

ABSTRACT

Title of Dissertation: ADAPTIVE MECHANISMS OF AN ESTUARINE *SYNECHOCOCCUS* BASED ON GENOMICS, TRANSCRIPTOMICS, AND PROTEOMICS

David Wilfred Marsan, Doctor of Philosophy, 2016

Dissertation directed by: Dr. Feng Chen, Professor, Institute of Marine and Environmental Technology

Picocyanobacteria are important phytoplankton and primary producers in the ocean. Although extensive work has been conducted for picocyanobacteria (i.e. *Synechococcus* and *Prochlorococcus*) in coastal and oceanic waters, little is known about those found in estuaries like the Chesapeake Bay. *Synechococcus* CB0101, an estuarine isolate, is more tolerant to shifts in temperature, salinity, and metal toxicity than coastal and oceanic *Synechococcus* strains, WH7803 and WH7805. Further, CB0101 has a greater sensitivity to high light intensity, likely due to its adaptation to low light environments. A complete and annotated genome sequence of CB0101 was completed to explore its genetic capacity and to serve as a basis for further molecular analysis. Comparative genomics between CB0101, WH7803, and WH7805 show that CB0101 contains more genes involved in regulation, sensing, and stress response. At the transcript and protein level, CB0101 regulates its metabolic pathways, transport systems, and sensing mechanisms when nitrate and phosphate are limited. Zinc toxicity led to oxidative stress and a global down regulation of photosystems and the translation

machinery. From the stress response studies seven chromosomal toxin-antitoxin (TA) genes, were identified in CB0101, which led to the discovery of TA genes in several marine *Synechococcus* strains. The activation of the *relB*²/*relE*¹ TA system allows CB0101 to arrest its growth under stressful conditions, but the growth arrest is reversible, once the stressful environment dissipates. The genome of CB0101 contains a relatively large number of genomic island (GI) genes compared to known marine *Synechococcus* genomes. Interestingly, a massive shutdown (255 out of 343) of GI genes occurred after CB0101 was infected by a lytic phage. On the other hand, phage-encoded host-like proteins (*hli*, *psbA*, *ThyX*) were highly expressed upon phage infection. This research provides new evidence that estuarine *Synechococcus* like CB0101 have inherited unique genetic machinery, which allows them to be versatile in the estuarine environment.

ADAPTIVE MECHANISMS OF AN ESTUARINE *SYNECHOCOCCUS* BASED
ON GENOMICS, TRANSCRIPTOMICS, AND PROTEOMICS

by

David Wilfred Marsan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Feng Chen, Chair
Professor Tsvetan Bachvaroff
Professor Rosemary Jagus
Professor Brian Palenik
Professor James Sullivan

© Copyright by
David Wilfred Marsan
2016

Dedication

To Emily and my family for their unconditional love and support

Acknowledgements

First and foremost, I am incredibly fortunate to have Feng Chen as an amazing scientific mentor during these past few years. He allowed me the freedom to explore my ideas and formulate experiments on numerous new topics. I am incredibly lucky to have had the benefit of his insightful conversations, guidance, and fast responses throughout my doctoral research.

Further, I'd like to thank my other committee members for their guidance and expertise: Dr. Tsvetan Bachvaroff for his enlightening insight and guidance with all things bioinformatics; Dr. Brian Palenik for scientific discussions that helped steer my project; Dr. James Sullivan for scientific discussions and helping to keep my perspective on the significance of my work; Dr. Rosemary Jagus for her scientific and funding support, which allowed me to focus fully on my research.

I would like to thank Dr. Rosemary Jagus and the NOAA Living Marine Resources Cooperative Science Center (LMRCSC) for providing not only a fellowship throughout my doctoral research, but also numerous mentoring and professional development opportunities.

Additionally, I would like to thank Mrs. Carole Ratcliffe and the Philip E. and Carole R. Ratcliffe Foundation for supporting my research and career development through the Ratcliffe Environmental Entrepreneurs Fellowship (REEF) program. This fellowship allowed me to not only cultivate my leadership and business skills, but also gain invaluable experience with the FDA.

Thanks to all the members of the Chen lab. A special thanks to Yuanchao Zhan for her great scientific advice and for always entertaining my questions on Chinese

culture. Thanks to Dan Fucich, Mengqi Sun, and Zhao Zhao for being helpful lab mates in so many ways. I wish each of you luck on your future projects and know you will leave the Chen lab with the same feeling of preparedness for your professional careers as I do!

Finally, none of this would be possible without the support of my family. Especially for my wife Emily for her constant love, support (including grammar revisions!), and ability to always pull me out of my work for relaxing Longwood walks. I am so grateful to my parents and brothers for their support, love, and always providing such an inspiring example to follow.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	v
List of Tables	viii
List of Figures	ix
Chapter 1: Introduction	1
1.1 Ecological significance of Picocyanobacteria	2
1.2 Current understanding of subcluster 5.2 and estuarine <i>Synechococcus</i>	4
1.3 Genomics, Transcriptomics, and Proteomics: What it all means	8
1.4 Scope of this dissertation	11
1.5 References	19
Chapter 2: Estuarine <i>Synechococcus</i> are more tolerant to changing environments compared to coastal and open ocean <i>Synechococcus</i>	29
2.1 Abstract	31
2.2 Introduction	32
2.3 Methods	34
2.3.1 Growth conditions of <i>Synechococcus</i> CB0101, WH7803, and WH7805	34
2.4 Results	35
2.4.1 Growth comparison of estuarine, coastal, and open-ocean <i>Synechococcus</i> in different conditions	35
2.5 Discussion	37
2.5.1 Environmental tolerance of estuarine <i>Synechococcus</i>	37
2.6 Conclusion	39
2.7 References	44
Chapter 3: Understanding the adaptive nature of estuarine <i>Synechococcus</i> using a multi- omics approach	48
3.1 Abstract	50
3.2 Introduction	51
3.3 Results	53
3.3.1 Genome sequencing of <i>Synechococcus</i> CB0101	53
3.3.2 Transcriptomic responses of CB0101 to nutrient depletion and zinc toxicity	54
3.3.3 Proteomic responses of CB0101 to nutrient depletion and zinc toxicity	56
3.4 Discussion	57
3.4.1 Stress handling capability unveiled by comparative genomics	57
3.4.2 Responses of CB0101 to stresses at the omics level	60
3.5 Conclusion	62
3.6 Experimental Procedures	64
3.6.1 Culture conditions for growth of CB0101	64
3.6.2 Genome sequencing of CB0101	64
3.6.3 Genome annotation and comparisons	65
3.6.4 Extraction of mRNA	66
3.6.5 Protein extraction and mass spectrometry analysis	66
3.6.6 Protein identification	67

2.5 E

3.6.7 Functional annotation and semi-quantitative analysis.....	68
3.7 References.....	76
Chapter 4: Surviving in a highly variable environment – the novel toxin-antitoxin systems in the estuarine <i>Synechococcus</i> strain CB0101.....	82
4.1 Abstract.....	84
4.2 Introduction.....	85
4.3 Methods.....	88
4.3.1 Cyanobacterial strains and stress experiments.....	88
4.3.2 Extraction of mRNA and qPCR reaction.....	89
4.3.3 TA system identification in CB0101 and completely sequenced <i>Synechococcus</i> and <i>Prochlorococcus</i> genomes.....	90
4.3.4 Western Analysis.....	90
4.4 Results.....	91
4.4.1 Annotated toxin/antitoxins in CB0101.....	91
4.4.2 Responses of <i>Synechococcus</i> TA systems to the environmental stresses.....	92
4.4.3 Production of <i>relE</i> ¹ arrests the growth of CB0101.....	94
4.4.4 Reversible <i>relB</i> ² / <i>relE</i> ¹ expression when the stress is removed.....	94
4.4.5 Light response for <i>relB</i> ² / <i>relE</i> ¹	95
4.5 Discussion.....	97
4.5.1 Why do <i>Synechococcus</i> spp. contain TA systems?.....	97
4.5.2 The <i>relB</i> / <i>relE</i> system.....	98
4.5.3 Ecological significance and oxidative response of <i>relB</i> / <i>relE</i>	99
4.5.4 Presence of TA systems in broader picocyanobacteria.....	100
4.6 Conclusion.....	101
4.7 References.....	110
Chapter 5: Phage infection of <i>Synechococcus</i> leads to massive shutdown of genomic island genes.....	117
5.1 Abstract.....	120
5.2 Introduction.....	121
5.3 Materials and Methods.....	123
5.3.1 <i>Synechococcus</i> culture growth.....	123
5.3.2 Sampling time point determination: One-step growth curve.....	123
5.3.3 RNA-Seq.....	124
5.3.4 Normalization of read counts, calculation of RPKM and heat map representation.....	125
5.3.5 Sample clustering.....	125
5.3.6 Differential expression analyses.....	126
5.3.7 Functional group assignment to differentially express host genes.....	126
5.3.8 Protein extraction and mass spectrometry analysis.....	126
5.3.9 Protein identification.....	127
5.3.10 Functional annotation and semi-quantitative analysis.....	128
5.3.11 Genomic Island type gene identification.....	128
5.4 Results.....	128
5.4.1 Infection cycle.....	128
5.4.2 Host response to phage infection.....	129
5.4.3 Phage response during infection.....	130

5.5 Discussion	131
5.5.1 Rapid shutdown of host GI system.....	131
5.5.2 Up-regulation of GI genes and proteins	133
5.5.3 Host core gene and protein response	134
5.5.4 Phage responses	135
5.6 Conclusion	138
5.7 References.....	145
Chapter 6: Conclusions and future directions.....	152
6.1 Conclusions.....	153
6.2 Future Directions	158
6.2.1 Still a lot more to learn from Chesapeake Bay <i>Synechococcus</i>	158
6.2.2 How large of a role do TA systems play in picocyanobacteria?	161
6.2.3 GI response: A factor only in CB0101 infection or a wider trend	162
References.....	164
Appendices.....	181

List of Tables

- Table 3.1. Differentially expressed transcripts and proteins for three stress conditions, nitrogen and phosphate depletion and zinc toxicity in the estuarine *Synechococcus* CB0101
- Table 4.1. The outcome of protein BLAST search of the 7 CB0101 TA pairs..
- Table 4.2. Expression of 7 CB0101 toxin/antitoxin pairs under various stress conditions detected by RNA-Seq and qPCR.
- Table 5.1. Transcript and protein counts of differentially expressed total core and genomic islands
- Table 5.2. CB0101 Genomic island residing genes, differentially expressed transcripts and proteins under different stressors

List of Figures

- Figure 1.1. Marine Food Web and the Carbon Cycle
- Figure 1.2. Schematic phylogenetic tree summarizing the genetic and ecological diversity of marine *Synechococcus* and molecular assays that have been developed to measure particular clades in the oceans.
- Figure 2.1. Cell culture growth apparatus for physiology experiments of *Synechococcus* CB0101.
- Figure 2.2. Comparison of specific growth rates of an estuarine (CB0101), coastal (WH7803), and open ocean (WH7805) *Synechococcus* under different salinities, temperatures, and light intensities.
- Figure 2.3. Comparison of specific growth rates between an estuarine (CB0101), coastal (WH7803), and open ocean (WH7805) *Synechococcus* under different concentrations of nitrate, urea, ammonium and cyanate
- Figure 2.4. Comparison of specific growth rates between an estuarine (CB0101), coastal (WH7803), and open ocean (WH7805) *Synechococcus* under different concentrations of phosphate, zinc sulfate heptahydrate, copper sulfate, and iron.
- Figure 3.1. Genomic characteristics of estuarine *Synechococcus* CB0101..
- Figure 3.2. Stress responses of the transcript expression program in the estuarine *Synechococcus* CB0101.
- Figure 3.3. Protein expression pattern in response to stress in the estuarine *Synechococcus* CB0101.
- Figure 3.4. Molecular mechanisms of nitrogen or phosphate deplete and zinc toxic conditions of the estuarine *Synechococcus* CB0101. .

Figure 4.1. Genetic organization of the seven TA systems of *Synechococcus* CB0101. Schematic representation of the 7 TA pairs (highlighted in purple) and associated genes. Toxin is represented by bold and antitoxin by italics.

Figure 4.2. Growth rate of *Synechococcus* CB0101 during a 72 h period under different stress conditions: control medium, nitrogen depletion, phosphate depletion, and zinc toxicity (n=3).

Figure 4.3. Global transcript expression of *Synechococcus* CB0101 to three stressors measured using RNA-Seq (p<0.01 and minimum 2 fold).

Figure 4.4. Time series detection of RelB²/RelE¹ proteins by Western blot analysis during zinc toxicity and subsequent releasing at 30 h using constant cell equivalents. A) Toxin RelE¹; B) Antitoxin/toxin dimer RelB²/RelE¹; C) Pixel density analysis to determine relative protein abundance; D) Growth rate of *Synechococcus* CB0101 during the time series with black arrow representing media change (release of zinc toxic conditions) at 35.5 h.

Figure 4.5. Photo-oxidative stress caused by different light intensities induces TA systems and arrests growth of CB0101.

Figure 5.1. Infection dynamics of the S-CBP1 phage in an estuarine *Synechococcus* CB0101 host.

Figure 5.2. The transcript expression program of *Synechococcus* CB0101.

Figure 5.3. The protein expression program of *Synechococcus*.

Figure 5.4. Temporal expression dynamics of S-CBP1 phage proteins during infection of *Synechococcus* CB0101. .

Chapter 1: Introduction

1.1 Ecological significance of Picocyanobacteria

Bacteria living in diverse habitats must sense, respond, and adapt to versatile environmental conditions. No other oxygenic photosynthetic organism has survived for longer and adapted to such diverse environments as cyanobacteria (1). As one of the oldest living phyla on Earth with fossil records dating back more than ~2.15 Gyr (2), cyanobacteria are credited with producing our present atmospheric oxygen.

Cyanobacteria are a fascinating and adaptive group that are widely distributed in habitats ranging from aquatic to terrestrial systems, tropical to polar regions, and hot springs to hypersaline ponds with immense biogeochemical influence (3). Among cyanobacteria, picocyanobacteria, whose abundance and global significance was only discovered thirty-seven years ago (4–7), exemplifies this adaptive heritage and plays a significant ecological role.

Small unicellular cyanobacteria less than 2-3 μm are usually called picocyanobacteria (8, 9). Marine picocyanobacteria contribute to roughly half of the global net primary production and play a key role in regulating global biogeochemical cycles (10) (Fig 1.1). The biological fixation of carbon dioxide by picocyanobacteria (as small phytoplankton in Fig 1.1) is the major driving force of the ocean's carbon cycling. Further, picocyanobacteria are the backbone of the marine food web due to not only supplying energy up the trophic levels but also being a major component of detritus to the deep ocean. It has recently been reported that marine picocyanobacteria can make up 40-60% of the digestible gut content of copepods from the Baltic Sea. This could have major implications on food web models and economic cost analysis for other fisheries as

copepods makeup important food sources for oysters, Blue Crabs, and Stripped Bass with a combined economic value of \$622 million per year.

Marine picocyanobacteria contain two major genera, *Prochlorococcus* and *Synechococcus* (8, 9). *Prochlorococcus* are small photosynthetic prokaryotes (diameter, 0.5-0.8 μm) that are ubiquitous in the euphotic zone from the surface to a depth of ~150m in the open ocean between 40°N and 40°S. Within this range, the abundance of *Prochlorococcus* in warm surface waters is typically greater than 10^5 cells ml^{-1} , while the abundance declines beyond these latitudinal limits. *Prochlorococcus* are thought to be absent when the temperature of seawater is below 15 °C (11). Furthermore, the *Prochlorococcus* lineage is believed to be outcompeted by other phytoplankton in high-nutrient waters (12, 13). *Synechococcus*, another major genus of picocyanobacteria, is also abundant in the ocean. Compared to its sister genus *Prochlorococcus*, *Synechococcus* is more ancient, diverse, and adaptive to environments ranging from freshwater, estuarine, coastal, and oceanic waters (14–20). Although concentrations of *Synechococcus* are typically an order of magnitude lower than those of *Prochlorococcus*, its larger average cell size (0.6-2.1 μm), coupled with a larger range, makes it approximately equal in terms of global photosynthetic biomass, thus allowing them to fix an order of magnitude more carbon (21). *Synechococcus* is usually limited to the upper 100 m of the water column (13), but have a broader latitudinal distribution, withstanding temperatures as low as 2°C (22, 23).

As a key component of picophytoplankton, *Synechococcus* can greatly influence the global cycling of carbon (10) (Fig. 1.1). *Synechococcus* supplies photosynthetically fixed carbon to higher trophic levels, plays a central role in dictating the elemental

composition of seawater over evolutionary time (24), and influences the dynamics and availability of nitrogen, phosphorus, and other potentially limited nutrients in the oceans (25). Although nutrients and trace metals play an important role in *Synechococcus* abundance and distribution, key ecological determinants such as light, temperature, and the degree of physical perturbations also influence the biogeographic distribution of *Synechococcus* (20). Annual mean averages of *Synechococcus* cells are 7.0×10^{26} , which account for 16.7% of total net global primary production. Projections of future global niche models suggest that oceanic *Synechococcus* communities will experience an increase in abundance of at least 14 % by the end of the 21st century, assuming current levels of warming continue. These increases in cell abundance and large shifts in communities will likely occur both in current regions and result in an overall expansion over the next 50 to 100 years (26). Considering the high abundance and productivity of *Synechococcus* in the ocean, it becomes imperative to fully understand how they respond to environmental changes.

1.2 Current understanding of subcluster 5.2 and estuarine *Synechococcus*

Currently, marine *Synechococcus* are divided into three subclusters, 5.1, 5.2, and 5.3 (Fig. 1.2) (16, 18, 27–29), illustrated in Figure 1.2. These include at least 20 clades (23, 27, 30). *Synechococcus* displays geographic niche adaptation (20). More specifically, clade II *Synechococcus* is common in tropical open ocean waters, whereas clades I and IV are largely found in coastal and higher latitude regions (14, 20, 31). The *Synechococcus* found in the estuarine or coastal bays often contain unique genotypes distinct from those in the open oceans. For example, *Synechococcus* isolated from the Chesapeake Bay and East Sea is predominantly from subcluster 5.2 and clade CB4 or

CB5 (32, 33). CB4 and CB5 clade strains lack PUB and some are motile found in temperature coastal waters while, CB5 is unique in that they are prevalent in polar/subpolar waters. However, much less is known about the biogeography of subclusters 5.2 and 5.3 *Synechococcus* compared with subcluster 5.1. The responses of subcluster 5.1 *Synechococcus* to changing environments such as wavelengths of light (27, 29, 34), nitrate (NO₃⁻) concentrations (27, 35, 36), and organic phosphorus sources (30, 37) have been studied. Further, due to their ecological significance and functional divergence, 31 genomes of marine *Synechococcus* have been sequenced (38, 39). Of these 31 genomes, only three belong to subcluster 5.2. While 16 clades have been characterized for subcluster 5.1 *Synechococcus*, only clades CB4 and CB5 have been identified for subcluster 5.2 (23, 40).

Compared to coastal and oceanic *Synechococcus*, much less is known about the ecological function and genomic evolution of estuarine *Synechococcus*. Estuaries are highly productive ecosystems at the precipice of anthropogenic influences. These *Synechococcus* thrive in a highly dynamic environment with constantly changing conditions in which open ocean cyanobacteria would fail. In 2004, 13 strains of *Synechococcus* were isolated from the Chesapeake Bay (Chen et al. 2004). The phylogenetic analyses based on the partial RuBisCO gene sequences and internal transcribed spacer (ITS) sequences show that most of these Chesapeake Bay *Synechococcus* isolates are different from coastal and oceanic strains (19, 32). The clone library analyses based on the environmental *rbcL* and ITS sequences both showed that subcluster 5.2 are the dominant *Synechococcus* in the Chesapeake Bay (19, 32). Microscopic counting shows that picocyanobacteria (mostly *Synechococcus*) make up 20-

40 % of phytoplankton chlorophyll-a in the Chesapeake Bay during the summer (41). Phycoerythrin (PE) containing estuarine subcluster 5.2 *Synechococcus* contain a novel phycobilisome (PBS) gene cluster not seen in other subclusters (42). Nowadays, more subcluster 5.2 *Synechococcus* strains or sequences have been reported in other estuarine or bay environments such as the East Sea, Gulf of Mexico, Bering Sea, Chuckhi Sea, and Baltic Sea (23, 33, 42). Subcluster 5.2 lineages, although generally found in estuarine or coastal environments, have also been identified in Arctic and subarctic oceans. The potential niche adaptation of this *Synechococcus* sub-cluster requires further research.

Synechococcus strain CB0101, a well-referenced strain of subcluster 5.2, represents the most common and abundant *Synechococcus* in the upper Chesapeake Bay (40, 43). As the largest estuarine ecosystem in the U.S., the Chesapeake Bay is known for its strong environmental gradients and seasonality. It has distinct upper and lower sections; its upper part above the Potomac (38° 1'11.8 N; 76° 20'8.7 W) is a deep drowned river channel sided by clay cliffs which have the tendency to become sub-anoxic at depth during the summer time. The lower section is heavily influenced by tides. This hydrology, as well as chemical and physical influences, plays a hefty role in determining *Synechococcus* communities. Microorganisms living in estuaries need to cope with rapid changes to salinity, temperature, light intensity, nutrients, and trace elements. From previous studies it was found that CB0101 could grow under a wide range of salinities (0-35 ppt) and temperatures (4-30 °C), reflecting its adaption to the strong hydrodynamics in the Bay (44). Nutrients in the Chesapeake Bay vary greatly over time and space (45). During the winter and spring, concentrations of nitrate, ammonium, and phosphate are high due to large amounts of precipitation and river runoff. These

nutrients can be used up quickly in the summer and fall when phytoplankton biomass increases dramatically (45–48), then causing seasonal nutrient limitation that suppresses excess algal and cyanobacterial growth (45, 49, 50). Cyanobacteria are heavily influenced in this environment across gradients of light, nutrients, and salinity. Considering the dynamic nature of estuarine ecosystems, *Synechococcus* species living in the Chesapeake Bay experience rapid environmental changes over a short period of time. CB0101 is often used as a phylogenetic reference for subcluster 5.2 (6, 15, 17, 42). Therefore, CB0101 represents a good model strain for studying the ecology and molecular functional capability of estuarine or subcluster 5.2 *Synechococcus*.

Cyanobacterial viruses or cyanophages can influence the abundance, diversity, and productivity of *Synechococcus spp.* in the ocean. Further, it has been predicted that 20-40% of all marine picocyanobacteria killed each day are due to viral lysis (51–53). All known marine cyanophages are tailed double-stranded DNA viruses, belonging to three well-defined bacteriophage families: *Myoviridae*, *Podoviridae*, and *Siphoviridae*. Cyanomyovirus infects a broad range of hosts, while cyanopodoviruses and cyanosiphoviruses are host specific (41, 52–54). Many viruses that infect coastal and open-ocean *Synechococcus* have been isolated from various marine environments. The isolation of subcluster 5.2 *Synechococcus* from Chesapeake Bay allows the further isolation of different viruses, which infect Chesapeake Bay *Synechococcus* (55). Unlike the widely present myoviruses (broad host range) infecting open-ocean *Synechococcus*, viruses infecting estuarine *Synechococcus* are predominantly host-specific podoviruses and siphoviruses (55). The predominance of cyanopodoviruses in the estuarine environment is evident based on the metagenomic recruitments (56). Cyanopodoviruses

are grouped into two phylogenetic clusters (MPP-A and MPP-B) and based on the DNA *pol* gene phylogeny (55, 56).

Recent studies suggest that MPP-B podoviruses are more abundant than MPP-A podoviruses (55–59), but few (12) of their genomes have been completed. Of the cyanopodovirus genomes sequenced, five were isolated using *Synechococcus* CB0101. CB0101's susceptibility to phage infection could have consequences for the cyanobacteria's ecological role due to its relative abundance and similarity to a number of other *Synechococcus spp.* found in the Chesapeake Bay. One of these podoviruses is S-CBP1, which is phylogenetically clustered in MPP-B. S-CBP1 is a podovirus with a capsid size of 52 ± 3 , 48 kb genome, burst size of 86, and a fast latent period of just 6-8 h (55). Another unique characteristic of S-CBP1 is the presence of the cyanobacterial photosynthetic core gene (*psbA*), which has been a point of interest for over a decade (56, 60–62). With the available host (CB0101) and phage (S-CBP1) cultures and their respective genome sequences, we can apply transcriptomics and proteomics to study the gene regulation of both host and phage by following the infection cycle of phage.

1.3 Genomics, Transcriptomics, and Proteomics: what it all means

In recent years, genomic approaches have been used to study questions of ecological importance ranging from adaptation of organisms in changing environments to evolution of complex phenotypes (63, 64). High-throughput sequencing, lowers costs, and methodological advancements have allowed next generation sequencing (NGS) to proliferate. Consequently, omic approaches have now been used to study *Synechococcus* genetic adaptation in rapidly evolving systems and variation among populations and

species. The advent of molecular biology and DNA sequencing methodologies has greatly extended our knowledge on microbial diversity in nature.

Since the first report of a completed genome sequence of marine *Synechococcus* in 2003 (65), a considerable number of partial or complete cyanobacterial genome sequences (100 cyanobacteria, of which 35 are *Synechococcus*) are now publicly available (66). This increasing number of sequenced genomes provides new opportunities for understanding microbial diversity, genomic evolution, metabolic organization, and ecological adaptation in diverse environments. For example, genome sequences of a coastal *Synechococcus* shows greater capacity to sense and respond to environmental stresses compared to an open ocean *Synechococcus* (67). The genomes of marine *Synechococcus* range in size from 2.2 to 2.9 Mbp with 52-67 % of their genes (~1570) common to all strains (38, 40). Extensive work has been done comparing *Synechococcus* gene pigment types and associated phycobilisome structures, leading to ecotype models for adaptation to various light niches (40, 68–70). There are at least 30 genome sequences of coastal and open ocean *Synechococcus* strains in the database (39). However, only two *Synechococcus* genomes (CB0101 and CB0205) come from the estuarine environment. *Synechococcus* WH5701 comes from a mud flat, and its genome sequence has been suggested to be a contaminated sequence which contains partial genome sequences from two different strains (40). In this dissertation, I used three complementary sequencing tools to gain a better understanding of how the estuarine *Synechococcus* CB0101 adapts to its highly variable environment.

Genomics studies the entire set of genes as a dynamic system over time to determine how they interact and influence biological pathways, networks, and physiology

within an organism. There are numerous sequencing platforms (454 pyrosequencing, PacBio, Illumina Hi-Seq, Illumina Mi-Seq, and Ion torrent), each with advantages and disadvantages. In my study, the genome of CB0101 was sequenced using two types of sequencers: 454 GS-FLX and PacBio. The advantages of these sequencing methods have been compared (71). Genomics is the bedrock upon which further functional studies into the transcriptome and proteome can be built (72).

The transcriptome is the complete set of transcripts in a cell and their quantity for a specific developmental stage or physiological condition (73). Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells (74). The key aims of transcriptomics are: to catalogue all species of transcript, including mRNAs, non-coding RNAs, and small RNAs; to determine the transcriptional structure of genes, in terms of their start sites, 5' and 3' ends, splicing patterns and other post-transcriptional modifications; and to quantify the changing expression levels of each transcript during development and under different conditions (73). Transcriptomics of *Synechococcus* has cataloged their responses to environmentally relevant stresses (75, 76), confirm putative non-coding transcripts (77), identify adaptation responses (78), explore cyanophage-host interactions (79), and explore time-variable synchronous patterns of gene expression (80). In this dissertation, I applied RNA-Seq to explore the transcriptional response of CB0101 to various environmental stresses and phage infection.

The term proteome was first coined to describe the set of proteins encoded by the genome (81). The study of the proteome, called proteomics, now evokes not only the proteins in any given cell, but also the set of all protein isoforms and modifications, the

interaction between them, the structural description of proteins, and their higher-order complexes (72, 82, 83). Proteomics complements other functional genomic approaches, including RNA-Seq based expression profiles and systematic genetics. Integration of these datasets through bioinformatics serves as a powerful reference of protein properties and functions (72, 82). Proteome work on *Synechococcus* is not as prevalent as genomics or transcriptomics, but recent studies have identified posttranslational modifications, response to nutrient stress (84), daily rhythms of cyanobacteria (85), cell division studies (86), and *Synechococcus-Roseobacter* interactions (87).

Physiological studies of *Synechococcus spp.* are important in understanding how a strain responds to specific growth conditions. However, by themselves, such studies do not provide information on biological control and regulation at the molecular level. The integration of molecular techniques and physiological tests is a powerful approach to predict function, interaction, competition for nutrients, symbioses, and other processes that determine the overall roles of the distinct marine organisms. Towards this goal, I combined physiological data derived from ‘model strain’ isolates with functional responses via omics (genomic, transcriptomic, proteomic, and metabolomics analyses) to explore how *Synechococcus* regulatory networks underpin marine ecological distinctness in estuarine environments.

1.4 Scope of this dissertation

Current knowledge of *Synechococcus spp.* is largely built upon open-ocean and coastal ecosystems studies. Little effort has been made to understand the physiological and ecological traits associated with estuarine *Synechococcus*, specifically those of subcluster 5.2. Being at the precipice of anthropogenic influence and its ability to thrive

in the dynamic estuarine ecosystem, *Synechococcus* CB0101 presents a perfect opportunity to explore adaptive mechanisms of this poorly understood group. Coupled with the immense depth and vast data produced using the latest techniques in NGS, further exploration will allow for an unprecedented examination of molecular functions of estuarine picocyanobacteria. This dissertation has helped bring into focus the poorly understood *Synechococcus* subcluster 5.2 and opened the door to understanding novel molecular drivers of *Synechococcus* adaptation.

The dissertation is divided into 6 chapters, all addressing the ecological adaptation of *Synechococcus* CB0101. The term adaptation, by its general nature, encompasses a wide range of mechanisms and evolutionary processes that lead to enhanced fitness and survival of individuals. My research integrates physiology, comparative genomics, and gene expression at the mRNA and protein levels. The work will focus on the following questions: 1) Does estuarine *Synechococcus* strain CB0101 respond to environmental stresses differently compared to coastal and oceanic *Synechococcus* strains? 2) Do CB0101 gene regulatory networks underpin estuarine ecological distinctness? 3) What differences in regulatory gene and protein expression for stress responses are present for CB0101? 4) How does CB0101 respond to phage infection at the omics level?

The three major findings in my dissertation include:

1) *Estuarine Synechococcus CB0101 is highly adaptable*

When compared with coastal and open ocean *Synechococcus* (WH7803 and WH7805), estuarine *Synechococcus* CB0101 demonstrated its higher adaptability to changing environments, such as heavy metals, salinity, and temperature. Comparative genomics between CB0101, WH7803, and WH7805 also showed that CB0101 has an

increased genetic capacity to sense, respond, and adapt to variable environments compared to its coastal and oceanic counterparts (Chapter 3).

2) *Toxin-antitoxin systems aid Synechococcus in responding to variable environments*

Toxin-antitoxin (TA) systems have been found in many bacteria but not previously in marine picocyanobacteria. TAs are stress response systems that allow microbial cells to survive under harsh environmental conditions. We demonstrated that TA systems, specifically *relB/relE* in marine *Synechococcus*, respond to environmental stresses and play an active role in conferring niche adaptation. This growth modulator induces a reversible persist state to enhance fitness and competitiveness. As a result, picocyanobacteria can persist during prolonged stress conditions and revive when suitable conditions arise. Further, eight other *Synechococcus* strains were found to contain TA systems, suggesting the role they play in the diversity and function of picocyanobacteria is still unknown.

3) *Genomic-Island genes play a major role during phage infection*

Synechococcus spp. are abundant in the marine estuarine ecosystem and subjected to frequent viral infection. What interactions occur at the molecular level during infection are not well defined. Thus, we investigated the whole transcriptome and proteome expression of the estuarine *Synechococcus* (CB0101) upon infection of podovirus S-CBP1. Strikingly, most expression occurred within genomic islands (GI), which were massively shutdown within 30 min post phage infection. Throughout the infection, phage-encoded host-like proteins (*hli*, *psbA*, *ThyX*) were highly expressed. Such a strong response in the GI system was not observed when CB0101 was exposed to nutrient or heavy metal stresses. Podovirus S-CBP1 seems to shut down the host GI system promptly

to prevent the development of host resistance, meanwhile promoting the activities of some "shared genes" between phage and host. The global shutdown of GI genes in estuarine *Synechococcus* could be an important phage-regulated process for a successful lytic infection.

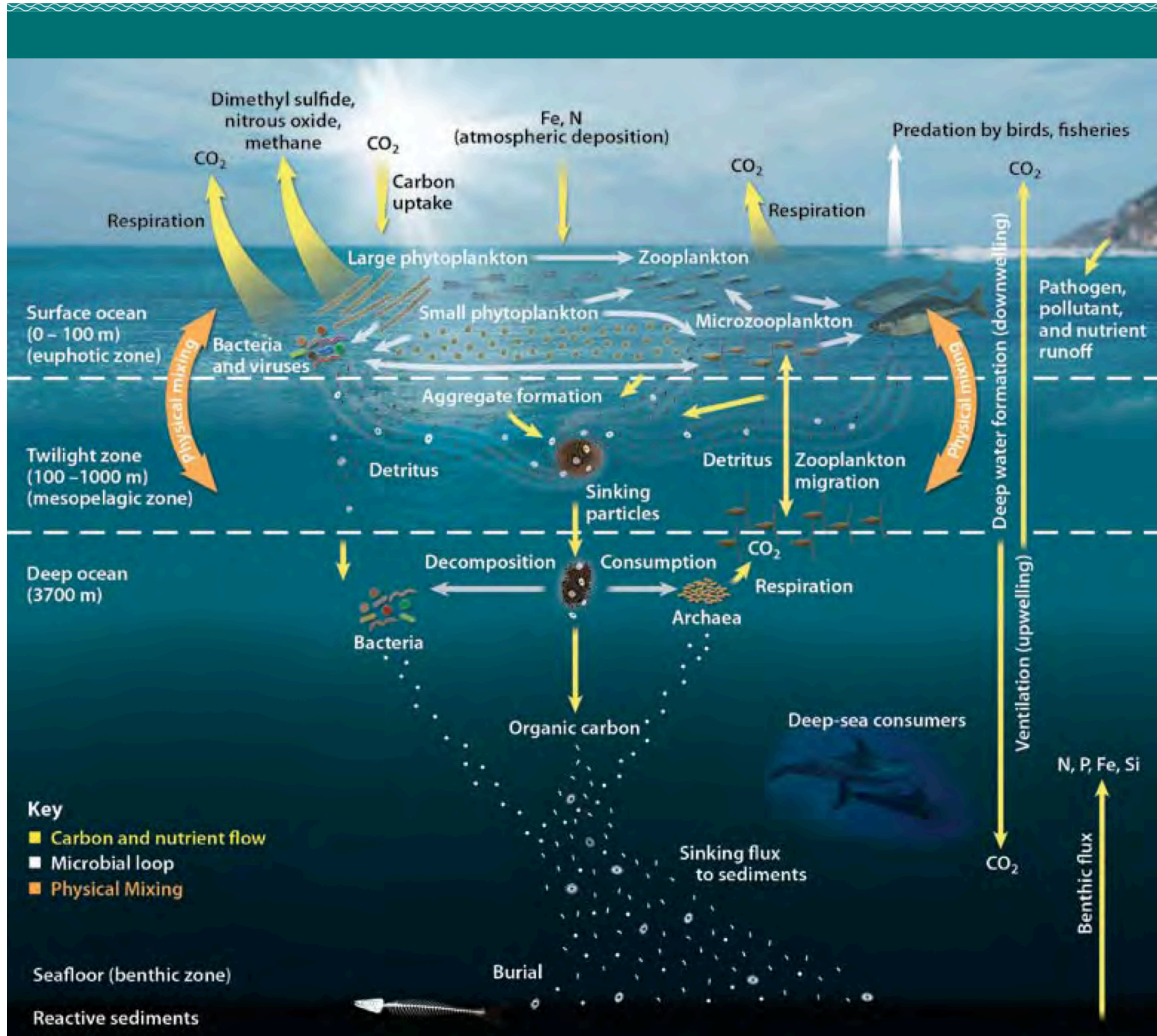


Fig. 1.1. Marine Food Web and the Carbon Cycle: Small photosynthetic organisms, including picocyanobacteria (i.e. *Synechococcus*) lie at the heart of the marine carbon cycle. Sunlit surface waters teem with picocyanobacteria that convert inorganic carbon dissolved in surface waters to organic carbon – which forms the basis of the marine food web – and accounts for about half of all primary production on earth. CO₂ fixed during photosynthesis in the upper ocean can be transferred to the depths via three major processes: passive sinking of particles, physical mixing of particulates and dissolved organic matter through currents, and active transport by zooplankton migrating to deeper

water. Climate change and ocean acidification significantly alter marine ecosystems functions, the efficiency of this biologically mediated ocean carbon export may change, leading to an effect on the net annual uptake of carbon.

Phytoplankton, such as those described above, are grazed upon by marine heterotrophs known as zooplankton. Viruses, which act as predators in oceanic food chains by infecting and lysing marine bacteria, also play an important but still poorly understood role in marine carbon turnover. The overall efficiency with which organic carbon is exported to the deep ocean depends on the type of phototrophic cells that create the organic material and the efficacy with which heterotrophic organisms respire it.

Carbon fixed in small phytoplankton (i.e. *Synechococcus*) eventually enter the water column as either particulate or dissolved organic carbon through directed exudation, consumption by grazing zooplankton, viral lysis, or cell death. Subsequently, most of this carbon material is degraded by heterotrophic bacteria, resulting in particulate solubilization and conversion of organic carbon back to CO₂. CO₂ fixed during photosynthesis by small phytoplankton (i.e. *Synechococcus*) in the upper ocean can be transferred to the depths via three major processes: passive sinking of particles, physical mixing, and active transport by zooplankton. However, if climate change and ocean acidification significantly alter marine ecosystems' functions, the efficiency of this biologically mediated ocean carbon export may change, leading to an indirect effect on the net annual uptake of carbon. For full description see; (88), adapted with permission from authors and publisher.

