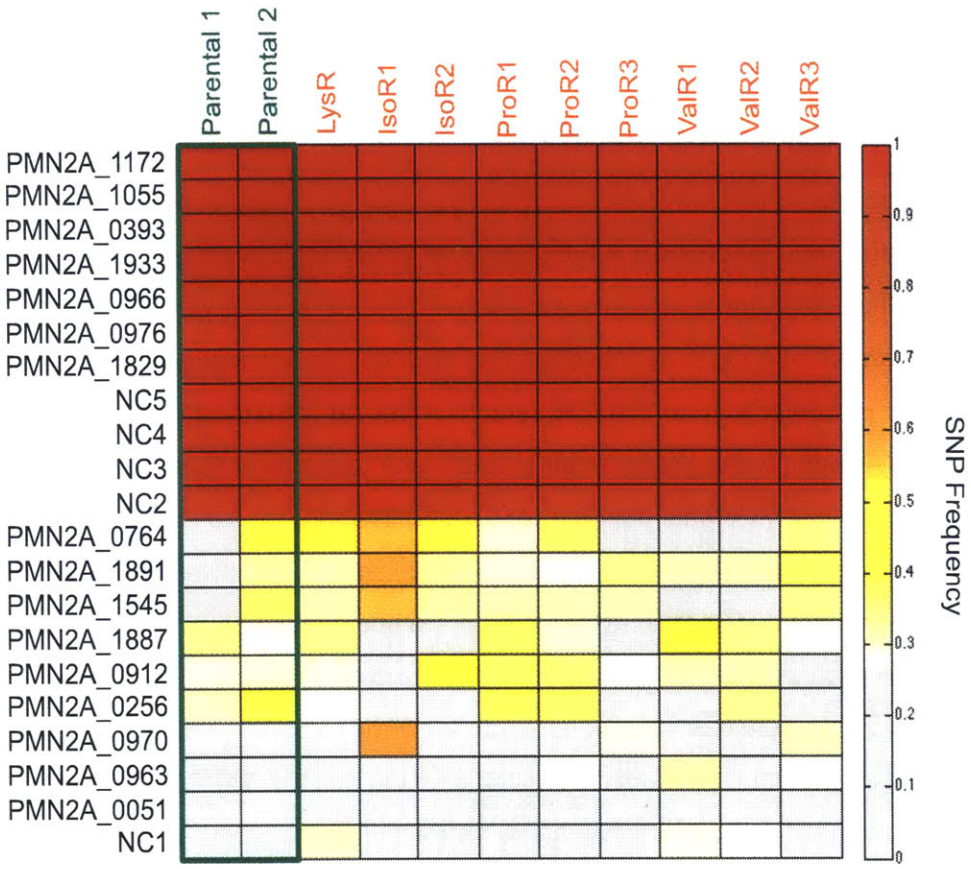

Figure 4-9: Single nucleotide polymorphism frequency in parental and amino acid-tolerant strains of *Prochlorococcus* NATL2A.
Columns represent each one of the strains that were re-sequenced and rows are labeled according to the name of the gene where the polymorphism was found with respect to the reference genome of *Prochlorococcus* NATL2A. SNPs that fell in non-coding regions were labeled NC. See Table 4-1 for details. The color scale represents the frequency of the variant in the sequence population in each genome.

Name	Nucleotide Change	Amino Acid Change	Polymorphism Type	Average Frequency in Resistant Strains	Average Frequency in Parental Strains	Description
PMN2A_1933	(A)2 -> (A)3	Frame Shift	Insertion (tandem repeat)	96.6%	94.7%	Hypothetical protein
PMN2A_1891	(T)4 -> (T)3	Frame Shift	Deletion (tandem repeat)	34.6%	25.4%	ATPase
PMN2A_1887	G -> A	S -> F	SNP (transition)	27.4%	30.0%	Conserved Hypothetical Protein
PMN2A_1829	A -> G	I -> T	SNP (transition)	99.7%	99.7%	Sigma-70 factor
PMN2A_1545	G -> A	D -> N	SNP (transition)	32.5%	31.6%	Isochorismate synthase
PMN2A_1172	C -> G	G -> A	SNP (transversion)	100.0%	100.0%	(S)-2-hydroxy-acid oxidase
PMN2A_1055	T -> C	V -> A	SNP (transition)	100.0%	100.0%	Urease. Metallo peptidase. MEROPS family M38
PMN2A_0976	G -> A	M -> I	SNP (transition)	99.8%	100.0%	Putative nicotinate-nucleotide adenylyltransferase
PMN2A_0970	G -> A	G -> S	SNP (transition)	30.6%	<25%	ATP synthase F1, beta subunit
PMN2A_0966	C -> A	A -> E	SNP (transversion)	99.8%	100.0%	Phosphoglycerate mutase
PMN2A_0963	G -> T	P -> Q	SNP (transversion)	<25%	29.3%	DNA-directed RNA polymerase (omega chain)
PMN2A_0912	C -> T	T -> M	SNP (transition)	27.9%	28.6%	Multi-sensor signal transduction histidine kinase
PMN2A_0764	C -> A	L -> I	SNP (transversion)	34.6%	31.0%	Acetolactate synthase, small subunit
PMN2A_0393	G -> A	P -> L	SNP (transition)	100.0%	100.0%	Delta(12)-fatty acid dehydrogenase
PMN2A_0256	C -> A	V -> F	SNP (transversion)	26.7%	40.2%	ATPase
PMN2A_0051	C -> T	S -> F	SNP (transition)	26.2%	<25%	2-phosphosulfolactate phosphatase
NC5	C -> T	Non-coding region	SNP (transition)	99.9%	100.0%	Intergenic region
NC4	(T)8 -> (T)7	Non-coding region	Deletion (tandem repeat)	96.0%	96.3%	Intergenic region
NC3	C -> T	Non-coding region	SNP (transition)	100.0%	100.0%	Intergenic region
NC2	C -> T	Non-coding region	SNP (transition)	100.0%	100.0%	Intergenic region
NC1	(T)10 -> (T)11	Non-coding region	Insertion (tandem repeat)	28.30%	<25%	Intergenic region

Table 4-1: Summary of single nucleotide polymorphisms found in the sequenced strains with respect to the reference genome of *Prochlorococcus* NATL2A.

A total of 21 SNPs were found in the sequenced strains, of which 11 seem to be fixed in the both the wild type and mutant populations as the frequency is higher than 96%. The rest of the SNPs were found in a frequency lower than 35% in both the wild type and the mutant population, indicating mixed populations of ancestral alleles. None of the mutations were significantly more abundant in the tolerant population.