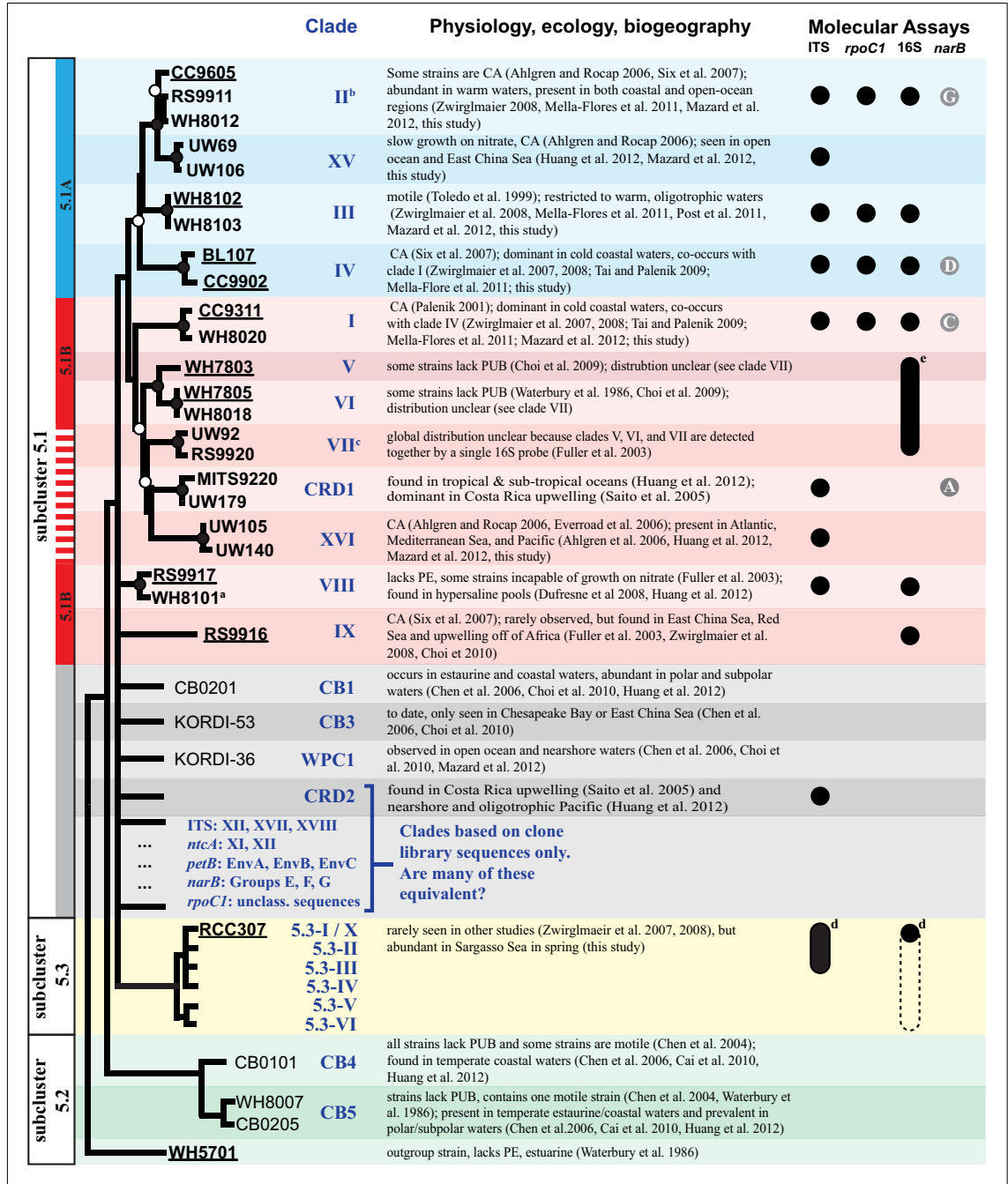


Fig. 1.2 Schematic phylogenetic tree summarizing the genetic and ecological diversity of marine *Synechococcus* and molecular assays that have been developed to measure particular clades in the oceans. The depicted tree is approximately based on a four loci concatenated tree. Examples of commonly studied strains in each clade are shown. Strains in bold were selected for multiple gene loci analysis, and strains with

sequenced genomes are underlined. Branch tips without taxa names indicated clades that are only represented by environmental sequences. Nodes with bootstrap support of >55% with circles. Polytomies indicate cases where the placement of clades is uncertain because of low bootstrap support or the absence of sequences available for multiple loci. Marine *Synechococcus* are classified into three subclusters 5.1, 5.2, and 5.3. Chromatic adaptation (CA) is the ability to shift the ratio of accessory pigments phycourobilin (PUB) and phycoerythrobilin (PEB) according to the spectral quality of light. Circles indicate clades for which quantitative molecular assays have been designed to measure their abundance in the environment. See original article for detailed description (17), adapted with permission from authors and publishers.

References:

1. Herrero A, Flores E (2008) The Cyanobacteria: Molecular Biology, Genomics and Evolution | Book. *Caister Acad Press* (February):484.
2. Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* 455(7216):1101–1104.
3. Whitton BA, Potts M (2000) The Ecology of Cyanobacteria: Their Diversity in Time and Space. *Kluwer Acad Publ Netherlands*:37–59.
4. Johnson PW, Sieburth JM (1979) Chroococcoid cyanobacteria in the sea: A ubiquitous and diverse phototrophic biomass. *Limnol Oceanogr* 24(5):928–935.
5. Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979) Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293–294.
6. Shih PM, et al. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* 110(3):1053–8.
7. Calteau A, et al. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics* 15(1):977.
8. Partensky F, Blanchot J, Vaultot D (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters : a review. *Bull l'Institut océanographique* 19(19):457–475.
9. Herdman M, Castenholz R, Waterbury J, Rippka R. (2012) Form-genus XIII. *Synechococcus*. *Bergey's Manual of Systematic Bacteriology*, pp 508–512.
10. Field CB (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281(5374):237–240.

11. Johnson ZI, et al. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311(5768):1737–1740.
12. Jiao N, et al. (2005) Dynamics of autotrophic picoplankton and heterotrophic bacteria in the East China Sea. *Cont Shelf Res* 25(10):1265–1279.
13. Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63(1):106–127.
14. Toledo G, Palenik B (1997) *Synechococcus* diversity in the California Current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. *Appl Environ Microbiol* 63(11):4298–4303.
15. Rocap G, Distel DL, Waterbury JB, Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68(3):1180–1191.
16. Fuller NJ, Marie D, Vaulot D, Post AF, Scanlan DJ (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* 69(5):2430–2443.
17. Ahlgren NA, Rocap G (2012) Diversity and distribution of marine *Synechococcus*: Multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* 3(JUN). doi:10.3389/fmicb.2012.00213.
18. Penno S, Lindell D, Post AF (2006) Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-

- regulatory gene *ntcA*. *Environ Microbiol* 8(7):1200–1211.
19. Chen F, et al. (2004) Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* 36(2):153–164.
 20. Zwirgmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
 21. Jardillier L, Zubkov M V, Pearman J, Scanlan DJ (2010) Significant CO₂ fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* 4(9):1180–1192.
 22. Shapiro LP, Haugen EM (1988) Seasonal distribution and temperature tolerance of *Synechococcus* in Boothbay Harbor, Maine. *Estuar Coast Shelf Sci* 26(5):517–525.
 23. Huang S, et al. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 6(2):285–97.
 24. Redfield AC (1958) The biological control of chemical factors in the environment. *Am Sci* 46(3):230A–221.
 25. Falkowski PG, Barber RT, Smetacek V (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* 281:200–206.
 26. Flombaum P, et al. (2013) Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA* 110(24):9824–9.
 27. Ahlgren NA, Rocap G (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N

- physiologies. *Appl Environ Microbiol* 72(11):7193–7204.
28. Muhling M, et al. (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using rpoC1-based restriction fragment length polymorphism. *Aquat Microb Ecol* 45(3):263–275.
 29. Six C, et al. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus spp.*: a comparative genomics study. *Genome Biol* 8(12):R259.
 30. Mazard S, Ostrowski M, Partensky F, Scanlan DJ (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* 14(2):372–386.
 31. Ferris MJ, Palenik B, Ferris (1998) Niche adaptation in ocean cyanobacteria. *Nature* 396:226–228.
 32. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* 72(3):2239–2243.
 33. Choi DH, Noh JH (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* 69(3):439–448.
 34. Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67(2):991–994.
 35. Fuller NJ, Tarran G a., Yallop M, Orcutt KM, Scanlan DJ (2006) Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol Oceanogr* 51(6):2515–2526.
 36. Moore LR, Post AF, Rocap G, Chisholm SW (2002) Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and

- Synechococcus*. *Limnol Oceanogr* 47(4):989–996.
37. Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T (2005) Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* 39(3):257–269.
 38. Dufresne A, et al. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9(5):R90.
 39. Federhen S (2015) Type material in the NCBI Taxonomy Database. *Nucleic Acids Res* 43(1):1086–1098.
 40. Scanlan DJ, et al. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 73(2):249–299.
 41. Wang K, Wommack KE, Chen F (2011) Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. *Appl Environ Microbiol* 77(21):7459–7468.
 42. Larsson J, et al. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 8(9):1892–903.
 43. Wang K (2007) Biology and ecology of *Synechococcus* and their viruses in the Chesapeake Bay. doi:10.1017/CBO9781107415324.004.
 44. Marsan D, Wommack KE, Ravel J, Chen F (2014) Draft genome sequence of *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay estuary. *Genome Announc* 2(1):e01111–13.
 45. Murphy RR, Kemp WM, Ball WP (2011) Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts* 34(6):1293–1309.

46. Sommer SE, Pyzik AJ (1974) Geochemistry of middle Chesapeake Bay sediments from upper Cretaceous to Present. *Chesap Sci* 15(1):39–44.
47. Sinex SA, Helz GR (1981) Regional geochemistry of trace elements in Chesapeake Bay sediments. *Environ Geol* 3(6):315–323.
48. Buchanan C, Lacouture R V., Marshall HG, Olson M, Johnson JM (2005) Phytoplankton reference communities for Chesapeake Bay and its tidal tributaries. *Estuaries* 28(1):138–159.
49. Fisher TR, Peele ER, Ammerman JW, Harding LW (1992) Nutrient limitation of phytoplankton in Chesapeake Bay. *Mar Ecol Prog Ser* 82:51–63.
50. Fisher TR, et al. (1999) Spatial and temporal variation of resource limitation in Chesapeake Bay. *Mar Biol* 133(4):763–778.
51. Proctor LM, Fuhrman J a. (1990) Viral mortality of marine bacteria and cyanobacteria. *Nature* 343:60–62.
52. Suttle C a., Chan AM, Cottrell MT (1990) Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* 347(6292):467–469.
53. Waterbury JB, Valois FW (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* 59(10):3393–3399.
54. Marston MF, Sallee JL (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol* 69(8):4639–4647.
55. Wang K, Chen F (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* 10(2):300–312.

56. Huang S, Zhang S, Jiao N, Chen F (2015) Comparative genomic and phylogenomic analyses reveal a conserved core genome shared by estuarine and oceanic cyanopodoviruses. *PLoS One* 10(11):e0142962.
57. Huang S, Wilhelm SW, Jiao N, Chen F (2010) Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *ISME J* 4(10):1243–1251.
58. Marston MF, et al. (2013) Marine cyanophages exhibit local and regional biogeography. *Environ Microbiol* 15(5):1452–1463.
59. Labrie SJ, et al. (2013) Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ Microbiol* 15(5):1356–1376.
60. Mann NH, et al. (2003) Bacterial photosynthesis genes in a virus. *Nature* 424:741–742.
61. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438(7064):86–9.
62. Sullivan MB, Coleman ML, Weigle P, Rohwer F, Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol* 3(5):0790–0806.
63. Wagner A (2011) The molecular origins of evolutionary innovations. *Trends Genet* 27(10):397–410.
64. Harrison PW, Wright AE, Mank JE (2012) The evolution of gene expression and the transcriptome-phenotype relationship. *Semin Cell Dev Biol* 23(2):222–229.
65. Palenik B, et al. (2003) The genome of a motile marine *Synechococcus*. *Nature*

- 424(6952):1037–1042.
66. Tatusova T, Ciufu S, Fedorov B, O’Neill K, Tolstoy I (2014) RefSeq microbial genomes database: New representation and annotation strategy. *Nucleic Acids Res* 42(D1). doi:10.1093/nar/gkt1274.
 67. Palenik B, et al. (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci* 103(36):13555–13559.
 68. Schwarz D, Orf I, Kopka J, Hagemann M (2013) Recent applications of metabolomics toward cyanobacteria. *Metabolites* 3(1):72–100.
 69. Coutinho F, Tschoeke DA, Thompson F, Thomson C (2015) Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus*. *PeerJ*:e–1522.
 70. Kelly L, Ding H, Huang KH, Osburne MS, Chisholm SW (2013) Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J* 7(9):1827–41.
 71. Loman NJ, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30(5):434–9.
 72. Tyers M, Tyers M, Mann M, Mann M (2003) From genomics to proteomics. *Nature* 422:193–7.
 73. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63.
 74. Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12(2):87–98.
 75. Ludwig M, Bryant DA (2012) *Synechococcus* sp. Strain PCC 7002 transcriptome:

- Acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. *Front Microbiol* 3(OCT):1–14.
76. Ludwig M, Bryant DA (2012) Acclimation of the global transcriptome of the cyanobacterium *Synechococcus sp.* strain PCC 7002 to nutrient limitations and different nitrogen sources. *Front Microbiol* 3(APR):1–15.
77. Vijayan V, Jain IH, O’Shea EK (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biol* 12(5):R47.
78. Billis K, Billini M, Tripp HJ, Kyrpides NC, Mavromatis K (2014) Comparative transcriptomics between *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 provide insights into mechanisms of stress acclimation. *PLoS One* 9(10):1–10.
79. Doron, S; Fedida A (2016) Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* 10:1437–1455.
80. Ottesen E a, et al. (2013) Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc Natl Acad Sci U S A* 110(6):E488–97.
81. Wilkins MR, et al. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13:19–50.
82. Bantscheff M, Lemeer S, Savitski MM, Kuster B (2012) Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. *Anal Bioanal Chem* 404(4):939–965.
83. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422(6928):198–207.

84. Cox AD, Saito MA (2013) Proteomic responses of oceanic *Synechococcus* WH8102 to phosphate and zinc scarcity and cadmium additions. *Front Microbiol* 4:1–17.
85. Guerreiro ACL, et al. (2014) Daily rhythms in the cyanobacterium *Synechococcus* elongatus probed by high-resolution mass spectrometry based proteomics reveals a small-defined set of cyclic proteins. *Mol Cell Proteomics*:13(8):1–37.
86. Koksharova O a, Klint J, Rasmussen U (2007) Comparative proteomics of cell division mutants and wild-type of *Synechococcus* sp. strain PCC 7942. *Microbiology* 153:2505–17.
87. Christie-Oleza JA, Scanlan DJ, Armengaud J (2015) “You produce while I clean up”, a strategy revealed by exoproteomics during *Synechococcus-Roseobacter* interactions. *Proteomics* 15(20):3454–3462.
88. DOE US (2008) Carbon Cycling and Biosequestration: Report from the March 2008 Workshop. *US Dep Energy Off Sci* (DOE/SC-108).

Chapter 2: Estuarine *Synechococcus* are more tolerant to changing environments compared to coastal and open ocean *Synechococcus*

Estuarine *Synechococcus* are more tolerant to changing environments compared to coastal and open ocean *Synechococcus*

David Marsan¹ and Feng Chen^{1, *}

¹Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD, USA

Running Title: Estuarine *Synechococcus*

Keywords: *Synechococcus*, stress response, estuary, comparison, subcluster 5.2

2.1 Abstract

Estuaries are among the most productive and dynamic aquatic ecosystems. Mixing of fresh and marine waters provide strong environmental gradients to the cyanobacteria living in these ecosystems. *Synechococcus* are the dominant picocyanobacteria in the estuarine environment; however, little is known about their niche adaptation in relevance to their coastal and open ocean counterparts. In this study, we compared the growth performance of representative estuarine, coastal, and oceanic *Synechococcus* strains (CB0101, WH7803, and WH7805) under different chemical and physical conditions, such as varied nutrients (nitrate, urea, ammonium, cyanate, and phosphate), metals (copper, zinc, and iron), salinity, temperature, and light intensity. We found that, estuarine *Synechococcus* CB0101 has a broader salt and temperature tolerance. In addition, CB0101 is less sensitive to metal toxicity but more sensitive to high light as compared to WH7803 and WH7805. Nitrogen and phosphate utilization was statistically the same throughout the three strains. Our results suggest that estuarine *Synechococcus* grow in a much wider range of environmental gradients compared to the coastal and open ocean strains represented by WH7803 and WH7805, respectively. This physiological comparison based on CB0101 is the first step towards predicting functions, interaction, competition for nutrients, and other processes that determine the overall roles of estuarine *Synechococcus*.

2.2 Introduction

Synechococcus are widely distributed and abundant picocyanobacteria which contribute to primary production in the surface waters of global oceans (1). Marine *Synechococcus* are phylogenetically and phenotypically diverse (2) and are divided into 3 subclusters: 5.1, 5.2 and 5.3 (3, 4). Of these, subcluster 5.1 (predominantly found in open ocean environments) represents a well-studied group of marine *Synechococcus*. Subcluster 5.1 contains many *Synechococcus* cultures isolated from coastal and open ocean, which can be divided into at least 20 clades (2). Extensive studies have been done to understand the physiology, diversity, genomics and evolution of subcluster 5.1 *Synechococcus* (2, 5–10). In contrast, our knowledge on subcluster 5.2 and 5.3 *Synechococcus* is limited (11–13).

Subcluster 5.2 strains are phycocyanin-enriched (PC-type) euryhaline strains widely distributed in coastal and estuarine waters (2, 12, 14). The establishment of subcluster 5.2 began with the isolation and characterization of many *Synechococcus* from the Chesapeake Bay (14, 11). As the largest estuarine ecosystem in the U.S., the Chesapeake Bay is known for its strong environmental gradients and being seasonally influenced. Two major clades (CB4 and CB5) within subcluster 5.2 have been identified in the Chesapeake Bay (11, 12, 15). Subcluster 5.2 *Synechococcus* dominates the upper Chesapeake Bay where the salinity is relatively low, while subcluster 5.1 becomes more abundant in the lower bay where tidal influences bring high salinity waters of the Atlantic Ocean (11, 15). The composition of subcluster 5.2 *Synechococcus* varies seasonally in the bay (16). Subcluster 5.2 *Synechococcus* are prevalent in summer, decreases in the winter,

but clade CB4 is still present even during lows of 4°C (15). Recently, the presence of subcluster 5.2 *Synechococcus* was reported in other coastal estuaries, such as East Sea (17), Gulf of Mexico, Bering Sea, Chuckhi Sea(12), and Baltic Sea(13).

Despite the wide presence of subcluster 5.2 *Synechococcus* in coastal estuaries, little is known on their environmental tolerance, particularly in comparison with coastal and open ocean *Synechococcus*. Numerous studies have been conducted to compare the growth of coastal and open ocean *Synechococcus* under different nutrient, light, temperature, and salinity conditions (6, 10, 18–22). Considering that *Synechococcus* living in estuaries need to cope with rapid changes to salinity, temperature, light intensity, nutrients, and trace elements, we hypothesize that estuarine *Synechococcus* can handle variable environmental conditions compared to coastal and oceanic *Synechococcus*.

Synechococcus strain CB0101, a well-referenced estuarine strain of subcluster 5.2, represents the most common and abundant *Synechococcus* in the upper Chesapeake Bay (2, 23). CB0101-like *Synechococcus* have been found throughout the Chesapeake Bay under a wide range of salinities (0 to 35 ppt) and temperatures (4 to 30 °C), reflecting its adaption to the hydrodynamics in the Bay (11). Nutrients in the Chesapeake Bay vary greatly over time (seasonally) and space (upper to lower Bay) (24). In the winter-spring time, concentrations of nitrate, ammonium, and phosphate are usually high due to large amounts of precipitation and river runoff. Consequently, these nutrients are used up more quickly in summer-fall seasons when phytoplankton biomass increases dramatically (24–27). The concentrations of nitrogen and phosphorus in the Chesapeake Bay, water can be low enough to seasonally limit excess algal and cyanobacterial growth (24, 28, 29). Considering the dynamic nature of estuarine ecosystems, *Synechococcus*

species living in the Chesapeake Bay experience rapid environmental changes over a short period of time, within a small spatial distance.

In order to fully understand the niche adaptation of estuarine *Synechococcus*, we propose to compare the growth performance of representative strains of estuarine, coastal, and oceanic *Synechococcus* in response to different chemical and physical settings. In this study, we compared the growth of three representative *Synechococcus* strains, CB0101 (estuarine), WH7803 (coastal), and WH7805 (open ocean), under various growth conditions, which include salinity, temperature, nutrient, light, and trace metals.

2.3 Methods

2.3.1 Growth conditions of Synechococcus CB0101, WH7803, and WH7805

CB0101 was cultivated in SN medium (30), with 15 ppt and vitamin B12 (10 μ g/L) (referred as SN15 hereafter). WH7803 and WH7805 were grown in accordance to previously published optimal conditions (A+ and SN30 media) (31, 32). In the laboratory, all cultures were grown at 23°C temperature with 15 μ E m⁻² s⁻¹ continuous light. Filtered (0.2 μ m pore size) air was bubbled into individual 500 mL cell culture flasks. From these main cultures, exponentially growing cells, at an optical density (OD) of 0.6, were separated into triplicate 25 mL cell culture flasks and subjected to the various conditions for at least 72 hours.

To compare the stress responses between estuarine, coastal and oceanic *Synechococcus*, estuarine strain CB0101, coastal strain WH7803, and open-ocean strain WH7805 were used in this study. WH7803 and WH7805 were selected because they have been used for extensive physiological and molecular work and are well represented

as coastal and oceanic ecotype strains (2, 4, 9, 31). The growth rates of CB0101, WH7803, and WH7805 under different concentrations and sources of nutrients and metals were compared. Cell counts were collected for each strain under a range of salinity (0-35 ppt), temperature (4-35 °C), and light intensity (15-400 $\mu\text{E m}^{-2} \text{s}^{-1}$) with their growth rates calculated (Fig. 2.2). For the nutrient test, triplicate cultures were subjected to near 0, 0, 1 nM, 1 μM , 1 mM, and 1 M of nitrate, urea, cyanate, ammonium, and phosphate, respectively (Fig. 2.3). For the metal tolerance experiment, CB0101, WH7803, and WH7805 were cultivated at near 0, 1 nM, 1 μM , 1 mM, and 1 M of zinc, copper, and iron, respectively (Fig. 2.4). Triplicate samples for every treatment were taken at each time point. The growth of cultures was monitored by counting cells using an Accuri C6 flow cytometer. The specific growth rate was calculated from the logarithmic change $\mu = \frac{\ln X_2 - \ln X_1}{t_2 - t_1}$ where X1 and X2 are densities at times t1 and t2.

2.4 Results

2.4.1 Growth comparison of estuarine, coastal, and open-ocean *Synechococcus* in different conditions

CB0101 was able to grow in salinities of near 0, 10, 15, 20, 25, 30, and 35 ppt, with the optimal growth at 15 ppt (Fig. 2.1A). WH7803 and WH7805 maintained high growth rates when salinity was higher than 20 ppt. However, WH7803 and WH7805 growth decreased dramatically when the salinity dropped from 20 to 10 ppt (Fig. 2.1A) and nearly ceased to grow below 10 ppt. When cultivated at 7 different temperatures (4, 10, 17, 20, 25, 30, and 35 °C), CB0101 showed optimal growth at 25 °C, while WH7803 and WH7805 reached optimal growth at 30 °C (Fig. 2.1B). All three strains were greatly inhibited when temperatures were above 35 °C or below 15 °C; however, CB0101 was

able to maintain a relatively high growth rate compared to WH7803 and WH7805. It is noteworthy that CB0101 was able to grow slowly at 4 °C (with a specific growth rate $\sim 0.22 \text{ d}^{-1}$), while the growth of WH7803 and WH7805 were completely inhibited at 4 °C (Fig. 2.2B). CB0101 responded to light changes differently from WH7803 and WH7805 (Fig. 2.2C). The growth rate of CB0101 decreased with increasing light intensity (15 to $400 \mu\text{E m}^{-2}\text{s}^{-1}$), but the growth rates of WH7803 and WH7805 increased with increasing light intensity. WH7805 grew slightly faster than WH7803 under the high light conditions ($200\text{-}400 \mu\text{E m}^{-2}\text{s}^{-1}$). In general, CB0101 grew best at lower light intensities ($15\text{-}100 \mu\text{E m}^{-2}\text{s}^{-1}$), while WH7803 and WH7805 grew optimally at higher light intensities ($>150 \mu\text{E m}^{-2}\text{s}^{-1}$) (Fig. 2.1C).

Similar growth responses for CB0101, WH7803, and WH7805 were seen under all nitrogen sources and concentrations, except for WH7803, which was unable to grow with urea (Fig. 2.3A, B, C, D). CB0101 had the highest growth rate with $1 \mu\text{M}$ - 1 mM nitrate, or 1 nM urea, cyanate, or ammonium. WH7803 and WH7805 showed a similar growth pattern with these nitrogen sources and concentrations. WH7803 was not able to grow with urea (Fig. 2.2B) as a nitrogen source but was able to utilize cyanate, while WH7805 displayed the highest growth rate at 1 nM of urea and cyanate (Fig. 2.2B and 2.3D). CB0101 outperformed WH7803 and WH7805 at 1 mM ammonium (Fig. 2.2C).

When comparing growth under different phosphate concentrations, both CB0101 and WH7803 exhibited the highest growth rate at 1 mM phosphate, while WH7805 grew best at $1 \mu\text{M}$ phosphate (Fig. 2.3A). CB0101 had a lower growth rate compared to WH7803 and WH7805 with phosphate concentrations less than $1 \mu\text{M}$. This data shows

that CB0101 does not grow well in environments with lower concentration of phosphate and prefers to grow in phosphate rich environment.

Zinc, copper, and iron are essential micronutrients for cyanobacteria, but at higher concentrations they can be toxic to *Synechococcus*. CB0101 was able to maintain high growth rates ($\sim 0.95 \text{ d}^{-1}$) with zinc and copper at concentrations between 1 nM and 10 mM, while the growth of WH7803 and WH7805 were inhibited when zinc and copper concentrations were higher than 1 nM (Fig. 2.3B & 2.3C). The growth of CB0101 was inhibited when zinc and copper concentrations exceeded 1 μM . CB0101 showed the highest growth rate at 1 mM iron and proliferated in a wide range of iron concentrations (1 nM to 1 M) (Fig. 2.3D). However, the growth of WH7803 and WH7805 were inhibited when iron concentrations were higher than 1 mM. Taken together, our results show that CB0101 has a wider tolerance to metal toxicity than WH7803 and WH7805.

2.5 Discussion

2.5.1 Environmental tolerance of estuarine Synechococcus

Comparative physiology tests show that estuarine *Synechococcus* CB0101 is more flexible in response to changing nutrient and trace elements compared to coastal and oceanic *Synechococcus*, represented by WH7803 and WH7805, respectively. CB0101 was originally isolated from the Chesapeake Bay (15, 33). As the largest estuarine ecosystem in the U.S., the Chesapeake Bay is well known for its strong environmental gradients and seasonality. Microorganisms living in estuaries need to cope with rapid changes to salinity, temperature, light intensity, nutrients, and trace elements. CB0101 can grow in a wide range of salinities (0 to 35 ppt) and temperatures (4 to 30 °C), reflecting its adaption to the strong hydrodynamics in the Bay. In contrast, coastal and

oceanic strains WH7803 and WH7805 failed to grow in lower salinity (<10 ppt) and temperature (<10-15 °C). It is noteworthy that CB0101 is more sensitive to higher light intensity ($> 150 \mu\text{E m}^{-2} \text{s}^{-1}$) compared to coastal and oceanic *Synechococcus* (Fig. 2.1C). The upper Chesapeake Bay is dominated by high turbidity water that causes light to be attenuated quickly(27). CB0101 outperformed WH7803 and WH7805 in higher concentrations of phosphate (>1 mM) but is less competitive in the lower end (phosphate <1nm). We did not observe a significant difference in growth among CB0101, WH7803, and WH7805 under different concentrations of nitrate and cyanate. Marine *Synechococcus* seem to share a large portion of common metabolism capabilities for nitrogen and phosphate sources (2, 9). CB0101 grew faster than WH7803 and WH7805 at 1 mM of ammonium, but the growth of all three strains was inhibited at 1 M of ammonium. Nutrients in the Chesapeake Bay vary greatly over time and space (24). In the winter-spring time, concentrations of nitrate, ammonium, and phosphate are usually high due to increased precipitation and river runoff, while these nutrients can be used up quickly in summer-fall seasons when phytoplankton biomass increases dramatically(24–27). The concentrations of nitrogen and phosphorus in the Chesapeake water can be low enough to seasonally limit excess algal and cyanobacterial growth (24, 28, 29).

It is clear that CB0101 has a higher tolerance for metal toxicity such as zinc and copper compared to WH7803 and WH7805. In estuarine environments, a sharp spike or decline in concentrations of metals can occur after a heavy rain or strong tidal movement (25, 26). Considering the dynamic nature of estuarine ecosystem, *Synechococcus* species living in the Chesapeake Bay can experience rapid environmental changes over a short period of time or within a small spatial distance. It can be hypothesized that due to the

variability of the environment CB0101 has gained genes, which would aid in adapting to such conditions.

2.6 Conclusion

Our results show that *Synechococcus* CB0101 is more adapted to grow at low light, high concentrations of nutrients and metals, and wider salinity or temperature ranges relative to WH7803 and WH7805. The results here are the first step towards understanding the ecological adaptation of estuarine *Synechococcus*. Additional studies based on genomics and functional genomics could provide more insight into specific genes or pathways that are responsive to the changing environment.

Figures:

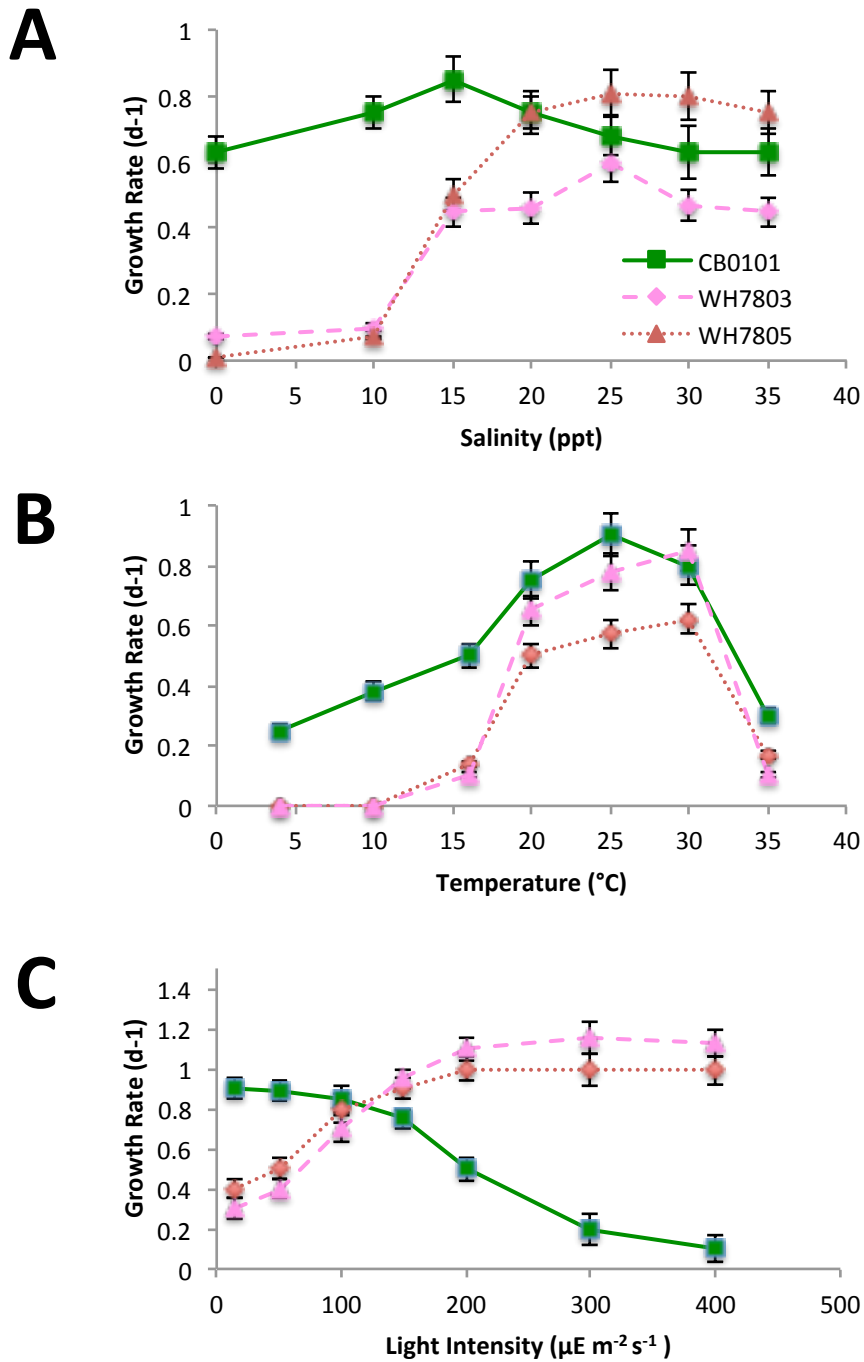


Figure 2.1. Comparison of specific growth rates between an estuarine (CB0101), coastal (WH7803), and open ocean (WH7805) *Synechococcus* under different A) salinities, B) temperatures, and C) light intensities.

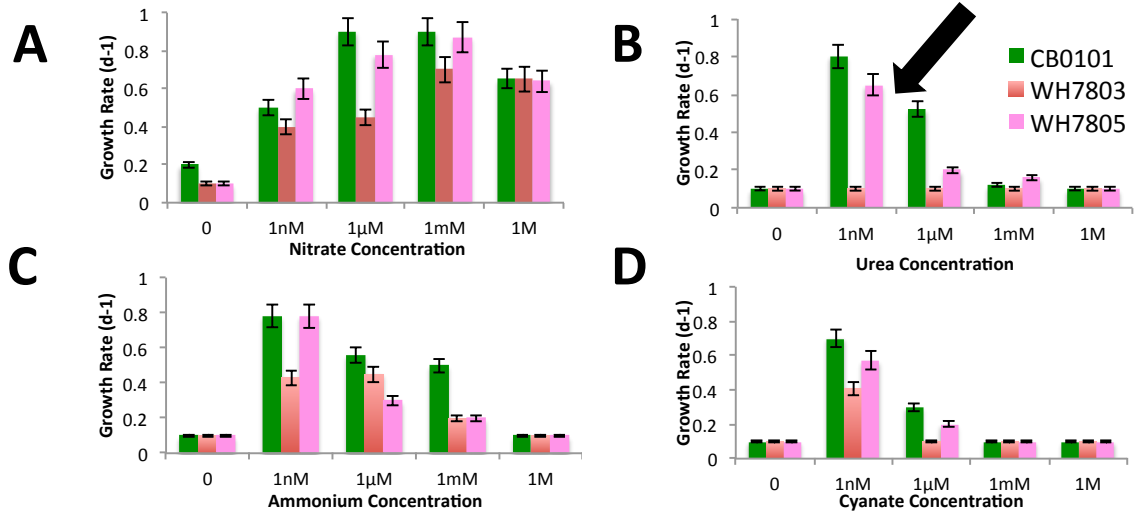


Figure 2.2. Comparison of specific growth rates between an estuarine (CB0101), coastal (WH7803), and open ocean (WH7805) *Synechococcus* under different concentrations of A) nitrate, B) urea, C) ammonium, D) and cyanate.

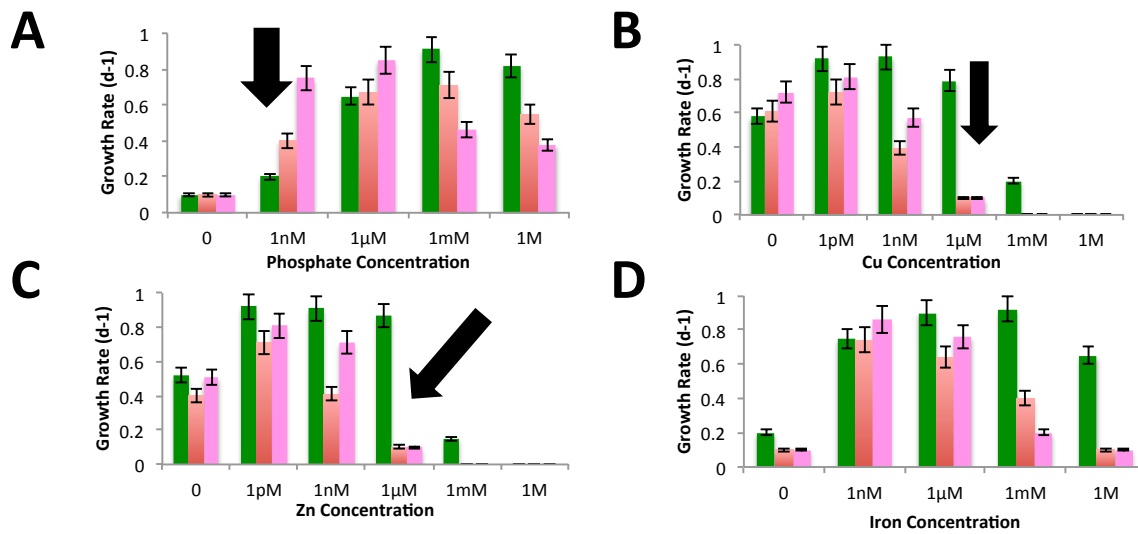


Figure 2.3. Comparison of specific growth rates between an estuarine (CB0101), coastal (WH7803), and open ocean (WH7805) *Synechococcus* under different concentrations of A) phosphate, B) zinc sulfate heptahydrate, C) copper sulfate, and D) iron.

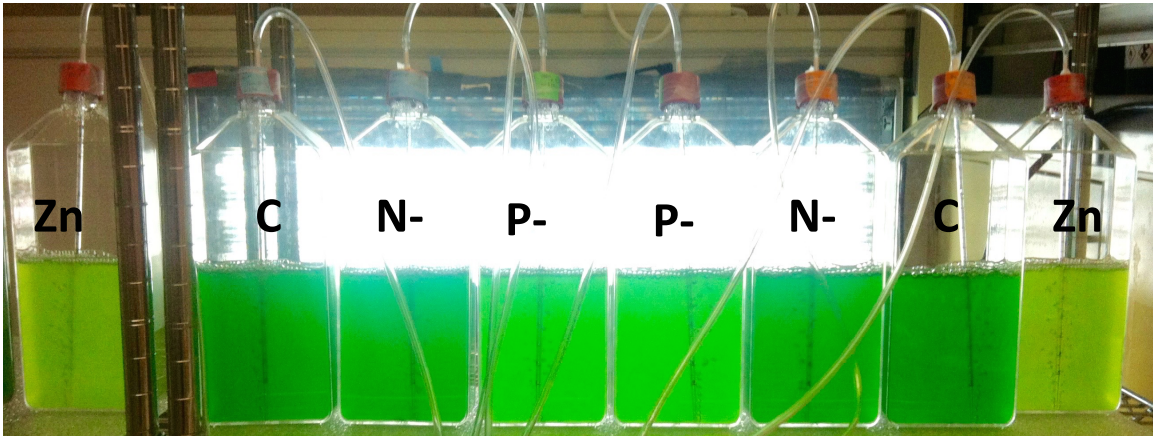


Figure 2.4. Cell culture growth apparatus for physiology experiments of *Synechococcus* CB0101. 'C' represents control, 'N-' nitrate limited, 'P-' phosphate deplete, and 'Zn' zinc toxic conditions. Cultures were grown under constant light with $15 \mu\text{E m}^{-2} \text{s}^{-1}$ continuous light at 23°C temperature and SN15 growth medium.

References:

1. Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979) Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293–294.
2. Scanlan DJ, et al. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 73(2):249–299.
3. Everroad RC, Wood AM (2012) Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* 64(3):381–392.
4. Fuller NJ, Marie D, Vaultot D, Post AF, Scanlan DJ (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *AEM* 69(5):2430–2443.
5. Palenik B, Ren Q, Tai V, Paulsen IT (2009) Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* 11(2):349–359.
6. Palenik B, et al. (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci* 103(36):13555–13559.
7. Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67(2):991–994.
8. Ferris MJ, Palenik B, Ferris (1998) Niche adaptation in ocean cyanobacteria. *Nature* 396(November):226–228.
9. Dufresne A, et al. (2008) Unraveling the genomic mosaic of a ubiquitous genus of

- marine cyanobacteria. *Genome Biol* 9(5):R90.
10. Ahlgren NA, Rocap G (2012) Diversity and distribution of marine *Synechococcus*: Multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* 3(JUN). doi:10.3389/fmicb.2012.00213.
 11. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *AEM* 72(3):2239–2243.
 12. Huang S, et al. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 6(2):285–97.
 13. Larsson J, et al. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 8(9):1892–903.
 14. Chen F, et al. (2004) Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* 36(2):153–164.
 15. Cai H, Wang K, Huang S, Jiao N, Chen F (2010) Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. *Appl Environ Microbiol* 76(9):2955–2960.
 16. Wang K, Wommack KE, Chen F (2011) Abundance and distribution of *Synechococcus spp.* and cyanophages in the Chesapeake Bay. *Appl Environ Microbiol* 77(21):7459–7468.
 17. Choi DH, Noh JH (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* 69(3):439–448.

18. Billis K, Billini M, Tripp HJ, Kyrpides NC, Mavromatis K (2014) Comparative transcriptomics between *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 provide insights into mechanisms of stress acclimation. *PLoS One* 9(10):1–10.
19. Muhling M, et al. (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using rpoC1-based restriction fragment length polymorphism. *Aquat Microb Ecol* 45(3):263–275.
20. Tetu SG, et al. (2009) Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* 3(7):835–849.
21. Zwirgmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
22. Sohm J a, et al. (2015) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* 10(2):1–13.
23. Wang K (2007) Biology and ecology of *Synechococcus* and their viruses in the Chesapeake Bay. doi:10.1017/CBO9781107415324.004.
24. Murphy RR, Kemp WM, Ball WP (2011) Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts* 34(6):1293–1309.
25. Sommer SE, Pyzik AJ (1974) Geochemistry of middle Chesapeake Bay sediment from upper Cretaceous to present. *Chesap Sci* 15(1):39–44.
26. Sinex SA, Helz GR (1981) Regional geochemistry of trace elements in Chesapeake Bay sediments. *Environ Geol* 3(6):315–323.