CONCLUSION

At this point we can only speculate about the mechanisms that lead to amino acid toxicity. It is probable that the toxicity is related with the inability of the streamlined regulatory system of *Prochlorococcus* to respond to rapid changes in nutrient availability (Garcia-Fernandez et al., 2004). Dissolved free amino acids in the ocean are usually present at concentrations in the low nanomolar range and they represent a very labile fraction of the dissolved organic matter (DOM) that undergoes rapid turnover rates in ocean (Suttle et al., 1991). Thus, it is likely that *Prochlorococcus* never observes drastic changes in amino acid concentrations in the oligotrophic environment where it lives. Also, it has been reported that *Prochlorococcus* displays a high-rate of uptake of organic compounds in the ocean that is 10-fold higher than marine *Synechococcus* (Zubkov et al., 2003). Therefore, it is possible that when extremely high concentrations of amino acids are present in the culture media, *Prochlorococcus* might not be able to efficiently regulate the uptake of amino acids overloading downstream metabolic processes, something that could lead to toxicity. Comparing the amino acid uptake rates of the parental and amino acid-tolerant populations might bring support to this hypothesis.

Independently of what might be the exact mechanism of amino acid toxicity, here we demonstrate that *Prochlorococcus* is able to overcome the toxicity of the amino acids by generating phenotypic variants. These variants are characterized by an extended-tolerance to the toxicity of several amino acids, a trait that might be transferred throughout the generations by epigenetic inheritance. Further investigation is required to confirm the nature of the tolerance. For instance, the determination of the methylation state of the whole-genome will bring insights into the possible role of DNA modifications in this process. Also, the comparison of the gene expression dynamics between the parental and the amino acid-tolerant strains might shed the light on the physiological differences between the phenotypic variants.

MATERIALS AND METHODS

Nitrogen starvation rescue experiments

Prochlorococcus NATL2A was grown in 200 ml Pro99 medium under continuous light at an irradiance of 20μE $m^{-2}s^{-1}$ and a temperature of 24°C. When the cultures reached mid-exponential phase of growth, the cells were collected by centrifugation at 7500xg and washed twice with filtered and autoclaved Sargasso Sea water. Cells were resuspended in 200 ml Pro99 medium lacking NH_4Cl and incubated for 48h to ensure that the natural nitrogen sources present in the sea water are consumed and nitrogen-starvation is completely achieved. At this point the culture was aliquoted into 10 ml glass tubes containing 5 ml of the N-starved cells and the different amino acid treatments and controls were added in triplicate. The concentration of each amino acid treatment was 0.8 mM to match the nitrogen quota of the standard formulation of the Pro99 media (Moore et al., 2007). To demonstrate that the cells are still viable after the starvation period NH_4Cl is added as a positive control. Rescue dynamics of the treatments are compared to the NH_4Cl and the no nitrogen addition control. Fluorescence is used as a proxy for growth. Two independent experiments were performed to test the full set of 20 proteinogenic amino acids.

Amino acid addition experiments

Prochlorococcus NATL2A was grown in 200 ml Pro99 medium under continuous light at an irradiance of 20μE $m^{-2}s^{-1}$ and a temperature of 24°C. When the cultures reached mid-exponential phase of growth, the different amino acid treatments were added at a final concentration of 0.8 mM in triplicate. The growth dynamics for the different amino acid treatments were compared to the untreated control.

Whole-genome re-sequencing

DNA from the parental and the amino acid tolerant strains of *Prochlorococcus* NATL2A was isolated using the QIAamp DNA mini kit (Qiagen). Genomic DNA was sequenced using the MiSeq Illumina Platform at the BioMicro Center of MIT using the protocol previously described in Biller *et al.* 2014 (Biller et al., 2014). Low quality regions of sequencing reads were removed from the raw Illumina data using quality_trim from CLC with the default settings in which at least 50% of the read must have a minimum quality of 20 (V3.2, from the CLC Assembly Cell package; CLC bio). Mapping and SNP identification were performed in the Genious 8 software (Kearse et al., 2012).

REFERENCES

Bierne, H., Hamon, M., and Cossart, P. (2012). Epigenetics and bacterial infections. Cold Spring Harb Perspect Med *2*, a010272.

Biller, S.J., Berube, P.M., Berta-Thompson, J.W., Kelly, L., Roggensack, S.E., Awad, L., Roache-Johnson, K.H., Ding, H., Giovannoni, S.J., Rocap, G., *et al.* (2014). Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. Sci Data *1*, 140034.

Casadesus, J., and Low, D. (2006). Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev *70*, 830-856.

Church, M.J., Ducklow, H.W., and Karl, D.M. (2004). Light dependence of [3H]leucine incorporation in the oligotrophic North Pacific ocean. Appl Environ Microbiol *70*, 4079-4087.

Garcia-Fernandez, J.M., de Marsac, N.T., and Diez, J. (2004). Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. Microbiol Mol Biol Rev *68*, 630-638.

Kaufmann, B.B., Yang, Q., Mettetal, J.T., and van Oudenaarden, A. (2007). Heritable stochastic switching revealed by single-cell genealogy. PLoS Biol *5*, e239.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., *et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., *et al.* (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet *3*, e231.

Moore, L., Coe, A., Zinser, E.R., Saito, M.A., Sullivan, M.B., Lindell, D., Frois-Moniz, K., Waterbury, J., and Chisholm, S.W. (2007). Culturing the marine cyanobacterium *Prochlorococcus*. Limnology and Oceanography: Methods *5*, 353-362.

Suttle, C., Chan, A., and Fuhrman, J. (1991). Dissolved free amino acids in the Sargasso Sea: uptake and respiration rates, turnover times, and concentrations. Marine Ecology Progress Series *70*, 189-199.

Veening, J.W., Smits, W.K., and Kuipers, O.P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol *62*, 193-210.

Zubkov, M.V., Fuchs, B.M., Tarran, G.A., Burkill, P.H., and Amann, R. (2003). High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. Appl Environ Microbiol *69*, 1299-1304.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

The prochlorosins from marine picocyanobacteria represent an outstanding example of combinatorial biosynthesis that enables the production of an unprecedented number of peptide secondary metabolites in microorganisms with small genomes sizes. Our biogeographic analysis of the distribution, abundance and diversity of lanthipeptide genes in natural populations of marine picocyanobacteria, arguably, establishes prochlorosins as the most diverse family of natural products described to date.

The work presented here represents steps towards the understanding of fundamental aspects about the ecology, evolution and biological function of these outstanding secondary metabolites. The principal findings include the following:

- The prochlorosin trait has a limited distribution in the marine picocyanobacterial phylogeny: it has been found so far only in the most deeply branching clade of *Prochlorococcus* – the one most closely related to *Synechococcus* – and in 4 diverse clades of *Synechococcus*. In the natural environment, the prochlorosin biosynthesis genes are distributed over broad stretches in the oceans, but patchily distributed across different environmental gradients and over seasonal cycles.

- The prochlorosin biosynthesis pathway displays a unique plasticity in its genomic organization: multiple prochlorosin precursor peptides are located throughout the genome and the rest of the biosynthetic genes are not always clustered in the same locus. This genetic organization contrasts with the paradigmatic view of secondary metabolite pathways as traits encoded by biosynthetic gene clusters that behave as single evolutionary units.

- The lanthipeptides in marine picocyanobacteria are undergoing a process of evolutionary diversification. The *procA* genes undergo a process of expansion, diversification and elimination that result in a suite of lanthipeptides with potential structures that are completely different to each other within genomes as well as between genomes of lanthipeptide-encoding strains. This diversification is also observed at the population level: natural populations of marine picocyanobacteria from different oceanic regions harbor largely dissimilar

collections of *procA* genes, which suggest that the selective pressures on the prochlorosins promote the diversification of lanthipeptide structures rather than the convergence of structures with similar ring topologies.

- The prochlorosin precursor peptide genes undergo atypical molecular evolution. The leader peptide region evolves through small changes to maintain its structure and remain as a client of the lanthionine synthetase. On the other hand, the core peptide evolves through the action of large sequence polymorphisms that result in a rapid change in its structure. This observation further supports the evolutionary diversification hypothesis, in which the trait that is being selected in the prochlorosins is structural diversity rather than structural similarity. Also, this unique evolutionary process might explain the impossibility of reconciling the phylogeny of the prochlorosin leader and core peptides.

- There is a unique evolutionary interplay between the components of prochlorosin biosynthesis pathway; while the peptide substrates independently expand and diversify within the genome, the catalytically promiscuous lanthionine synthetase evolves under a strong purifying selection that maintains its substrate tolerant state. This relationship indicates that the lanthipeptide production trait in marine picocyanobacteria might find its evolutionary advantage in the plasticity of the production of multiple cyclic peptides with diverse ring topologies.

- Prochlorosins are likely bioactive molecules that can trigger specific physiological responses upon contact with the producer cells. The whole-transcriptome response to prochlorosins is dominated by the differential expression of genes related with peptide transport and metabolism and cell-wall processes. Also, the production of prochlorosins is not autoregulatory. In contrast to some lanthipeptide pathways from gram-positive bacteria, the sensing of its own lanthipeptides during growth does not modulate the expression of its biosynthesis genes in the producer strain.

- The addition of a subset of four prochlorosins to a small panel of *Prochlorococcus* and heterotrophic bacterial strains did not affect their growth dynamics. Finding ecologically relevant prochlorosin-target pairs is necessary to further test a possible role of prochlorosins in antibiosis.

- The prochlorosins might serve as nutrient sources for members of the abundant SAR11 clade of heterotrophic bacteria, which co-occur with *Prochlorococcus* in the oceans. The amino acid composition of the prochlorosins make them suitable compounds that could supply the unique requirements for reduced sulphur and organic nitrogen of the SAR11 group. We demonstrated that a cultured representative of the SAR11 clade can utilize prochlorosin 2.11 as the only sulphur source to support growth. This finding opens the possibility that prochlorosins could mediate metabolic interactions between groups of oligotrophic bacteria.

One of the most interesting questions in terms of the molecular evolution of the prochlorosins is: how do marine picocyanobacteria produce extreme sequence variations in small cassettes (core regions) while the rest of the gene remains constant? In Appendix A, I propose a model for the molecular mechanisms that might underlie the diversification of the core regions of the prochlorosins. The experimental validation of this model and the determination of whether this recombination-mediated mechanism is used in other families of ribosomal natural products such as the cyanobactins, is an interesting avenue of research to pursue.

The dynamic nature of the prochlorosin trait poses many challenges to study the biological function of the prochlorosins in the laboratory. For this reason, future efforts to elucidate the function of the prochlorosins should be carried out in the natural context where these molecules happen: the oligotrophic ocean. The combination of metagenomic and metatranscriptomic approaches might help to elucidate the changes in the structure and in the expression profile of the entire microbial population in response to the addition of prochlorosins. This will help in the identification of the microbial groups that are responsive to the presence of lanthipeptides and will serve as the basis for the formulation

of more targeted experiments in the laboratory. We have taken the first steps in this direction by conducting a field experiment in the North Pacific Subtropical Gyre to evaluate the effect of prochlorosins in endogenous microbial communities (see Appendix B for the experimental design). This is the first experiment where a microbial community is probed against an endemic secondary metabolite and therefore is expected to bring insights into the biological function and mechanism of action of these peptides in their natural context.

One of the most interesting avenues stemming from this work is the possible role of lanthipeptides as compounds that could mediate metabolic interactions between abundant groups of oligotrophic bacteria. In Chapter 3, we showed evidence that the prochlorosin 2.11 is able to supply the reduced sulphur nutrient for an isolate of the SAR11. To further elucidate the nature of lanthipeptides as nutrient sources in the ocean and the implications in microbial interactions the following questions need to be addressed:

Do lanthionine-containing peptides represent a better form of reduced sulfur and organic nitrogen compared to their equivalent linear peptide or free amino acid forms? Additional prochlorosin amendment experiments are needed to confirm and further elucidate the results described in Chapter 3. Two additional control treatments will be included to test the nutrient source hypothesis. First, a linear version of the peptide (with Cys residues present) will be needed to test if a cyclic peptide structure is required for uptake and utilization. Second, a suite of amino acids in equimolar concentrations to those present in the prochlorosins will be added in order to compare the effects of the prochlorosin additions to that of free amino acids. Also, assays with varying concentrations of Pcn 2.11 (and other prochlorosin structures) should be conducted to gain a better understanding of their ability to alleviate reduced sulfur and organic nitrogen limitation in SAR11, and to check for dose-dependent responses.

Do sulfate-assimilating bacteria also metabolize lanthipeptides? To determine if the prochlorosins are acting to establish specific interactions with specific groups, we will test whether prochlorosins are readily metabolized by any microbial group or if the presence of lanthionines or (methyl-lanthionines) render these peptides refractory to assimilation unless a dedicated catabolic –likely co-evolved– pathway is present.

Do prochlorosin-encoding strains of Prochlorococcus establish more robust co-cultures with SAR11 compared to other Prochlorococcus strains? SAR11 strain HTCC7211 will be co-cultured with various closely related strains of LL-IV *Prochlorococcus* that do and do not produce prochlorosins. These co-cultures will be monitored for growth effects on both organisms to examine potential bilateral interactions.

What genes are associated with the uptake and metabolism of lanthipeptides in SAR11? Identification of genes and metabolic pathways implicated in the transport and metabolism of prochlorosins may provide hints regarding the mechanism for lanthipeptide utilization as well as indicate other clades of heterotrophic bacterioplankton that might have the same capacity. What other genes are responsive to the presence of lanthipeptides? Investigation of all genes found to be differentially expressed during the utilization of lanthipeptides as a nutrient source (for either SAR11 alone, or both SAR11 and *Prochlorococcus* in co-culture) could provide leads as to the types of interactions being mediated by these compounds.