27. Buchanan C, Lacouture R V., Marshall HG, Olson M, Johnson JM (2005) Phytoplankton reference communities for Chesapeake Bay and its tidal tributaries. *Estuaries* 28(1):138–159.
28. Fisher TR, Peele ER, Ammerman JW, Harding LW (1992) Nutrient limitation of phytoplankton in Chesapeake Bay. *Mar Ecol Prog Ser* 82:51–63.
29. Fisher TR, et al. (1999) Spatial and temporal variation of resource limitation in Chesapeake Bay. *Mar Biol* 133(4):763–778.
30. Waterbury JB, Watson SW, Valois FW, Franks DG (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Photosynthetic Picoplankton. Can. Bull. Fish. Aquatic Sci.* (Department of Fisheries and Oceans, Ottawa, Canada), pp 71–120.
31. Collier JL, Brahamsha B, Palenik B (1999) The marine cyanobacterium *Synechococcus* sp. WH7805 requires urease (urea amiohydrolase, EC 3.5. 1.5) to utilize urea as a nitrogen source: molecular-genetic and biochemical analysis of the enzyme. *Microbiology* 145(1999):447–459.
32. Lindell D, Padan E, Post AF (1998) Regulation of ntcA expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J Bacteriol* 180(7):1878–1886.
33. Marsan D, Wommack KE, Ravel J, Chen F (2014) Draft genome sequence of *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay estuary. *Genome Announc* 2(1):e01111–13.

Chapter 3: Understanding the adaptive nature of estuarine *Synechococcus*
using a multi-omics approach

Understanding the adaptive nature of estuarine *Synechococcus* using a multi-omics approach

David Marsan¹, Zhangxian Xie², Lianzhong Luo³, Dazhi Wang² and Feng Chen^{1,*}

¹Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD, USA

²State Key Laboratory of Marine Environmental Science/College of the Environment and Ecology, Xiamen University, Xiamen, 361005, China

³Xiamen Medical College, Xiamen, 361008, China

Running Title: Estuarine *Synechococcus* highly adaptive

Keywords: *Synechococcus*, multi-omics, stress response, estuary, nitrogen deplete, phosphate deplete, zinc toxicity, RNA-Seq, proteomics

3.1 Abstract

Estuaries are among the most productive and dynamic marine environments. Estuarine *Synechococcus* are abundant and key drivers of carbon and nitrogen cycling, however little is known on how they thrive and adapt to such a dynamic system. The majority of estuarine *Synechococcus* belong to subcluster 5.2, which is poorly studied in terms of molecular mechanisms underlying their niche adaptation. *Synechococcus* CB0101, a member of subcluster 5.2, is commonly present in the Chesapeake estuary. In this study, we applied multi-omics (genomics, transcriptomics, and proteomics) analyses to investigate how CB0101 responds to selected environmental stressors. Our data shows that CB0101 seamlessly adjusts its metabolic pathways and transport systems when faced with nitrate and phosphate limitation. Oxidative stress caused by zinc toxicity resulted in down regulation of photosystems and translational machinery. CB0101 contains a suite of genes involved in toxin-antitoxin regulation (stress response), zinc excess sensors, phosphate starvation, polyphosphate utilization, and five different nitrogen metabolism pathways combined with multiple copies of transporter systems, which enables it to respond quickly to the dynamic estuarine environments.

3.2 Introduction

Cyanobacteria are an ancient and diverse group of prokaryotes found in many different habitats, which reflects their ability to adapt to variable or extreme environments (1). In the marine environment, unicellular cyanobacteria *Prochlorococcus* and *Synechococcus* are the two major prokaryotic primary producers. Compared to their sister genus *Prochlorococcus*, *Synechococcus* are more ancient and diverse in varying geographic environments ranging from estuarine, coastal and oceanic waters (2–8). *Synechococcus* can be temporally or regionally important to carbon fixation (9) and their larger cell size allows them to fix an order of magnitude more carbon than *Prochlorococcus* cells (10). Currently, marine *Synechococcus* are divided into three subclusters, 5.1, 5.2, and 5.3 (4, 6, 11–13) which include at least 20 clades (11, 14, 15). The vast majority of coastal and oceanic *Synechococcus* belong to subcluster 5.1. The adaptive potential of subcluster 5.1 *Synechococcus* to changing environments such as wavelengths of light (11, 13, 16), nitrate (NO₃⁻) concentrations (11, 17, 18), and organic phosphorus sources (15, 19) is well defined.

Molecular functional analysis of subcluster 5.1 *Synechococcus* now allows for both *in situ* and *in vivo* studies providing valuable insights into their biogeochemical functions and factors controlling their distributions (8, 17, 20–22). A number of subcluster 5.1 *Synechococcus* genomes have provided insights into environmental adaptations (22–25). Specific gene differences have explained *in vivo* variations in pigment composition and nitrogen utilization (13, 24–26). In addition, strains have been shown to contain different complements of iron sensing and acquisition genes (24, 26). Transcriptomic studies of subcluster 5.1 *Synechococcus* have allowed researchers to

catalog responses to environmentally relevant stresses, confirm putative non-coding transcripts, and identify adaptation responses (22, 27–29).

Synechococcus strain CB0101, a well-referenced strain of subcluster 5.2, represents the most common and abundant *Synechococcus* in the upper Chesapeake Bay (24, 30). As the largest estuarine ecosystem in the U.S., Chesapeake Bay is well known for its strong environmental gradients and seasonality. Nutrients in the Chesapeake Bay vary greatly over time and space (31). In the winter-spring time, concentrations of nitrate, ammonium and phosphate are usually high due to high precipitation and river runoff, while these nutrients can be used up quickly in summer-fall seasons when phytoplankton biomass increases dramatically (31–34). The concentrations of nitrogen and phosphorus in the Chesapeake water can be low enough to seasonally limit excess algal and cyanobacterial growth (31, 35, 36). Considering the dynamic nature of estuarine ecosystems, *Synechococcus* species living in the Chesapeake Bay can experience rapid environmental changes over a short period of time or within a small spatial distance.

CB0101 is often used as a phylogenetic reference for subcluster 5.2 (3, 5, 37, 38), therefore, CB0101 is a good model strain for studying the ecology and molecular functional capability of estuarine or subcluster 5.2 *Synechococcus*. In this study, we applied a multi-omics approach (genomics, transcriptomics, and proteomics) to explore stress responses of the estuarine *Synechococcus* strain CB0101 under nitrate depletion, phosphate depletion, and zinc toxicity. The goal is to understand how *Synechococcus* adapt to the estuarine environment by studying their stress responses at the molecular level. Our study provides the first comprehensive analysis of molecular functions of

estuarine or subcluster 5.2 *Synechococcus* in response to different environmental conditions.

3.3 Results

3.3.1 Genome sequencing of Synechococcus CB0101

The complete genome sequence of CB0101 was determined using a hybrid approach combining 454 GS-FLX Titanium 8-kb paired-end and PacBio libraries. The genome sequence of CB0101 consists of 2,789,627 bp (63.4 % G+C content) with 3,176 genes and 46 tRNAs (Fig. 3.1A). Roughly 36 % (1,132 genes) of the genome can be assigned to a known functional cluster of orthologous groups (COG) while 64 % (1331 hypothetical and 710 known genes) were unable to be grouped. The assigned genes were filtered into 27 subsystem features. Those genes associated with an increased capacity to sense and respond to changes in the environment were 35 membrane transporters, 35 regulation and cell signaling, 58 stress response, 33 virulence, disease and defense and 3 dormancy.

Further, a genome comparison between CB0101 and other strains from the coastal and open ocean environments was conducted (Fig. 3.1B). Stress response related COGs were compared and found that the three strains share 36 genes while 6 were specific to CB0101. These six genes consisted of various oxidative stress sensing genes. Information gleaned from the complete genome showed an ability for CB0101 to sense and respond to environmental conditions such as nitrogen or phosphate depletion and zinc metal excess. Merely, identifying genes and pathways does little to understand the functional response of an organism. Thus, a genome wide profiling of the transcriptome and proteome expression under all three environmental stressors was conducted.

3.3.2 Transcriptomic responses of CB0101 to nutrient depletion and zinc toxicity

To understand how CB0101 adapts and responds to nutrient and metal stress at the molecular level three culture experiments were designed. CB0101 was exposed to media with limited nutrients (nitrate or phosphate) or high concentrations of zinc 50 μM . We selected these conditions for two reasons 1) to obtain a strong response at the mRNA and protein level in order to clearly observe expression dynamics; 2) these conditions correspond to those found seasonally in the Chesapeake Bay, thus meaning CB0101 must be adapted to the potential stressors. The growth of CB0101 under each condition is shown in Fig. 3.S1. Without nitrate or phosphate, CB0101 grew much slower compared to the control (Fig. 3.S1 where is this?). The growth inhibition of CB0101 with a lack of phosphate was more severe than that with a lack of nitrate. CB0101 growth was almost fully inhibited with 50 μM zinc in the culture medium. This experiment shows distinct growth responses of CB0101 to nutrient starvation (or stress) or metal toxicity can be seen after 2-3 days.

Transcriptional (mRNA) response of CB0101 exposed to three stress conditions for 72 h was shown in Fig. 3.2A. Transcriptional responses with nitrogen and phosphate stress displayed a similar global response of gene regulation as depicted in the heat map (Fig. 3.2A). The gene expression pattern for zinc toxicity was distinct from those of nitrogen and phosphorus limitation (Fig. 3.2A). Based on a 2-fold change of mRNA, 843 (27 %), 368 (12 %), and 1,195 (38 %) differentially expressed genes were identified from CB0101 (3,173 total genes) ($p < 0.01$ and minimum 2-fold) for nitrogen, phosphate, and zinc treatments, respectively (Table 1). Based on a 4-fold change of mRNA, 108, 15, and 370 differentially expressed genes ($p < 0.01$ and minimum 4 fold) were observed for

nitrogen, phosphate, and zinc treatments, respectively (Table 1). Transcript wide response of CB0101 treated with zinc greatly exceeded that of nitrogen or phosphorus depletion (Table 1).

When filtered into cluster of orthologous groups (COGs), an interesting response for each individual stressor was observed (Fig. 3.2B). Response to nitrogen depletion genes the urea cycle highly expressed. This is represented by the up-regulated COG categories of: amino acids (urea metabolism pathways), respiration and electron transport system (urea metabolism i.e. cytochrome/ferredoxin), co-factors (ATP and energy pathways), carbohydrates (TCA cycle and glycogen), nitrogen metabolism (ammonium metabolism), and stress response (protection from oxidative stress caused by using urea). In contrast, down-regulated genes involved nitrate metabolism including transporters, RNA metabolism and folate biosynthesis (gene transcript and protein fold differences for CB0101 under every experiment described in this dissertation is included in the Appendix). The major upregulated gene and protein expression included UreaABC, UreA, GlnA, and GltB.

Phosphate depletion resulted in less differentially expressed transcripts compared with the other treatments (Table 1). Of the differentially expressed genes in the phosphate treatment, the up-regulated responses consisted of pathways to extract phosphate from other molecules and sources (Fig. 3.2B). The highest represented COGs included respiration (cytochrome and electron carriers), phosphorous metabolism (ABC phosphate transporters, polyphosphate kinase, alkaline phosphatase (cleaves phosphate from other molecules i.e. nucleotides, proteins, and alkaloids), carbohydrate metabolism (pentose phosphate pathway), and nitrogen metabolism (ammonia and nitrite transporters, and

cyanate hydratase) (Fig. 3.2B). The major gene pathways included P-specific ABC transporters, PhoR, and PPX. Conversely, the down-regulated genes were dominated by RNA metabolism and protein metabolism.

Lastly, under zinc toxicity, mRNA expression at the 2-fold level displayed over 1,195 differentially expressed genes, among which 525 genes were up-regulated and another 670 were down-regulated (Table 1). Under a high concentration of zinc, the genes involved in photosynthesis, protein metabolism, carbohydrate metabolism, and respiration were down-regulated significantly (Fig. 3.2B). On the other hand, the genes responsible for cofactors/vitamin/prosthetic metabolisms, DNA, RNA, and protein metabolisms were up-regulated (Fig. 3.2B). The major upregulated genes and proteins included RND, ATPase, ZiaA, ZIP, ZnuABC, ZiaR, ZUR, COG0523 and SmtB.

3.3.3 Proteomic responses of CB0101 to nutrient depletion and zinc toxicity

At the protein level 522 total unique proteins with peptides that matched were identified over all treatments. A total of 103 and 63 differentially expressed proteins ($p < 0.01$ and minimum 2-fold) were identified from CB0101 in nitrogen and phosphate treatments, respectively (Table 1). We were unable to obtain protein data from CB0101 in the zinc treatment due to potential translational inhibition caused by toxin-antitoxin. In general, proteomics detected fewer differentially expressed genes compared to transcriptomics.

The protein expression patterns of CB0101 under nitrate and phosphate depletion were displayed in the heat map (Fig. 3.3). The proteomic responses for both treatments were similar in the upper cluster, but distinct responses were seen in the lower clusters (Fig. 3.3A). CB0101 under phosphate depletion showed more protein down-regulation

compared to nitrogen depletion. These proteins participated in pathways for the metabolism of carbohydrates and amino acids, potentially to source phosphate. Cellular processes also slowed, represented by down-regulation of photosynthesis, cell division, and protein production (Fig. 3.3B). The number of down-regulated proteins was comparable for both treatments except that the respiration of CB0101 under phosphate depletion was more down regulated than that under nitrate depletion (Fig. 3.3B). However, only a fraction of all differentially expressed genes were detected by the proteomic analysis. This markedly different response between the transcriptomic and proteomic analyses could be caused by a delay in translation of protein, along with posttranslational regulation occurring. Another factor, and likely the main cause could be that the protein isoforms may be too lowly abundant to be detected, unlike the greater sequencing depth of the RNA-Seq work.

3.4 Discussion

3.4.1 Stress handling capability unveiled by comparative genomics

The genomic sequence of CB0101 represents the first complete genome for subcluster 5.2 *Synechococcus*. Based on the specific genes and gene clusters carried by CB0101, we hypothesized that CB0101 had an increased genetic capacity to sense and respond to changes in the estuarine environment. To test that hypothesis, we compared CB0101's genome content to that of two other well-understood coastal and oceanic *Synechococcus* strains (WH7803 and WH7805). It is not the intention of this study to conduct a comprehensive comparative genomics between all known *Synechococcus* genomes (see (24, 25) for comparative genomics of marine *Synechococcus*). These previous studies identified specific genome structures of *Synechococcus*, which included

core and variable gene content. Our work focuses on identifying ecologically significant genes in estuarine *Synechococcus* to understand how they respond to environmental stressors.

Insights from the comparative genomics (CB0101, WH7803 and WH7805) reveal nutrient adaptability and a metal-intensive ecological strategy, as indicated by the many more genes involved in metal homeostasis as well as metal enzymes, sensor kinases and various metabolic pathways for nitrogen, phosphate, and amino acid metabolism that were not present in oligotrophic strains WH7803 and WH7805. Unique genes within CB0101 (as compared to WH7803 and WH7805) were predominately found within the functional categories of amino acids and derivatives, stress response, membrane transport, virulence/defense, and regulation/cell signaling. A number of heavy metal ABC transporters (Ni, Co, Cu, Zn, Fe) and high affinity branched chain amino acid transport systems are present. CB0101 contains unique *phoH*, alkaline phosphatase, and an additional ammonium transporter not seen in the other strains. The *phoH* gene is from the phosphate starvation inducible family of proteins that responds under phosphate limited conditions and is thought to have a function in regulating the cellular response to phosphate limitation (21, 26). Alkaline phosphatase is used to obtain phosphate from other organic sources (23). The additional ammonium transporter (*amt*) likely allows for greater scavenging uptake of ammonium (39). These genes were up regulated in CB0101 and might play a key role during nitrate and phosphate deplete conditions (Fig. 3.2). Interestingly, CB0101 contains 15 toxin-antitoxin (TA) genes, was found to be associated with stress response; specifically the *relE/relB* pair regulates oxidative stress by arresting

translation (Chapter 4). This is likely one of many mechanisms that CB0101 has developed to adapt to the estuarine environment.

CB0101, WH7803, and WH7805 differ in aspects of their membrane transporter complement that may reflect alterations in the nutrients they are exposed to in different ecosystems. Interestingly, CB0101 has a unique set of 7 ABC transporters for branched chain-amino acid transport, arguably for metabolism of organic nutrients, but determining the actual activity will require future nutrient uptake studies. Unique to CB0101 metal enzymes, cofactors, and metal transporters suggests a greater capacity to store, move, and utilize metals. This is consistent with higher metal quotas for copper and zinc. Interestingly, it has been shown that *Synechococcus* WH8016, which contains similar transporters was more resistant to copper than oligotrophic strains(40).

Fifty-eight stress response genes were identified in CB0101, with COG systems of heat shock, oxidative stress, and detoxification being most numerous. Of these 58 genes, 8 were unique to CB0101 when compared to WH7803 and WH7805. Thirty-four genes are involved in virulence defense including resistance to antibiotics and toxic compounds, and 6 of them are related to Co-Zn-Cd resistance. Within stress response type genes CB0101 contains a unique set of 5 oxidative stress genes, but lacks 5 other oxidative stress genes contained in WH7803 and WH7805 (Table S2). Interestingly, CB0101 has manganese and iron superoxide dismutases (SOD), while WH7803 and WH7805 contain copper and zinc SOD. Cu and Zn-SOD have been found in many marine *Synechococcus*, but many freshwater and estuarine cyanobacteria have only Fe or Mn SOD. Marine *Prochlorococcus* possess only Ni-SOD (41). CB0101 does not contain 6 osmotic stress genes found in WH7803 and WH7805, suggesting that estuarine

Synechococcus may not suffer similar osmotic stress like their coastal and oceanic counterparts.

3.4.2 Responses of CB0101 to stresses at the omics level

CB0101 was inhibited the most with zinc toxicity, moderately with phosphate depletion, and the least with nitrogen depletion (Fig. 3.S1). The genetic capacities of CB0101 led us to hypothesize a greater ability to cope with metal stresses and nitrogen and phosphate. We now wanted to understand how CB0101 regulates its transcript and protein expression in response to these stresses.

When cells were grown with a depleted nitrogen-source, CB0101 adjusts metabolism pathways to the available N-source. Transcript and protein levels for genes associated with urea metabolism, urea transporters, cytochrome/ferredoxin, TCA cycle, glycogen, ammonium metabolism, and stress response protection from oxidative stress caused by metabolizing urea were all highly expressed. These pathways worked to transport urea in order to metabolize it to ammonium, then to Gln and Glu (Fig. 3.4). Using urea has the ill effect of oxidative stress, which causes stress response pathways to become active. At the same time nitrate metabolism pathways including transporters, RNA metabolism and folate biosynthesis were downregulated. These responses were in line with previous work showing that most cyanobacteria can use nitrate, nitrite, and ammonia as primary N-sources, although urea and organic N-compounds can also be used in some cases (42). CB0101 is capable of using all of those N sources, and when nitrate is removed from the medium it adjusts accordingly to scavenge urea (Fig. 3.4).

When starved of phosphorus, CB0101 cells significantly increased phosphate scavenging with a 3-fold increase of numerous ABC phosphate transport. Furthermore,

periplasmic alkaline phosphatases, for example *phoA*, release phosphate from various compounds, making it available for the transport systems. These enzymes may provide the ability to scavenge phosphate from a large variety of substrates, as suggested by its in vitro activity. Sensor kinase and response regulators comprising a two-component signal transduction pathway governs the response to phosphorus limitation. Transcripts of phosphate regulon (Pho regulon), three clusters of genes exhibited substantial induction upon phosphate starvation. One such cluster contains a periplasmic protein, *urtA*, which was highly repressed under phosphate-limited conditions (Fig. 3.4).

Strong transcriptional responses were evident when CB0101 was exposed to zinc. CB0101 contains numerous zinc related genes to cope with zinc toxicity or homeostasis, which are absent in WH7803 and WH7805. Some of these genes, which are upregulated during toxic conditions are *ZIP* (zinc transporter family protein), *ZnuB* (ABC transporter inner membrane), *ZnuC* (ABC transporter ATP-bind protein), *ZnuA* (two copies, ABC transporter periplasmic-binding protein), COG0523 (zinc homeostasis GTPase family), *ZUR* (zinc uptake regulation protein), zinc finger SWIM domain, ATP-dependent zinc metalloprotease, and *czc* system (mediates low-level metal ion resistance). These genes in combination are thought to aid cyanobacteria in coping with metal induced stress and were all upregulated at the mRNA and protein level (Fig. 3.4).

In order to deal with excess zinc, *Synechocystis* sp. PCC6802 has a zinc specific efflux pump, *ZiaA* (43), which is similar to other P1-type ATPase metal ion transporters responsible for transporting Zn^{2+} from the cytoplasm to the periplasmic space. Expression of *ZiaA* is induced by zinc and is regulated by *ZiaR* (43), a zinc-specific repressor protein. CB0101 contains similar zinc efflux pumps, which are upregulated during zinc

toxicity (Fig. 3.4). Genes for zinc excess sensors, *smtB* and *ziaA*, found in CB0101 are absent in WH7803 and WH7805, both of which are upregulated. The *smt* locus of CB0101 contains a metal-regulated gene (*smtA*), which encodes a class II metallothionein, and a divergently transcribed gene, *smtB*, which encodes a repressor of *smtA* transcription (44). Together with at least one ZnuABC system and a putative ZIP transporter for uptake, as well as at least one zinc-related CO0523 protein, this strain utilizes the suite of genes to cope well with both zinc excess and scarcity. It has been suggested that metallothioneins in *Synechococcus* may be under the control of Zur, which is in line with the response of CB0101 in our experiments. It suggests that, metallothioneins are not just devices to combat metal toxicity, but may play a more central role in essential zinc homeostasis (Fig. 3.4). By integrating multi-omics molecular techniques to predict functions (i.e. response to nutrient depletion or metal toxicity) with ecological data derived from ‘model strain’ isolates we now have a better understanding of the distinct responses of CB0101 to the Chesapeake Bay environment. We believe similar mechanisms are at play with other estuarine subcluster 5.2 *Synechococcus* strains.

3.5 Conclusion

The adaptive nature of estuarine *Synechococcus* is reflected by additional stress responding genes in CB0101 and the expression of these genes and proteins. The omics studies highlighted the flexibility of different nitrogen metabolism sources, with CB0101 seamlessly switching from nitrate to urea. While under phosphate limiting conditions upregulation of additional transporters and breakdown of polyphosphate were confirmed at the mRNA and protein level. Finally, zinc toxicity caused almost total shutdown of mRNA and protein expression, however genes associated with zinc sensing and

transporters were active. The combination of multi-omics analyses provides a deeper understanding on the niche adaptation of *Synechococcus* in the estuarine environment. Omics studies based on other environmental settings or other members of subcluster 5.2 *Synechococcus* can be developed in the future to explore broader environmental responses of estuarine picocyanobacteria.

3.6 Experimental Procedures

3.6.1 Culture conditions for growth of CB0101

For the multi-omics experiments, CB0101 was cultivated in SN medium with 15 ppt salinity and vitamin B12 (10 µg/L) (referred to as SN15 medium hereafter), and duplicate cultures were subjected to 4 different conditions for 72 h: (1) Control, where the culture was grown in SN15 medium; (2) Nitrogen-deplete condition, where the culture was cultivated in the SN15 medium that did not contain sodium nitrate; (3) Phosphate-deplete condition, where the culture was grown in the SN15 medium that lacked dipotassium phosphate; (4) Zinc toxic condition, where the culture was grown in the SN15 medium amended with 50 µM of zinc sulfate heptahydrate. Triplicate samples for every treatment were taken at each time point. The growth of cultures was monitored by counting cells using an Accuri C6 flow cytometer. Samples for mRNA and protein analysis were immediately extracted at 0 h and after 72 h. The specific growth rate was calculated from the logarithmic change $\mu = \frac{\ln X_2 - \ln X_1}{t_2 - t_1}$ where X1 and X2 are densities at times t1 and t2.

3.6.2 Genome sequencing of CB0101

The partial genome of CB0101 was previously described and sequenced using 454 (45). To span difficult repeat regions, we obtained long Pacific Biosciences sequences. Cells were collected by centrifugation (10,000 x g, 10 min), the pellet transferred into a 2 ml tube and immediately processed. We obtained DNA from 25 ml cultures using the Epicentre Masterpure kit (Epicentre) and sequenced the genome at the Institute of Genome Sciences (IGS). Total reads resulted in 121,716 with an average read

length of 6,295 bp, resulting in ~40X coverage. Low quality regions of sequencing data were removed from the raw PacBio data using the HGAP assembly program. Contigs were ordered into putative scaffolds based on their similarity to closely related closed *Synechococcus* genomes, as determined by Mauve (46). As the cultures sequenced were known to contain heterotrophs, we identified the most '*Synechococcus*-like' contigs from non-axenic cultures by searching each resulting contig against a custom database of sequenced marine microbial genomes using BLAST (47). Contigs with a best match to a non-*Synechococcus* genome were removed from the assembly.

3.6.3 Genome annotation and comparisons

The assembled complete genome for CB0101 was annotated using the RAST server (48) and MANATEE (49) programs. Subsequently, the annotations of each were compared and any discrepancies were manually explored and assigned. Clusters of orthologous groups of proteins (COGs) were computed, as described in each program. To facilitate comparisons among other picocyanobacteria genomes, we re-annotated *Synechococcus* WH7803 and WH7805 with the RAST and MANATEE pipelines as described above; this ensured that a uniform methodology for gene calling and functional annotation was used. For genome analysis, a combination of three gene modeling programs- CLC, Manatee, and RAST were used in the determination of potential coding sequences. These assignments were further checked manually. A revised gene/protein set was searched against the KEGG genes, Pfam, PROSITE, PRINTS, ProDom and COGs databases, in addition to BLASTP against the NCBI non-redundant database. From these results, categorizations were developed using the KEGG and COGs hierarchies. Transfer

RNAs were identified using CLC. To identify regions of atypical nucleotide composition, the trinucleotide composition was determined.

3.6.4 Extraction of mRNA

At the designated time points (0 and 72 h post media change) samples were collected and immediately processed for mRNA and protein analysis. RNA and protein were separated using Trizol extraction. mRNA was extracted from the total RNA using Ambion MICROBExpress kit (Life technologies, AM1905), modified with 4 specific oligo primers for rRNA depletion of CB0101 ribosome. rRNA removal evaluation was conducted using the Agilent 2100 Bioanalyzer resulting in 98 % removal of rRNA and confirmed the presence of high quality mRNA. The NEXTflex RNA-Seq kit (BIOO Scientific, #5129-01) was then used to prepare mRNA libraries for sequencing using the Illumina Hi-Seq. RNA libraries were created and barcodes assigned to each sample for multiplexing of a single Hi-Seq lane. Using a Hi-Seq with 100 bp read length, RNA-Seq sequencing was carried out by the IGS and resulted in 144,403,530 read sequences. RNA-Seq analyses were carried out using the CLC platform R, Bioconductor, top-hat/cufflinks (Table S1).

3.6.5 Protein extraction and mass spectrometry analysis

Proteins were extracted using the same sample for RNA extraction. The bottom layer containing proteins was transferred to a new tube and the extraction procedure followed the extraction procedure previous described (50). Protein quantification was carried out according to the instruction of 2D Quant kit (GE Healthcare, 80-6482-56). Equivalent of proteins (200 µg) in each sample were loaded on a prepared SDS-PAGE (manufactory) for in gel digestion (51). The tryptic peptides were extracted with solvent

containing 50% acetonitrile (ACN) and 0.1% trifluoroacetic acid, and then vacuum dried using a SpeedVac for further mass spectrometry (MS) analysis.

Samples were introduced into a Waters nanoACQUITY UPLC using a C18 trap column (1.7 μm particle size, 100 μm x 100 mm internal diameter) fitted with a C18 trap column. (5 μm particle size, 180 μm x 20 mm internal diameter). Peptides were eluted at a flow rate of 300 nl/min using an acidified (formic acid, 0.1 % v/v) water-acetonitrile gradient (2 %-35 % acetonitrile in 135 min, full-time 185 min). The peptides were subjected to nanoelectrospray ionization followed by tandem mass spectrometry (MS/MS) in a Q EXACTIVE (ThermoFisher Scientific, San Jose, CA) coupled online to the UPLC. Data-dependent scans were conducted by precursor ion selection in the Orbitrap followed by high-energy collision induced dissociation. One full m/z scan between 350 and 2,000 Da was followed by 15 MS/MS scans with a following Dynamic Exclusion duration of 15 s. The electrospray voltage applied was 1.6 kV and the heated capillary was at 280 °C. Automatic gain control (AGC) was used to optimize the spectra generated by the Orbitrap.

3.6.6 Protein identification

The instrument data files (.raw) were interpreted using Proteome Discoverer (ver. 1.4.0.288; ThermoFisher Scientific, San Jose, CA). MS/MS spectra were searched with SEQUEST engine against a protein database containing 15,527 translated protein-coding sequences predicted from genome of *Synechococcus* CB0101 and its phage genome as well as 6,774 contaminated bacterial sequences and 247 contaminant sequences such as keratin and trypsin. Database searching was restricted to tryptic peptides. For protein identification, a mass tolerance of 10 ppm was permitted for intact peptide masses and

0.05 Da for fragmented ions, with allowance for one missed cleavages in the trypsin digests and carbamidomethyl (C) as the fixed modification, oxidation (M) as the variable modification. Peptides having a Delta Cn better or equal to 0.05 were used for Percolator. Peptide spectral matches were further validated based on q-value of decoy database search at a level of 1% false discovery rate (FDR). Peptide identifications were grouped into proteins with Proteome Discoverer according to the law of parsimony and filtered to 1% FDR. Finally, proteins matched with at least two peptides were accepted as confident identifications for further bioinformatics analysis.

3.6.7 Functional annotation and semi-quantitative analysis

Functional annotations of proteins identified were derived from their annotations of RNA or DNA sequences. For semi-quantitative analysis(52), spectral counts were normalized by the total number of spectra identified in each sample as well as by the length of the matching protein. The resulting values were used for statistical significance test and represented as protein expression level.

Table 3.1. Differentially expressed (A) 2 fold $p < 0.01$ and (B) 4 fold $p < 0.01$ transcripts and proteins for three stress conditions, nitrogen and phosphate limited and zinc toxicity of the estuarine *Synechococcus* CB0101 (N/A= no data available). Differential expression was determined as each condition compared to the control expression. Percentages are based on the total number of coding genes found in *Synechococcus* CB0101 (3,173).

A

2 Fold change ($p < 0.01$)	Transcriptomics			Proteomics		
	Nitrogen	Phosphate	Zinc	Nitrogen	Phosphate	Zinc
Total responsive genes/proteins	843 (27%)	368 (12%)	1195 (38%)	103 (3%)	63 (2%)	N/A
Total up-regulated genes/proteins	400	222	525	64	23	N/A
Total down-regulated genes/proteins	443	146	670	39	40	N/A

B

4 Fold change ($p < 0.01$)	Transcriptomics			Proteomics		
	Nitrogen	Phosphate	Zinc	Nitrogen	Phosphate	Zinc
Total responsive genes/proteins	108	15	370	19	12	N/A
Total up-regulated genes/proteins	41	10	43	16	4	N/A
Total down-regulated genes/proteins	67	5	327	3	8	N/A

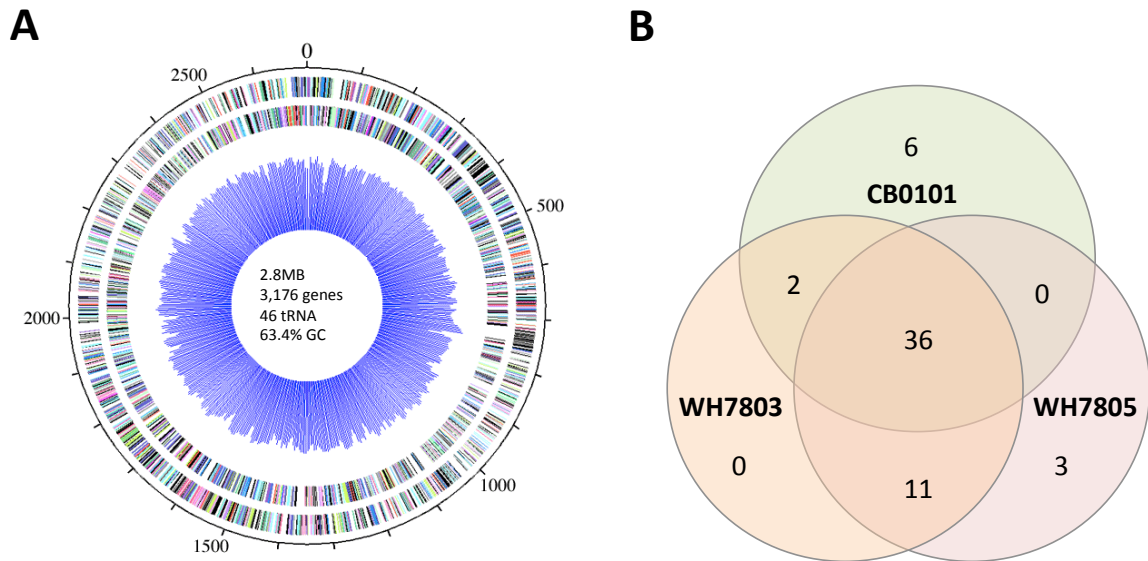


Figure 3.1. Genomic characteristics of estuarine *Synechococcus* CB0101. A) Circular representation of *Synechococcus* sp. CB0101 complete genome structure. Genome stats are: 2.8MB, 3,176 coding genes, 46tRNA, and a GC content of 63.4%. The outer scale designates coordinated in base pairs. The first circle shows predicted coding regions on the plus strand, color-coded by role categories. The second circle shows predicted coding regions on the minus strand color coded by role categories. The inner circle represents coverage of the genome, which averaged 40x. B) COG comparison between estuarine (CB0101), coastal (WH7803), and open ocean (WH7805) *Synechococcus* genomes. Gene categories related to stress response, sensing, transport, storage and export were compared. COG categories were determined using the RAST server.

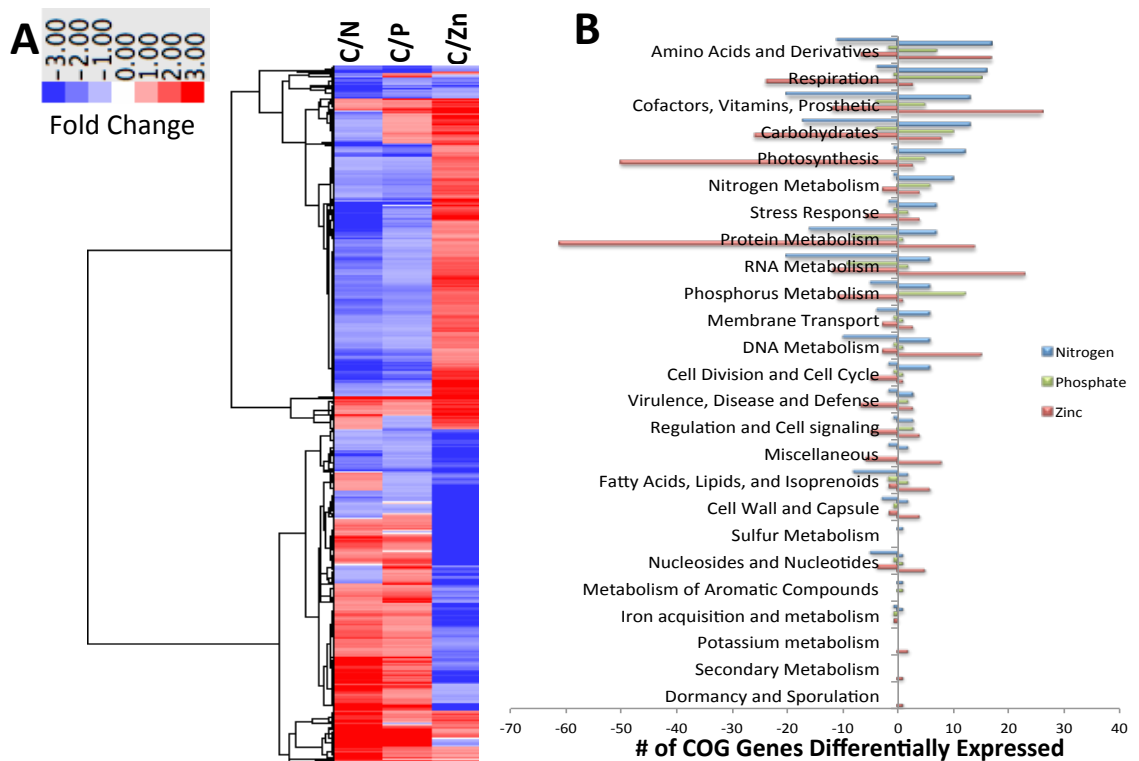


Figure 3.2. The transcript expression program to stress responses in the estuarine *Synechococcus* CB0101. A) Heat map and clustering of differentially expressed genes by their expression patterns. Stress conditions are all compared to Control, with C/N representing nitrogen depletion, C/P phosphate depletion, and C/Zn zinc toxicity. Dark red equates to a strong increase (>3 fold), while dark blue to a strong decrease (>3 fold) in gene expression. (B) Differentially expressed genes filtered into COGs for each of the three stress conditions. The left hand side denotes the number of downregulated genes differentially expressed while the right is the number of upregulated genes.

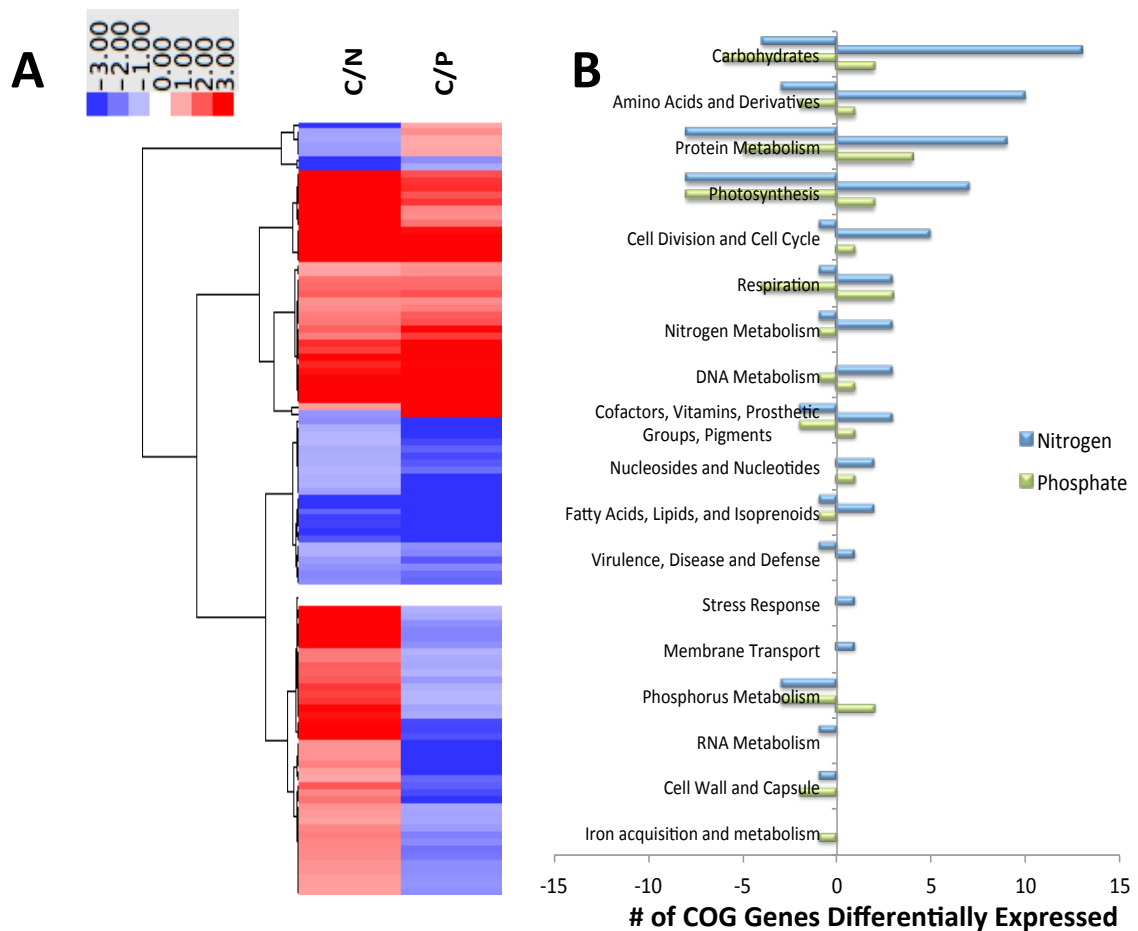


Figure 3.3. The protein expression program to stress responses in the estuarine *Synechococcus* CB0101. A) Heat map and clustering of differentially expressed proteins by their expression patterns. Stress conditions are all compared to Control, with C/N representing nitrogen depletion, C/P phosphate depletion. Protein was unable to be analyzed from the Zinc samples. Dark red equates to a strong increase (>3 fold), while dark blue to a strong decrease (>3 fold) in gene expression. (B) Differentially expressed genes filtered into COGs for each of the stress conditions. The left hand side denotes the number of downregulated genes differentially expressed while the right is the number of upregulated genes.