APPENDIX A

Proposed Molecular Mechanism for the Expansion and Diversification of Prochlorosins

SUMMARY

One of the hallmarks of the prochlorosins is the startling variation found in the core peptide region (Chapter 2, Fig. 2-1); arguably they are the most diverse group of microbial secondary metabolites. What is the molecular mechanism that enables the generation of hypervariable core peptide regions? To gain a better understanding of the mechanisms that enable this diversification, I studied the diversity and structure of the *procA* biosynthetic gene clusters to identify specific molecular features that could enable the generation of a large number of different lanthionine-containing natural products. Here I present a model that integrates elements of the SOS-induced mutagenesis pathway with specific molecular features of prochlorosin gene clusters to explain the localized hypervariability in the core regions of the precursor peptide, and the susceptibility of genes clusters to undergo recombination and changes in copy number.

MOTIVATING QUESTIONS AND BACKGROUND

Results from Chapter 2 suggest that the precursor peptide genes are undergoing an atypical process of molecular evolution. The striking physical demarcation between a conserved and a hypervariable region within the same coding sequence (Chapter 2, Fig. 2-1C) is indicative of a possible targeted mutagenesis mechanism. But, how do marine picocyanobacteria produce extreme sequence variations in small cassettes (core regions) while the rest of the gene remains constant? **Is it possible that core regions of prochlorosins experience increased levels of mutagenesis?** A well-known pathway responsible for the increase in the mutation rate is SOS-induced mutagenesis (Patel et al., 2010; Tippin et al., 2004) The SOS response is a global response to DNA damage in which genome replication is arrested and DNA repair and mutagenesis are induced. The early phase of the SOS response is mostly dominated by accurate DNA repair. If the use of error-free pathways cannot complete repair and re-start replication, the mutagenic phase of SOS is triggered. This phase is mediated by DNA polymerases that replicate past lesions in a process termed translesion DNA synthesis (TLS) (Patel et al., 2010). There are three SOS-induced DNA polymerases known in bacteria, Pol II, IV and V, Pol V being the lowest fidelity polymerase and the one largely responsible for the increase in chromosomal mutations. Pol V is strictly regulated in the cell by RecA nucleoprotein filaments formed by RecA binding to single-stranded DNA and ATP. This complex essential to activate Pol V-catalyzed TLS and therefore prevents genomic mutation overload (Patel et al., 2010). Importantly, previous work in our lab has demonstrated the this pathway is also present in *Prochlorococcus*, as it was shown that after 30 seconds of UV irradiation the mutation frequency increased 100 fold compared to the spontaneous mutation frequency (Osburne MS, unpublished results). **How does stress-induced mutagenesis participate in the expansion and diversification of prochlorosins?**

PROPOSED MODEL

The model proposes that the presence of a putative crossover hotspot instigator (*chi*) site at the 3' end of the leader peptide region **(Figure A-1)** serves as the molecular signal that increases the chances of recombination in the prochlorosin precursor peptide genes, and targets an error-prone polymerase to the beginning of the core region of the complementary DNA strand during recombination. In this scenario, during events of recombination repair, the downstream segment of leader peptide region would be preferentially copied by an error-prone polymerase that could introduce multiple changes in a very short segment of DNA **(Figure A-2A)**. Importantly, and in contrast to other bacteria, *Prochlorococcus* and *Synechococcus* lack two of the three known translesion repair polymerases (pols II and IV), and encode only one error-prone polymerase (Pol V), which is known to copy undamaged DNA with extremely poor fidelity. This fact exacerbates the effect of SOS-induced mutagenesis because Pol V can introduce base substitutions as well as -1 and -2 frameshifts (Tippin et al., 2004). This type of localized mutagenesis has the potential of creating a diverse set of alleles of core peptide regions that do not share any type of phylogenetic relationship among each other, consistent with what we observed in our phylogenetic analyses of the prochlorosin genes.

Another important consequence of having a putative *chi* site within the precursor peptide genes is that prochlorosin gene clusters become a hotspot of recombination. The *chi* - dependent generation of the 3' ssDNA invasive end has the potential of reaching several homologous regions at different locations within the gene cluster on a sister chromatid; this random strand invasion process might result in duplication or deletion events **(Figure A-2B)**. Prochlorosin gene clusters are enriched in regions of high similarity that correspond to conserved parts of the leader peptide region and the 5' UTR upstream regions, this feature, and the presence of multiple neighboring *chi* sites make *procA* loci recombination prone regions **(Figure A-3)**. In this context, the leader peptide region of the prochlorosin genes has evolved molecular features that are important for evolutionary processes; the 5'-GCTGGNGGN motif not only encodes the double glycine signal for proteolytic cleavage that demarks the end of the leader peptide and the beginning of the core peptide, but also serves at the DNA level as a molecular signal that delimits the start

point for error prone replication and increased recombination. Accordingly, prochlorosin gene clusters might have evolved specific structural features that allow them to invoke two powerful drivers of diversity: error prone replication and recombination.

CONCLUSION

Our sequence analyses of prochlorosin gene clusters suggest that these natural products might evolve through a mechanism that involves intramolecular DNA recombination and targeted error-prone replication, which in turn could lead to the expansion and diversification of prochlorosin gene clusters. Our model proposes a novel evolutionary mechanism that enables the rapid diversification of peptide-based secondary metabolites and sheds light on how nature creates and evolves novel bioactive molecules. Further work is required towards the experimental validation of our model. A key point is the demonstration of the *in vitro chi* activity –attenuation of the endonuclease activity of the RecBCD complex– of the 5'-GCTGGNGG motif in *Prochlorococcus*.

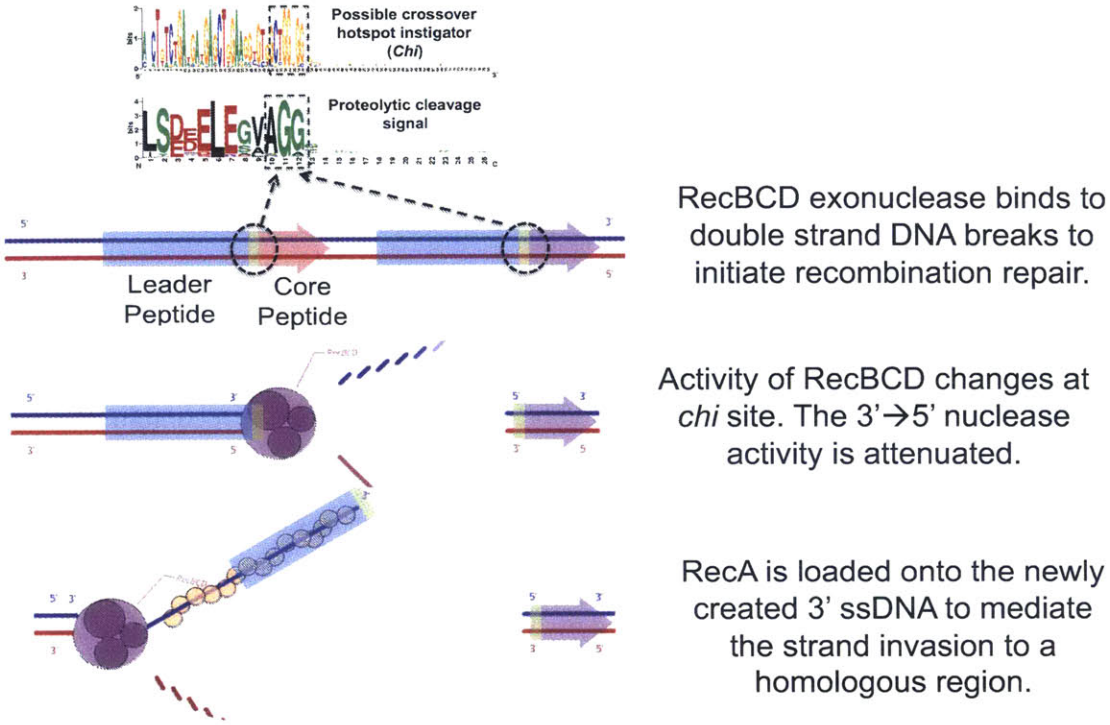

Figure A-1[3]. Molecular signals present in the precursor peptide and the precursor peptide gene of prochlorosins

The sequence logo shows the conservation of sites at the protein and the DNA level. In *E. coli* the *chi* sequence is 5'-GCTGGTGG-3', in *Prochlorococcus* and *Synechococcus* the proposed *chi* is 5'- GCTGGNGG-3'. The *chi* sequence is asymmetric and is recognized by the RecBCD enzyme only when the enzyme is traveling through the dsDNA from the 3'-side of *chi*. Interaction with *chi* results in an attenuation of the nuclease, but not the helicase, activity of RecBCD enzyme, permitting the enzyme to continue unwinding without further DNA degradation. A consequence of this *chi* -dependent attenuation is that ssDNA with a 3'-end in the vicinity of *chi* is created. This ssDNA is the substrate for invasion of the homologous dsDNA promoted by RecA and SSB proteins. Both the *chi* -dependent generation of the 3' ssDNA invasive end and the *chi*-dependent attenuation of the RecBCD enzyme nuclease activity are sufficient to explain the recombination hotspot activity of *chi* (Dillingham and Kowalczykowski, 2008).

[3] This figure uses elements from the general recombination model presented in http://sites.fas.harvard.edu/~biotext/animations/GeneralRecombination.html, which have been adapted to illustrate the recombination events in the context of a *procA* gene cluster.

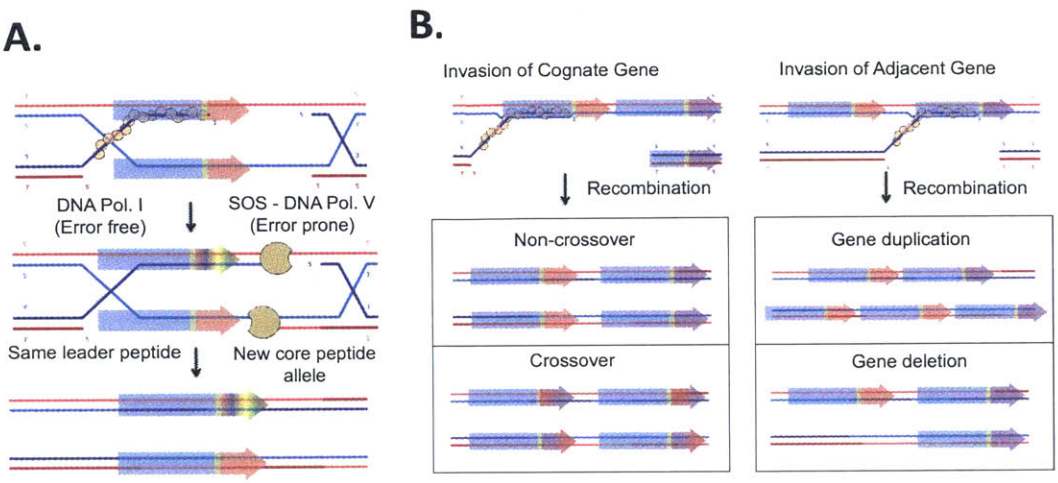

Figure A-2. [4]: Model of the molecular mechanism that enables the diversification of core peptide regions and expansion of *procA* genes

A. The *procA* gene is composed of a leader peptide region (blue) and a core peptide region (red). A putative *chi* site is present in the boundary between leader and core peptide regions (green). The RecA nucleoprotein filament (yellow circles) catalyzes the strand invasion of a sister chromatid. RecA mediates a homology search -in this case conserved DNA regions in the leader peptide- through the formation of a triple-stranded intermediate, aligning the ssDNA with its homologous target regions on the duplex DNA. The 3' ssDNA invasive end provides a priming site for DNA polymerases (brown pac-man) to fill out the gaps. Under the conditions of the SOS-response, PolV is recruited to the site and is activated by the RecA nucleoprotein filament where it could introduce multiple substitutions and frame shifts. After the completion of the other steps in recombination repair and the resolution of the Holiday junction (not shown), one new version of the *procA* gene will be created.

B. Involvement of *chi* sites in recombination and gene duplication/deletion. Physical structure of a schematic *procA* gene cluster, each *procA* gene is composed of a leader peptide region (blue) and a core peptide region (red or purple). If recombination is resolved in non-crossover event the dsDNA break will be repaired and the locus will remain intact. In contrast, if the recombination resolves in a crossover event, diversity will be introduced in the locus. Also, strand invasion of other (non-cognate) homologous regions present in the cluster by the RecA nucleoprotein filaments can result in changes in the gene copy number by means of duplication/deletions events.

[4] This figure uses elements from the general recombination model presented in http://sites.fas.harvard.edu/~biotext/animations/GeneralRecombination.html, which have been adapted to illustrate the recombination events in the context of a *procA* gene cluster.