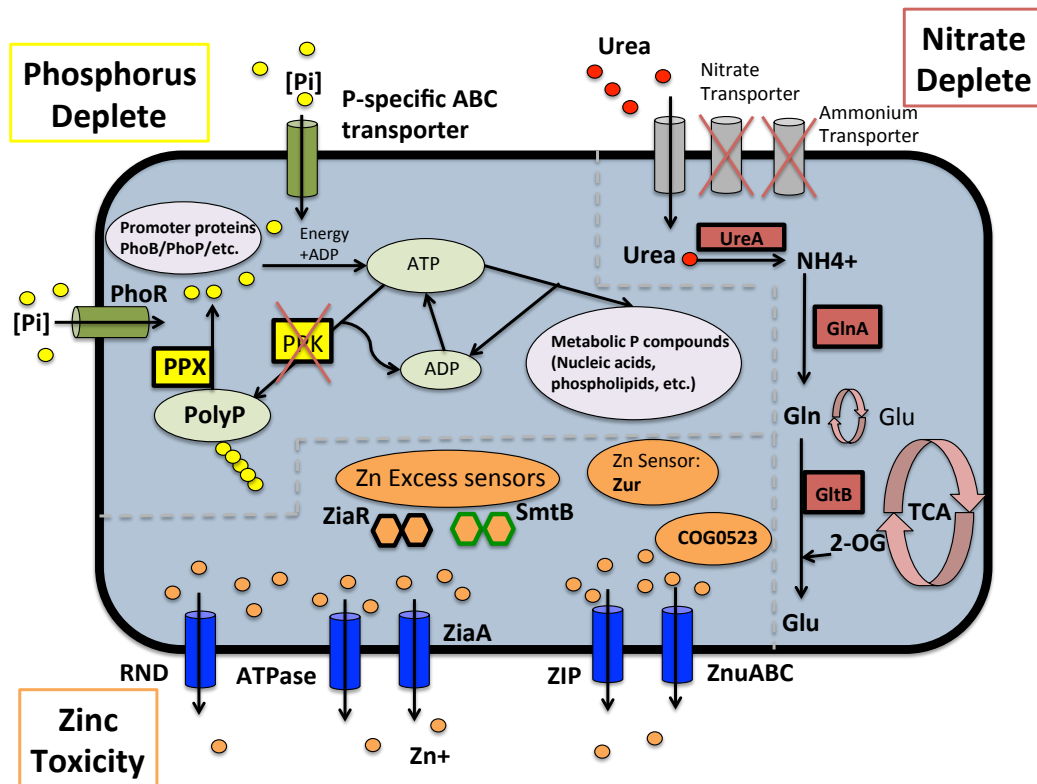


Figure 3.4. Molecular mechanisms of nitrogen or phosphate deplete and zinc toxic conditions of the estuarine *Synechococcus* CB0101. This diagram represents a composite of upregulated mRNA and protein (shown in bold text) and down regulated (red X) pathways. Black arrows indicate pathways by which each condition transported or metabolized the nutrient. Cylinders indicate transporter protein complexes, circles – specific nutrient pools, squares- protein or enzymes seen being differentially expressed, oval- promoters or sensors differentially expressed. Nitrogen depletion conditions causes nitrate and ammonium transporters to be down regulated which ureaABC transporters were upregulated. CB0101 transports urea into the cell, which is then metabolized to ammonium, by *ureA*. The pathway continues with $GlnA + NH_4^+$ producing Gln. The TCA cycle is upregulated and 2-OG along with Gln is used to produce Glu, all of which is

aided by GltB. Under phosphate deplete conditions, upregulation of P-specific ABC transporters and PhoR scavenge for any available phosphate. PPX is highly expressed, resulting in the metabolism of PolyP to create [Pi] in order to produce other metabolic P compounds (Nucleic acids, phospholipids). Zinc toxicity is dominated by efflux transporters (RND, ATPase, ZiaA, ZIP, ZnuABC), which are controlled by various zinc excess sensors (ZiaR, SmtB, Zur, and COG0523).

References:

1. Tandeau de Marsac N, Houmard J (1993) Adaptation of cyanobacteria to environmental stimuli: new steps towards molecular mechanisms. *FEMS Microbiol Lett* 104(1-2):119–189.
2. Toledo G, Palenik B (1997) *Synechococcus* diversity in the California Current as seen by RNA polymerase (*rpoCI*) gene sequences of isolated strains. *Appl Environ Microbiol* 63(11):4298–4303.
3. Rocap G, Distel DL, Waterbury JB, Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68(3):1180–1191.
4. Fuller NJ, Marie D, Vaultot D, Post AF, Scanlan DJ (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Society* 69(5):2430–2443.
5. Ahlgren NA, Rocap G (2012) Diversity and distribution of marine *Synechococcus*: Multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* 3. doi:10.3389/fmicb.2012.00213.
6. Penno S, Lindell D, Post AF (2006) Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* 8(7):1200–1211.
7. Chen F, et al. (2004) Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO)

- large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* 36(2):153–164.
8. Zwirgmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
 9. Flombaum P, et al. (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA* 110(24):9824–9.
 10. Jardillier L, Zubkov M V, Pearman J, Scanlan DJ (2010) Significant CO₂ fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* 4(9):1180–1192.
 11. Ahlgren NA, Rocap G (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* 72(11):7193–7204.
 12. Muhling M, et al. (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using *rpoC1*-based restriction fragment length polymorphism. *Aquat Microb Ecol* 45(3):263–275.
 13. Six C, et al. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus spp.*: a comparative genomics study. *Genome Biol* 8(12):259-270.
 14. Huang S, et al. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 6(2):285–97.
 15. Mazard S, Ostrowski M, Partensky F, Scanlan DJ (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* 14(2):372–386.

16. Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67(2):991–994.
17. Fuller NJ, Tarran G a., Yallop M, Orcutt KM, Scanlan DJ (2006) Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol Oceanogr* 51(6):2515–2526.
18. Moore LR, Post AF, Rocap G, Chisholm SW (2002) Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr* 47(4):989–996.
19. Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T (2005) Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* 39(3):257–269.
20. Tai V, Palenik B (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* 3(8):903–915.
21. Tetu SG, et al. (2009) Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* 3(7):835–849.
22. Cox AD, Saito MA (2013) Proteomic responses of oceanic *Synechococcus* WH8102 to phosphate and zinc scarcity and cadmium additions. *Front Microbiol* 4:1–17.
23. Palenik B, et al. (2003) The genome of a motile marine *Synechococcus*. *Nature* 424(6952):1037–1042.
24. Scanlan DJ, et al. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 73(2):249–299.
25. Dufresne A, et al. (2008) Unraveling the genomic mosaic of a ubiquitous genus of

- marine cyanobacteria. *Genome Biol* 9(5):R90.
26. Palenik B, et al. (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci* 103(36):13555–13559.
 27. Vijayan V, Jain IH, O’Shea EK (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biol* 12(5):R47.
 28. Ludwig M, Bryant DA (2012) Acclimation of the global transcriptome of the cyanobacterium *Synechococcus* sp. strain PCC 7002 to nutrient limitations and different nitrogen sources. *Front Microbiol* 3:1–15.
 29. Billis K, Billini M, Tripp HJ, Kyrpides NC, Mavromatis K (2014) Comparative transcriptomics between *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 provide insights into mechanisms of stress acclimation. *PLoS One* 9(10):1–10.
 30. Wang K (2007) Biology and ecology of *Synechococcus* and their viruses in the Chesapeake Bay. doi:10.1017/CBO9781107415324.004.
 31. Murphy RR, Kemp WM, Ball WP (2011) Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts* 34(6):1293–1309.
 32. Sommer SE, Pyzik AJ (1974) Geochemistry of middle Chesapeake Bay sediments from upper Cretaceous to present. *Chesap Sci* 15(1):39–44.
 33. Sinex SA, Helz GR (1981) Regional geochemistry of trace elements in Chesapeake Bay sediments. *Environ Geol* 3(6):315–323.
 34. Buchanan C, Lacouture R V., Marshall HG, Olson M, Johnson JM (2005) Phytoplankton reference communities for Chesapeake Bay and its tidal tributaries. *Estuaries* 28(1):138–159.

35. Fisher TR, Peele ER, Ammerman JW, Harding LW (1992) Nutrient limitation of phytoplankton in Chesapeake Bay. *Mar Ecol Prog Ser* 82:51–63.
36. Fisher TR, et al. (1999) Spatial and temporal variation of resource limitation in Chesapeake Bay. *Mar Biol* 133(4):763–778.
37. Shih PM, et al. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* 110(3):1053–8.
38. Larsson J, et al. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 8(9):1892–903.
39. Muro-Pastor MI, Reyes JC, Florencio FJ (2005) Ammonium assimilation in cyanobacteria. *Photosynth Res* 83(2):135–150.
40. Brand LE, Sunda WG, Guillard RRL (1983) Limitation of marine phytoplankton reproductive rates by zinc, manganese, and iron. *Limnol Oceanogr* 28(6):1182–1198.
41. Priya B, et al. (2007) Comparative analysis of cyanobacterial superoxide dismutases to discriminate canonical forms. *BMC Genomics* 8:435.
42. Flores E, Herrero A (1995) Assimilatory nitrogen metabolism and its regulation. *The Molecular Biology of Cyanobacteria*, pp 487–517.
43. Thelwell C, Robinson NJ, Turner-Cavet JS (1998) An SmtB-like repressor from *Synechocystis* PCC 6803 regulates a zinc exporter. *Proc Natl Acad Sci* 95(18):10728–10733.
44. Morby AP, Turner JS, Huckle JW, Robinson NJ (1993) SmtB is a metal-dependent repressor of the cyanobacterial metallothionein gene *smtA*: identification of a Zn inhibited DNA-protein complex. *Nucleic Acids Res* 21(4):921–925.

45. Marsan D, Wommack KE, Ravel J, Chen F (2014) Draft genome sequence of *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay estuary. *Genome Announc* 2(1):e01111–13.
46. Darling AE, Tritt A, Eisen JA, Facciotti MT (2011) Mauve assembly metrics. *Bioinformatics* 27(19):2756–2757.
47. Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
48. Glass EM, Meyer F (2011) The metagenomics RAST server: A public resource for the automatic phylogenetic and functional analysis of metagenomes. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, pp 325–331.
49. Galens K, et al. (2011) The IGS standard operating procedure for automated prokaryotic annotation. *Stand Genomic Sci* 4(2):244–51.
50. Wang DZ, et al. (2011) Homology-driven proteomics of dinoflagellates with unsequenced genomes using MALDI-TOF/TOF and automated de Novo sequencing. *Evidence-based Complement Altern Med* 2011. doi:10.1155/2011/471020.
51. Dong H-P, Wang D-Z, Dai M, Chan LL, Hong H-S (2009) Shotgun proteomics: Tools for analysis of marine particulate proteins. *Limnol Oceanogr Methods* 7(DEC):865–874.
52. Knief C, et al. (2012) Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J* 6(7):1378–1390.

Chapter 4: Surviving in a highly variable environment - the novel toxin-antitoxin systems in the estuarine *Synechococcus* strain CB0101

**Surviving in a highly variable environment - the novel toxin-antitoxin systems in the
estuarine *Synechococcus* strain CB0101**

David Marsan¹, Allen Place¹, and Feng Chen^{1*}

¹Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD, USA

Keywords: *Synechococcus*, Toxin-Antitoxin system, persister cell, stress response, *relB/relE*

Significance Statement

Toxin-antitoxin (TA) systems have been found in many bacteria, but not in marine picocyanobacteria. TAs are stress response systems that allow microbial cells to survive under harsh environmental conditions. We demonstrated that TA systems, especially *relB/relE* in marine *Synechococcus* respond to environmental stresses, and play an active role in conferring niche adaptation. This growth modulator induces a reversible persister state to enhance fitness and competitiveness. As a result, picocyanobacteria can persist during long periods of time under stress conditions and revive when suitable conditions arise. Further, eight other *Synechococcus* strains were found to contain TA systems, suggesting the role they play in the diversity and function of picocyanobacteria is still unknown.

4.1 Abstract

Bacterial toxin-antitoxin (TA) systems are genetic elements composed of a toxin gene and its cognate antitoxin, with the ability to regulate growth. TA systems have not previously been reported in marine *Synechococcus* or *Prochlorococcus*. Here we report the finding of seven TA system pairs (Type II) in the estuarine *Synechococcus* CB0101, and their response to five different stressors; nitrogen and phosphate starvation, phage infection zinc toxicity, and photo-oxidative stress. Database searches discovered that 8 other marine *Synechococcus* strains also contain at least one TA pair but none in *Prochlorococcus*. We demonstrated that the *relB/relE* TA pair was active and arrested translation through RNA degradation when CB0101 was under oxidative stress caused by either zinc toxicity or high light intensities, but the arrest was reversible when the stress event was removed. We propose that having TA systems such as *relB/relE* allows for niche adaptation to the low light environments of the Chesapeake Bay enabling *Synechococcus* like CB0101 to thrive. Consequently, other picocyanobacteria should be able to utilize their TA systems to cope with rapidly changing environments (e.g. nutrient or oxidative stress) by regulating cell growth, thus conferring niche adaptation and enhancing fitness.

4.2 Introduction

Cyanobacteria are an ancient and diverse group of prokaryotes found in habitats worldwide, which reflects their considerable ability to adapt to variable and extreme environmental factors such as nutrient availability, light intensity, and metal toxicity [1]. Estuaries, which are among the most productive yet variable aquatic environments on Earth, are one such habitat. Mixing of fresh and marine waters provides strong environmental gradients to the cyanobacteria living in these ecosystems. Estuarine cyanobacteria (e.g. *Synechococcus*) contribute significantly to global primary production and are at the interface of direct anthropogenic affects [2–4]. In estuaries, the bioavailability of nutrients (e.g. N and P), light intensity, and trace metals (e.g. Fe and Zn) are highly variable. Due to the complex geochemistry of these compounds and physical dynamics, understanding cyanobacterial response to the estuarine environment is important towards understanding the function of ecosystems.

Indeed, in their long evolutionary history, cyanobacteria have developed highly refined response strategies shared with other prokaryotes under various stressors [5,6]. One common successful survival strategy is the ability to undergo reversible growth arrest [5,6] using chromosomal toxin-antitoxin (TA) systems. TA systems have been documented in many organisms, especially *Escherichia coli*, to cope with various stresses by reducing growth, inhibiting growth or killing a subpopulation of cells [7–9]. The freshwater filamentous cyanobacteria *Anabaena sp.* PCC7120 contains numerous chromosomal TA systems, some of which are predicted to promote survival under particular stresses [10].

TA systems, also referred to as addiction or suicide modules, typically consist of an auto-regulated operon encoding a stable toxin and a labile antitoxin [9,11,12]. Based on the antitoxin nature and mode of action, TA systems are grouped into type I, II, III, IV, and V classes [13–15]. Antitoxins of type II, the most abundant type of TA, are proteins that inactivate toxins by forming TA complexes [11]. The vast majority of type II toxins are mRNA specific endoribonucleases, also called mRNA interferases, such as *mazF*, *relE*, and *hicA* toxins [16]. Almost all described type II TA systems share this mode of regulation: tight expression auto-regulated by a TA complex binding to the promoter region and toxin neutralization by formation of a TA complex. All apparently result in translation inhibition.

Type II systems have been found in plasmids and chromosomes of free-living bacteria [17–19]. Type II TA system prevalence has led to the proposal that chromosomal TA systems are stress-response elements contributing to prokaryotes' adaptation to stressful environments [9,11]. According to this hypothesis, under unfavorable conditions antitoxins would be degraded by stress-induced proteases, which cause relief of transcriptional repression of TA operons and release of toxins from TA complexes [16]. As a consequence, the free toxins would induce reversible growth arrest or cell death by inhibiting an essential cellular process such as protein synthesis or DNA replication [16]. This hypothesis has been supported by a series of recent experimental findings [20–23]. For instance, in *E. coli*, activation of TA systems is triggered by various stresses, and ectopic expression of nearly all characterized chromosomal toxins could improve the ability to confer resistance to stress [20].

Despite the ubiquity of TA systems in other prokaryotic organisms, they have not previously been reported in marine picocyanobacteria such as *Synechococcus* or *Prochlorococcus*. Previous studies have hypothesized that a lack of TA systems in some prokaryotes is related to their small genome sizes (<3Mb) and relatively simple life style in stable environments [17]. Currently, more than 60 marine picocyanobacteria genomes have been sequenced (NCBI search as of Jan 2016). In general, genome sizes of marine *Synechococcus* and *Prochlorococcus* are smaller than 3Mb [24–26]. Recently, we identified seven putative chromosomal TA pairs in the genome (2.8Mb) of *Synechococcus* strain CB0101, which was isolated from the Chesapeake Bay [27].

Marine *Synechococcus* can be divided into three major subclusters, 5.1, 5.2 and 5.3 [24,25,28]. Subclusters 5.1 and 5.3 *Synechococcus* are present in coastal and oceanic waters, while subcluster 5.2 *Synechococcus* mainly occupies estuaries [29]. Compared to coastal and oceanic *Synechococcus*, much less is known about the ecological function and genomic evolution of estuarine *Synechococcus*. Subcluster 5.2 *Synechococcus* dominate the estuarine environment [30] making up 20-40 % of phytoplankton chlorophyll-a during the summer [31] and contain a novel pigment gene cluster not seen in other subclusters [29]. *Synechococcus* strain CB0101 (a member of subcluster 5.2) is able to grow in a wide range of salinities (0-30 ppt) and temperatures (4 to 32 °C), and is often subjected to viral infection [31–33]. The presence of TA systems in the genome of CB0101 leads to the hypothesis that estuarine cyanobacteria similar to CB0101 retain TA systems to aid in adaptation to strong environmental gradients in the estuary.

In this study, we identified and characterized the chromosomal TA systems found in *Synechococcus* strain CB0101 [27]. The expression of TA systems in CB0101 in

response to stressful environmental conditions (nutrient, metal toxicity, light intensity, and phage infection) were analyzed using RNA-Seq, confirmed through qPCR, and further confirmed through Western blotting of extracts. Time series experiments displayed the activation and shutdown of the toxin and antitoxin proteins when a stressful environment (i.e. Zn toxicity or photo-oxidative stress caused by high light intensity) was encountered and subsequently relieved. Our results suggest that by inducing a reversible growth arrest, TA toxins allow stressed cells to remain in a dormant or non-growing stress-tolerant state until more favorable environmental conditions return. Genomic searches led to the finding of TA systems in 8 other marine *Synechococcus*. This is the first identification, characterization, and functional confirmation of TA systems in marine picocyanobacteria.

4.3 Methods

4.3.1 Cyanobacterial strains and stress experiments

Cultures of CB0101 were grown in SN medium with 15 ppt salinity and vitamin B12 (10 μ g/L) (referred to as SN15 medium hereafter) at 23 °C under 15 μ E m⁻² s⁻¹ continuous light. Filtered (0.2 μ m pore size) air was bubbled into individual 500 mL cell culture flasks. For the RNA-Seq analysis, CB0101 cultures were initially grown under the above conditions and duplicates were subjected to 4 different conditions for 72 h: (1) Control, where the culture was grown in SN15 medium; (2) Nitrogen depleted condition, where the cultures were cultivated in the SN15 medium that did not contain sodium nitrate; (3) Phosphate deplete condition, where the cultures were grown in the SN15 medium that lacked dipotassium phosphate; (4) Zinc toxic condition, where the cultures were grown in the SN15 medium amended with 50 μ M of zinc sulfate heptahydrate. The

growth of cultures was monitored by counting cells using an Accuri C6 flow cytometer.

The specific growth rate was calculated from the logarithmic change $\mu = \frac{\ln X_2 - \ln X_1}{t_2 - t_1}$ where X_1 and X_2 are densities at times t_1 and t_2 . Samples were taken at 0 min and after 72 h, and mRNA was extracted immediately after sampling.

For the phage infection experiment the length of the lytic cycle was determined by one-step growth curve. Once the lytic cycle was determined, three time points were identified to correlate with attachment (30 min), latent (5 h), and lysis (12 h). Triplicate cultures of CB0101 were grown to exponential phase (10^8 cells ml^{-1}), and then infected with S-CBP1 [33] podovirus (10^8 infective phage particles ml^{-1}) for a multiplicity of infection (MOI) of 1. Control cultures were amended with sterilized SN15 medium. At the designated time points mRNA was extracted (Supplemental Table 4.S2).

For the light intensity experiment, exponentially growing CB0101 culture (in SN15 medium) was split into five 25-ml cell culture flasks, which were exposed to five different light intensities (15, 50, 100, 150, and 200 $\mu\text{E m}^{-2} \text{s}^{-1}$). The entire culture was extracted and mRNA and protein were separated for qPCR and Western blot analysis, respectively.

4.3.2 Extraction of mRNA and qPCR reaction

Samples were collected and immediately processed for mRNA analysis. RNA was separated using Trizol extraction. mRNA was extracted from the total RNA using Ambion MICROBExpress kit (Life technologies, AM1905), modified with 4 specific oligo primers for rRNA depletion of CB0101 ribosomes. rRNA removal evaluation was conducted using the Agilent 2100 Bioanalyzer resulting in 98 % removal of rRNA and confirmed the presence of high quality RNA. The NEXTflex RNA-Seq kit (BIOO

Scientific, #5129-01) was then used to prepare mRNA libraries for sequencing using the Illumina Hi-Seq. RNA libraries were created and barcodes assigned to each sample for multiplexing of a single Hi-Seq lane. RNA-Seq was carried out by the Institute of Genome Sciences (IGS) using a Hi-Seq with 100 bp-paired end read length. As a further confirmation, qPCR was used to validate the expression of each TA pair transcript expression (Supplemental Table 4.S2).

4.3.3 TA system identification in CB0101 and completely sequenced Synechococcus and Prochlorococcus genomes

BLASTCLUST (-L 0.75 -S1.0) searches for all identified TA genes grouped by family and cluster was used to identify TA genes within *Synechococcus* CB0101. These TA systems were further validated using the RASTA-Bacteria and TA finder programs [34,35]. Further, this method was used with representatives of each cluster used as queries in a BLAST search (e-value 0.01) against ~60 completely sequenced *Synechococcus* and *Prochlorococcus* genomes available on the NCBI Microbial Genomes website at the time of this analysis (January 2016) [36]. Significant hits among proteins encoded in these genomes were classified as toxins or antitoxins; in the case of multiple matches to different TA families, the protein was assigned according to the highest-scoring match TA query. Co-directed genes with adjacent chromosome locations belonging to different toxin/antitoxin families were recorded as a TA pair.

4.3.4 Western Analysis

Western blot analysis was conducted using a procedure described by Mahmood and Yang [37]. Amino acid sequences for all 7 TA pairs were provided to GenScript for epitope analysis. Results were analyzed using their Optimum Antigen design tool and the

relB²/relE¹ pair was selected due to transcript expression under each stress and favorable antigen design. Antibodies were raised for one TA pair, *relB²/relE¹* using GenScript service. The antibody epitope sequences used were TARLPDDLTAELDAC and CVLVVRVGHHRKEVYR. Western blot analysis was performed against samples from the resulting cultures using a primary rabbit polyclonal anti-FLAG antibody and a goat anti-rabbit HRP conjugated secondary antibody as directed by the manufacturer. Constant whole cell equivalents ($\sim 2 \times 10^7$ cells) mixed in running buffer were used for each Western. NuPAGE Novex gels with 4-12% Bis-tris gel with MES running buffer was used. Gels were further processed using ImageJ and image lab software for the Bio-Rad ChemiDoc imager. Pixel densities and predicted molecular weights were calculated for each gel.

4.4 Results

4.4.1 Annotated toxin/antitoxins in CB0101

Genome annotation revealed the presence of 7 chromosomal pairs of TAs, homologous to *E. coli* types II (Table 4.1 and Fig. 4.1). Each TA pair was analyzed using InterProScan 5 and ExPASy for motif, PI, and molecular weight prediction (Supplemental Table 4.S1). The type II TA pairs included *yefM¹/yoeB¹*, *mazE¹/mazF¹*, *relB²/relE¹*, *vapB¹/vapC¹*, *relB¹/vapC¹*, *phd¹/doc¹*, and *phd²/doc²* all of which use protein-to-protein interactions and arrest translation.

The amino acid sequences of each TA pair from CB0101 were BLAST searched in GenBank to identify their closest homologs. The closest neighbors to the TA pairs of CB0101's are shown in Table 4.1. The DNA sequence identities of these closest hits varied from 50 to 93 % among the 7 TA pairs. Two-thirds of TA genes were related to

the TA genes in picocyanobacteria (mainly marine *Synechococcus* and *Cyanobium*), while one-third of CB0101 TA genes seem to arise from heterotrophic bacteria, i.e. *Collimonas fungivorans*, *Thioalkalivibrio sulfidophilus*, and *Polaromonas glacialis* (Table 4.1). Interestingly, the TA systems were found in diverse ecotypes of marine *Synechococcus*, i.e. KORDI-49 (coastal), WH8103 (open ocean), WH5701 (estuarine), and CB0205 (estuarine). Also of note is the consistent presence of toxin and antitoxin pairs from the same organism i.e. TA pair #3 from *Cyanobium sp.* PCC7001 (Table 4.1). Most pairs exhibit highly basic or acidic PI values strengthening the case for their covalent binding with each other (Supplemental Table 4.S1). Many of the pairs contain 1-3 nucleotide overlaps suggesting transcriptional coupling (Fig. 4.1). For most of the TA pairs the coding regions on either side contain palindromic sequences (Fig. 4.1). The genome location of each TA pair falls within regions defined as genomic islands. Genomic islands (GIs) are large regions (more than eight kilobases long) of non-conserved, non-core genes that are sporadically distributed among the genome. Genomic islands were identified using the IslandViewer program.

4.4.2 Responses of *Synechococcus* TA systems to the environmental stresses

The growth rate of CB0101 under different stress conditions (nitrogen depletion, phosphorus depletion, and zinc toxicity) is shown in Fig. 4.2. CB0101 grew more slowly compared to the control under all conditions (Fig. 4.2). However, the growth of CB0101 under phosphate depletion and zinc toxicity was significantly slower than the growth of control ($p < 0.05$). High zinc exposure nearly suppressed the growth of CB0101 completely (t-test value 2.7, $p = 0.09$), which is significantly different than the control.

RNA-Seq analysis was conducted to determine the expression response of CB0101 to four stress conditions (nitrate depletion, phosphorous depletion, zinc toxicity, and phage infection) (Table 4.2 and Supplemental Table 4.S2). The transcripts of toxin *relE¹* were progressively upregulated 2.3, 3.1, and 10.5 fold ($p < 0.01$ and minimum 2 fold) in nitrate depletion, phosphate depletion, and zinc toxicity, respectively (TA pair #5, Table 4.2). *relE¹* was also upregulated 2 fold at 30min post infection of phage S-CBP1. The antitoxin gene *relB²* was downregulated by 2.5 fold when CB0101 was grown under nitrate depletion (TA pair #5, Table 4.2).

Other upregulated toxin genes included *doc¹* (2.7 fold in zinc toxicity) and another *yoeB¹* (2.2 and 2.1 fold in both nitrate depletion and zinc toxicity). The antitoxin *phd¹* was upregulated 2.5 fold only under nitrate deplete conditions. *relB²* was only downregulated (2.5 fold) under nitrogen deplete conditions. While *relB¹* was downregulated 2.6-fold under zinc toxic conditions.

In order to confirm the RNA-Seq expression data, PCR primers were made for each TA pair and the quantity of specific TA gene transcripts was measured using qPCR (Supplemental Table 4.S3). The same samples were used from the RNA-Seq experiments to verify results ($p < 0.01$ and minimum 2 fold). The qPCR further confirmed the RNA-Seq expression data that the TA systems were active under the different stress experiments (Table 4.2). To demonstrate the actual activity of TA systems at the protein level, we selected one TA pair (*relB²/relE¹*) to further study the expression of their proteins based on the detection of specific antibodies. This TA pair was selected because the toxin was expressed at the transcript level, showed the largest expression difference and optimally designed antibodies could be made to both pairs. The degradation of the

tRNAs and ribosomal RNAs of CB0101 under high Zn conditions was observed, consistent with *relE* toxin translation inhibition (Supplemental Fig. 4.S1).

4.4.3 Production of *relE*¹ arrests the growth of CB0101

The *relE*¹ and *relB*² genes on CB0101's chromosome encode RelE and RelB homologous to the *E. coli* RelE (24 % identity) and RelB (41 % identify), respectively. RelE is predicted to be an 84-residue sequence at 9.53 kDa, while RelB is predicted at 73 residues and 8.08 kDa (Supplemental Table 4.S1). The protein structures of *relE*¹ and *relB*² were predicted using Phyre2 (Supplemental Fig. 4.S2). Expression of mRNA for the *relB*²/*relE*¹ pair in response to nitrate depletion, phosphate depletion and zinc toxicity were shown in Fig. 4.3A. The mRNA expression of the *relB*²/*relE*¹ pair has been described above. The antibodies were used to confirm the presence of each protein (Fig. 4.3B-C). The expressed RelB²/RelE¹ proteins in all stress conditions were detected using specific antibodies loaded with constant cell equivalents (Fig. 4.3B-C). Under the normal growth condition (control) a dimer of RelB²/RelE¹ was detected at 18 kDa (Fig. 4.3C), while the toxin protein RelE¹ was not detected at its predicted weight 9 kDa (Fig. 4.3B). When CB0101 became stressed in all three treatments, toxin protein RelE¹ (band 9kDa) became more prevalent and the dimer of antitoxin/toxin protein RelB²/RelE¹ (band 18 kDa) declined (Fig. 4.3B-C).

4.4.4 Reversible *relB*²/*relE*¹ expression when the stress is removed

After confirming the presence of both toxin and antitoxin proteins (RelB²/ RelE¹) in all stress treatments, a time series experiment was set up to determine the progression of the TA pair when CB0101 was exposed to zinc toxicity (Fig. 4.4). In some characterized TA systems, the toxin can be counteracted by subsequent production of

antitoxin. It would be interesting to know if the stress responses of toxin and antitoxin (*relB²/relE¹*) are reversible when the stress level is reduced. In this experiment, CB0101 was exposed to the same zinc toxicity (as the previous experiment), and subsamples were collected at 9 different time points (0, 0.5, 3, 12, 24, 36, 38, 50, 62, and 72 h). CB0101 was grown in the presence of zinc for 36 h; the culture was then exposed to fresh non-toxic media and allowed to grow for another 36 h (Fig. 4.4). When the zinc was added to the culture the toxin (RelE¹) began to accumulate over time (Fig. 4.3A; 0-24 h). Upon removal of zinc at 36 h, toxin protein began to dissipate gradually over time (Fig. 4.4A; 36-72 h). The opposite trend was observed with the dimer of antitoxin/toxin. Upon exposure to zinc the RelB²/RelE¹ complex decreased gradually and was fully degraded within 24 h (Fig. 4.4B). When the culture was released from zinc toxicity at 36 h, the dimer of RelB²/RelE¹ recovered almost instantaneously (Fig. 4.4B). Pixel densities shows the inverse synchronization of the toxin and antitoxin proteins (Fig. 4.4C). The growth rate of CB0101 was inhibited with zinc exposure in the first 24 h, and recovered after the release of zinc stress (Fig. 4.4D).

4.4.5 Light response for *relB²/relE¹*

Since zinc toxicity is inducing *relB²/relE¹* and this type of stress induces oxidative stress, we believe a better representation of niche adaptation would be responses to photo-oxidative stress. A major source of oxidative stress which photosynthetic organisms deal with is the production of reactive oxygen species (ROS) [38] at high light intensities. CB0101 grew rapidly at low light intensities (15-50 $\mu\text{E m}^{-2} \text{s}^{-1}$) and the

growth was severely inhibited when light intensity exceeded $200 \mu\text{E m}^{-2} \text{s}^{-1}$ (Fig. 4.5A). The level of photo bleaching increased with increasing light intensity (Fig. 4.5B).

The upper Chesapeake Bay, in which CB0101 thrives in, is a low light environment due to increased particulate influxes. The qPCR analysis of each TA pair showed the expression of *relE*¹ was upregulated (2.7, 3.9 and 5.6 fold) when CB0101 was exposed to light at 100, 150, and $200 \mu\text{E m}^{-2} \text{s}^{-1}$, respectively (Fig. 4.5C). The transcripts of two other toxins, *yoeB*¹ and *doc*^{1, 2}, were upregulated (2.4 and 2.6 fold, respectively) at the highest light intensity. Lastly, the mRNA of two antitoxins, *relB*¹ and *phd*² were downregulated at the highest light intensity. The Western blot of RelB²/RelE¹ showed the presence of a dimer (18kDa), which progressively decreased as the light intensity increased (Fig. 4.5D). Conversely, the RelE¹ protein became visible at $100 \mu\text{E m}^{-2} \text{s}^{-1}$ and increased at 150 and $200 \mu\text{E m}^{-2} \text{s}^{-1}$.

4.5 Discussion

4.5.1 Why do *Synechococcus* spp. contain TA systems?

The presence of 7 pairs of TA systems in the chromosome of CB0101 is the first description in marine cyanobacteria (Table 4.1 and Fig. 4.1). The number and type of TA systems in CB0101 are comparable to other prokaryotes [12]. TA systems have been found in prokaryotic plasmids and chromosomes. While the plasmid-encoded TAs are known to stabilize plasmids in cells, the chromosome TA systems have multiple functions [39]. Four possible mechanisms for the presence of TA systems in chromosomes have previously been identified: 1) TA systems on chromosomes may fulfill a similar function to plasmid types and mediate stabilization of important genetic regions [19]; 2) Protection against invading DNAs such as plasmids and phages [20,40]; 3) Formation of bacterial persister cells [41]; 4) Regulation of biofilm formation or global regulators of translation [7].

We believe that at least three of the four above-mentioned mechanisms are applicable to TA systems in CB0101. First, TA systems of CB0101 are in close proximity to important genes such as Type IV pilus, DNA repair, rod shape-determining proteins (*Mre*), SOS-response repressor and protease *LexA* (mechanism 1). All of which are involved in functions to boost niche ecotype or overall survival. Secondly, the up-regulation of the *relE*¹ toxin of the *relB*²/*relE*¹ pair occurred within 30 min of cyanophage infection (Supplemental Table 4.S2). Expression of *relE*¹ may cause the cell to arrest translation, thus slowing or stopping the phage from replicating (mechanism 2). Thirdly, TA systems may allow CB0101 to develop dormancy or persister cells in response to environmental stresses (mechanism 3). This is reflected by the ability for CB0101 to

arrest translation and growth during stressful environmental conditions by regulating its TA systems such as *relB²/relE¹* (Fig. 4.4). CB0101 was able to resume its growth when the stress was removed, likely undergoing a persister state (Fig. 4.4). A bacterial persister state is a quasisdormancy state where the cells may recover and proliferate if enough antitoxin is produced to neutralize the toxin [40,41]. Formation of persister cells regulated by TA systems could be an important mechanism for *Synechococcus* species living in dynamic or unstable environments.

4.5.2 The *relB/relE* system

The expression of antitoxins and toxin genes, *relB²/relE¹*, were tightly regulated by the various stress conditions (Fig. 4.4 & 4.5). The mechanisms behind how *relB²/relE¹* respond to stress could be homologous to other prokaryotic systems such as *E. coli* [40]. Typically transcription of the TA operon is auto-regulated by binding of the antitoxin or by the TA system to the promoter [42]. Depending on the stoichiometric ratio of antitoxin to toxin, several types of complexes may be formed with distinct affinities to the promoter [43–46]. This coincides with the time-series experiment where the antitoxin RelB¹ rapidly recovered after the stress was released (Fig. 4.4).

Our studies show that RelB²/RelE¹ growth inhibition of CB0101 could be completely eliminated by subsequently ending the stress condition allowing RelB² to recover almost instantaneously (Fig. 4.4). The *relB²/relE¹* TA pair was active and arrested translation when under zinc oxidative stress, and the arrest was reversible when the stress event was removed (Fig. 4.1 & 4.4). This result is in accordance with the previous demonstration that the overproduction of TA toxin caused reversible growth arrest [43,45]. Therefore, it is reasonable to propose that the chromosomal *relB²/relE¹*

system of CB0101 represents a growth modulator that induces a persister state to enhance fitness and competitiveness under particular stress conditions, such as metal or oxidative stress. As a result, cyanobacteria with this TA system can persist for a long period of time under stressful conditions and revive when suitable conditions arise.

4.5.3 Ecological significance and oxidative response of *relB/relE*

Oxidative stress caused by high light intensities is known to be one of the greatest forms of stress caused to photosynthetic organisms. The penetration of sunlight into water places constraints on the survival and spatial distribution of marine cyanobacteria. Light penetration through the water column is controlled by the amount and kinds of materials that are dissolved and suspended in the water [47,48]. Water turbidity in the Chesapeake Bay is highly variable with highly turbid conditions present in the upper bay. This turbidity creates a low light environment of $(10-100 \mu\text{E m}^{-2} \text{ s}^{-1})$ within 0.15 m and 1 m [47]. Estuarine *Synechococcus* CB0101 is low light adapted relative to coastal and oceanic *Synechococcus* (Fig. 4.5). In the event of upwelling, lower turbidity, or clearing water, CB0101 becomes stressed due to higher light intensities. In order to limit that stress, CB0101 is able to induce *relB²/relE¹* TA system, which arrests growth and translation (Fig. 4.5).

Irradiance has long been known to exert a major control on cyanobacteria proliferation; excessive light can cause photo-oxidative cell death and *Prochlorococcus* and *Synechococcus* have different strategies to minimize this potential light-induced death [49,50]. The phototrophic nature of these organisms means that they not only need to manage the oxidative stress generated by oxygen reduction in the same way as heterotrophic organisms, but also produce oxidative oxygen during photosynthetic

electron transport [38]. Reactive oxygen species (ROS) are byproducts of aerobic metabolism and potent agents that cause oxidative damage [38]. In oxygenic photosynthetic organism such as cyanobacteria, ROS are inevitably generated by photosynthetic electron transport. Photosynthetic organism are able to overcome photodamage by the rapid and efficient repair of PSII [38,51]. Based on these findings, along with the observation of persister state effect of the *relB²/relE¹* system, it seems reasonable to propose that the activation of the *relB²/relE¹* system might redirect the energy normally utilized for growth to aid in repair mechanisms caused by brief amounts of high light intensities. In this model, the *relB²/relE¹* system is envisioned to delay programmed cell death and instead undergo a persister state. With TA systems found within GIs we can predict that they were acquired through horizontal gene transfer from other bacteria, perhaps from freshwater ecotypes where TA systems are common.

4.5.4 Presence of TA systems in broader picocyanobacteria

Based on a search of *Synechococcus* and *Prochlorococcus* genomes publically available on NCBI (described in methods). Eight *Synechococcus* strains ranging from estuarine, coastal, and open ocean (CB0205, WH5701, WH8102, WH8103, RS9916, RS9917, KORDI-49, and KORDI-51) with putative hits to at least one pair of TA were identified. Hits to current *Prochlorococcus* genomes were not found. Our results suggest that TA systems in marine *Synechococcus* are probably more widely dispersed than thought. Although it is not our focus here to conduct a thorough investigation of the distribution of TAs in marine picocyanobacteria the diversity and function of TA systems deserves further investigation.

It has been proposed that the absence of TA systems in prokaryotes with small genomes could be due to their relatively simple life style in stable environmental conditions [17]. Makarova et al., [18] argued that the lack of TAs in small genomes is a consequence of the general “laws” of scaling of differential functional categories of genes with genome size. Although it has been predicted (with 95 % confidence) that no TAs are present in microbial genomes, which contain less than 3,100 genes [18]. This is clearly not the case since TAs are present in marine *Synechococcus* strains CB0101, CB0205, WH5701, WH8102, WH8103, RS9916, RS9917, KORDI-49, and KORDI-51, whose genomes contain less than 3,100 genes. Further work is needed to understand the number, type, diversity, evolution and ecological significance of TAs in marine picocyanobacteria. It is clear that TA systems play a major role in regulation of *Synechococcus*.