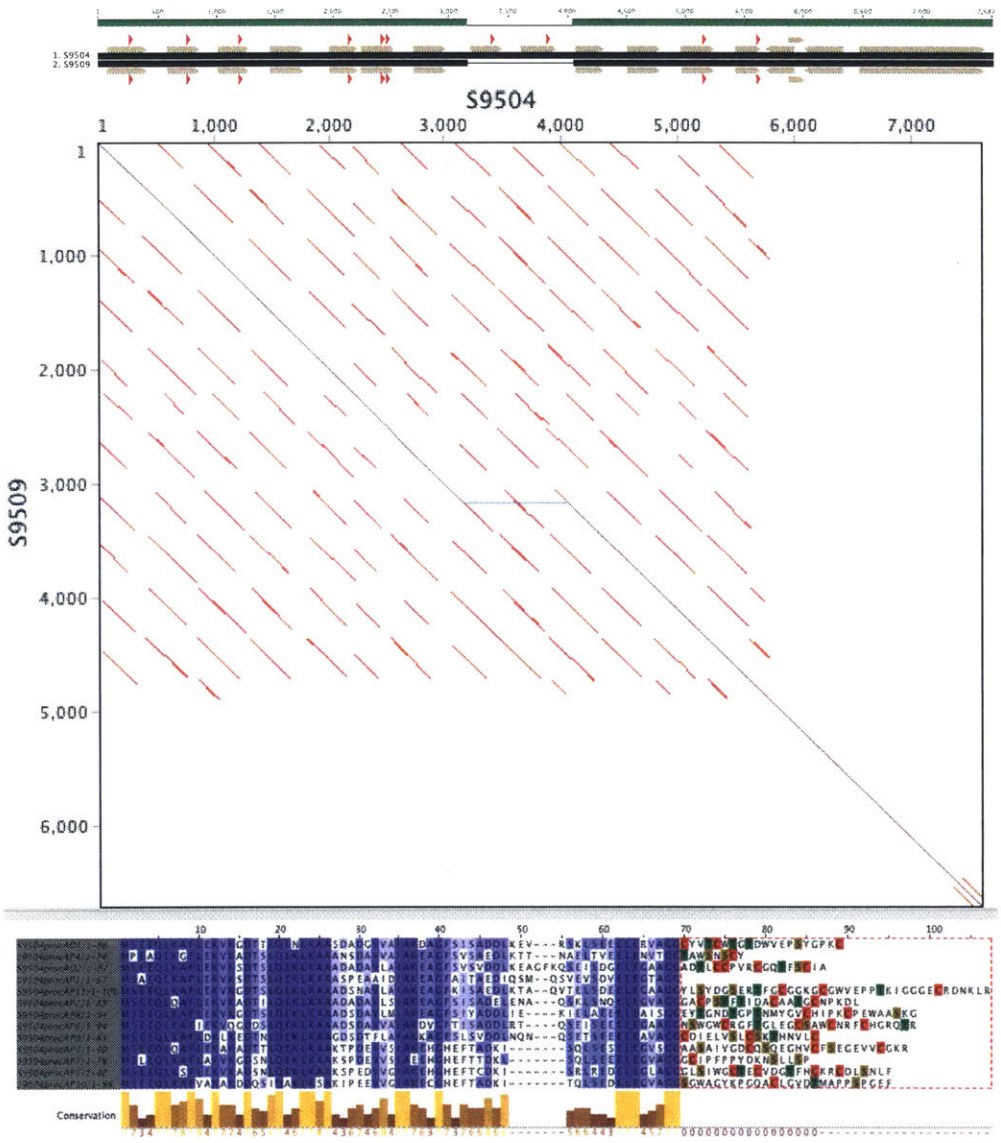

Figure A-3: Prochlorosin gene duplication/deletion in two closely related *Synechococcus* genomes.

The panel above depicts the pairwise alignment of a homologous prochlorosin gene clusters from *Synechococcus* strains S9504 (top) and S9509 (bottom). Presence of a putative *chi* sites (5'- GCTGGNGG-3') are shown as red arrowheads. Below is a Dotplot analysis that demonstrates the presence of multiple high-similarity sequences within the gene cluster that along with high density of putative *chi* sites might mediate possible internal recombination events. The repeated regions correspond to DNA segments that comprise the 5' UTR and the leader peptide region of the precursor peptide genes. The bottom panel shows a multiple sequence alignment for the amino acid sequence of the precursor peptides encoded in this cluster. The genes in this cluster might have formed by gene duplication events and were expected to have some level of similarity. However, the

precursor peptides contain multiple substitutions in the leader peptide region and are completely different in the core peptide region. This result supports the idea that the high degree of variation within prochlorosin gene clusters might arise during recombination events and not through the sequential accumulation of multiple single substitutions within each peptide.

REFERENCES

Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev *72*, 642-671.

Patel, M., Jiang, Q., Woodgate, R., Cox, M.M., and Goodman, M.F. (2010). A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V. Crit Rev Biochem Mol Biol *45*, 171-184.

Tippin, B., Pham, P., and Goodman, M.F. (2004). Error-prone replication for better or worse. Trends Microbiol *12*, 288-295.

APPENDIX B

Hawaii Ocean Experiment: Prochlorosin Amendment

SUMMARY

Station ALOHA represents an ideal site in the oligotrophic Ocean to study the ecological role of the lanthipeptide production trait. Here I present the experimental design and preliminary results of a field experiment conducted at this site aimed at characterizing the effect of prochlorosins in natural microbial communities.

BACKGROUND AND MOTIVATION

Station ALOHA located at 22°45′N, 158°W, approximately 100 km north of Oahu, Hawaii, has been a remarkable site for the study of marine microbes for the past three decades (Karl and Church, 2014). The study of the diversity of prochlorosin precursor peptide genes in samples from station ALOHA (Chapter 2, Fig. 2-2) demonstrate that there are more than 600 different prochlorosins residing in the *Prochlorococcus* and *Synechococcus* populations that dwell in this site in the North Pacific Subtropical Gyre. In addition, the distribution of the *procA* genes along the water column indicate that abundance of lanthipeptide-encoding cells of picocyanobacteria reaches its highest at depths between 125 and 175m, where there can be up to 10,000 copies of the *procA* gene per milliliter of sea water. Also, prochlorosin-encoding strains of *Prochlorococcus* have been isolated from the station, which are useful to link the observations in the field with laboratory experiments. Consequently, station ALOHA represents an ideal site for the study of the biological role of prochlorosins.

The use of metatranscriptomic approaches has proven useful for identification of the marine microbial community response during perturbation studies. This approach has already been used at Station ALOHA to identify the members of the community and their possible physiological responses to the addition of dissolved organic matter (McCarren *et al.,* 2010; Sharma *et al.*, 2014). Thus, a metatranscriptomic approach might be useful for the determination of the effect of the addition of prochlorosin to the microbial communities for which these metabolites are indigenous. Here I present an outline of the design of an experiment performed to this end.

EXPERIMENTAL DESIGN & PRELIMINARY RESULTS

The prochlorosin amendment experiment was conducted during the Hawai'i Ocean Experiment - Phosphorous Rally HOE-PhoR expedition in the summer of 2013 (http://hahana.soest.hawaii.edu/hoephor/hoephor.html). The global idea of the experiment is to determine the whole-community transcriptional response to the addition of purified recombinant prochlorosins. For this, sea water samples from depth where the prochlorosin-encoding cells are usually abundant were amended with prochlorosins and appropriate controls under the following conditions:

Sea water samples and incubation conditions: Approximately 200 liters of seawater from 150 m depth were collected using a CTD rosette. The environmental conditions of the water column at the moment of sampling are presented in Fig. B-1 and Fig. B-2. According to the parameters observed for the 150m depth, the laboratory of the RV Kilo Moana was setup at a temperature 23 °C and the light levels of the incubation site at 10 μE to mimic the natural conditions.

Treatments and Controls: The sea water collected was carefully transferred to 20L acid-washed and autoclaved polycarbonate carboys in an experimental setup showed in Fig. B-3. Triplicates of the following treatments were performed: 1) Prochlorosin mix (Pcn1.7, Pcn 2.11, Pcn 3.2 and Pcn 3.3) at final concentration of ~50 nM, an environmental-relevant concentration for dissolved organic molecules. 2) Linear Mix (L-Pcn 2.11, Pcn 3.2 and Pcn 3.3) at final concentration of ~50 nM. These linear peptides have Cys to Ser substitutions to avoid spontaneous formation of disulfide bonds and 3) Untreated control. Sampling for DNA, RNA and flow cytometry was performed as shown in Fig. B-4

Preliminary Results: Flow cytometry analysis of the samples from the 36h time point shows differences in the cytometry signal of the total microbial community between the samples treated with prochlorosins and the linear peptide and untreated control (Fig. B-5). The principal change observed is the increase in the low-DNA/small-cell-size population, a signal that is interestingly similar to the flow cytometry signature of the SAR11 cells (Fig. B-5). Further experiments will reveal the nature of these changes.

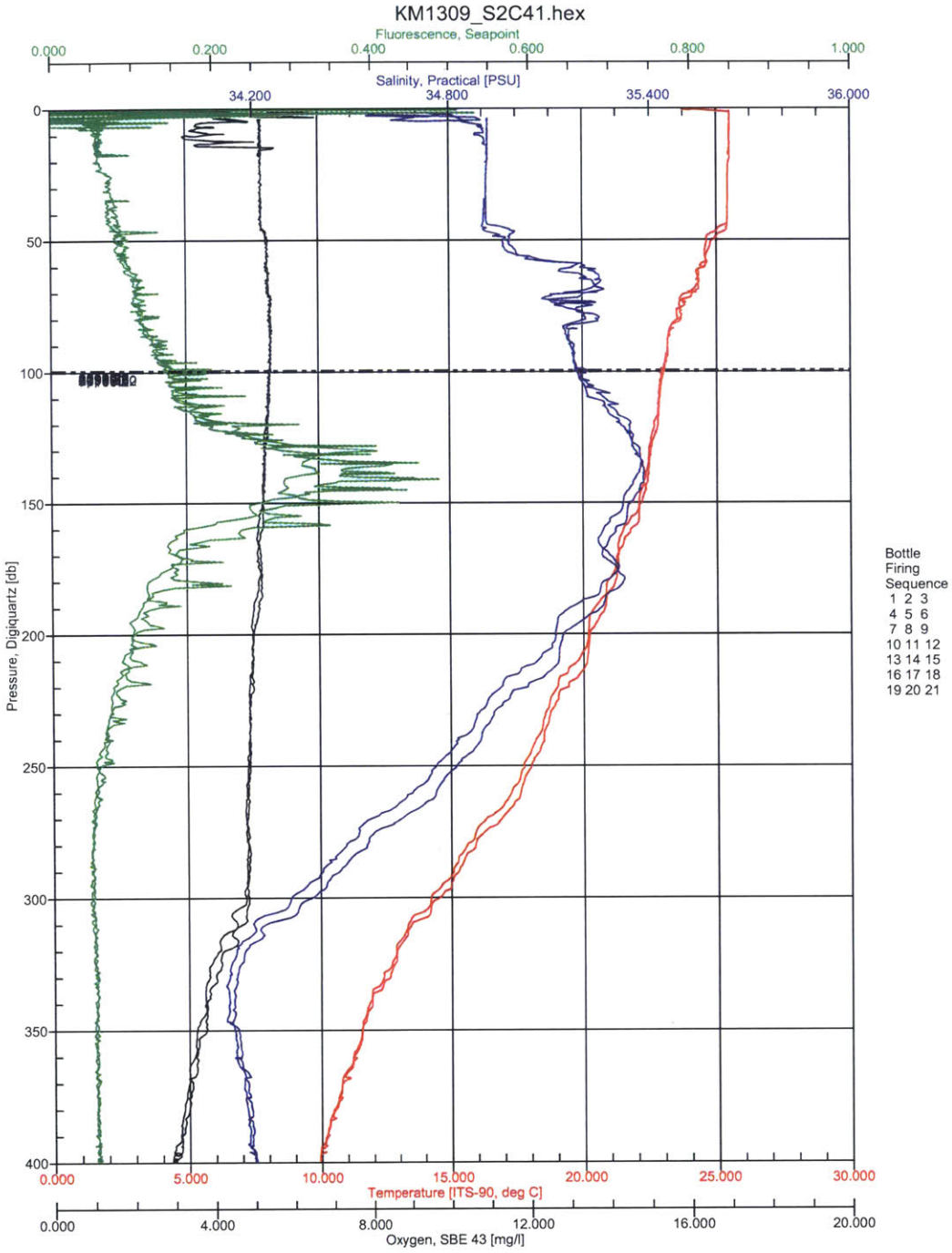

Figure B-1. Temperature, oxygen, chlorophyll and salinity conditions of the water column sampled

Water samples for the experiment were taken at 150 m depth.