4.6 Conclusion

Here we described the first TA systems in marine picocyanobacteria. *Synechococcus* strain CB0101, a member of marine *Synechococcus* subcluster 5.2, contains 7 chromosomal toxin-antitoxin gene pairs which were expressed in response to five different stressors; nitrogen or phosphate starvation, zinc toxicity, phage infection and photo-oxidative stress. Based on the TA pair *relB*²/*relE*¹ we demonstrated that the TA system arrests the growth and translation under stress conditions, and revives growth when the adverse event is removed. Activation of the *relB*²/*relE*¹ system allows *Synechococcus* like CB0101 to regulate growth in response to oxidative stress caused by variable light intensities in the Chesapeake Bay. Therefore, the presence of numerous TA systems in the chromosome of CB0101 may contribute to its ability to acclimate to

variable conditions in the estuarine ecosystem. The growth regulation of chromosomal type II TA systems may promote marine *Synechococcus* to cope with stressful environments.

The BLAST search on known marine *Synechococcus* and *Prochlorococcus* genomes led to the finding of TA pairs present in eight different marine *Synechococcus* strains isolated from estuarine, coastal, and open ocean environments. Intriguingly, none were found within *Prochlorococcus*. The presence of TA genes in diverse *Synechococcus* spp. suggests that we know very little about the diversity, evolution and ecological functions of TA genes in marine picocyanobacteria. Evaluating TA system-mediated growth regulation of marine picocyanobacteria in response to various environmental changes, i.e. temperature and salinity is important towards understanding the effect of climate change on ecosystem function.

Table 4.1. The outcome of protein BLAST search of the 7 *Synechococcus* CB0101 TA gene pairs. Toxin is represented by bold and antitoxin by italics. All matches represent a best fit to length (>90%).

TA Pair #	TA Family	Closest Neighbor	Identity	E value
1	<i>yefM</i> ¹	<i>Synechococcus</i> sp. KORDI-49	76%	3E-38
1	yoeB ¹	<i>Synechococcus</i> sp. WH8103	81%	1E-43
2	<i>phd</i> ¹	<i>Cyanobium</i> sp. CACIAM 14	87%	2E-36
2	doc ¹	<i>Collimonas fungivorans</i>	50%	4E-33
3	<i>vapB</i> ¹	<i>Cyanobium</i> sp. PCC7001	75%	2E-33
3	vapC ¹	<i>Cyanobium</i> sp. PCC7001	74%	7E-63
4	<i>relB</i> ¹	<i>Synechococcus</i> WH5701	72%	1E-14
4	vapC ²	<i>Synechococcus</i> sp. WH5701	81%	4E-67
5	<i>relB</i> ²	<i>Thioalkalivibrio sulfidophilus</i>	55%	4E-19
5	relE ¹	<i>Thioalkalivibrio sulfidophilus</i>	64%	5E-28
6	<i>mazE</i> ¹	<i>Thauera phenylacetica</i>	59%	2E-24
6	mazF ¹	<i>Polaromonas glacialis</i>	71%	5E-63
7	<i>phd</i> ²	<i>Synechococcus</i> sp. CB0205	93%	0E+00
7	doc ²	<i>Cyanobium</i> sp. CACIAM 14	67%	5E-32

Table 4.2. Expression of 7 *Synechococcus* CB0101 toxin/antitoxin pairs under various stress conditions detected by RNA-Seq and qPCR. Upregulation (red) or down regulation (blue) is based on expression of the control sample ($p < 0.01$ and minimum 2 fold). Toxin is represented by bold and antitoxin by italics.

TA Pair #	TA Family	Toxin Activity	Nitrogen RNA-Seq	Nitrogen qPCR	Phosphate RNA-Seq	Phosphate qPCR	Zinc RNA-Seq	Zinc qPCR
1	<i>yefM</i> ¹	mRNA interferase or inhibitor of translation initiation	-1.6	-1.9	-1.4	-1.7	-1.3	-1.3
1	yoeB ¹		2.2	2.6	1.5	1.8	2.1	2.4
2	<i>phd</i> ¹	Binds to the 30S ribosomal subunit	2.5	2.8	1.3	1.5	-1.2	-1.7
2	doc ¹		1.8	2.0	1.2	1.5	2.7	2.6
3	<i>vapB</i> ¹	Cleavage of tRNA	1.8	2.0	1.6	1.6	-1.7	-1.7
3	vapC ¹		1.4	1.8	-1.2	-1.5	1.7	1.6
4	<i>relB</i> ¹	Cleavage of tRNA	-1.0	-1.5	-1.0	-1.3	-2.6	-2.9
4	vapC ²		1.1	1.2	1.3	1.6	-1.1	-1.2
5	<i>relB</i> ²	Cleavage of ribosome-bound mRNA	-2.5	-3.0	1.2	1.6	1.1	1.3
5	relE ¹		2.3	2.5	3.1	3.5	10.6	11.2
6	<i>mazE</i> ¹	Ribosome-independent mRNA cleavage and cleavage of 23S rRNA	-1.2	-1.6	-1.2	-1.3	1.4	1.3
6	mazF ¹		-1.1	-1.3	-1.4	-1.5	1.7	1.7
7	<i>phd</i> ²	Binds to the 30S ribosomal subunit	1.4	1.5	1.2	1.5	-2.0	-2.2
7	doc ²		-1.5	-1.7	-1.3	-1.6	1.1	1.3

TA pair:

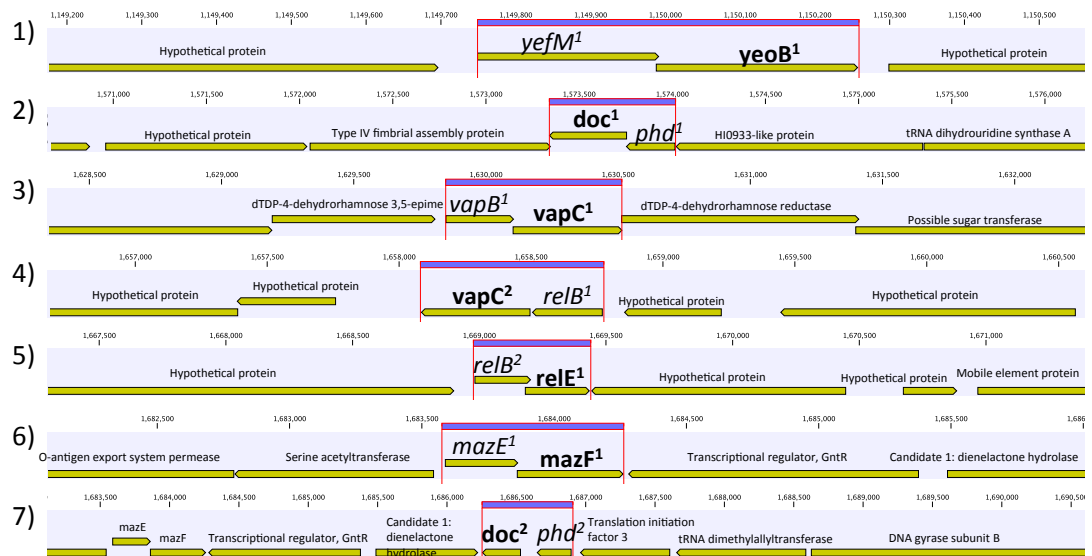


Figure 4.1. Genetic organization of the 7 TA systems of *Synechococcus* CB0101.

Schematic representation of the 7 TA pairs (highlighted in purple) and associated genes.

Toxin is represented by bold and antitoxin by italics.

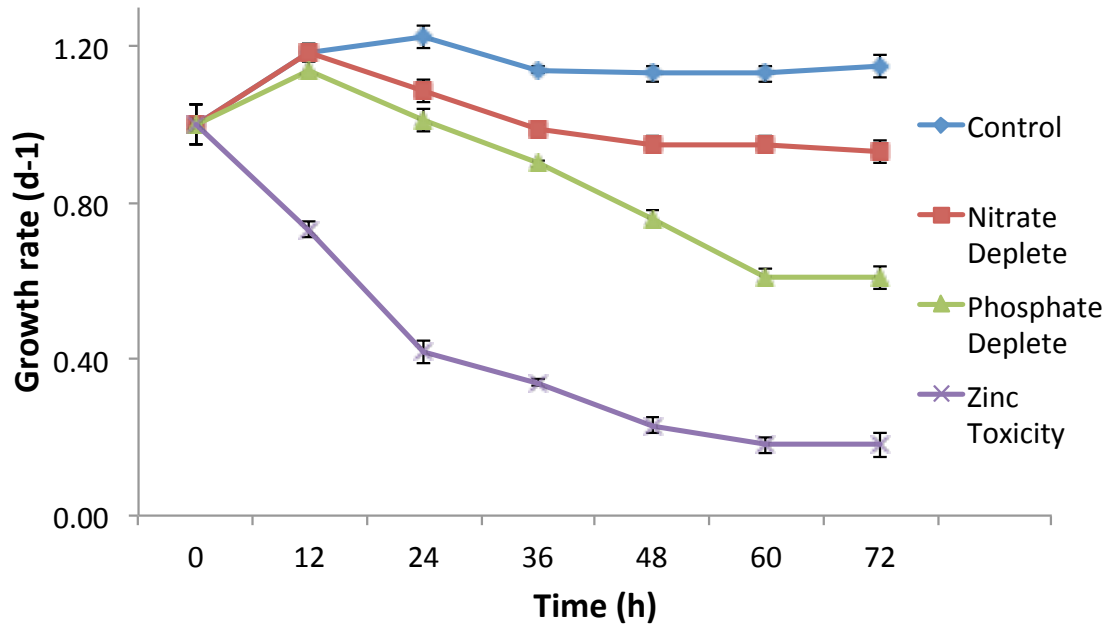


Figure 4.2. Growth rate of *Synechococcus* CB0101 during a 72 h period under different stress conditions: control medium, nitrogen limitation, phosphate limitation, and zinc toxicity (n=3). See Chapter 3 for detailed explanation on *Synechococcus* CB0101 response to these conditons.

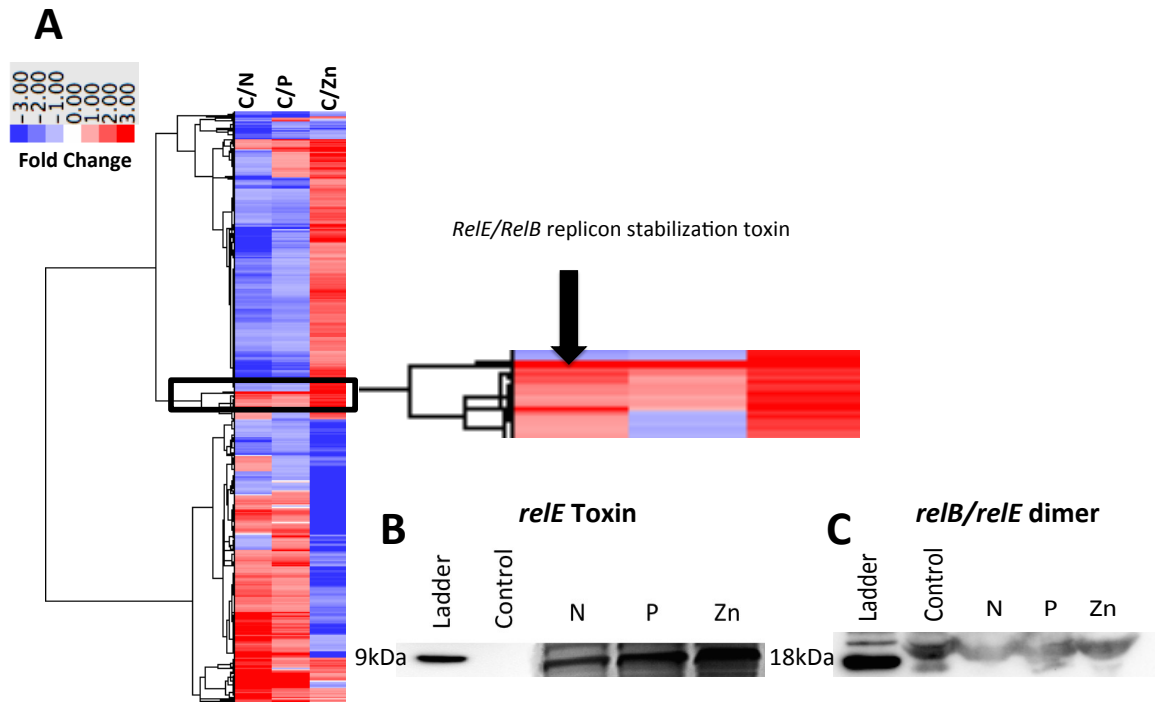


Figure 4.3. Global transcript expression of *Synechococcus* CB0101 to three stressors measured using RNA-Seq ($p < 0.01$ and minimum 2 fold) and compared to a control expression. A) Heat map and clustering of differentially expressed genes. Stress conditions are all compared to Control, with C/N representing nitrogen depletion, C/P phosphate depletion, and C/Zn zinc toxicity representing 843 (27%), 368 (12%), and 1,195 (38%) differentially expressed genes identified from 3,173 total genes. Dark red equates to a strong increase (>3 fold), while dark blue to a strong decrease (>3 fold) in gene expression. Zoomed image of the expression from the toxin-antitoxin pair *relB²/relE¹*. B) Detection of the toxin RelE¹ and C) antitoxin/toxin dimer RelB²/RelE¹ proteins by Western blot analysis from each stress condition including nitrogen (N) and phosphate (P) depletion, and zinc toxicity (Zn). Constant cell equivalents were used.

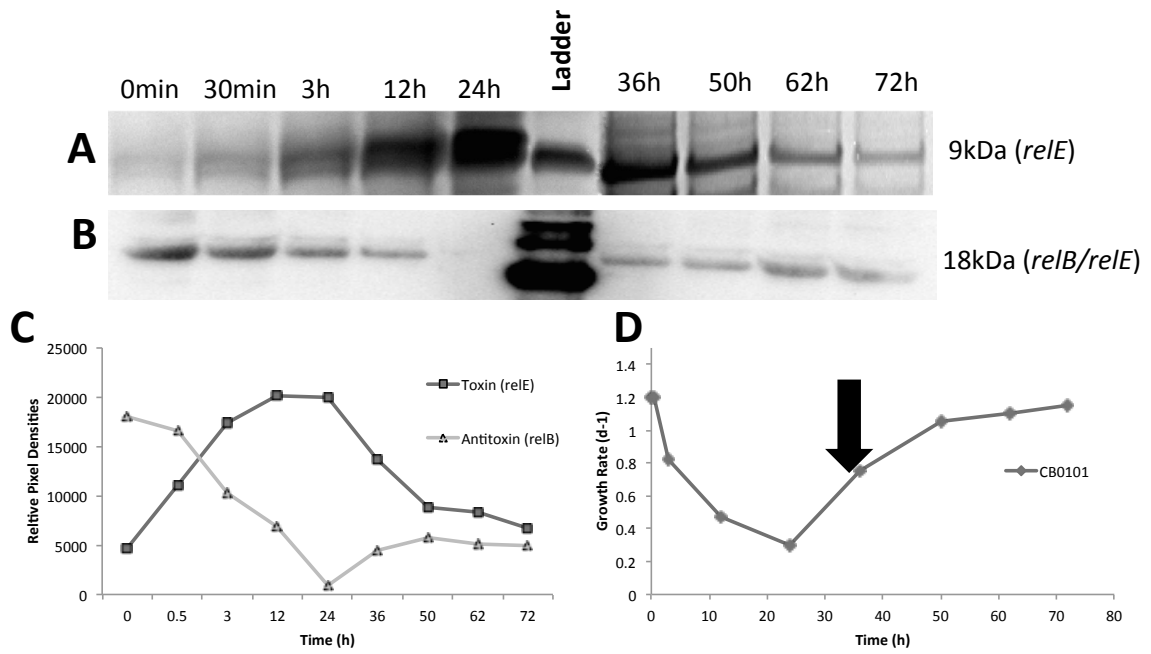


Figure 4.4. Time series detection of RelB²/RelE¹ proteins by Western blot analysis during zinc toxicity and subsequent releasing at 30 h using constant whole cell equivalents ($\sim 2 \times 10^7$ cells) mixed in running buffer. NuPAGE Novex gels with 4-12% Bis-tris gel with MES running buffer was used. A) Toxin RelE¹; B) Antitoxin/toxin dimer RelB²/RelE¹; C) Pixel density analysis to determine relative protein abundance; D) Growth rate of *Synechococcus* CB0101 during the time series with black arrow representing media change (release of zinc toxic conditions) at 35.5 h.

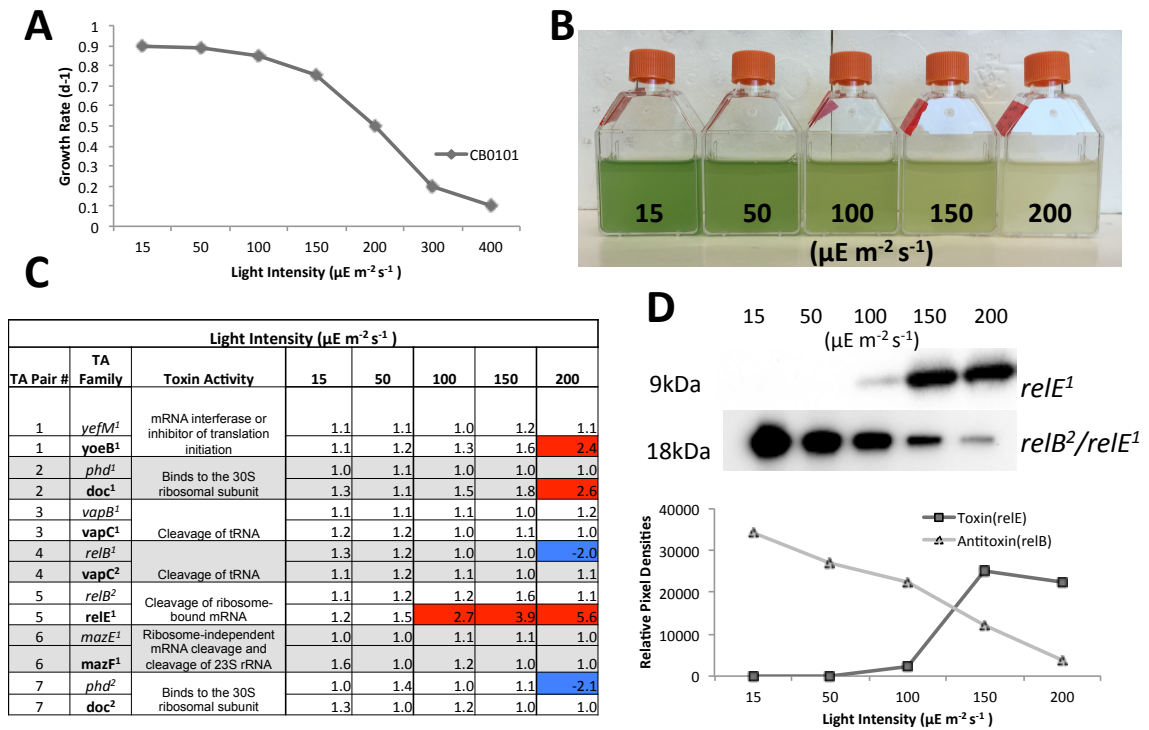


Figure 4.5. Photo-oxidative stress caused by different light intensities induces TA systems and arrests growth of CB0101. A) Growth rate of *Synechococcus* CB0101 under different light intensities; B) CB0101 cultures exposed to different light intensities 15, 50, 100, 150, and 200 $\mu\text{E m}^{-2} \text{s}^{-1}$ for 48 h; C) Expression of each CB0101 TA system responding to different light intensities detected by qPCR; D) Western blot analysis of RelE¹ toxin at 9kDa and RelB²/RelE¹ dimer at 18kDa under different light intensities with pixel densities measured.

References:

1. Tandeau de Marsac N, Houmard J. Adaptation of cyanobacteria to environmental stimuli: new steps towards molecular mechanisms. *FEMS Microbiol Lett.* 1993;104: 119–189. doi:10.1016/0378-1097(93)90506-W
2. Affronti LF, Marshall HG. Diel abundance and productivity patterns of autotrophic picoplankton in the lower Chesapeake Bay. *J Plankton Res.* 1993;15: 1–8. doi:10.1093/plankt/15.1.1
3. Ray RT, Hass LW, Sieracki ME. Autorophic picoplankton dyanmics in a Chesapeake Bay sub-estuary. *Mar Ecol Progress Ser.* 1989;52: 273–285.
4. Iriarte A, Purdie DA. Size distribution of chlorophyll a biomass and primary production in a temperate estuary (Southampton Water): the contribution of photosynthetic picoplankton. *Mar Ecol Prog Ser.* 1994;115: 283–297.
5. Schwarz R, Forchhammer K. Acclimation of unicellular cyanobacteria to macronutrient deficiency: Emergence of a complex network of cellular responses. *Microbiology.* 2005;151: 2503–2514. doi:10.1099/mic.0.27883-0
6. Lewis K. Persister cells, dormancy and infectious disease. *Nat Rev Microbiol.* 2007;5: 48–56. doi:10.1038/nrmicro1557
7. Kolodkin-Gal I, Verdinger R, Shlosberg-Fedida A, Engelberg-Kulka H. A differential effect of E. coli toxin-antitoxin systems on cell death in liquid media and biofilm formation. *PLoS One.* 2009;4. doi:10.1371/journal.pone.0006785
8. Van Melderen L, De Bast MS. Bacterial toxin-antitoxin systems: More than selfish entities? *PLoS Genetics.* 2009. doi:10.1371/journal.pgen.1000437
9. Yamaguchi Y, Park J-H, Inouye M. Toxin-Antitoxin Systems in Bacteria and

- Archaea. *Annu Rev Genet.* 2011;45: 61–79. doi:10.1146/annurev-genet-110410-132412
10. Ning D, Jiang Y, Liu Z, Xu Q. Characterization of a chromosomal type II toxin-antitoxin system *mazEaFa* in the cyanobacterium *Anabaena sp.* PCC 7120. *PLoS One.* 2013;8. doi:10.1371/journal.pone.0056035
 11. Gerdes K, Christensen SK, Løbner-Olesen A. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol.* 2005;3: 371 – 382. doi:10.1038/nrmicro1147
 12. Schuster CF, Bertram R. Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. *FEMS Microbiol Lett.* 2013;340: 73–85. doi:10.1111/1574-6968.12074
 13. Fozo EM, Hemm MR, Storz G. Small toxic proteins and the antisense RNAs that repress them. *Microbiol Mol Biol Rev.* 2008;72: 579–89. doi:10.1128/MMBR.00025-08
 14. Gerdes K, Wagner EGH. RNA antitoxins. *Current Opinion in Microbiology.* 2007. pp. 117–124. doi:10.1016/j.mib.2007.03.003
 15. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci USA.* 2009;106: 894–9. doi:10.1073/pnas.0808832106
 16. Yamaguchi Y, Inouye M. Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat Rev Microbiol.* Nature Publishing Group; 2011;9: 779–790. doi:10.1038/nrmicro2651
 17. Pandey DP, Gerdes K. Toxin-antitoxin loci are highly abundant in free-living but

- lost from host-associated prokaryotes. *Nucleic Acids Res.* 2005;33: 966–976.
doi:10.1093/nar/gki201
18. Makarova KS, Wolf YI, Koonin E V. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct.* 2009;4: 19. doi:10.1186/1745-6150-4-19
 19. Leplae R, Geeraerts D, Hallez R, Guglielmini J, Drze P, Van Melderen L. Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 2011;39: 5513–5525.
doi:10.1093/nar/gkr131
 20. Hazan R, Sat B, Engelberg-Kulka H. *Escherichia coli mazEF*-mediated cell death is triggered by various stressful conditions. *J Bacteriol.* 2004;186: 3663–3669.
doi:10.1128/JB.186.11.3663-3669.2004
 21. Prysak MH, Mozdierz CJ, Cook AM, Zhu L, Zhang Y, Inouye M, et al. Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage. *Mol Microbiol.* 2009;71: 1071–1087. doi:10.1111/j.1365-2958.2008.06572.x
 22. Singletary LA, Gibson JL, Tanner EJ, McKenzie GJ, Lee PL, Gonzalez C, et al. An SOS-regulated type 2 toxin-antitoxin system. *J Bacteriol.* 2009;191: 7456–7465. doi:10.1128/JB.00963-09
 23. Christensen-Dalsgaard M, Jørgensen MG, Gerdes K. Three new RelE-homologous mRNA interferases of *Escherichia coli* differentially induced by environmental stresses. *Mol Microbiol.* 2010;75: 333–348. doi:10.1111/j.1365-

2958.2009.06969.x

24. Scanlan DJ, Ostrowski M, Mazard S, Dufresne a, Garczarek L, Hess WR, et al. Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev.* 2009;73: 249–299. doi:10.1128/MMBR.00035-08
25. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik B, et al. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 2008;9: R90. doi:10.1186/gb-2008-9-5-r90
26. Shi T, Falkowski PG. Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc Natl Acad Sci USA.* 2008;105: 2510–2515. doi:10.1073/pnas.0711165105
27. Marsan D, Wommack KE, Ravel J, Chen F. Draft genome sequence of *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay estuary. *Genome Announc.* 2014;2: e01111–13. doi:10.1128/genomeA.01111-13
28. Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J.* 2012;6: 285–97. doi:10.1038/ismej.2011.106
29. Larsson J, Celepli N, Ininbergs K, Dupont CL, Yooseph S, Bergman B, et al. Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J.* Nature Publishing Group; 2014;8: 1892–903. doi:10.1038/ismej.2014.35
30. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *AEM.* 2006;72: 2239–2243.

doi:10.1128/AEM.72.3.2239

31. Wang K, Wommack KE, Chen F. Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. *Appl Environ Microbiol.* 2011;77: 7459–7468. doi:10.1128/AEM.00267-11
32. Huang S, Wang K, Jiao N, Chen F. Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environ Microbiol.* 2012;14: 540–558. doi:10.1111/j.1462-2920.2011.02667.x
33. Wang K, Chen F. Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol.* 2008;10: 300–312. doi:10.1111/j.1462-2920.2007.01452.x
34. Sevin EW, Barloy-Hubler F. RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biol.* 2007;8: R155. doi:10.1186/gb-2007-8-8-r155
35. Shao Y, Harrison EM, Bi D, Tai C, He X, Ou HY, et al. TADB: A web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Res.* 2011;39. doi:10.1093/nar/gkq908
36. Tatusova T, Ciuffo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: New representation and annotation strategy. *Nucleic Acids Res.* 2014;42. doi:10.1093/nar/gkt1274
37. Mahmood T, Yang PC. Western blot: Technique, theory, and trouble shooting. *N Am J Med Sci.* 2012;4: 429–434. doi:10.4103/1947-2714.100998
38. Latifi A, Ruiz M, Zhang CC. Oxidative stress in cyanobacteria. *FEMS Microbiol*

- Rev.* 2009;33: 258–278. doi:10.1111/j.1574-6976.2008.00134.x
39. Brantl S. Bacterial type I toxin-antitoxin systems. *RNA Biol.* 2012;9: 1488–90. doi:10.4161/rna.23045
 40. Christensen SK, Mikkelsen M, Pedersen K, Gerdes K. RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc Natl Acad Sci USA.* 2001;98: 14328–33. doi:10.1073/pnas.251327898
 41. Lewis K. Persister cells. *Annu Rev Microbiol.* 2010;64: 357–372. doi:10.1146/annurev.micro.112408.134306
 42. Bukowski M, Rojowska A, Wladyka B. Prokaryotic toxin-antitoxin systems--the role in bacterial physiology and application in molecular biology. *Acta Biochim Pol.* 2011;58: 1–9. doi:20111979 [pii]
 43. Overgaard M, Borch J, Jørgensen MG, Gerdes K. Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity. *Mol Microbiol.* 2008;69: 841–857. doi:10.1111/j.1365-2958.2008.06313.x
 44. Li GY, Zhang Y, Inouye M, Ikura M. Structural Mechanism of transcriptional autorepression of the *Escherichia coli* RelB/RelE antitoxin/toxin module. *J Mol Biol.* 2008;380: 107–119. doi:10.1016/j.jmb.2008.04.039
 45. Boggild A, Sofos N, Andersen KR, Feddersen A, Easter AD, Passmore LA, et al. The crystal structure of the intact *E. coli* RelBE toxin-antitoxin complex provides the structural basis for conditional cooperativity. *Structure.* 2012;20: 1641–1648. doi:10.1016/j.str.2012.08.017
 46. Taylor P, Unterholzner SJ, Poppenberger B, Rozhon W. Toxin-antitoxin systems. *Bioengineered.* 2014;5: 37–41. doi:10.4161/mge.26219

47. Batiuk R, Bergstrom P, Kemp M, Koch E, Murray L, Stevenson J, et al. Chesapeake Bay submerged aquatic vegetation water quality and habitat-based requirements and restoration targets: A second technical synthesis. United States Environmental Protection Agency for the Chesapeake Bay Program. 2000. doi:10.1001/archderm.136.8.1011
48. Kemp WM, Boynton WR, Adolf JE, Boesch DF, Boicourt WC, Brush G, et al. Eutrophication of Chesapeake Bay: Historical trends and ecological interactions. *Mar Ecol Prog Ser.* 2005;303: 1–29. doi:10.3354/meps303001
49. Ting CS, Rocap G, King J, Chisholm SW. Cyanobacterial photosynthesis in the oceans: The origins and significance of divergent light-harvesting strategies. *Trends Microbiol.* 2002;10: 134–142. doi:10.1016/S0966-842X(02)02319-3
50. Franklin DJ, Brussaard CPD, Berges J a. What is the role and nature of programmed cell death in phytoplankton ecology? *Eur J Phycol.* 2006;41: 1–14. doi:10.1080/09670260500505433
51. Ohnishi N, Allakhverdiev SI, Takahashi S, Higashi S, Watanabe M, Nishiyama Y, et al. Two-step mechanism of photodamage to photosystem II: Step 1 occurs at the oxygen-evolving complex and step 2 occurs at the photochemical reaction center. *Biochemistry.* 2005;44: 8494–8499. doi:10.1021/bi047518q
52. Zeller T, Klug G. Thioredoxins in bacteria: Functions in oxidative stress response and regulation of thioredoxin genes. *Naturwissenschaften.* 2006. pp. 259–266. doi:10.1007/s00114-006-0106-1
53. Lee JW, Helmann JD. Functional specialization within the fur family of metalloregulators. *BioMetals.* 2007. pp. 485–499. doi:10.1007/s10534-006-9070-7

54. López-Gomollón S, Hernández JA, Pellicer S, Angarica VE, Peleato ML, Fillat MF. Cross-talk between iron and nitrogen regulatory networks in *Anabaena* (*Nostoc*) *sp.* PCC 7120: Identification of overlapping genes in FurA and NtcA regulons. *J Mol Biol.* 2007;374: 267–281. doi:10.1016/j.jmb.2007.09.010
55. Bagchi SN, Bitz T, Pistorius EK, Michel KP. A *Synechococcus elongatus* PCC 7942 mutant with a higher tolerance toward the herbicide bentazone also confers resistance to sodium chloride stress. *Photosynth Res.* 2007;92: 87–101. doi:10.1007/s11120-007-9176-y

Chapter 5: Phage infection of *Synechococcus* leads to massive shutdown of genomic island genes

Phage infection of *Synechococcus* leads to massive shutdown of genomic island genes

David Marsan¹, Zhangxian Xie², Dazhi Wang², and Feng Chen^{1,*}

¹Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD, USA

²State Key Laboratory of Marine Environmental Science/College of the Environment and Ecology, Xiamen University, Xiamen, 361005, China

Key words: *Synechococcus*, phage-host, genomic islands, host-like metabolic genes

5.1 Abstract

Synechococcus spp. are abundant in the marine estuarine ecosystem and subjected to frequent viral infection. What interactions occur at the molecular level during infection is not well defined. Thus, we investigated the whole transcriptome and proteome expression of the estuarine *Synechococcus* (CB0101) upon infection of podovirus S-CBP1. Strikingly, most differential expression occurred within genomic islands (GI), which were massively shutdown within 30 min post phage infection. Throughout the infection, phage-encoded host-like proteins (*hli*, *psbA*, *ThyX*) were highly expressed. Such a strong response in the GI system was not observed when CB0101 was exposed to nutrient or heavy metal stresses. Podovirus S-CBP1 seems to shut down the host GI system promptly to prevent the development of host resistance, meanwhile promoting the activities of some "shared genes" between phage and host. The global shutdown of GI genes in estuarine *Synechococcus* could be an important phage-regulated process for a successful lytic infection.

5.2 Introduction

Picocyanobacteria of the genus *Synechococcus* are ubiquitous and abundant in the world's oceans (1, 2). As a major primary producer, *Synechococcus* contributes significantly to photosynthetic carbon fixation (3, 4). Marine *Synechococcus* are genetically diverse and exhibit geographic niche adaptation (5–14). Due to *Synechococcus spp.* ecological significant and genetic diversity, over 31 genomes have been sequenced (11, 15). Genomes range in size from 2.2 to 2.9 Mbp with 52-67% of their genes (~1570) common to all strains (11, 13). These core genes are often interrupted at discrete locations by genomic islands (GIs) - large regions (more than eight kilobases long) of non-conserved, non-core genes that are sporadically distributed among members of the genus (11, 13).

Hypervariable GI regions are often differentially expressed in response to environmental stressors (16, 17) and viral infection (18). GI genes are involved in niche differentiation, such as nutrient scavenging, stress response, and phage resistance (17, 19, 20). However, they are also characterized by large numbers of novel genes of unknown function, such as those annotated as phage-like genes. The number of GIs is strongly correlated with genome size, often representing 10-31% of the total genome sequence of *Synechococcus* (11).

Synechococcus in the marine environment are subjected to frequent viral infection (21–23). Many viruses infecting marine *Synechococcus* have been isolated, and all are known to belong to three major tailed phage families (*Myoviridae*, *Podoviridae* and *Siphoviridae*) (24–28). Several lines of evidence indicate that GIs are part of the co-evolutionary process between marine cyanobacteria and their phages. Previous work with

Prochlorococcus has identified important phage-resisting machinery (20) and the dominance of up-regulated genes within GIs during podovirus infection (29). Recently, Doran et al., (2015) explored the transcript response of three oceanic *Synechococcus* strains infected by a broad host cyanomyovirus. They found within all three hosts that the vast majority of expression was located in GIs. Will estuarine *Synechococcus* follow the same GI responding pattern found in oceanic *Synechococcus* upon phage infection?

In general, *Synechococcus* living in various marine environments can be subdivided into three major subclusters, 5.1, 5.2 and 5.3 (11, 13, 30). Subclusters 5.1 and 5.3 *Synechococcus* are present in coastal and oceanic waters, while subcluster 5.2 mainly occupies estuaries. Compared to oceanic *Synechococcus*, much less is known about the ecological function and genomic evolution of estuarine species. Subcluster 5.2 *Synechococcus* dominate the estuarine environment (10) making up 20-40 % of phytoplankton chlorophyll-a during the summer (31) and contain a novel pigment cluster (32). Viruses infecting subcluster 5.2 *Synechococcus* have been isolated from the Chesapeake Bay, one such is the podovirus S-CBP1 (28), which infects the well-studied subcluster 5.2 strain *Synechococcus* CB0101 (here-after CB0101). The genome of CB0101 has been sequenced with a genome size of 2.8Mbp and 3,176 genes, of which 343 are GI genes (33).

In this study, we analyzed the genome wide mRNA and protein expression of subcluster 5.2 strain CB0101 during the infection of a lytic phage, S-CBP1, in tandem. Surprisingly, massive down-regulation of host GI genes occurred upon phage infection. Among 129 differentially expressed GI genes, 83 % were down regulated 30 min post

phage infection. This type of rapid and extensive shutdown of host GI systems during infection has not been previously reported.

5.3 Materials and Methods

5.3.1 Synechococcus culture growth

Synechococcus CB0101 was grown at 23 °C under 15 $\mu\text{E m}^{-2} \text{s}^{-1}$ continuous light in SN medium with 15 ppt salinity and vitamin B12 (referred to as SN15 medium hereafter). Filtered (0.2 μm pore size) air was bubbled into individual 500 mL cell culture flasks. Six liters of CB0101 culture were grown to exponential phase at a density of 1×10^8 cells/mL as determined by counting use an Accuri C6 flow cytometer. The culture was then split into 500 mL cell culture flasks to begin the phage host experiments.

5.3.2 Sampling time point determination: One-step growth curve

S-CBP1 growth, burst size, and latent time has previously been characterized by (28), building upon this analysis a one-step growth curve was performed before the RNA-Seq and proteomic analysis to determine sampling points. The one-step growth curve was measured following the method described by Jiang and colleagues (34) with some modifications. In our modified protocol, flow cytometry counting of released viral particles, instead of conventional plaque assay method, was used to avoid plating and shorten the observation time. Phage lysates (approximately 10^9 PFU ml^{-1}) were inoculated into the exponentially growing host cultures (approximately 1×10^8 cells ml^{-1} , with a doubling time of 24 h, as measured by cell counts) with a multiplicity of infection of 1.2-1.5 for 1 h at 25 °C. The control received the same amount of microwave-killed phage lysates. The mixture was then diluted in fresh SN15 media by 100-fold to

minimize the further adsorption of phage to host cells. After inoculation (taken as T0), a subsample of suspension was withdrawn from each culture periodically for up to 6 days. The released virus-like particles (VLPs) were counted using the Accuri C6 flow cytometer. The latent period was estimated by the time interval of first wave of significant increase in VLPs (3-fold increases as threshold) observed in the samples and found to be 8 h for S-CBP1. From these results the sampling time points of 0.5, 5, and 12 h were selected to roughly correlate to attachment, latent, and cell lysis.

From the one-step growth curve samples time points were determined to best capture the attachment, latent and rise period of infection. 25 mL was sampled at 0.5, 5, and 12 h to roughly correlate with our desired infection periods. Triplicate cultures of CB0101 were grown to exponential phase (10^8 cells ml^{-1}), and then infected with S-CBP1 podovirus (10^8 infective phage particles ml^{-1}) for a multiplicity of infection (MOI) of 1. Triplicate uninfected control cultures were amended with 50mL of microwaved killed S-CBP1 in SN15 medium and sampled at designated time points. Counting of *Synechococcus* cells and viruses were conducted using the above described flow cytometry method throughout the infection period 0 to 60 h.

5.3.3 RNA-Seq

At the designated time points 25 mL samples were collected and immediately processed for mRNA and protein analysis. Samples were centrifuged at 10,000 rpm then separated using Trizol extraction for RNA and protein. mRNA was extracted from the total RNA using Ambion MICROBExpress kit (Life technologies, AM1905), modified with specific primers for rRNA depletion of CB0101 ribosome. rRNA removal evaluation was conducted using the Agilent 2100 Bioanalyzer resulting in 98 % removal

of rRNA and confirmed the presence of high quality RNA. The NEXTflex RNA-Seq kit (BIOO Scientific, #5129-01) was then used to prepare mRNA libraries for sequencing using the Illumina Hi-Seq. RNA libraries were created and barcodes assigned to each sample for multiplexing of a single Hi-Seq lane. RNA-Seq sequencing was carried out by the Institute of Genome Sciences (IGS) using a Hi-seq with 100bp paired-end read length. Resulting in 144,403,530 read sequences. RNA-seq and exhaustive statistical analysis were carried out using the CLC platform, R, Bioconductor, top-hat/cufflinks, described in detail below (Supplemental Table 5.S1 and Fig. 5.S1-5.S3).

The 100bp reads were aligned separately for each sample to the reference genomes of CB0101 and S-CBP1, using bowtie2. Reads from the whole-transcriptome library were counted for each gene, normalized to uninfected control host transcript levels and to gene length. The normalized data were used for clustering the phage genes by their expression profile and are shown in the supplemental figures 5.S1-5.S11. A summary of the RNA-sequencing data (number of reads, number of mapped reads and so on) is presented in Supplementary Table 5.1.

5.3.4 Normalization of read counts, calculation of RPKM and heat map representation

Read counts of phage and host were normalized separately using the R software package *edgeR* in Bioconductor (35). The results matrices were used for calculating the RPKM values (36) for all genes and for differential expression analysis.

5.3.5 Sample Clustering

After normalization and calculation of RPKM, host-infected and control samples were hierarchically clustered via Pearson's correlation using the *pvcust* package (37) in R. Node support was calculated through multi-scale resampling (10,000 bootstraps).

5.3.6 Differential Expression Analyses

Differential expression was calculated between infected and control host samples at every time point, between time points of host control samples and between time points of phage infected samples. Statistical package *edgeR* (35) was used and genes with a false discovery rate <0.01 were considered as differentially expressed. Further, genes were “up-regulated” if normalized RPKM >2 and “down-regulated” if <-2 .

5.3.7 Functional group assignment to differentially expressed host genes

Host genes found differentially expressed between infected and control samples at every time point were placed into one of 14 functional categories based on MG-RAST (38) by grouping similar functions into the same category using R.

5.3.8 Protein extraction and mass spectrometry analysis

Proteins were extracted using the same sample for RNA extraction. The bottom layer containing proteins was transferred to a new tube and follow the extraction procedure previous described (39). Protein quantification was carried out according to the instruction of 2D Quant kit (GE Healthcare, 80-6482-56). Equivalent of proteins (200 μg) in each sample were loaded on a prepared SDS-PAGE for in gel digestion (40). The generated tryptic peptides were extracted with solvent containing 50 % acetonitrile (ACN) and 0.1 % trifluoroacetic acid, and then vacuum dried using a SpeedVac for further mass spectrometry (MS) analysis.

Samples were introduced into a Waters nanoACQUITY UPLC using a C18 trap column (1.7 μm particle size, 100 μm x 100 mm internal diameter) fitted with a C18 trap column. (5 μm particle size, 180 μm x 20 mm internal diameter). Peptides were eluted at a flow rate of 300 nl/min using an acidified (formic acid, 0.1 % v/v) water-acetonitrile

gradient (2%-35% acetonitrile in 135 min, full-time 185 min). The peptides were subjected to nanoelectrospray ionization followed by tandem mass spectrometry (MS/MS) in a Q EXACTIVE (ThermoFisher Scientific, San Jose, CA) coupled online to the UPLC. Data-dependent scans were conducted by precursor ion selection in the Orbitrap followed by high-energy collision induced dissociation. One full m/z scan between 350 and 2,000 Da was followed by 15 MS/MS scans with a following Dynamic Exclusion duration of 15 s. The electrospray voltage applied was 1.6 kV and the heated capillary was at 280 °C. Automatic gain control (AGC) was used to optimize the spectra generated by the Orbitrap.

5.3.9 Protein identification

The instrument data files (.raw) were interpreted using Proteome Discoverer (ver. 1.4.0.288; ThermoFisher Scientific, San Jose, CA). MS/MS spectra were searched with SEQUEST engine against a protein database containing 15,527 translated protein-coding sequences predicted from genome of *Synechococcus* CB0101 and its phage genome as well as 6,774 contaminated bacterial sequences and 247 contaminant sequences such as keratin and trypsin. Database searching was restricted to tryptic peptides. For protein identification, a mass tolerance of 10 ppm was permitted for intact peptide masses and 0.05 Da for fragmented ions, with allowance for one missed cleavages in the trypsin digests and carbamidomethyl (C) as the fixed modification, oxidation (M) as the variable modification. Peptides having a Delta Cn better or equal to 0.05 were using for Percolator. Peptide spectral matches were further validated based on q-value of decoy database search at a level of 1 % false discovery rate (FDR). Peptide identifications were grouped into proteins with Proteome Discoverer according to the law of parsimony and

filtered to 1 % FDR. Finally, proteins matched with at least two peptides were accepted as confident identifications for further bioinformatics analysis.

5.3.10 Functional annotation and semi-quantitative analysis

Functional annotations of identified proteins were derived from their annotations of RNA or DNA sequences. For semi-quantitative analysis (41), spectral counts were normalized by the total number of spectra identified in each sample as well as by the length of the matching protein. The result values were used for statistical significance test and represented as protein expression level.

5.3.11 Genomic Island type gene identification

Genomic island's and genes were identified using the IslandViewer 3 website and interface (42). Following the directions described by the developers CB0101 genome was uploaded for analysis. The resulting information identified 343 potential GI genes present in the genome, with at 4 GI regions confirmed.

5.4 Results

5.4.1 Infection Cycle

When CB0101 was challenged with podovirus S-CBP1, cell counts began declining 8 h after infection, ultimately reaching complete lysis at 48 h. Viral counts began to increase 8 h after infection with progeny first released (Fig 5.1). mRNA and protein samples were taken at 30 min, 5 h and 12 h, respectively, representing attachment, latent, and maturation of phage infection cycle. Statistical analysis confirmed that extensive coverage of both CB0101 and S-CBP1 mRNA were accomplished with 15 to 3 million (~7x coverage) reads mapping to CB0101 and 1.5 to 12 million (~240x

coverage) to S-CBP1. Reproducibility between duplicates were statistically significant (Supplemental Table 5.S1 and Fig. 5.S1-5.S3).

5.4.2 Host response to phage infection

CB0101 contains a total of 3,176 known genes including 2,833 core and 343 GI genes. At 30 min post infection, 5 % (131 genes) of core CB0101 genes were down-regulated, and 3 % (77 genes) were up-regulated as compared to the baseline uninfected control sample (Table 5.1). A continuation of down-regulation was observed 12 h post infection, 9 % (268 genes) of CB0101 core genes while, 2 % (67 genes) were up-regulated ($p < 0.01$ and minimum 2 fold) (Table 5.1). The general trend over the infection is more genes being down-regulated while those that are up-regulated stay constant from 30 min to 12 h of infection (Fig 5.2A). 31 % (107 genes) of all GI genes were down-regulated at 30 min post infection, while just 6 % (22 genes) were up-regulated (Table 5.1, Fig. 5.2B). The GI genes were then progressively down-regulated as the infection continued to 12 h, accounting for 69% (235 of 343) of all differentially expressed genes (Table 5.1). When differentially expressed genes are filtered into clusters of orthologous groups (COG), the massive down-regulation of GI genes evident (Fig 5.2B).

Interestingly, 20 of the GI genes were constantly up-regulated throughout the entire infection. Among these 20 GI genes, some can be identified and included those coding for high-light-inducible stress response (*hli*), transcription (*rpoC1*, *rpoA*), ribosome (*rplO*), translocase (*secF*, *secD*), DNA-binding (*stpA*), menaquinone (menB; electron transport), enolase (*eno*), accelerated folding of proteins associated with photosynthesis (*TLP40*), sulfide quinone reductase (SQRDL: electron acceptor), and transcriptional regulator (*vidC*).

For the core genes of CB0101, protein metabolism transcripts increased during the early (30 min) infection but gradually decreased (5 h and 12 h) (Fig 5.2B). These protein metabolism genes are almost exclusively ribosomal subunits responsible for translation. Other COG categories that were continuously up-regulated were associated with energy production (i.e. CO₂ fixation, ATP synthesis) (Fig 5.2B). In contrast, a set of photosynthetic genes was progressively down-regulated (photosystem II associated genes *psbL*, *psbD*, *psbZ*, *psbJ*, *psbL*, *psbK*, *psbI*, *psbY*, and *psbX*).

At the proteomic level, a total of 825 proteins (26% of all predicted) from CB0101 were identified (Fig. 5.3A). At 30 min, 435 proteins were identified, with 112 (4%) core proteins being differentially expressed (Table 5.1). At 5 h post infection, 397 proteins were found, and 60 (2%) were differentially expressed. At 12 h post infection, 362 proteins were identified, and 52 (14%) were differentially expressed. Overall, the downward trend in the number of identified proteins from 30 min to 12 h can also be seen in the COG data (Fig 5.3B). The GI proteins (Hli, RpoC1, RpoA, RplO, SecF, SecD, StpA, MenB, Eno, TLP40, SQRDL, and VidC) were up-regulated (by 2 fold) during the entire lytic cycle. Because large portions of GI genes were down-regulated at the mRNA level, proteins of these GI genes were not detectable due to low abundance or from rapid degradation.. Proteins associated with energy production and various forms of metabolism (i.e. CO₂ fixation and ATP synthesis) dominated as the most abundant up-regulated protein categories, coinciding with the transcript response (Fig 5.2B and 5.3B).

5.4.3 Phage response during infection

Phage S-CBP1 was linearly transcribed resulting in a lack of differentially expressed transcripts ($p < 0.01$ and minimum 2 fold). However, expression clusters can be determined, showing attachment, replication, and maturation stages of infection (Supplemental Fig 5.1-5.3). Nevertheless, differential expression at the protein level was observed, likely due to differences in folding speed (Fig 5.4A). We were able to detect proteins from 55% of the phage genome throughout the infection. Of the 28 phage proteins identified, 12 were differentially expressed (Fig 5.4A), while the remaining 16 were relatively constant over the infection. Three proteins, *hli*, hypothetical protein 15, and *psbA* ranked among the most abundant in the phage proteome (Fig 5.4B). These proteins clustered together in the genome of S-CBP1 at a location labeled “hot spot” (Fig 5.4C). Of the 2,265 phage protein counts, 68-78% belonged to the “hot spot” region proteins (Fig 5.4B & 5.4C). These proteins were constant throughout the entire infection. Moreover, the “hot spot” proteins including ssDNA and ThyX are host-like genes with homologues found in the genome of CB0101.

5.5 Discussion

5.5.1 Rapid shutdown of host GI system

This study is the first to explore the transcriptional and proteomic program of a subcluster 5.2 *Synechococcus* infected by a podovirus. Our findings indicate that GI genes in estuarine CB0101 are far more involved in lytic infection than the host core genes. At 30 min post infection, 129 of 337 differentially expressed transcripts were from GI regions. Among them, 83 % (107 out of 129 genes) were down regulated (Table 5.1). The transcript data is reinforced by the proteomic expression, where a lack of GI proteins is observed (Fig 5.3B), indicating that GI response could be similar to an on-off switch

rather than a gradual change at the protein level. The percentile (83 %) of down-regulated GI genes is much higher than those (35-51%) reported from the responses of marine *Synechococcus* WH7801, WH8102 and WH8109 to the infection of myovirus Syn9 (18). The marked difference in GI gene response to phage infection between estuarine and open ocean *Synechococcus* strains was unexpected. It is unclear why GI genes in estuarine CB0101 are massively shutdown in response to the infection of a podovirus. Such a dramatic response in the GI system was also not observed when marine *Prochlorococcus* MED4 was infected by podovirus P-SSP7 (29).

One potential explanation could be the genomic difference between estuarine (subcluster 5.2) and open ocean *Synechococcus* (subcluster 5.1). It has been known that the number of GI genes varies greatly among *Synechococcus* strains sequenced (11). Estuarine strains like CB0101, CB0205, and WH5701 have relatively large genomes and contain more GI genes than their open ocean counterparts. It is thought that the increased number of GI genes aids in sensing and adapting to the rapid changes for each specific ecotype. Rapid shutdown of GI genes in estuarine CB0101 could be a function that lytic podovirus S-CBP1 use to quickly turn off the GI system to reduce host resistance ability. Further study is needed to understand whether such a GI shutdown during infection is a common feature among estuarine *Synechococcus* or other subcluster 5.2 species.

Another explanation for the unexpected massive shutdown of GI genes could be the superior detection of low abundance transcripts and dynamic range of RNA-seq. This broader dynamic range allows for the detection of more differentially expressed genes (43). This means for RNA-seq, if you sequence deep enough, you can get the same dynamic range as the actual RNA molecules in the sample. While most research papers

have >10 million mapped reads/sample on average, RNA-seq should provide higher sensitivity than microarray, thus allowing for a much clearer picture of whole genome expression.

GI residing genes are often differentially expressed in response to environmental stresses (16, 17, 44). However, in a separate experiment when CB0101 was exposed to nutrient starvation or oxidative stress, only a few genes within the GI regions were differentially expressed (Table 5.2). These results suggest that the GI systems respond differently to phage infection and to these environmental stresses, at least for estuarine CB0101.

5.5.2 Up-regulation of GI genes and proteins

Despite the massive shutdown, 20 GI genes remained up-regulated at the transcriptomic and proteomic level throughout the infection period. These GI genes were coded for high-light-inducible stress response (*hli*), transcription (*rpoCI*, *rpoA*), ribosome (*rplO*), translocase (*secF*, *secD*), DNA-binding (*stpA*), menaquinone (*menB*; electron transport), enolase (*eno*), accelerated folding of proteins associated with photosynthesis (*TLP40*), sulfide quinone reductase (*SQRDL*: electron acceptor), transcriptional regulator (*vidC*). Of these up-regulated genes, similar homologues were also found expressed within *Prochlorococcus* and other *Synechococcus* GI regions (18, 29).

Perhaps the most compelling evidence for host GI genes being linked to a phage infection tactic is the numerous homologues of host-like genes found in the phage genome. The homologues of *hli*, *psbA*, *ssDNA*, and *ThyX* can be found within GI regions, which are expressed during infection of CB0101 (Fig 5.2 and 5.3). Furthermore, homologues of similar host genes are prevalent in phage genomes infecting

Synechococcus and *Prochlorococcus* spp. (16, 45). Thus, there seems to be a connection between genes up-regulated during infection, their position in the host genome, and the presence of homologues in phage for both *Synechococcus* and *Prochlorococcus*.

Additionally, not only did these same patterns occur in two different host genera, but also in response to infection by phages belonging to two very different phage families, T7-like (29) and T4-like cyanophages (18). It is clear that GI genes play an important role during cyanophage infection(13, 18, 20, 29). We hypothesize that it is possible some of these host responses reflect an attempt at defense against infection, as inferred from the identity of some of the response genes. Alternatively, these genes may be induced by the phage to aid in the replication process, with the immediate response induced by phage proteins entering into the cell together with phage DNA.

5.5.3 Host core gene and protein response

Transcript and protein COG categories display how evidently the host metabolic processes shifted towards phage needs over time (Fig. 5.2B and 5.3B). Upon phage infection (30 min), we observed a much lower percentage of host core gene expression compared to the GI genes. However, the up-regulation of genes associated with phage reliance is evident by increased host ribosomal subunit expression within 30 min of infection. The expression of transcripts and proteins for ribosomal subunits peaked at 5 h and declined during the maturation phase at 12 h. Such a response pattern is consistent with the classical expression phases for other T7-like phage (46). Other transcripts and proteins up-regulated at all time points were associated to energy production (i.e. CO₂ fixation, ATP synthesis). In contrast, a set of photosynthesis genes were progressively down-regulated over the time series, specifically photosystem II associated genes *psbL*,

psbD, *psbZ*, *psbJ*, *psbL*, *psbK*, *psbI*, *psbY*, and *psbX*. This is particularly of note because phage replication is a function of host growth rate, which has been linked with a dependence on photosynthesis(47). A decline in host photosystem II transcript and protein could be caused by the slow shutdown owing to cell lysis or a factor of phage photosystem proteins replacing/supplementing host factors.

Until about a decade ago, it was generally thought that phage infection led to a complete shutdown of host transcription (48). However, the employment of whole-genome transcriptional analyses to assess responses to infection has shown this not to be the case, with a range of expression in different host-phage systems (18, 29, 49, 50) (this study). First, common across all systems is an immediate brief expression followed by a subsequent response to infection. Furthermore, some but not all bacteria experience a massive decline in host gene expression upon infection. However, our findings can be generalized across subcluster 5.2 *Synechococcus* (this study), marine *Synechococcus* (18), and *Prochlorococcus* (29), with both transient increases in transcript levels accompanied by an overall decline in host expression.

5.5.4 Phage responses

The transcriptomic data shows that the genome of phage S-CBP1 is linearly transcribed, resulting in no differential expression at the mRNA level. Rather, the proteomic data displays a robust response with a set of three highly expressed proteins. Interestingly, a cluster of three phage genes, the “hot spot” region (as described in this paper), made up of *hli*, *psbA*, and a hypothetical protein, were the most active proteins identified throughout the entire infection (Fig 5.4B). The “hot spot” region made up 68-78% of the total protein counts and was constant through the attachment, replication, and

maturation phases. These important phage genes are thought to be significant for continued host photosynthetic activity, production of ATP, and reducing equivalents needed for phage DNA replication and nucleotide biosynthesis (18, 29, 47, 51).

S-CBP1 protein abundance was highly skewed towards host-like genes, specifically Hli, PsbA, ssDNA, and ThyX (Fig 5.4B). The photosynthesis genes found in this cluster are thought to be involved in the production of energy, while ThyX is an alternative type of thymidylate synthase which synthesizes the essential DNA precursor, thymidylate (dTMP), from uridylate (dUMP) (52). Among the sequenced cyanophage genomes, *thyX* is only found in cyanopodoviruses, which infect *Synechococcus* (53). We hypothesize that the “hot spot” cluster and *thyX* in podovirus S-CBP1 are involved in redirection of host metabolism. Together, these findings suggest that these genes form a functional unit to produce energy and deoxynucleotide carbon substrates necessary for cyanophage DNA replication.

The genome content and architecture of the cyanophage S-CBP1 are similar to that of the *Escherichia coli*-infecting T7 podovirus. As in T7, the S-CBP1 genome was transcribed linearly (46) from left to right of the genome map (Fig 5.4C), with three distinct infection phases discerned (Supplemental Fig 5.1-5.7). The first phase includes the *marR* gene, which suggests a role in redirecting transcription from the host towards the phage. The second phase is the expression of genes participating in DNA metabolism and replication, which allow for RNA transcription and/or DNA replication. The third phase consists of genes involved in phage particle formation and DNA maturation. Thus, the three expression phases in S-CBP1 are analogous to T7 and T4-like phage including those infecting *Synechococcus*'s sister genera, *Prochlorococcus* (18, 29).

Regardless of the mechanism of shutdown, we hypothesize that S-CBP1 has evolved to not only make use of host up-regulated proteins (such as RpoC1, RpoA, RplO, SecF, SecD, StpA, Eno, TLP40, SQRDL, and VidC), but also shutdown potential host defenses within GIs. Certainly, phages are known to exploit host stress-response proteins during infections (54, 55). As well, *Prochlorococcus* have shown phage resistance through shuffling of genes within GI regions (20). We believe that rapid response, within the first 30 mins, to phage infection is common among marine cyanobacteria ((18, 29); this study).

The presence and expression of host-like genes that are involved in the maintenance of photosynthesis, phosphate stress, and mobilization of carbon stores during infection is not a new observation (16–18, 29, 44, 45). However, a central question as to what type of benefit and why they are linked to host GI regions remains unknown. We do know that a statistically significant number of phage host-like genes have consistently been shown to be present in GI regions (16–18, 29, 44). It could be that phages have evolved to use these gene products to gain a fitness advantage. Incorporation of these genes into their own genomes would enable phage to regulate their expression, conferring a fitness advantage, and leading to preferential retention. This retention would increase the probability of transfer back to the host in genome islands, unsuccessful infection, and those genes beneficial to the host would remain in the host genome. Analysis of CB0101's GI regions provides an interesting illustration in support of this scenario. GI genes possibly associated with host defense are down-regulated, while GI gene homologues to phage “host-like” genes are up-regulated in response to phage

infection aiding the host metabolic shift towards phage production. Further work should be conducted to see how GI genes and host-like genes in phages could be linked.

5.6 Conclusion

Estuarine *Synechococcus* strain CB0101 (subcluster 5.2) exhibited a strong shutdown of GI genes in response to infection by podovirus S-CBP1. It is striking to see such a massive shutdown of host GI genes upon infection, compared to a relatively small portion of down-regulation in the non-GI genes. Although GIs are known to be involved in response to environmental stress, we did not observe such a dramatic down regulation when CB0101 was grown under nutrient starvation or exposed to heavy metal toxicity. It could be that podovirus S-CBP1 is able to prevent the development of host resistance by quickly shutting down the GI system as is seen in other T7-like phage.

At the transcriptional level, phage S-CBP1 was linearly transcribed, but three distinct expression clusters (attachment, replication, and maturation) were observed. Several phage-encoded 'host-like' genes (e.g. *hli*, *psbA*, and *thyX*) were highly expressed at the protein level. These "hot spot" phage genes are often found in the GI region, suggesting a mutual benefit for cyanophage and cyanobacteria to maintain ecologically important genes. Studying the role of GI genes in response to phage infection is an important step towards understanding the ecological interaction and co-evolution between cyanobacteria and cyanophage.

Table 5.1. Transcript and protein counts of differentially expressed total core (2,833 genes, core genes are any gene not contained in a genome island) and genomic islands (343 genes) ($p < 0.01$ and minimum 2 fold). Identified proteins were 435, 397, and 362 for 30min, 5h, and 12h, respectively.

	Transcriptomics			Proteomics		
	0.5h	5h	12h	0.5h	5h	12h
Total core responsive genes/proteins	208 (7%)	340 (12%)	335 (12%)	112 (4%)	60 (2%)	52 (2%)
Core up-regulated genes/proteins	77 (3%)	132 (5%)	67 (2%)	106 (4%)	52 (2%)	49 (2%)
Core down-regulated genes/proteins	131 (5%)	208 (7%)	268 (9%)	6 (0.2%)	8 (0.2%)	3 (0.1%)
Total responsive GI genes/proteins	129 (38%)	236 (69%)	255 (74%)	21 (6%)	10 (3%)	9 (3%)
# of up-regulated GI genes/proteins	22 (6%)	54 (16%)	20 (6%)	18 (5%)	7 (2%)	8 (2%)
# of down-regulated GI genes/proteins	107 (31%)	182 (53%)	235 (69%)	3 (1%)	3 (1%)	1 (1%)

Table 5.2. CB0101 Genomic island residing gene transcript and protein differentially expressed under different stressors (p<0.01 and minimum 2 fold).

	Viral Infection	Nitrogen Deplete	Phosphate Deplete	Zinc Toxicity
Total responsive GI transcripts	255	74	25	92
Total responsive GI proteins	21	20	10	N/A
# of up-regulated GI transcripts	54	35	15	43
# of up-regulated GI proteins	18	19	7	N/A
# of down-regulated GI genes	235	39	10	49
# of down-regulated GI proteins	3	1	-	N/A

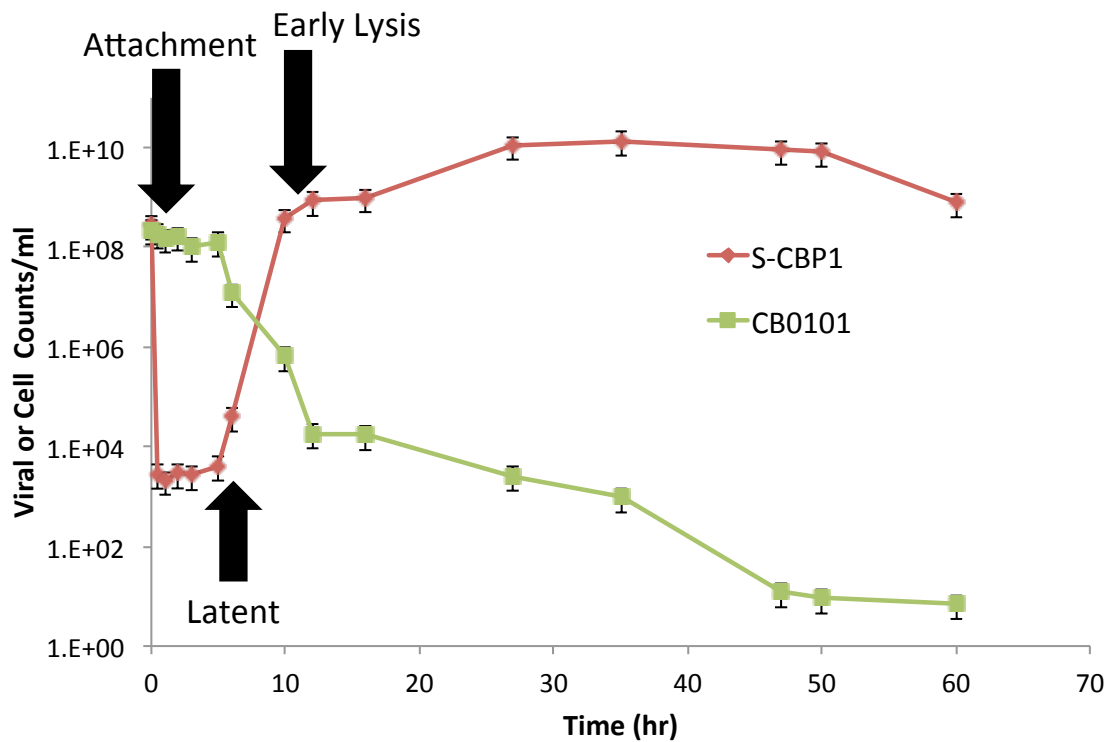


Figure 5.1. Infection dynamics of the S-CBP1 phage in an estuarine *Synechococcus* CB0101 host. Timing of phage replication was determined by one-step growth curve, cell and viral counts. mRNA and protein were measured at 0, 30 min, 5 h, and 12 h to correspond to attachment, latent, and early lysis phases. Total host lysis occurs within 48 h of infection.

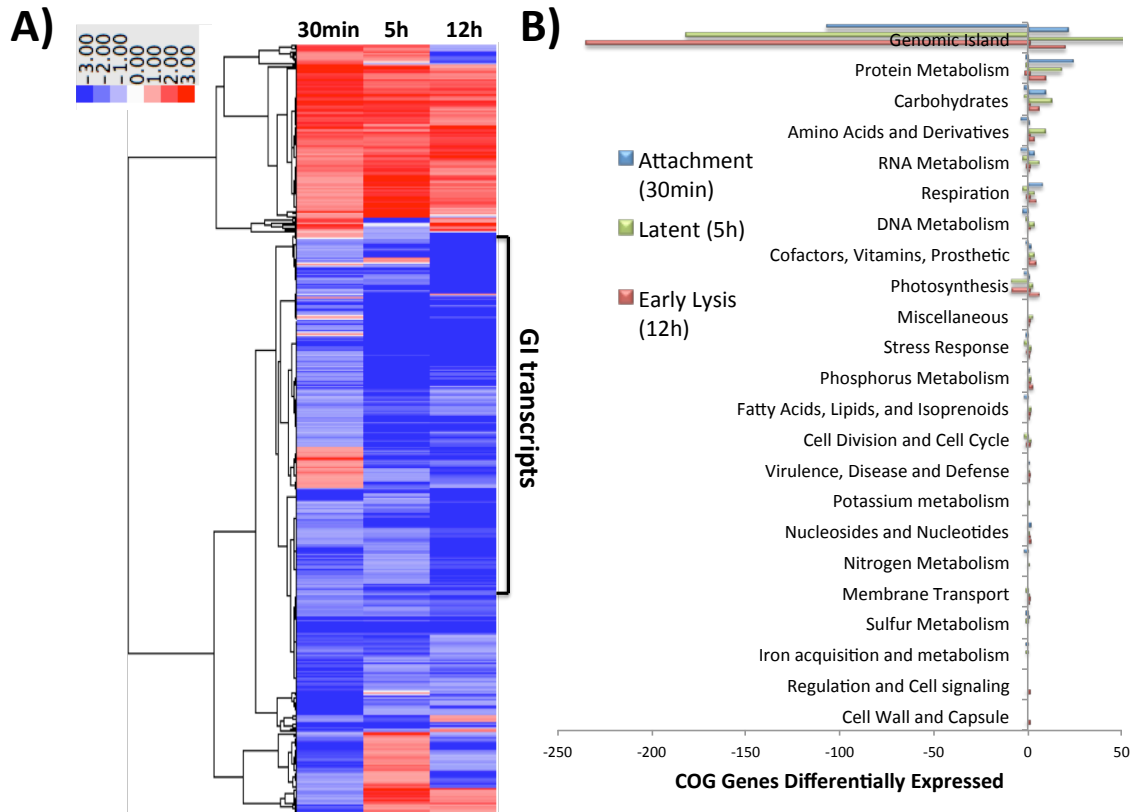


Figure 5.2. The transcript expression program of *Synechococcus* CB0101. (A) Heat map and clustering of the 590 unique differentially expressed genes by their expression patterns ($p < 0.01$ and minimum 2 fold). Dark red equates to a strong increase (>3 fold), while dark blue to a strong decrease (<-3 fold) in gene expression. (B) Differentially expressed genes filtered into 23 COGs. X-axis denotes the number of unique genes that are either up-regulated (positive side of graph) or down-regulated (negative side of graph).

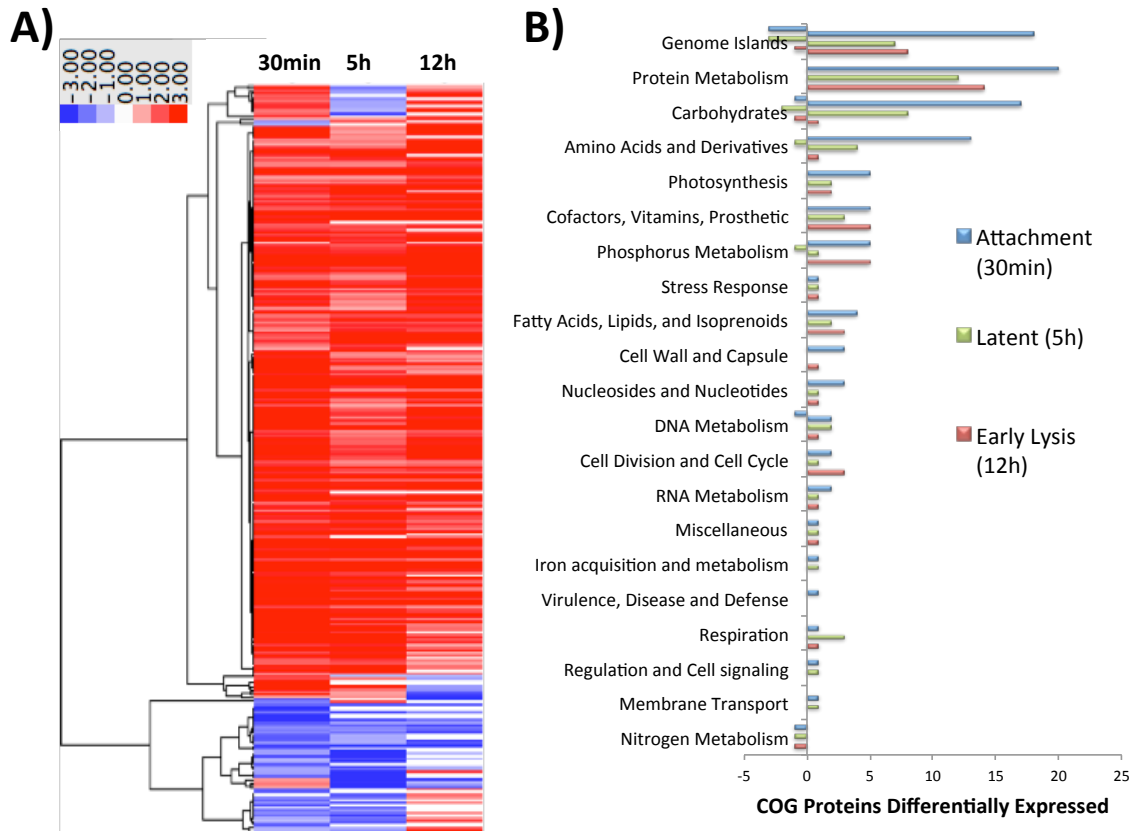


Figure 5.3. The protein expression program of *Synechococcus* CB0101: (A) Heat map and clustering of the 133 unique differentially expressed proteins by their expression patterns ($p < 0.01$ and minimum 2 fold). Dark red equates to a strong increase (>3 fold), while dark blue to a strong decrease (<-3 fold) in protein expression. (B) Differentially expressed proteins filtered into 21 COGs. X-axis denotes the number of unique genes that are either up-regulated (positive side of graph) or down-regulated (negative side of graph).

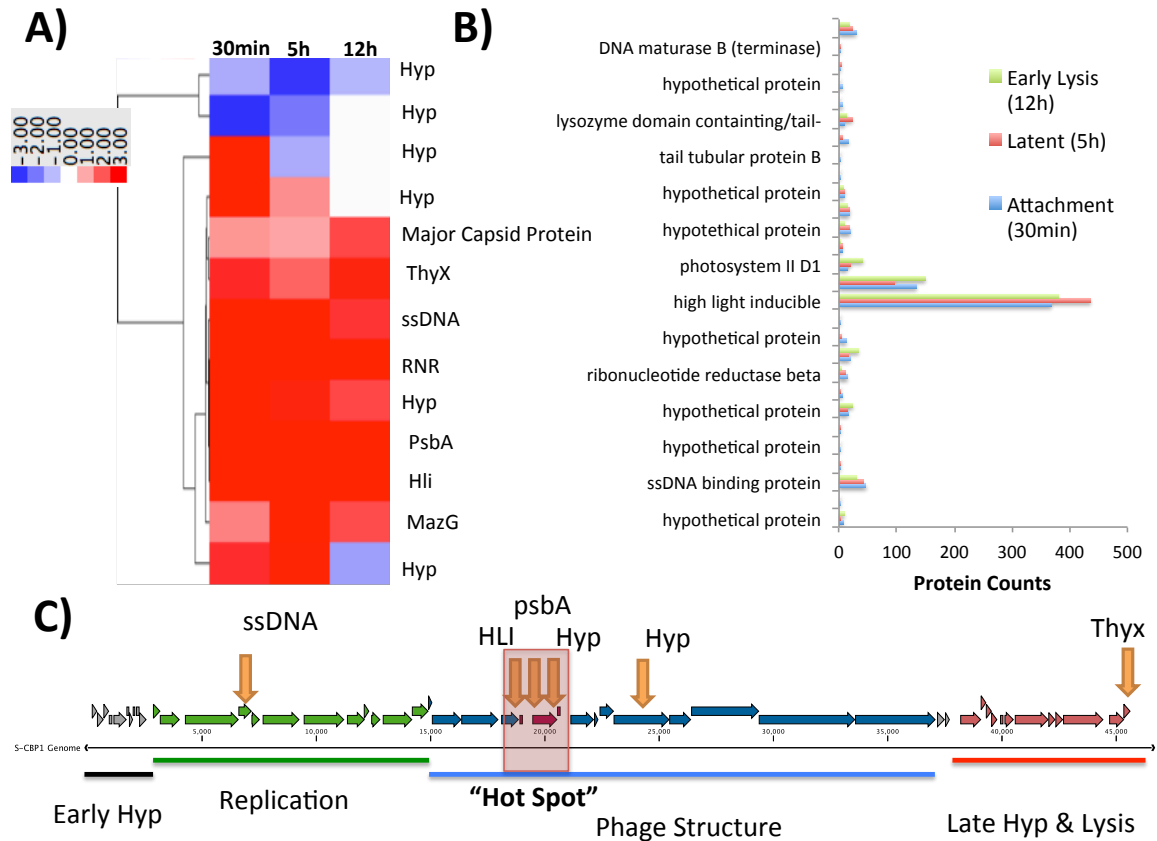


Figure 5.4. Temporal expression dynamics of S-CBP1 phage proteins during infection of *Synechococcus* CB0101. (A) Heat map and clustering of differentially expressed proteins by their expression patterns as compared to T0. (B) Protein counts mapping back to each individual phage protein over the three sampling points. (C) Genome map highlighting the location of the “hot spot” at the center of the genome and the other most expressed phage proteins.

References

1. Johnson PW, Sieburth JM (1979) *Chroococcoid* cyanobacteria in the sea: A ubiquitous and diverse phototrophic biomass. *Limnol Oceanogr* 24(5):928–935.
2. Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979) Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293–294.
3. Li WKW (1994) Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: Measurements from flow cytometric sorting. *Limnol Oceanogr* 39(1):169–175.
4. Richardson TL, Jackson G a (2007) Small phytoplankton and carbon export from the surface ocean. *Science* 315(5813):838–40.
5. Toledo G, Palenik B (1997) *Synechococcus* diversity in the California Current as seen by RNA polymerase (*rpoC1*) gene sequences of isolated strains. *Appl Environ Microbiol* 63(11):4298–4303.
6. Rocap G, Distel DL, Waterbury JB, Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68(3):1180–1191.
7. Fuller NJ, Marie D, Vaultot D, Post AF, Scanlan DJ (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* 69(5):2430–2443.
8. Ahlgren NA, Rocap G (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N

- physiologies. *Appl Environ Microbiol* 72(11):7193–7204.
9. Penno S, Lindell D, Post AF (2006) Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* 8(7):1200–1211.
 10. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* 72(3):2239–2243.
 11. Dufresne A, et al. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9(5):R90.
 12. Zwirgmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
 13. Scanlan DJ, et al. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 73(2):249–299.
 14. Huang S, et al. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 6(2):285–97.
 15. Federhen S (2015) Type material in the NCBI Taxonomy Database. *Nucleic Acids Res* 43(1):1086–1098.
 16. Thompson LR, et al. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci* 108(39):E757–E764.
 17. Stuart RK, Brahamsha B, Busby K, Palenik B (2013) Genomic island genes in a coastal marine *Synechococcus* strain confer enhanced tolerance to copper and oxidative stress. *ISME J* 7(6):1139–49.

18. Doron, S; Fedida A (2016) Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* 10:1437–1455.
19. Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 19(12):2226–2238.
20. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011) Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* 474(7353):604–608.
21. Proctor LM, Fuhrman J a. (1990) Viral mortality of marine bacteria and cyanobacteria. *Nature* 343:60–62.
22. Suttle C a., Chan AM, Cottrell MT (1990) Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* 347(6292):467–469.
23. Mann NH, et al. (2003) Bacterial photosynthesis genes in a virus. *Nature* 424:741–742.
24. Waterbury JB, Valois FW (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* 59(10):3393–3399.
25. Wilson WH, Joint IR, Carr NG, Mann ' NH (1993) Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol* 59(11):3736–3743.
26. Lu J, Chen F, Hodson RE (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. *Appl Environ Microbiol* 67(7):3285–3290.
27. Marston MF, Sallee JL (2003) Genetic diversity and temporal variation in the

- cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol* 69(8):4639–4647.
28. Wang K, Chen F (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* 10(2):300–312.
 29. Lindell D, et al. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449(7158):83–86.
 30. Huang S, Wilhelm SW, Jiao N, Chen F (2010) Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *ISME J* 4(10):1243–1251.
 31. Wang K, Wommack KE, Chen F (2011) Abundance and distribution of *Synechococcus spp.* and cyanophages in the Chesapeake Bay. *Appl Environ Microbiol* 77(21):7459–7468.
 32. Larsson J, et al. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 8(9):1892–903.
 33. Marsan D, Wommack KE, Ravel J, Chen F (2014) Draft Genome Sequence of *Synechococcus sp.* Strain CB0101, Isolated From the Chesapeake Bay Estuary. *Genome Announc* 2(1):e01111–13.
 34. Jiang SC, Kellogg CA, Paul JH (1998) Characterization of marine temperate phage-host systems isolated from Mamala Bay, Oahu, Hawaii. *Appl Environ Microbiol* 64(2):535–542.
 35. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–40.

36. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628.
37. Suzuki R, Shimodaira H (2006) Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12):1540–1542.
38. Glass EM, Meyer F (2011) The metagenomics RAST server: A public resource for the automatic phylogenetic and functional analysis of metagenomes. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, pp 325–331.
39. Wang DZ, et al. (2011) Homology-driven proteomics of dinoflagellates with unsequenced genomes using MALDI-TOF/TOF and automated de Novo sequencing. *Evidence-based Complement Altern Med* 2011. doi:10.1155/2011/471020.
40. Dong H-P, Wang D-Z, Dai M, Chan LL, Hong H-S (2009) Shotgun proteomics: Tools for analysis of marine particulate proteins. *Limnol Oceanogr Methods* 7:865–874.
41. Knief C, et al. (2012) Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J* 6(7):1378–1390.
42. Dhillon BK, et al. (2015) IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res* 43(W1):W104–8.
43. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9(1). doi:10.1371/journal.pone.0078644.
44. Coleman ML, et al. (2006) Genomic islands and the ecology and evolution of

- Prochlorococcus*. *Science* 311(5768):1768–1770.
45. Sullivan MB, Coleman ML, Weigle P, Rohwer F, Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol* 3(5):0790–0806.
 46. Molineux I (2005) The Bacteriophages. *Oxford University Press*:277–301.
 47. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438(7064):86–9.
 48. Roucourt B, Lavigne R (2009) The role of interactions between phage and bacterial proteins within the infected cell: A diverse and puzzling interactome. *Environ Microbiol* 11(11):2789–2805.
 49. Poranen MM, et al. (2006) Global changes in cellular gene expression during bacteriophage PRD1 infection. *J Virol* 80(16):8081–8088.
 50. Lavigne R, et al. (2013) A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *MBio* 4(2). doi:10.1128/mBio.00061-13.
 51. Clokie MRJ, Mann NH (2006) Marine cyanophages and light. *Environ Microbiol* 8(12):2074–2082.
 52. Myllykallio H, et al. (2002) An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* 297(5578):105–7.
 53. Marston MF, et al. (2013) Marine cyanophages exhibit local and regional biogeography. *Environ Microbiol* 15(5):1452–1463.
 54. Tilly K, Murialdo H, Georgopoulos C (1981) Identification of a second *Escherichia coli* *groE* gene whose product is necessary for bacteriophage

morphogenesis. *Proc Natl Acad Sci U S A* 78(3):1629–33.