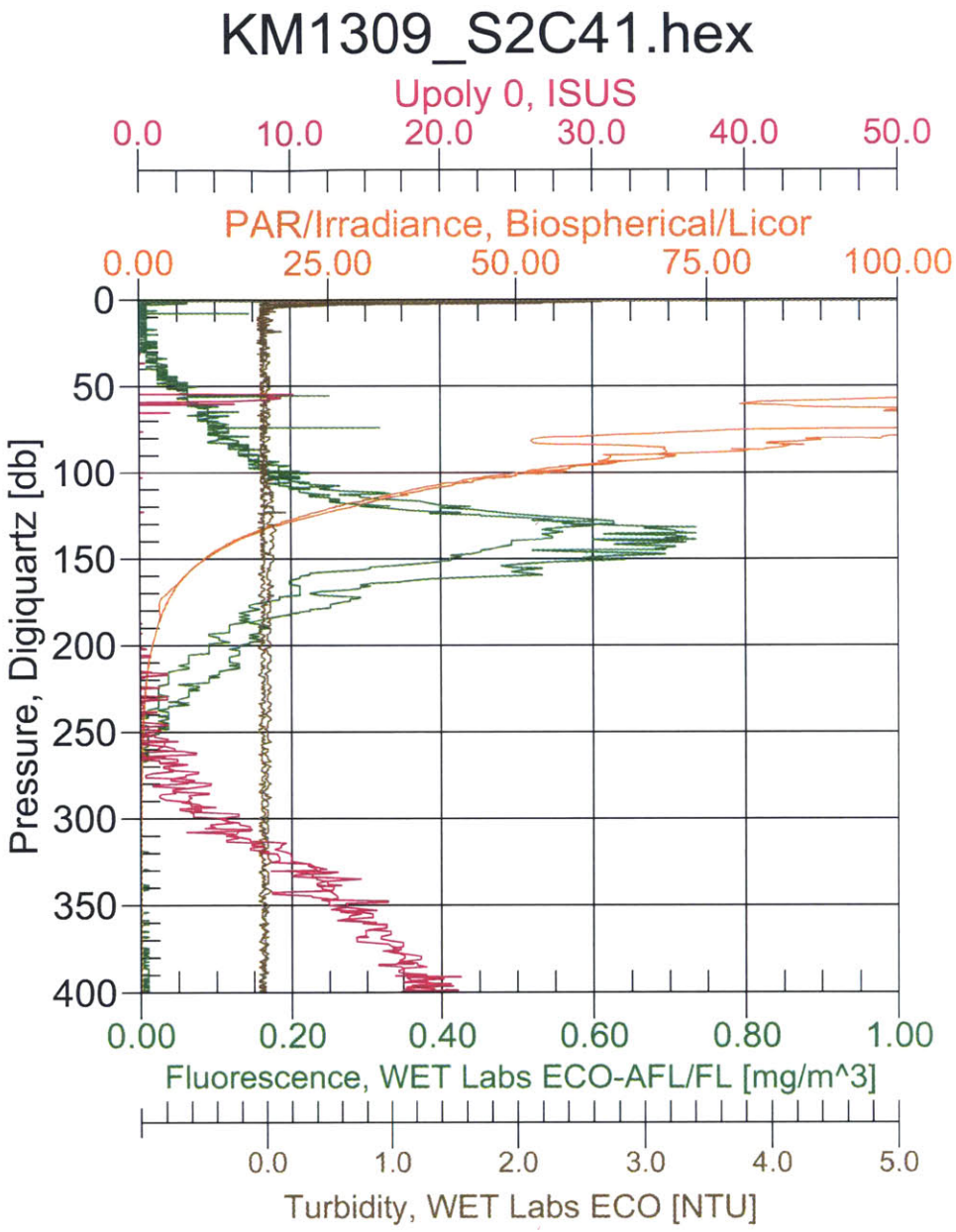

Figure B-2. Light irradiance, nitrate, chlorophyll and turbidity conditions of the water column sampled.
Water samples for the experiment were taken at 150 m depth. Photosynthetically active radiation (PAR).

A.

B.

Figure B-3. Experimental setup for the prochlorosin incubation experiment
The incubation setup is a closed system that enables the sampling through peristaltic pumps that minimizing the chance of contamination. **A.** During daytime the sea water samples were incubated at 10 μE of irradiance and using blue light screens so the light conditions are not drastically changed. **B.** During nighttime the carboys were covered with a sophisticated system of Linear Low Density Polyethylene that created suitable dark conditions.

		Seawater (Liters)						
Experiment	**Replicate**	**t=0 (DNA)**	**t=1h (RNA)**	**t=6h (RNA)**	**t=12h (RNA)**	**t=24h (RNA)**	**t=36h (RNA)**	**t=36h (DNA)**
Prochlorosin Treatment	A	10	2	2	2	2	2	10
	B	10	2	2	2	2	2	10
	C	10	2	2	2	2	2	10
Linear Peptide Treatment	A	-	2	2	2	2	2	10
	B	-	2	2	2	2	2	10
	C	-	2	2	2	2	2	10
Control	A	-	2	2	2	2	2	10
	B	-	2	2	2	2	2	10
	C	-	2	2	2	2	2	10
			Incubation at 23 °C and 10 µE in 20L Carboys					

Figure B-4. Sampling scheme of the prochlorosin incubation experiment
In addition to the sampling for DNA and RNA, triplicate samples for flow cytometry were taken at each time point.

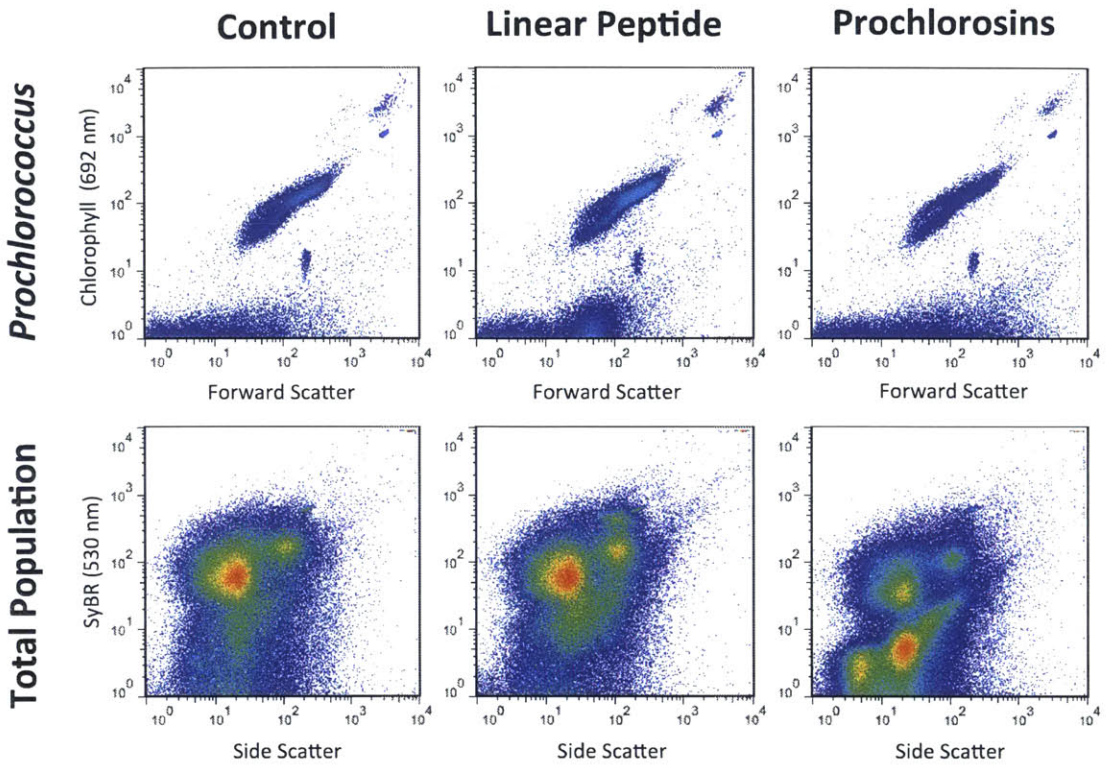

Figure B-5. Flow cytometry analysis of the microbial community in the prochlorosin incubation experiment.

Glutaraldehyde-fixed samples of the prochlorosin-treated, linear peptide-treated and the untreated incubations from the 36h time point were analyzed by flow cytometry. The top panel shows the signal from chlorophyll fluorescence and light scatter from which phytoplankton populations can be identified. There are no major changes in the phytoplankton populations between the treatments. In the bottom panel, the samples were stained with the DNA-binding dye SyBR green I to identify populations based on their DNA content and light scatter. Emergence of low-DNA/low-side-scatter population can be observed in the prochlorosin-treated samples.

REFERENCES

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

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. 2010 Sep 21;107(38):16420-7

Sharma AK, Becker JW, Ottesen EA, Bryant JA, Duhamel S, Karl DM, Cordero OX, Repeta DJ, DeLong EF. (2014). Distinct dissolved organic matter sources induce rapid transcriptional responses in coexisting populations of *Prochlorococcus*, Pelagibacter and the OM60 clade. Environ Microbiol. 2014 Sep;16(9):2815-30.