55. Ueno H, Yonesaki T (2004) Phage-induced change in the stability of mRNAs. *Virology* 329(1):134–141.

Chapter 6: Conclusions and Future Directions

6.1 Conclusions

As a representative strain of subcluster 5.2, estuarine *Synechococcus* CB0101 demonstrates its flexibility in response to strong environmental gradients. In comparison with coastal and oceanic *Synechococcus* strains (represented by WH7803 and WH7805, respectively), which are members of subcluster 5.1, estuarine strain CB0101 exhibits strong adaptive capability for handling chemical and physical changes. This is exemplified by the facts that: 1) CB0101 is able to grow in much wider range of salinity (0-35 ppt) compared to WH7803 and WH7805. CB0101 is not affected by salinity changes, while WH7803 and WH7805 are when salinity is lower than 20 ppt; 2) CB0101 can tolerate a wider temperature range (4-35 °C) compared to WH7803 and WH7805, especially when the temperature is below 15°C; 3) CB0101 can tolerate higher concentrations of heavy metals, such as copper and zinc, compared to WH7803 and WH7805. The high environmental tolerance of CB0101 is reflected by the strong seasonal and spatial variations in the Chesapeake estuary. My results show that CB0101 is more sensitive to higher light intensity compared to WH7803 and WH7805, suggesting that some subcluster 5.2 *Synechococcus* found in estuaries could be adapted to lower light availability in the turbid upper Chesapeake Bay. In general, no major difference was observed on the growth of CB0101, WH7803, and WH7805 in terms of nutrient utilization. This work was presented in Chapter 2.

The complete genome sequence of CB0101 was obtained by combining 454 and PacBio sequencing technologies. The genome size of CB0101 is 2.79 Mb, larger than that of WH7803 (2.37Mb) and WH7805 (2.65Mb). The GC content of CB0101 is 63.4%, higher than that of WH7803 (60.0 %) and WH7805 (57.0 %). Higher GC content is

known to increase tolerance and ability for organisms to grow under extreme conditions (1). CB0101 contains many genes involved in stress responses. For example, the chromosome of CB0101 contains 14 toxin-antitoxin genes, which are not found in WH7803 and WH7805. Chromosomal toxin-antitoxin systems are commonly found in prokaryotes and are involved in stress response by regulating the growth of cells or inducing cell death. The CB0101 genome contains 2,833 core genes of which 343 genes are found within genomic islands. The amount of genomic island genes in CB0101 is much higher than that found in coastal and open ocean *Synechococcus*. Genomic islands are known to be the locations where horizontal gene transfer occurs frequently. For example, acquisition of antibiotics or heavy metal resistant genes could enable microorganisms to be more competitive in stressful environments. In comparison with WH7803 and WH7805, CB0101 contains many more genes involved in metal homeostasis, sensor kinases, and various metabolic pathways for nitrogen and phosphorus metabolism. The genomic features of CB0101 suggest that estuarine subcluster 5.2 *Synechococcus* may inherit many stress responding systems, which allows them to thrive in the dynamic estuarine environment. This work was presented in Chapter 3.

In order to understand how CB0101 responds to changing environments at the genomic level, I applied transcriptomics and proteomics to study the expression of mRNA and proteins under different growth conditions. CB0101 was incubated under nitrate limitation, phosphate limitation, and high zinc exposure, respectively. The results show that CB0101 seamlessly adjusts its metabolic pathways, transport systems, and sensing mechanisms when faced with nitrate and phosphate limitation. Oxidative stress

caused by zinc toxicity resulted in a global down regulation of photosystems and translation machinery.

From this study a particularly intriguing translational regulator became apparent; the CB0101 genome contains seven pairs of toxin-antitoxin (TA) genes, which have not been found in known marine *Synechococcus* (and *Prochlorococcus*) genomes. Efforts were made to understand under stress conditions how the TA systems in CB0101 were regulated at the mRNA and protein level. Among the 14 identified TA genes, 2 toxin genes (*voeB¹* and *relE¹*) and 1 antitoxin gene (*relB²*) were significantly up-regulated (> 2 fold) under nitrogen limitation. Three toxin genes (*voeB¹*, *doc¹* and *relE¹*) were up-regulated significantly (> 2 fold) under high zinc exposure. Under phosphate limitation only *relE¹* was up-regulated. It appears that *relE¹* is the most sensitive antitoxin gene among the 7 CB0101 toxin genes in response to nutrient and metal stresses as evidenced by high transcriptional fold change. One antitoxin gene *relB²* was down regulated (> 2 fold) under nitrogen limitation, and two antitoxin genes *relB¹* and *phd²* were down regulated (> 2 fold) under high zinc condition. The expressions of these TA genes were comparable between RNA-Seq and qRT-PCR analyses. We developed specific antibodies for the RelB²/RelE¹ TA pair in order to confirm the presence of these proteins. To understand the regulation of TA genes at the protein level, we applied Western blot analysis to observe protein expression of RelB²/RelE¹ TA pair under heavy metal and photo-oxidative light stresses. Under normal growth conditions (control), a protein dimer of RelB²/RelE¹ was detected, indicating strong binding between the inactivated toxin and antitoxin. When CB0101 became stressed by nutrient limitation or heavy metal exposure, the RelB²/RelE¹ dimer decreased, resulting in more RelE¹ toxin protein. This inverse

relationship is the result of the antitoxin being degraded, thus allowing the RelE¹ toxin protein to act on its intended target, cleaving ribosome bound mRNA, and causing translational inhibition.

When CB0101 was exposed to high zinc concentration, the growth was inhibited with the dimer RelB²/RelE¹ declining within 24 hours and resulting RelE¹ toxin increasing. Upon the removal of zinc toxicity, the growth resumed and the yield of dimer RelB²/RelE¹ increased, indicating that the arrest was reversible when the stress event was removed. A clear RelB²/RelE¹ response was observed when CB0101 was cultivated under a wide range of light intensities (15-200 $\mu\text{E m}^{-2} \text{s}^{-1}$). The growth of CB0101 was inhibited and the yield of toxin protein RelE¹ began to increase when the light intensity exceeded 100 $\mu\text{E m}^{-2} \text{s}^{-1}$. We conceived that TA systems, such as *relB/relE*, allow for niche adaptation to the low light environments of the Chesapeake Bay, enabling *Synechococcus* like CB0101 to thrive. We demonstrate that TA genes in CB0101 are regulated under different manipulated growth conditions at both mRNA and protein level. This TA system is one such example for how CB0101 has an enhanced capability for niche adaptation in a highly variable environment. The actual ecological function of TA system of estuarine *Synechococcus* in the natural environment still remains to be explored. The TA genes of CB0101 appear to originate from genes in diverse bacteria. The finding of TA system in CB0101 led to the discovery of TA systems in eight other marine *Synechococcus* strains including two estuarine strains (WH5701 and CB0205), two coastal strains (KORDI-49 and KORDI-51), and four open ocean strains (RS9916, RS9917, WH8102, and WH8103). The relationship between TA genes among these

marine *Synechococcus* should be further investigated. This work was presented in Chapter 4.

Synechococcus CB0101 is infected by many phages isolated from the Chesapeake Bay. How does CB0101 respond to a phage infection at the molecular level? S-CBP1 is a podovirus, which infects CB0101, and whose complete genome has been sequenced in our lab. I monitored the global gene expression of host CB0101 and phage S-CBP1 simultaneously using transcriptomics and proteomics, following a viral infection cycle, which included attachment, latent, and early lysis stages. Surprisingly, a massive shutdown of genes located in the genomics islands (GI) of CB0101 was observed upon the attachment of phage S-CBP1 (within 30 minutes of phage addition). The down-regulation of GI genes continued to increase at the latent and early lysis stages. Many genes located in the GI regions aid in the development of phage resistance. It is unclear why the infection of a cyanophage results in such dramatic shutdown of GI genes of host. A possible explanation is that podovirus S-CBP1 prevents the development of host resistance by quickly turning off the host GI system.

Many cyanophages are known to contain host-like genes, especially the genes related to photosynthesis. Despite the rapid shutdown of host GI genes upon phage infection, several phage-encoded host-like genes (e.g. *hli*, *psbA*, and *thyX*) in phage S-CBP1 were highly expressed at the protein level. This result suggests that expression of these host-like genes is critical during the phage infection. This work was presented in Chapter 5.

In summary, my dissertation greatly deepens our understanding of the physiology and ecology of the member of subcluster 5.2 *Synechococcus* CB0101. The application of

multi-omics (genomics, transcriptomics, and proteomics) provides new insight into the specific molecular mechanisms used by subcluster 5.2 *Synechococcus* for its niche partitioning.

6.2 Future Directions

Picocyanobacteria are abundant and ubiquitous in the aquatic ecosystem. As a major group of picocyanobacteria, *Synechococcus* species exhibit great adaptive ability and genetic diversity throughout the marine environment. Estuaries are highly variable ecosystems where nutrient rich freshwater meets seawater. The overall understanding of how *Synechococcus* subcluster 5.2 live and adapt to the estuarine environment is still limited. However, my dissertation research offers the first biological and molecular evidence on how subcluster 5.2 *Synechococcus* strain CB0101 responds to potential environmental stresses, and this is the first step towards understanding ecological adaptation and molecular evolution of subcluster 5.2 or estuarine *Synechococcus*.

6.2.1 Still a lot to learn from Chesapeake Bay *Synechococcus*

The Chesapeake Bay is an ideal estuary to investigate the biographic distribution of subcluster 5.2 *Synechococcus*. It has been shown that the population structure of *Synechococcus* changes greatly with salinity and temperature (2–4). In order to gain a comprehensive understanding of genetic regulation in subcluster 5.2 *Synechococcus* and the underlying physiological response variations, additional cultures should be isolated from different times and locations. The strains from subcluster 5.2 could then be compared at the physiological and omics level to further define the adaptability of subcluster 5.2 *Synechococcus*.

The finding that CB0101 has low light adaptability is an interesting feature. CB0101 was isolated from the upper Chesapeake Bay; therefore, will other subcluster 5.2 *Synechococcus* isolated from the Chesapeake Bay also be sensitive to light? Or is this merely a factor of niche adaptation for CB0101? Further, is it possible to see light partitioning of *Synechococcus* along the salinity gradient of Chesapeake Bay?

Additional genome sequences from subcluster 5.2 *Synechococcus* are needed. Currently, only one complete genome sequence (CB0101) is available for estuarine subcluster 5.2 *Synechococcus*. The other two estuarine isolates, *Synechococcus* WH5701 and CB0205, contain uncompleted genome sequences. I would recommend fully sequencing additional subcluster 5.2 members, which represent different clades and niche habitats of the Chesapeake Bay.

With the additional genomes an extensive comparative genomics analysis could be conducted. I anticipate this comparative genomics analysis will first focus on subcluster 5.2 *Synechococcus*, identifying unique genes between different niches and the creation of a broad picture of the genetic capacity of each clade. Furthermore, comparative genomics between winter and summer Chesapeake Bay *Synechococcus* could provide a better understanding on cold adaptation of *Synechococcus*. From this analysis, a comparison of freshwater, estuarine, coastal, and open ocean *Synechococcus* could be conducted to identify genes unique to estuarine and subcluster 5.2 strains, which allow for their increased adaptability to variable environments.

Another dataset to study is the meta-genomic and meta-transcriptomic data we collected in the Chesapeake Bay and Delaware Bay between 2013 and 2015. This DOE-supported sequencing effort includes viral and bacterial samples in both bays over

temporal and spatial scale. Once the final sequence data is analyzed, we can focus the sequence data on the *Synechococcus* communities. This survey will allow us to investigate how different types of *Synechococcus* are distributed along the estuarine gradient; what *Synechococcus* genes are highly regulated *in situ*; and how *Synechococcus* gene expression is related to environmental variation. Moreover, the interaction between *Synechococcus* and phage can be explored by examining both viral and bacterial meta-omics data.

The above analyses not only hold scientific importance, but also economic. The Chesapeake Bay creates an economic value between \$3.39 billion in sales and \$890 million in income. In order to better predict future output and protect this valuable resource a firm understanding throughout the entire food chain must be appreciated. Numerous studies have examined the importance of copepods to higher trophic level organism like oysters, Blue Crabs, and Stripped Bass. These organisms have an economic value in the Chesapeake Bay of \$44 million, \$78 million, and \$500 million respectively. The collapse of copepods would cause irreparable harm to the major fisheries. In order to fully appreciate what copepods food source we need to fully understand the picocyanobacteria communities in the Chesapeake Bay, which they feed on.

Recent research has discovered that marine picocyanobacteria are an important food source for copepods, making up between 40-60% of their digested gut content. *Synechococcus* can contribute as much as 40% of primary production during the summer months of the Chesapeake Bay. Among metazooplankton, appendicularians, cladocerans, rotifers, and bivalve larvae are known to substantially feed on picocyanobacteria. It has recently been shown that copepods not only ingest but also assimilated *Synechococcus*.

Therefore, grazing on *Synechococcus*, like CB0101, by crustacean zooplankton should be more appreciated in food web and productivity assessments. The grazing on *Synechococcus* may be a common year-round phenomenon, in aquatic environments, like the Chesapeake Bay. If copepods are capable to directly and efficiently utilize *Synechococcus*, this would facilitate a direct energy transfer from microbial to metazooplankton, surpassing the microbial loop. Thus, a full study on Chesapeake Bay picocyanobacteria community structure, abundance, and analysis of copepod digested picocyanobacteria should be conducted in order to ensure sustainability of this invaluable fishery.

6.2.2 How large of a role do TA systems play in picocyanobacteria?

Our work with TA systems in CB0101 demonstrates an exciting regulatory mechanism within picocyanobacteria. The presence of at least one TA in other *Synechococcus* strains allows for a much wider and comprehensive exploration of their potential function. One such example would be TA system cross-talk within CB0101. It has been shown that other bacterial TA systems mutually affect each other's expression and activity (5). Most bacteria have many different TA systems like CB0101. Although their functions are debatable, many TA toxins have similar activity and an inhibitory effect on bacteria. Therefore, an important question is whether TA systems are redundant or not. Another intriguing issue is whether distinctive TA systems are functionally connected and cross-talk in *Synechococcus*, as seen in other prokaryotes (6, 7). It was found in *E. coli* that such cross-reaction occurs with the *relB*²/*relE*¹-dependent transcriptional activation of *mazE/mazF* (5). As both TA systems are present in CB0101 it would be interesting for future work to see if such cross-reaction or cross-talk can

occur in *Synechococcus spp.* Examples of future studies that should be pursued include:

- 1) A comprehensive search of publically available partial and complete genomes from *Synechococcus* and *Prochlorococcus* for TA systems. This will allow us to identify the distribution of TA genes within picocyanobacteria and potentially understand the evolution of TA genes in picocyanobacteria.
- 2) *In situ* monitoring of the expression of *Synechococcus* TA genes is an important step for demonstrating the ecological relevance of TA system in the natural environment. For examples, we do not know how abundant TA genes are in the Chesapeake Bay. Further, if TAs are expressed, are there differences between the upper bay *Synechococcus* compared to those of the lower bay? Additionally, are different types of TA genes more prevalent in different niche habitats? Although meta-transcriptomics is a powerful tool to address this question, other specific detection methods such as qRT-PCR or antibody analysis can be used to quantify the expression of specific genes or proteins over various biogeochemical gradients. The major challenge will be to identify the suitable gene markers that are conserved among a defined *Synechococcus* group.

6.2.3 GI response: A factor only in CB0101 infection or a wider trend?

With the increased appreciation of GI regions in picocyanobacteria, it has become clear that these areas play important parts, not only in niche adaptation, but also in phage infection. However, as stated in the above section, many of the genes found within GIs are hypothetical or poorly understood due to a lack of homology to known functional proteins. Earlier studies, together with this work, have shown that many GI genes respond to phage infection, often with over a 2-fold response as compared to control. In order to fully understand what is occurring, many of these genes need to be elucidated. It is

possible that there is a cross-talk system between GI regions, which would confirm why there is a global shutdown of almost every GI region. Further, it is clear that host-like genes in the phage genome are playing a part during infection. It has been suggested that they are used to increase phage fitness during infection, but this hypothesis has not been fully confirmed. Future questions that still need to be answered include: 1) Is this GI response common among *Synechococcus* strains, or is this a variation from this single system? A single study, which includes a large deal of sequencing depth using RNA-Seq to test results, can be found in (8, 9). 2) If the *hli*, *psbA*, and *ThyX* genes are knocked out from the phage, how is the infectivity affected, if at all? 3) Do other podoviruses that infect CB0101 cause the similar GI response? How do the GI regions of CB0101 react to the infection of myoviruses or siphoviruses?

References:

1. Šmarda P, et al. (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci U S A* 111(39):E4096–102.
2. Chen F, et al. (2004) Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* 36(2):153–164.
3. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* 72(3):2239–2243.
4. Cai H, Wang K, Huang S, Jiao N, Chen F (2010) Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. *Appl Environ Microbiol* 76(9):2955–2960.
5. Kasari V, Mets T, Tenson T, Kaldalu N (2013) Transcriptional cross-activation between toxin-antitoxin systems of *Escherichia coli*. *BMC Microbiol* 13(1):45.
6. Fiebig A, Rojas CMC, Siegal-Gaskins D, Crosson S (2010) Interaction specificity, toxicity, and regulation of a paralogous set of ParE/RelE-family toxin-antitoxin systems. *Mol Microbiol* 77(1):236–251.
7. Maisonneuve E, Shakespeare LJ, Jørgensen MG, Gerdes K (2011) Bacterial persistence by RNA endonucleases. *Proc Natl Acad Sci U S A* 108(32):13206–13211.
8. Doron, S; Fedida A (2016) Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* 10:1437–1455.

9. Lindell D, et al. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449(7158):83–86.

Dissertation References:

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422(6928):198–207.

Affronti LF, Marshall HG. Diel abundance and productivity patterns of autotrophic picoplankton in the lower Chesapeake Bay. *J Plankton Res.* 1993;15: 1–8. doi:10.1093/plankt/15.1.1

Ahlgren NA, Rocap G (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* 72(11):7193–7204.

Ahlgren NA, Rocap G (2012) Diversity and distribution of marine *Synechococcus*: Multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* 3. doi:10.3389/fmicb.2012.00213.

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011) Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* 474(7353):604–608.

Bagchi SN, Bitz T, Pistorius EK, Michel KP. A *Synechococcus elongatus* PCC 7942 mutant with a higher tolerance toward the herbicide bentazone also confers resistance to sodium chloride stress. *Photosynth Res.* 2007;92: 87–101. doi:10.1007/s11120-007-9176-y

Bantscheff M, Lemeer S, Savitski MM, Kuster B (2012) Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. *Anal Bioanal Chem* 404(4):939–965.

Batiuk R, Bergstrom P, Kemp M, Koch E, Murray L, Stevenson J, et al. Chesapeake Bay submerged aquatic vegetation water quality and habitat-based requirements and restoration targets: A second technical synthesis. United States Environmental Protection Agency for the Chesapeake Bay Program. 2000. doi:10.1001/archderm.136.8.1011

Billis K, Billini M, Tripp HJ, Kyrpides NC, Mavromatis K (2014) Comparative transcriptomics between *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 provide insights into mechanisms of stress acclimation. *PLoS One* 9(10):1–10.

Boggild A, Sofos N, Andersen KR, Feddersen A, Easter AD, Passmore LA, et al. The crystal structure of the intact *E. coli* RelBE toxin-antitoxin complex provides the structural basis for conditional cooperativity. *Structure.* 2012;20: 1641–1648. doi:10.1016/j.str.2012.08.017

- Brand LE, Sunda WG, Guillard RRL (1983) Limitation of marine phytoplankton reproductive rates by zinc, manganese, and iron. *Limnol Oceanogr* 28(6):1182–1198.
- Brantl S. Bacterial type I toxin-antitoxin systems. *RNA Biol.* 2012;9: 1488–90. doi:10.4161/rna.23045
- Buchanan C, Lacouture R V., Marshall HG, Olson M, Johnson JM (2005) Phytoplankton reference communities for Chesapeake Bay and its tidal tributaries. *Estuaries* 28(1):138–159.
- Bukowski M, Rojowska A, Wladyka B. Prokaryotic toxin-antitoxin systems--the role in bacterial physiology and application in molecular biology. *Acta Biochim Pol.* 2011;58: 1–9. doi:20111979 [pii]
- Cai H, Wang K, Huang S, Jiao N, Chen F (2010) Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. *Appl Environ Microbiol* 76(9):2955–2960.
- Calteau A, et al. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics* 15(1):977.
- Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Chen F, et al. (2004) Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* 36(2):153–164.
- Chen F, Wang K, Kan J, Suzuki MT, Wommack KE (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* 72(3):2239–2243.
- Choi DH, Noh JH (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* 69(3):439–448.
- Christensen SK, Mikkelsen M, Pedersen K, Gerdes K. RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc Natl Acad Sci USA.* 2001;98: 14328–33. doi:10.1073/pnas.251327898
- Christie-Oleza JA, Scanlan DJ, Armengaud J (2015) “You produce while I clean up”, a strategy revealed by exoproteomics during *Synechococcus*-*Roseobacter* interactions. *Proteomics* 15(20):3454–3462.
- Clokier MRJ, Mann NH (2006) Marine cyanophages and light. *Environ Microbiol* 8(12):2074–2082.

- Coleman ML, et al. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 311(5768):1768–1770.
- Collier JL, Brahamsha B, Palenik B (1999) The marine cyanobacterium *Synechococcus* sp. WH7805 requires urease (urea amidohydrolase, EC 3.5. 1.5) to utilize urea as a nitrogen source: molecular-genetic and biochemical analysis of the enzyme. *Microbiology* 145(1999):447–459.
- Coutinho F, Tschoeke DA, Thompson F, Thomson C (2015) Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus*. *PeerJ*:e–1522.
- Cox AD, Saito MA (2013) Proteomic responses of oceanic *Synechococcus* WH8102 to phosphate and zinc scarcity and cadmium additions. *Front Microbiol* 4:1–17.
- Darling AE, Tritt A, Eisen JA, Facciotti MT (2011) Mauve assembly metrics. *Bioinformatics* 27(19):2756–2757.
- Dhillon BK, et al. (2015) IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res* 43(W1):W104–8.
- DOE US (2008) Carbon Cycling and Biosequestration: Report from the March 2008 Workshop. US Dep Energy Off Sci (DOE/SC-108).
- Dong H-P, Wang D-Z, Dai M, Chan LL, Hong H-S (2009) Shotgun proteomics: Tools for analysis of marine particulate proteins. *Limnol Oceanogr Methods* 7(DEC):865–874.
- Doron, S; Fedida A (2016) Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* 10:1437–1455.
- Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik B, et al. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 2008;9: R90. doi:10.1186/gb-2008-9-5-r90
- Everroad RC, Wood AM (2012) Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* 64(3):381–392.
- Falkowski PG, Barber RT, Smetacek V (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* 281:200–206.
- Federhen S (2015) Type material in the NCBI Taxonomy Database. *Nucleic Acids Res* 43(1):1086–1098.
- Ferris MJ, Palenik B, Ferris (1998) Niche adaptation in ocean cyanobacteria. *Nature* 396:226–228.

Field CB (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281(5374):237–240.

Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair.

Proc Natl Acad Sci USA. 2009;106: 894–9. doi:10.1073/pnas.0808832106
Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair.

Fisher TR, et al. (1999) Spatial and temporal variation of resource limitation in Chesapeake Bay. *Mar Biol* 133(4):763–778.

Fisher TR, Peele ER, Ammerman JW, Harding LW (1992) Nutrient limitation of phytoplankton in Chesapeake Bay. *Mar Ecol Prog Ser* 82:51–63.

Flombaum P, et al. (2013) Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA* 110(24):9824–9.

Flores E, Herrero A (1995) Assimilatory nitrogen metabolism and its regulation. *The Molecular Biology of Cyanobacteria*, pp 487–517.

Fozo EM, Hemm MR, Storz G. Small toxic proteins and the antisense RNAs that repress them. *Microbiol Mol Biol Rev*. 2008;72: 579–89. doi:10.1128/MMBR.00025-08

Franklin DJ, Brussaard CPD, Berges J a. What is the role and nature of programmed cell death in phytoplankton ecology? *Eur J Phycol*. 2006;41: 1–14. doi:10.1080/09670260500505433

Fuller NJ, Marie D, Vaultot D, Post AF, Scanlan DJ (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* 69(5):2430–2443.

Fuller NJ, Tarran G a., Yallop M, Orcutt KM, Scanlan DJ (2006) Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol Oceanogr* 51(6):2515–2526.

Galens K, et al. (2011) The IGS standard operating procedure for automated prokaryotic annotation. *Stand Genomic Sci* 4(2):244–51.

Gerdes K, Christensen SK, Løbner-Olesen A. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol*. 2005;3: 371 – 382. doi:10.1038/nrmicro1147

Glass EM, Meyer F (2011) The metagenomics RAST server: A public resource for the automatic phylogenetic and functional analysis of metagenomes. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, pp 325–331.

Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 19(12):2226–2238.

Guerreiro ACL, et al. (2014) Daily rhythms in the cyanobacterium *Synechococcus elongatus* probed by high-resolution mass spectrometry based proteomics reveals a small-defined set of cyclic proteins. *Mol Cell Proteomics*:13(8):1–37.

Harrison PW, Wright AE, Mank JE (2012) The evolution of gene expression and the transcriptome-phenotype relationship. *Semin Cell Dev Biol* 23(2):222–229.

Hazan R, Sat B, Engelberg-Kulka H. *Escherichia coli* mazEF-mediated cell death is triggered by various stressful conditions. *J Bacteriol.* 2004;186: 3663–3669. doi:10.1128/JB.186.11.3663-3669.2004

Herdman M, Castenholz R, Waterbury J, Rippka R. (2012) Form-genus XIII. *Synechococcus*. *Bergey's Manual of Systematic Bacteriology*, pp 508–512.

Herrero A, Flores E (2008) *The Cyanobacteria: Molecular Biology, Genomics and Evolution* | Book. Caister Acad Press (February):484.

Huang S, et al. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 6(2):285–97.

Huang S, Wang K, Jiao N, Chen F. Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environ Microbiol.* 2012;14: 540–558. doi:10.1111/j.1462-2920.2011.02667.x

Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J.* 2012;6: 285–97. doi:10.1038/ismej.2011.106

Huang S, Wilhelm SW, Jiao N, Chen F (2010) Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *ISME J* 4(10):1243–1251.

Huang S, Zhang S, Jiao N, Chen F (2015) Comparative genomic and phylogenomic analyses reveal a conserved core genome shared by estuarine and oceanic cyanopodoviruses. *PLoS One* 10(11):e0142962.

- Iriarte A, Purdie DA. Size distribution of chlorophyll a biomass and primary production in a temperate estuary (Southampton Water): the contribution of photosynthetic picoplankton. *Mar Ecol Prog Ser.* 1994;115: 283–297.
- Jardillier L, Zubkov M V, Pearman J, Scanlan DJ (2010) Significant CO₂ fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* 4(9):1180–1192.
- Jiang SC, Kellogg CA, Paul JH (1998) Characterization of marine temperate phage-host systems isolated from Mamala Bay, Oahu, Hawaii. *Appl Environ Microbiol* 64(2):535–542.
- Jiao N, et al. (2005) Dynamics of autotrophic picoplankton and heterotrophic bacteria in the East China Sea. *Cont Shelf Res* 25(10):1265–1279.
- Johnson PW, Sieburth JM (1979) Chroococcoid cyanobacteria in the sea: A ubiquitous and diverse phototrophic biomass. *Limnol Oceanogr* 24(5):928–935.
- Johnson ZI, et al. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311(5768):1737–1740.
- Kelly L, Ding H, Huang KH, Osburne MS, Chisholm SW (2013) Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J* 7(9):1827–41.
- Kemp WM, Boynton WR, Adolf JE, Boesch DF, Boicourt WC, Brush G, et al. Eutrophication of Chesapeake Bay: Historical trends and ecological interactions. *Mar Ecol Prog Ser.* 2005;303: 1–29. doi:10.3354/meps303001
- Knief C, et al. (2012) Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J* 6(7):1378–1390.
- Koksharova O a, Klint J, Rasmussen U (2007) Comparative proteomics of cell division mutants and wild-type of *Synechococcus* sp. strain PCC 7942. *Microbiology* 153:2505–17.
- Kolodkin-Gal I, Verdinger R, Shlosberg-Fedida A, Engelberg-Kulka H. A differential effect of *E. coli* toxin-antitoxin systems on cell death in liquid media and biofilm formation. *PLoS One.* 2009;4. doi:10.1371/journal.pone.0006785
- Labrie SJ, et al. (2013) Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ Microbiol* 15(5):1356–1376.
- Larsson J, Celepli N, Ininbergs K, Dupont CL, Yooseph S, Bergman B, et al. Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water

Baltic Sea. *ISME J*. Nature Publishing Group; 2014;8: 1892–903.
doi:10.1038/ismej.2014.35

Larsson J, et al. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 8(9):1892–903.

Latifi A, Ruiz M, Zhang CC. Oxidative stress in cyanobacteria. *FEMS Microbiol Rev*. 2009;33: 258–278. doi:10.1111/j.1574-6976.2008.00134.x

Lavigne R, et al. (2013) A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *MBio* 4(2). doi:10.1128/mBio.00061-13.

Lee JW, Helmann JD. Functional specialization within the fur family of metalloregulators. *BioMetals*. 2007. pp. 485–499. doi:10.1007/s10534-006-9070-7

Leplae R, Geeraerts D, Hallez R, Guglielmini J, Drze P, Van Melderen L. Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res*. 2011;39: 5513–5525. doi:10.1093/nar/gkr131

Lewis K. Persister cells, dormancy and infectious disease. *Nat Rev Microbiol*. 2007;5: 48–56. doi:10.1038/nrmicro1557

Li GY, Zhang Y, Inouye M, Ikura M. Structural Mechanism of transcriptional autorepression of the *Escherichia coli* RelB/RelE antitoxin/toxin module. *J Mol Biol*. 2008;380: 107–119. doi:10.1016/j.jmb.2008.04.039

Li WKW (1994) Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: Measurements from flow cytometric sorting. *Limnol Oceanogr* 39(1):169–175.

Lindell D, et al. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449(7158):83–86.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438(7064):86–9.

Lindell D, Padan E, Post AF (1998) Regulation of *ntcA* expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J Bacteriol* 180(7):1878–1886.

Loman NJ, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30(5):434–9.

López-Gomollón S, Hernández JA, Pellicer S, Angarica VE, Peleato ML, Fillat MF. Cross-talk between iron and nitrogen regulatory networks in *Anabaena* (*Nostoc*) sp. PCC 7120: Identification of overlapping genes in *FurA* and *NtcA* regulons. *J Mol Biol*. 2007;374: 267–281. doi:10.1016/j.jmb.2007.09.010

Lu J, Chen F, Hodson RE (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. *Appl Environ Microbiol* 67(7):3285–3290.

Ludwig M, Bryant DA (2012) Acclimation of the global transcriptome of the cyanobacterium *Synechococcus* sp. strain PCC 7002 to nutrient limitations and different nitrogen sources. *Front Microbiol* 3(APR):1–15.

Mahmood T, Yang PC. Western blot: Technique, theory, and trouble shooting. *N Am J Med Sci*. 2012;4: 429–434. doi:10.4103/1947-2714.100998

Makarova KS, Wolf YI, Koonin E V. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct*. 2009;4: 19. doi:10.1186/1745-6150-4-19

Mann NH, et al. (2003) Bacterial photosynthesis genes in a virus. *Nature* 424:741–742.

Marsan D, Wommack KE, Ravel J, Chen F (2014) Draft genome sequence of *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay estuary. *Genome Announc* 2(1):e01111–13.

Marston MF, et al. (2013) Marine cyanophages exhibit local and regional biogeography. *Environ Microbiol* 15(5):1452–1463.

Marston MF, Sallee JL (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol* 69(8):4639–4647.

Mazard S, Ostrowski M, Partensky F, Scanlan DJ (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* 14(2):372–386.

Molineux I (2005) *The Bacteriophages*. Oxford University Press:277–301.

Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T (2005) Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* 39(3):257–269.

Moore LR, Post AF, Rocap G, Chisholm SW (2002) Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr* 47(4):989–996.

Morby AP, Turner JS, Huckle JW, Robinson NJ (1993) SmtB is a metal-dependent repressor of the cyanobacterial metallothionein gene *smtA*: identification of a Zn inhibited DNA-protein complex. *Nucleic Acids Res* 21(4):921–925.

- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628.
- Muhling M, et al. (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using rpoC1-based restriction fragment length polymorphism. *Aquat Microb Ecol* 45(3):263–275.
- Muro-Pastor MI, Reyes JC, Florencio FJ (2005) Ammonium assimilation in cyanobacteria. *Photosynth Res* 83(2):135–150.
- Murphy RR, Kemp WM, Ball WP (2011) Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts* 34(6):1293–1309.
- Myllykallio H, et al. (2002) An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* 297(5578):105–7.
- Ning D, Jiang Y, Liu Z, Xu Q. Characterization of a chromosomal type II toxin-antitoxin system mazEaFa in the cyanobacterium *Anabaena* sp. PCC 7120. *PLoS One*. 2013;8. doi:10.1371/journal.pone.0056035
- Ohnishi N, Allakhverdiev SI, Takahashi S, Higashi S, Watanabe M, Nishiyama Y, et al. Two-step mechanism of photodamage to photosystem II: Step 1 occurs at the oxygen-evolving complex and step 2 occurs at the photochemical reaction center. *Biochemistry*. 2005;44: 8494–8499. doi:10.1021/bi047518q
- Ottesen E a, et al. (2013) Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc Natl Acad Sci U S A* 110(6):E488–97.
- Overgaard M, Borch J, Jørgensen MG, Gerdes K. Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity. *Mol Microbiol*. 2008;69: 841–857. doi:10.1111/j.1365-2958.2008.06313.x
- Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12(2):87–98.
- Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67(2):991–994.
- Palenik B, et al. (2003) The genome of a motile marine *Synechococcus*. *Nature* 424(6952):1037–1042.
- Palenik B, et al. (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci* 103(36):13555–13559.

Palenik B, Ren Q, Tai V, Paulsen IT (2009) Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* 11(2):349–359.

Pandey DP, Gerdes K. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* 2005;33: 966–976.
doi:10.1093/nar/gki201

Partensky F, Blanchot J, Vaultot D (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters : a review. *Bull l'Institut océanographique* 19(19):457–475.

Partensky F, Hess WR, Vaultot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63(1):106–127.

Penno S, Lindell D, Post AF (2006) Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* 8(7):1200–1211.

Poranen MM, et al. (2006) Global changes in cellular gene expression during bacteriophage PRD1 infection. *J Virol* 80(16):8081–8088.

Priya B, et al. (2007) Comparative analysis of cyanobacterial superoxide dismutases to discriminate canonical forms. *BMC Genomics* 8:435.

Proctor LM, Fuhrman J a. (1990) Viral mortality of marine bacteria and cyanobacteria. *Nature* 343:60–62.

Pryszak MH, Mozdierz CJ, Cook AM, Zhu L, Zhang Y, Inouye M, et al. Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage. *Mol Microbiol.* 2009;71: 1071–1087. doi:10.1111/j.1365-2958.2008.06572.x

Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* 455(7216):1101–1104.

Ray RT, Hass LW, Sieracki ME. Autorophic picoplankton dyanmics in a Chesapeake Bay sub-estuary. *Mar Ecol Progress Ser.* 1989;52: 273–285.

Redfield AC (1958) The biological control of chemical factors in the environment. *Am Sci* 46(3):230A–221.

Richardson TL, Jackson G a (2007) Small phytoplankton and carbon export from the surface ocean. *Science* 315(5813):838–40.

- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–40.
- Rocap G, Distel DL, Waterbury JB, Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 68(3):1180–1191.
- Roucourt B, Lavigne R (2009) The role of interactions between phage and bacterial proteins within the infected cell: A diverse and puzzling interactome. *Environ Microbiol* 11(11):2789–2805.
- Scanlan DJ, Ostrowski M, Mazard S, Dufresne a, Garczarek L, Hess WR, et al. Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev.* 2009;73: 249–299. doi:10.1128/MMBR.00035-08
- Schuster CF, Bertram R. Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. *FEMS Microbiol Lett.* 2013;340: 73–85. doi:10.1111/1574-6968.12074
- Schwarz D, Orf I, Kopka J, Hagemann M (2013) Recent applications of metabolomics toward cyanobacteria. *Metabolites* 3(1):72–100.
- Schwarz R, Forchhammer K. Acclimation of unicellular cyanobacteria to macronutrient deficiency: Emergence of a complex network of cellular responses. *Microbiology.* 2005;151: 2503–2514. doi:10.1099/mic.0.27883-0
- Sevin EW, Barloy-Hubler F. RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biol.* 2007;8: R155. doi:10.1186/gb-2007-8-8-r155
- Shao Y, Harrison EM, Bi D, Tai C, He X, Ou HY, et al. TADB: A web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Res.* 2011;39. doi:10.1093/nar/gkq908
- Shapiro LP, Haugen EM (1988) Seasonal distribution and temperature tolerance of *Synechococcus* in Boothbay Harbor, Maine. *Estuar Coast Shelf Sci* 26(5):517–525.
- Shi T, Falkowski PG. Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc Natl Acad Sci USA.* 2008;105: 2510–2515. doi:10.1073/pnas.0711165105
- Shih PM, et al. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* 110(3):1053–8.
- Sinex SA, Helz GR (1981) Regional geochemistry of trace elements in Chesapeake Bay sediments. *Environ Geol* 3(6):315–323.

Singletary LA, Gibson JL, Tanner EJ, McKenzie GJ, Lee PL, Gonzalez C, et al. An SOS-regulated type 2 toxin-antitoxin system. *J Bacteriol.* 2009;191: 7456–7465. doi:10.1128/JB.00963-09

Six C, et al. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* 8(12):R259.

Sohm J a, et al. (2015) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* 10(2):1–13.

Sommer SE, Pyzik AJ (1974) Geochemistry of middle Chesapeake Bay sediments from upper Cretaceous to Present. *Chesap Sci* 15(1):39–44.

Stuart RK, Brahamsha B, Busby K, Palenik B (2013) Genomic island genes in a coastal marine *Synechococcus* strain confer enhanced tolerance to copper and oxidative stress. *ISME J* 7(6):1139–49.

Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol* 3(5):0790–0806.

Suttle C a., Chan AM, Cottrell MT (1990) Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* 347(6292):467–469.

Suzuki R, Shimodaira H (2006) PvcIust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12):1540–1542.

Tai V, Palenik B (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* 3(8):903–915.

Tandeau de Marsac N, Houmard J (1993) Adaptation of cyanobacteria to environmental stimuli: new steps towards molecular mechanisms. *FEMS Microbiol Lett* 104(1-2):119–189.

Tatusova T, Ciufu S, Fedorov B, O’Neill K, Tolstoy I (2014) RefSeq microbial genomes database: New representation and annotation strategy. *Nucleic Acids Res* 42(D1). doi:10.1093/nar/gkt1274.

B, Rozhon W. Toxin-antitoxin systems. *Bioengineered.* 2014;5: 37–41. doi:10.4161/mge.26219

Taylor P, Unterholzner SJ, Poppenberger B, Rozhon W. Toxin-antitoxin systems. *Bioengineered.* 2014;5: 37–41. doi:10.4161/mge.26219

- Tetu SG, et al. (2009) Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* 3(7):835–849.
- Thelwell C, Robinson NJ, Turner-Cavet JS (1998) An SmtB-like repressor from *Synechocystis* PCC 6803 regulates a zinc exporter. *Proc Natl Acad Sci* 95(18):10728–10733.
- Thompson LR, et al. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci* 108(39):E757–E764.
- Tilly K, Murialdo H, Georgopoulos C (1981) Identification of a second *Escherichia coli* *groE* gene whose product is necessary for bacteriophage morphogenesis. *Proc Natl Acad Sci U S A* 78(3):1629–33.
- Ting CS, Rocap G, King J, Chisholm SW. Cyanobacterial photosynthesis in the oceans: The origins and significance of divergent light-harvesting strategies. *Trends Microbiol.* 2002;10: 134–142. doi:10.1016/S0966-842X(02)02319-3
- Toledo G, Palenik B (1997) *Synechococcus* diversity in the California Current as seen by RNA polymerase (*rpoC1*) gene sequences of isolated strains. *Appl Environ Microbiol* 63(11):4298–4303.
- Tyers M, and Mann M (2003) From genomics to proteomics. *Nature* 422:193–7.
- Ueno H, Yonesaki T (2004) Phage-induced change in the stability of mRNAs. *Virology* 329(1):134–141.
- Van Melderren L, De Bast MS. Bacterial toxin-antitoxin systems: More than selfish entities? *PLoS Genetics*. 2009. doi:10.1371/journal.pgen.1000437
- Vijayan V, Jain IH, O’Shea EK (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biol* 12(5):R47.
- Wagner A (2011) The molecular origins of evolutionary innovations. *Trends Genet* 27(10):397–410.
- Wang DZ, et al. (2011) Homology-driven proteomics of dinoflagellates with unsequenced genomes using MALDI-TOF/TOF and automated de Novo sequencing. *Evidence-based Complement Altern Med* 2011. doi:10.1155/2011/471020.
- Wang K (2007) Biology and ecology of *Synechococcus* and their viruses in the Chesapeake Bay. doi:10.1017/CBO9781107415324.004.
- Wang K, Chen F (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* 10(2):300–312.

- Wang K, Chen F. Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol.* 2008;10: 300–312. doi:10.1111/j.1462-2920.2007.01452.x
- Wang K, Wommack KE, Chen F (2011) Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. *Appl Environ Microbiol* 77(21):7459–7468.
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63.
- Waterbury JB, Valois FW (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* 59(10):3393–3399.
- Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979) Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293–294.
- Waterbury JB, Watson SW, Valois FW, Franks DG (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Photosynthetic Picoplankton. Can. Bull. Fish. Aquatic Sci.* (Department of Fisheries and Oceans, Ottawa, Canada), pp 71–120.
- Whitton BA, Potts M (2000) *The Ecology of Cyanobacteria: Their Diversity in Time and Space.* Kluwer Acad Publ Netherlands:37–59.
- Wilkins MR, et al. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13:19–50.
- Wilson WH, Joint IR, Carr NG, Mann ' NH (1993) Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol* 59(11):3736–3743.
- Yamaguchi Y, Inouye M. Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat Rev Microbiol.* Nature Publishing Group; 2011;9: 779–790. doi:10.1038/nrmicro2651
- Zeller T, Klug G. Thioredoxins in bacteria: Functions in oxidative stress response and regulation of thioredoxin genes. *Naturwissenschaften.* 2006. pp. 259–266. doi:10.1007/s00114-006-0106-1
- Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9(1). doi:10.1371/journal.pone.0078644.

Zwirgmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.

Appendices

Table 3.S1 Sequencing statistics from RNA-Seq experiments run on an Illumina Hi-Seq 100bp paired end reads.

Sample Name	Control		N Deplete		P Deplete		Zn Toxicity	
	1	2	1	2	1	2	1	2
Replicate								
Number of reads	16,475,981	15,255,472	21,253,553	17,331,616	20,081,184	12,868,239	25,707,352	15,430,133
Mapped reads	14,851,449	13,460,718	19,393,230	15,550,524	17,517,735	11,625,811	23,913,853	13,631,813
% mapped reads to CB0101	90%	88%	91%	90%	87%	90%	93%	88%

Table 2.S2 Stress response COG categories as compared between the genomes of *Synechococcus* CB0101, WH7803 and WH7805.

Category	Subcategory	Subsystem	Role	CB0101	WH7805	WH7803
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Chaperone protein DnaJ	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Chaperone protein DnaK	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Glutathione synthetase (EC 6.3.2.3)	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Heat shock protein GrpE	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Heat-inducible transcription repressor HrcA	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Hypothetical radical SAM family enzyme in heat shock gene cluster, similarity with CPO of BS HemN-type	X	X	-
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Nucleoside 5-triphosphatase RdgB (dHATP, dTTP, XTP-specific) (EC 3.6.1.15)	X	X	-
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Ribonuclease PH (EC 2.7.7.56)	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Ribosomal protein L11 methyltransferase (EC 2.1.1.-)	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	Translation elongation factor LepA	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	rRNA small subunit methyltransferase I	X	X	X
Stress Response	Heat shock	Heat shock dnaK gene cluster extended	tmRNA-binding protein SmpB	X	X	X
Stress Response	Osmotic stress	Betaine biosynthesis from glycine	Dimethylglycine N-methyltransferase	-	X	X
Stress Response	Osmotic stress	Betaine biosynthesis from glycine	Glycine N-methyltransferase (EC 2.1.1.20)	-	X	X
Stress Response	Osmotic stress	Betaine biosynthesis from glycine	Sarcosine N-methyltransferase	-	X	X
Stress Response	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)	-	X	X
Stress Response	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)	-	X	X
Stress Response	Osmotic stress	Choline and Betaine Uptake and Betaine Biosynthesis	L-proline glycine betaine binding ABC transporter protein ProX (TC 3.A.1.12.1)	-	X	X
Stress Response	Oxidative stress	Glutaredoxins	Glutaredoxin 3 (Grx1)	X	X	X
Stress Response	Oxidative stress	Glutaredoxins	Uncharacterized monothiol glutaredoxin ycf64-like	X	X	X
Stress Response	Oxidative stress	Glutathione: Biosynthesis and gamma-glutamyl cycle	Gamma-glutamyltranspeptidase (EC 2.3.2.2)	X	X	X
Stress Response	Oxidative stress	Glutathione: Biosynthesis and gamma-glutamyl cycle	Glutathione synthetase (EC 6.3.2.3)	X	X	X
Stress Response	Oxidative stress	Glutathione: Non-redox reactions	Hydroxyacylglutathione hydrolase (EC 3.1.2.6)	X	X	X
Stress Response	Oxidative stress	Glutathione: Non-redox reactions	Lactoylglutathione lyase (EC 4.4.1.5)	X	X	X
Stress Response	Oxidative stress	Glutathione: Redox cycle	Glutaredoxin 3 (Grx1)	X	X	X
Stress Response	Oxidative stress	Glutathione: Redox cycle	Glutathione peroxidase (EC 1.11.1.9)	X	X	X
Stress Response	Oxidative stress	Glutathione: Redox cycle	Glutathione reductase (EC 1.8.1.7)	X	X	X
Stress Response	Oxidative stress	Glutathione: Redox cycle	Uncharacterized monothiol glutaredoxin ycf64-like	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Alkyl hydroperoxide reductase subunit C-like protein	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Fe2+/Zn2+ uptake regulation proteins	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Ferric uptake regulation protein	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Ferric uptake regulation protein FUR	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Ferroxidase (EC 1.16.3.1)	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Iron-binding ferritin-like antioxidant protein	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Non-specific DNA-binding protein Dps	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Peroxide stress regulator	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Superoxide dismutase [Fe] (EC 1.15.1.1)	X	X	X
Stress Response	Oxidative stress	Oxidative stress	Distant similarity with leukotriene C4 synthase (microsomal glutathione S-transferase)	X	-	-
Stress Response	Oxidative stress	Oxidative stress	Glutathione S-transferase (EC 2.5.1.18)	X	-	-
Stress Response	Oxidative stress	Oxidative stress	Glutathione S-transferase family protein	X	-	-
Stress Response	Oxidative stress	Oxidative stress	Manganese superoxide dismutase (EC 1.15.1.1)	X	-	-
Stress Response	Oxidative stress	Oxidative stress	Ruberythrin	X	-	-
Stress Response	Oxidative stress	Oxidative stress	Rubredoxin	X	-	-
Stress Response	Oxidative stress	Ruberythrin	Rubredoxin	X	-	-
Stress Response	Oxidative stress	Cluster containing Glutathione synthetase	Putative Holliday junction resolvase YggF	-	X	X
Stress Response	Oxidative stress	Glutathione: Biosynthesis and gamma-glutamyl cycle	Similar to 5-oxoprolinase (EC 3.5.2.9) and Methylhydantoins A, B (EC 3.5.2.14)	-	X	X
Stress Response	Oxidative stress	Oxidative stress	Catalase (EC 1.11.1.6)	-	X	X
Stress Response	Oxidative stress	Oxidative stress	Peroxidase (EC 1.11.1.7)	-	X	X
Stress Response	Oxidative stress	Oxidative stress	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)	-	X	X
Stress Response	no subcategory	Hfl operon	GTP-binding protein HflX	X	X	X
Stress Response	no subcategory	SigmaB stress response regulation	RNA polymerase sigma factor SigB	X	X	X
Stress Response	no subcategory	SigmaB stress response regulation	Serine phosphatase RsbU, regulator of sigma subunit	X	X	X
Stress Response	Detoxification	Glutathione-dependent pathway of formaldehyde detoxification	S-(hydroxymethyl)glutathione dehydrogenase (EC 1.1.1.284)	-	-	X
Stress Response	Detoxification	Glutathione-dependent pathway of formaldehyde detoxification	S-formylglutathione hydrolase (EC 3.1.2.12)	-	-	X
Stress Response	Oxidative stress	Oxidative stress	Metallothionein	-	-	X

Table 4.S1 Protein description of the TA systems found within CB0101. Toxin represented by bold and anti-toxin by italics. *Bold amino acid sequences were used for peptide selection of antibodies.

TA Family	Amino Acid Sequences	PI	Mw
<i>yefM</i> ¹	MGSISASEARRRFLALIDEVRESHQPVEIHGKRGSAVLLAEQDWRAIQETLYLCAIPGMRESILEGLATPVEELSE DAGW	4.77	8891.07
yoeB ¹	MSWRVLFTRQAQKDARKLASSSPALKNKAQQLLELLAVDPFQQPPPEALVGDQGAYSRRINIQRHLVYAVDREA AVK VLRLLWSHYD	9.98	10215.81
<i>phd</i> ¹	MESSIGAFEAKAQLSRLLRAVEQGEHFTITVRGRPVADLVPHRSPSSEGVAAAIEALQAFPRIRGVSDADVT SFVAEGRR	4.95	14900.04
doc ¹	MTLVIDASMALAWVFERQQASDAQRASELLACCGQQAWVVPGLWHLEVANALLVAERRGVIAQDASDLFVRLSSLPIC TDS DAGPEQQPRLIALARAHGLSSYDATYDLAHLRGAALASFDRQLNQA AVAMGVPLT	6.92	9311.45
<i>vapB</i> ¹	MTASLPSRVFNNGNSQAVRIPAEFRLGTDQVQISRTPEGDLLIHPCPPQGRQALFNALAGFDADFAERLEQSRAEQPIQ TREG L	5.21	9319.42
vapC ¹	MIYLLDTNIIYLIKQRPPEVAERIDQLPGTAQLAMSFITWAELLQGAVGSSRRDAVERQLDHLARQVEVLYPEDS QICRHYA EQATALRRAGTPIGANDLWACHALAVDATALVTHNLREFTRMSGLSVVDWVQQP	5.48	15422.67
<i>relB</i> ¹	MHLLLDTHLLIWAMGSPQRLPNGLADMLEDPGNTPLFSVASLWELVIKQAPNKPDFNVQPALLRRALLECGWQELTITAN HALAVADLPPLHRDPFDRLLLAQAKADGLLITADEQLARYPGPIRWMAPLRPSEES	5.22	15254.73
vapC ²	MLLDTGMRQVNMHEAKTHLSRLVEEAAAGESFLICKAGKPMVRVTSIDQSDAPRPLRRLGLLEGQCSVPDDFDRF GSE AIADLFEGA	5.02	9614.99
<i>relB</i> ²	MAQVTARLPDDLAELDAVAQQLNRCRADVIRQAI EYYLDDIEDLRAGAASLRDPADPVL DWA EVRDVL LAAD	4.06	8081.04
relE ¹	MCCSLRIKRSAAKALAE LPNADRLRLVAADKLC E VPAAGSALKGEFGLRRLRVGRYRIVCEWQQELVVLVVRVGH RKEVYR	9.95	9533.32
<i>mazE</i> ¹	MTMRYDGEAVVRARLFMSGRSQALRLPARLRLRGPDVEIEPIGDGLWVQPCADPSEGLGDWLERFYADHPPLPVEFLED RQDQQPQERDWA	4.64	10520.84
mazF ¹	MRRTLDTNICSYVLRKRPVQVVERFRQLDRRQLWL SAIVAAELRF GA EKLGSRRFRGVSVEAWLSGFELRDWPLAATHHY ARLRAQLEAKGTPVGNLDMIAAHALAEDSVITNNAREFHRIPGVAVEEWQLD	9.5	15208.44
<i>phd</i> ²	MRTVNVEAKTHFSRLIDA AHAGETIVAKGGKPWARLVPLETPAPQRQPGVLAGQLQLPPEILLEALPEDELRAFEIPL P	5.86	8944.4
doc ²	MSAATAWELATKVR L GKLEIAEPLLSDL PCLLAAQGFELLSVDLRHGLRAGGYPHAH RDPFDRL LVAQAELES LTVSINAA LRDFPCRLLW	5.84	10133.81

Table 4.S2 Toxin-antitoxin pairs as identified in the genome of the estuarine

Synechococcus CB0101 during phage infection by S-CBP1. Toxin represented by bold and anti-toxin by italics.

TA Pair #	TA Family	Toxin Activity	Viral 30min RNA-Seq	Viral 30min qPCR	Viral 5hr RNA-Seq	Viral 5hr qPCR	Viral 12hr RNA-Seq	Viral 12hr qPCR
1	<i>yefM¹</i>	mRNA interferase or inhibitor of translation initiation	-1.2	-1.1	-1.1	-1.1	-1.2	-1.1
1	yoeB¹		-1.0	-1.0	-2.0	-1.8	-1.6	-1.4
2	<i>phd¹</i>	Binds to the 30S ribosomal subunit	-1.8	-1.9	-2.0	-2.0	-2.9	-2.7
2	doc¹		-1.8	-1.5	1.3	1.4	-1.3	-1.3
3	<i>vapB¹</i>	Cleavage of tRNA	1.1	1.0	1.0	1.0	-1.2	-1.1
3	vapC¹		-1.0	-1.0	1.4	1.4	-1.1	-1.1
4	<i>relB¹</i>	Cleavage of tRNA	1.0	1.0	-1.0	-1.0	-2.1	-2.0
4	vapC²		-1.1	-1.0	-1.1	-1.0	1.0	1.0
5	<i>relB²</i>	Cleavage of ribosome-bound mRNA	-1.0	-1.0	-1.1	-1.1	-1.3	-1.2
5	relE¹		2.5	2.6	-1.3	-1.2	1.7	1.7
6	<i>mazE¹</i>	Ribosome-independent mRNA cleavage and cleavage of 23S rRNA	-1.1	-1.1	-1.5	-1.4	-2.3	-2.2
6	mazF¹		-3.2	-3.1	1.3	1.2	-1.1	-1.3
7	<i>phd²</i>	Binds to the 30S ribosomal subunit	-1.0	-1.0	1.0	1.0	1.5	1.5
7	doc²		-1.4	-1.3	-1.4	-1.5	-1.3	-1.3

Table 4.S3 Primers developed for qPCR confirmation of RNA-Seq transcript expression. Toxin represented by bold and anti-toxin by italics.

TA Pair #	TA Family	Forward Primer	Reverse Primer
1	<i>yefM</i> ¹	cgttgagattcacggcaag	tcgctcagttcctccacag
1	yoeB ¹	caagctcgctagctcctcac	tcagtcgtagtggctccaga
2	<i>phd</i> ¹	gagtgagctttggcttgct	cagataggtggcgctcgtagc
2	doc ¹	acggttactgagggctgttg	cacatctgcatccgaaacac
3	<i>vapB</i> ¹	gtcattgcctagccgtgtgt	gattgctcaaacgttcagc
3	vapC ¹	caggtggaggtgctgtatcc	acccaatccaccacagacag
4	<i>relB</i> ¹	ctggacacacacctgctgat	agttcttgccagccacactc
4	vapC ²	accacactctctcgctagt	tacagaacattggcctcca
5	<i>relB</i> ²	aatcgatgccgagctgac	ctcagccaatccagaacag
5	relE ¹	acaagctttgcgaggttcc	gatacacctccttgcgatgc
6	<i>mazE</i> ¹	gatatgacggtgaggctgtg	gagggtgatcggcatagaag
6	mazF ¹	gctctccgggttgagttg	caatccaactgccactcctc
7	<i>phd</i> ²	ttccacattcagcaggacag	gttgatcgcgcacgtagat
7	doc ²	tctgagtgatctgccgtgtc	agcgtgagtgattccagctc

Table 5.S1 Sequencing statistics from RNA-Seq experiment run on Illumina Hi-Seq 100bp paired end read.

Time point	0min		30min		5h		12h	
	1	2	1	2	1	2	1	2
Replicate								
Number of reads	16,475,981.00	15,255,472.00	15,763,839.00	13,632,482.00	10,767,144.00	30,896,516.00	16,248,159.00	11,845,502.00
Mapped reads	15,518,810	14,588,719	14,758,335	13,102,290	10,557,414	27,372,056	14,974,210	11,443,627
% mapped reads	94%	96%	94%	96%	98%	89%	92%	97%
Mapped to phage	-	-	1,566,536	2,829,025	3,588,045	12,932,785	11,081,136	8,351,591
Mapped to host (non-rRNA regions)	15,518,810	14,588,719	13,191,799	10,273,265	6,969,369	14,439,271	3,893,074	3,092,036
% mapped to phage/mapped to non-rRNA	0%	0%	10%	21%	33%	42%	68%	71%
% mapped to host/mapped to non-rRNA	94%	96%	84%	75%	65%	47%	24%	26%

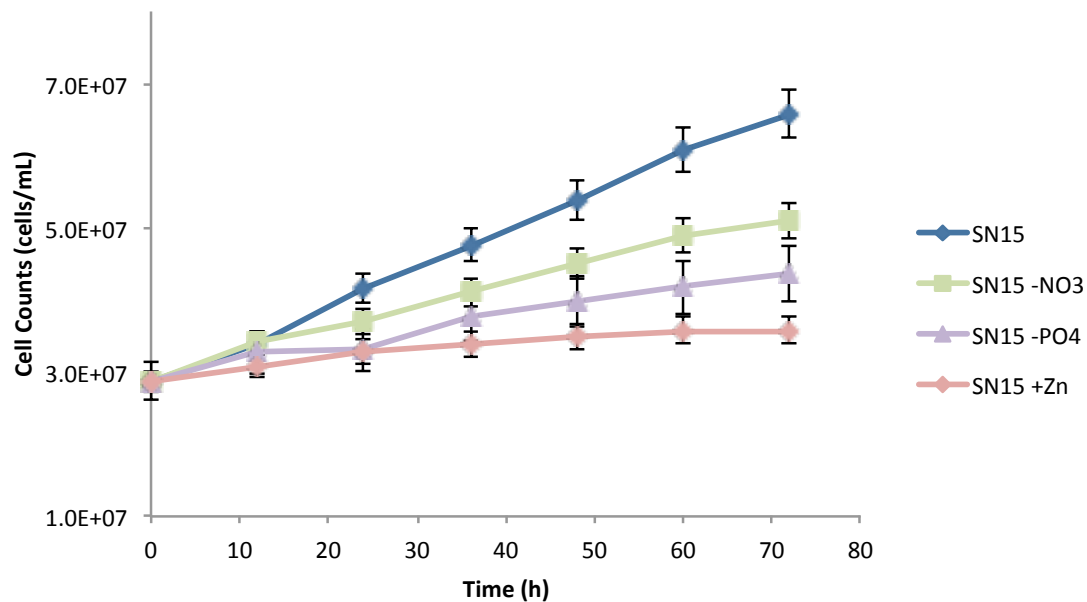


Figure 3.S1 Experimental setup for stress response of the estuarine *Synechococcus* CB0101. Depicts growth after 72 h under the different stress conditions with “C” control, “N-” nitrogen deplete, “P-” phosphate deplete, and “Zn” zinc toxicity.

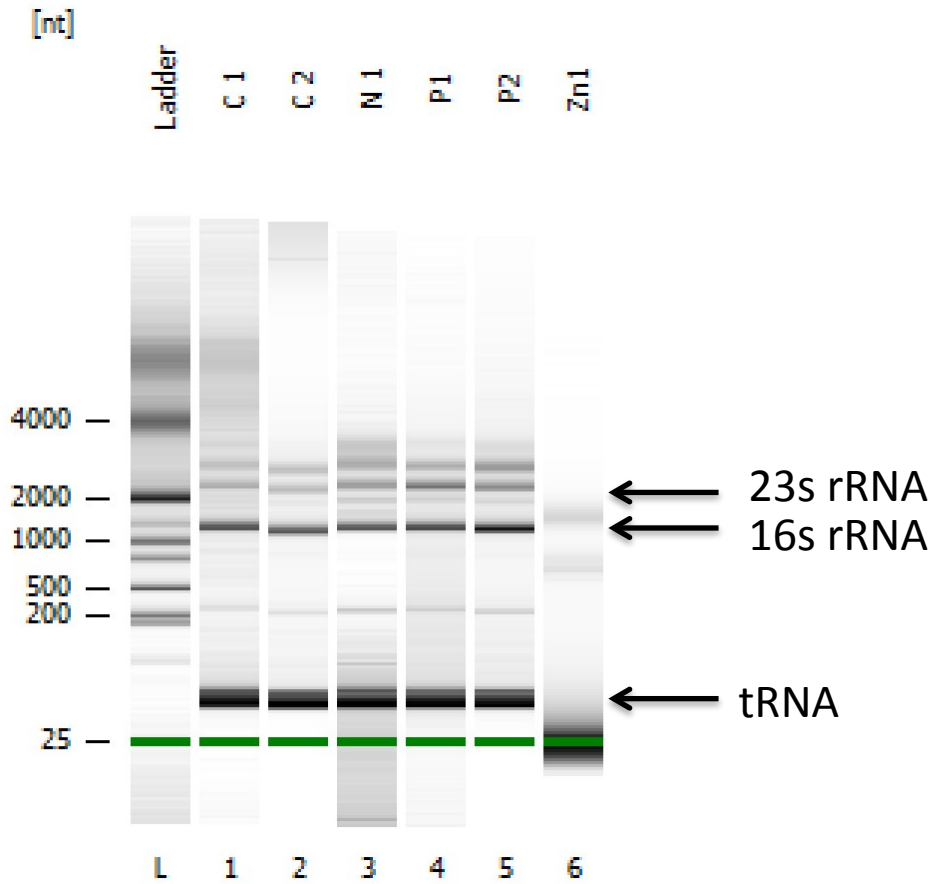
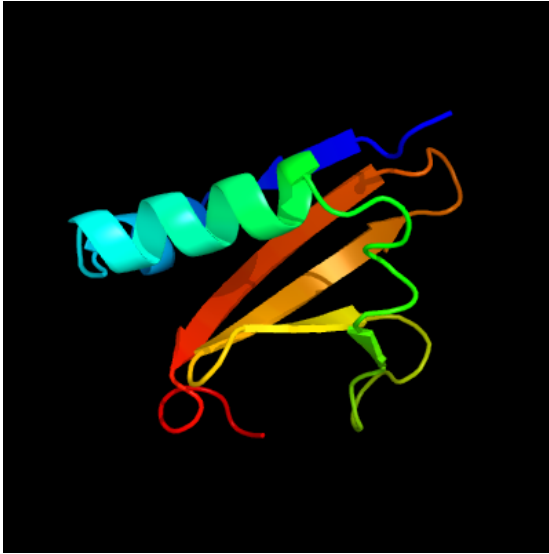


Figure 4.S1 Agilent 2100 Bioanalyzer analysis of total RNA controls and 5 treatments. Other lanes represent Control (C1 & C2), nitrogen deplete (N1 & N2), and phosphate deplete (P1 & P2) each with intact ribosomal subunits, while Zn is degraded.

relE



relB

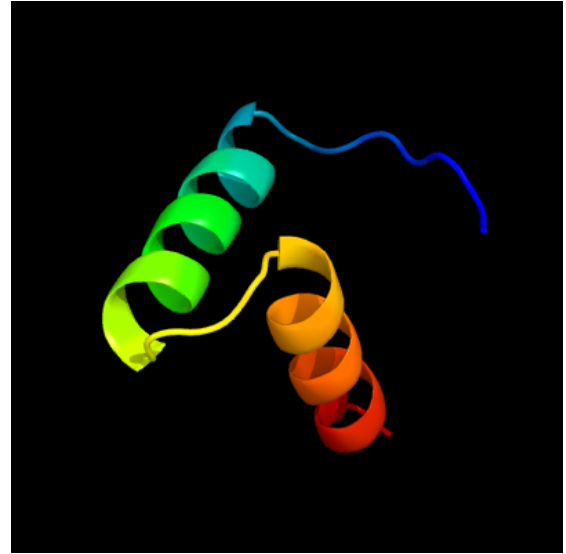


Figure 4.S2 Predicted protein structure of the *relE*¹ toxin protein and *relB*² antitoxin from CB0101 made using phyre2.

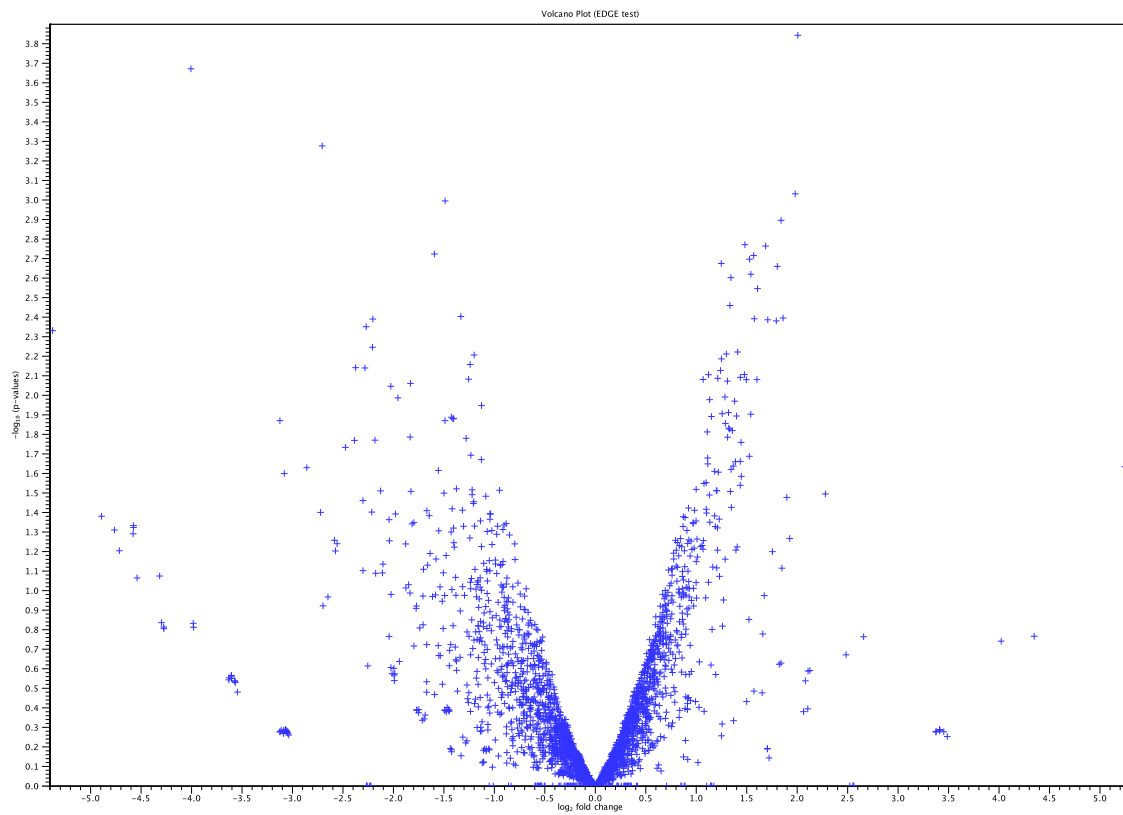


Figure 5.S1 CB0101 transcript statistical DGE test for 30 min post infection. Solid vertical lines indicate 2-fold change in gene expression compared to control sample. Each gene is represented with blue cross.

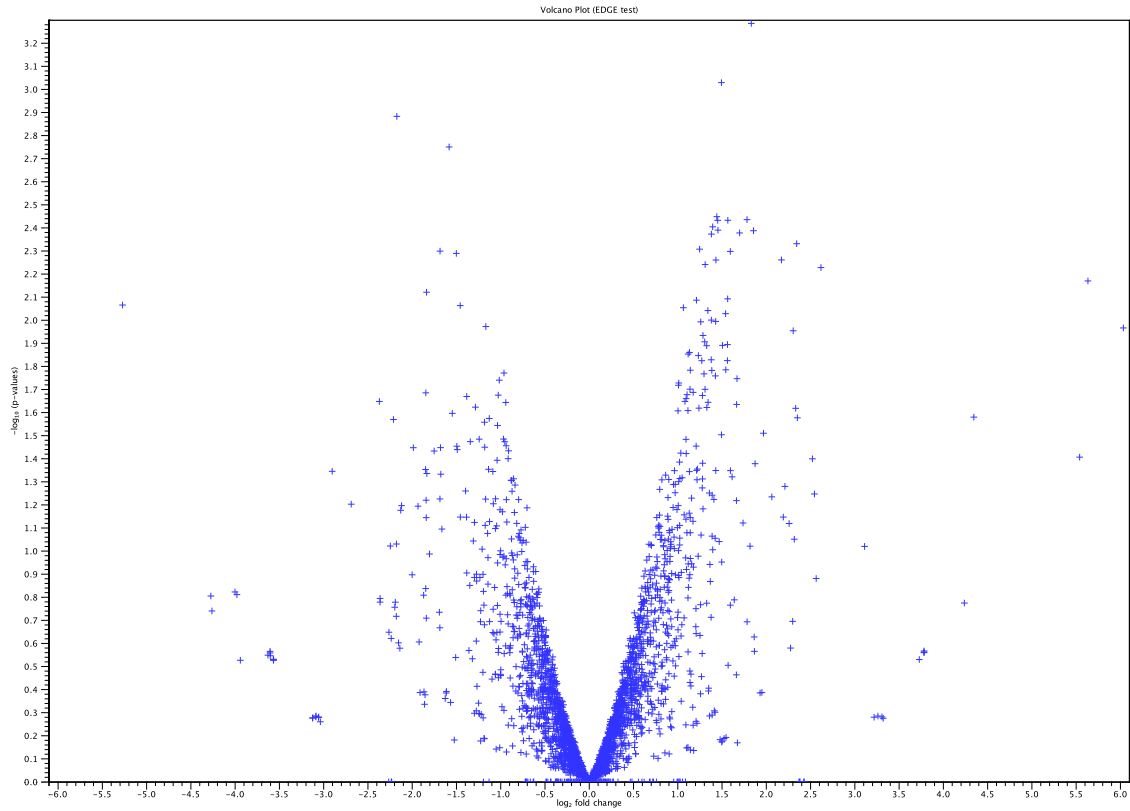


Figure 5.S2 CB0101 transcript statistical DGE test for 5 h post infection. Solid vertical lines indicate 2-fold change in gene expression compared to control sample. Each gene is represented with blue cross.

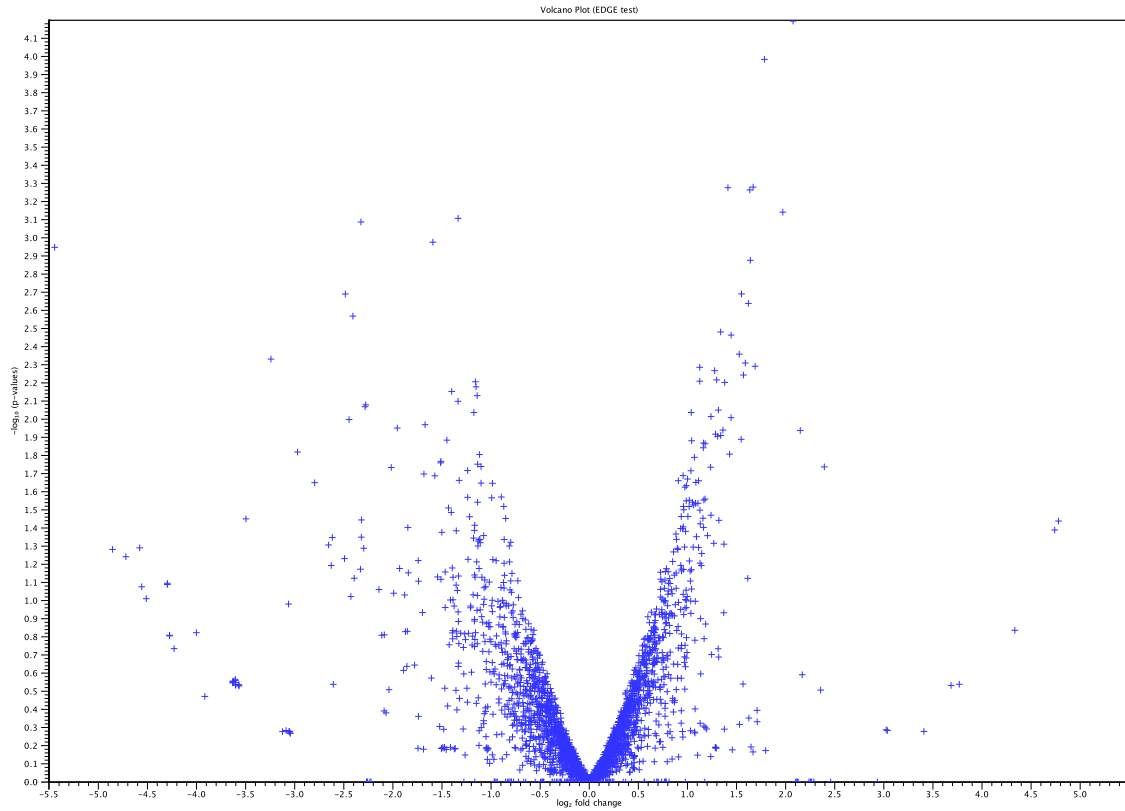


Figure 5.S3 CB0101 transcript statistical DGE test for 12 h post infection. Solid vertical lines indicate 2-fold change in gene expression compared to control sample. Each gene is represented with blue cross.

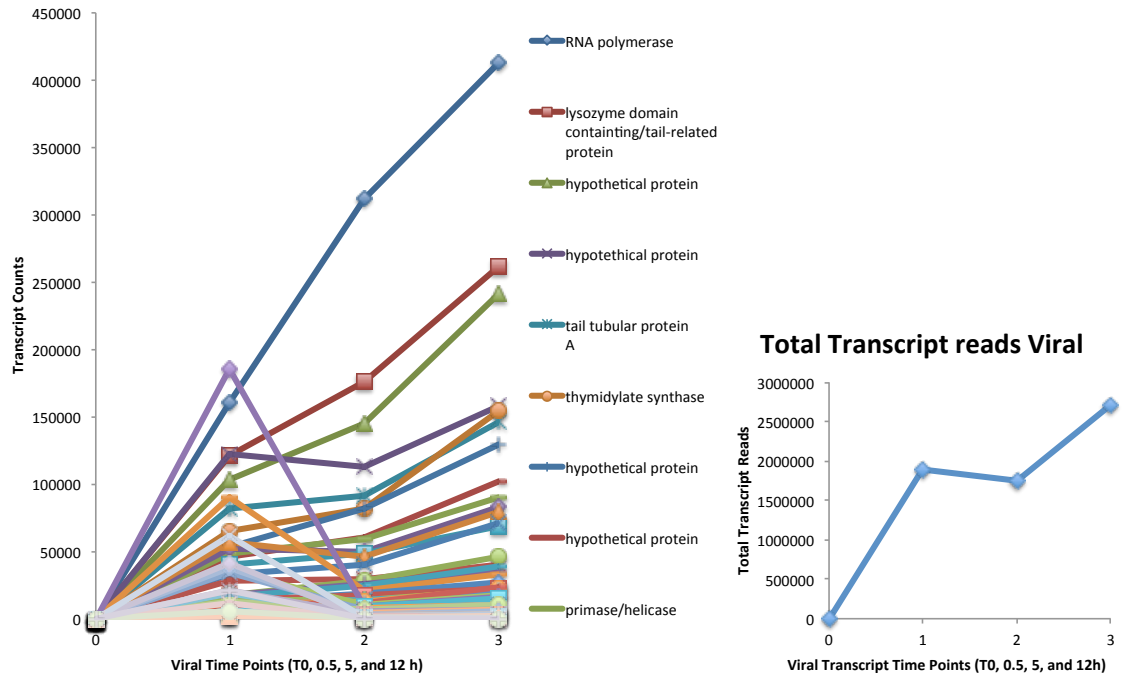


Figure 5.S4 Total viral transcripts broken down into clusters based on amount of reads mapping back to specific genes. Top nine most abundant viral genes are listed. Total reads mapping back to viral genes increased over the infection time series (0 min, 30 min, 5 h, and 12 h).

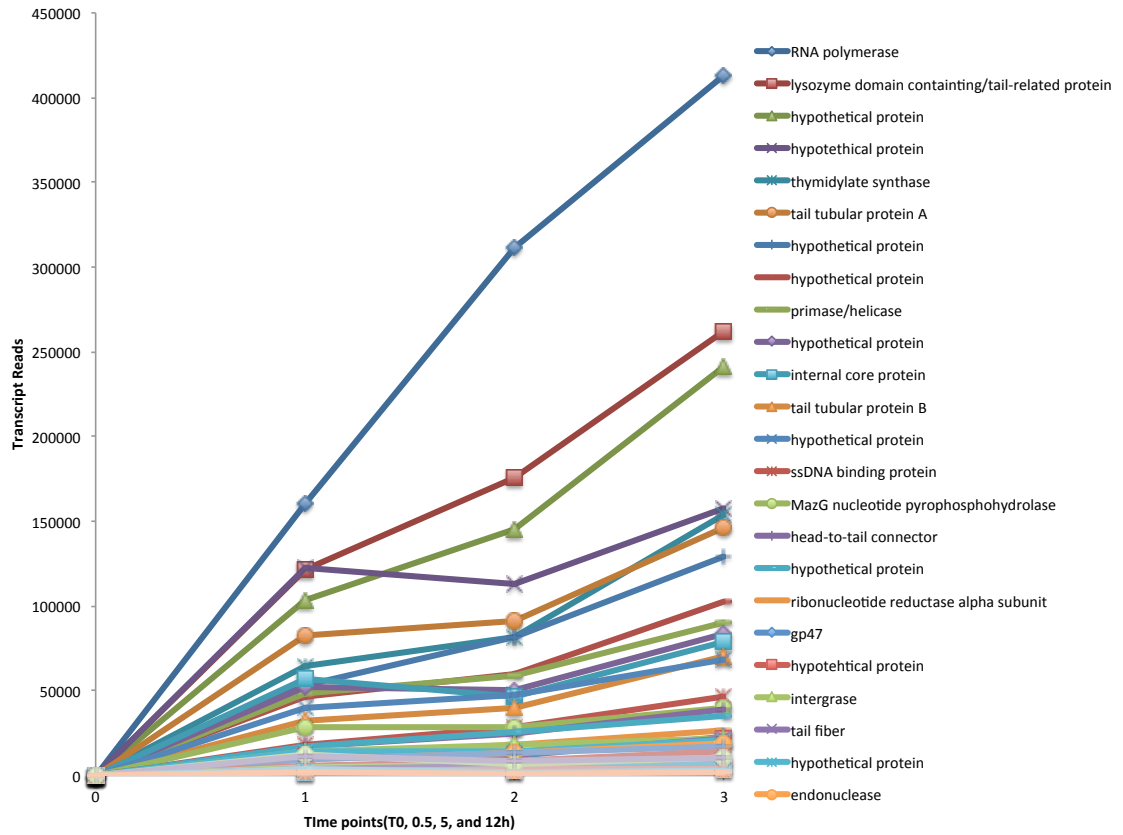


Figure 5.S5 Viral transcript cluster 1 reads mapping back to viral genes over the infection time series (0 min, 30 min, 5 h, and 12 h).

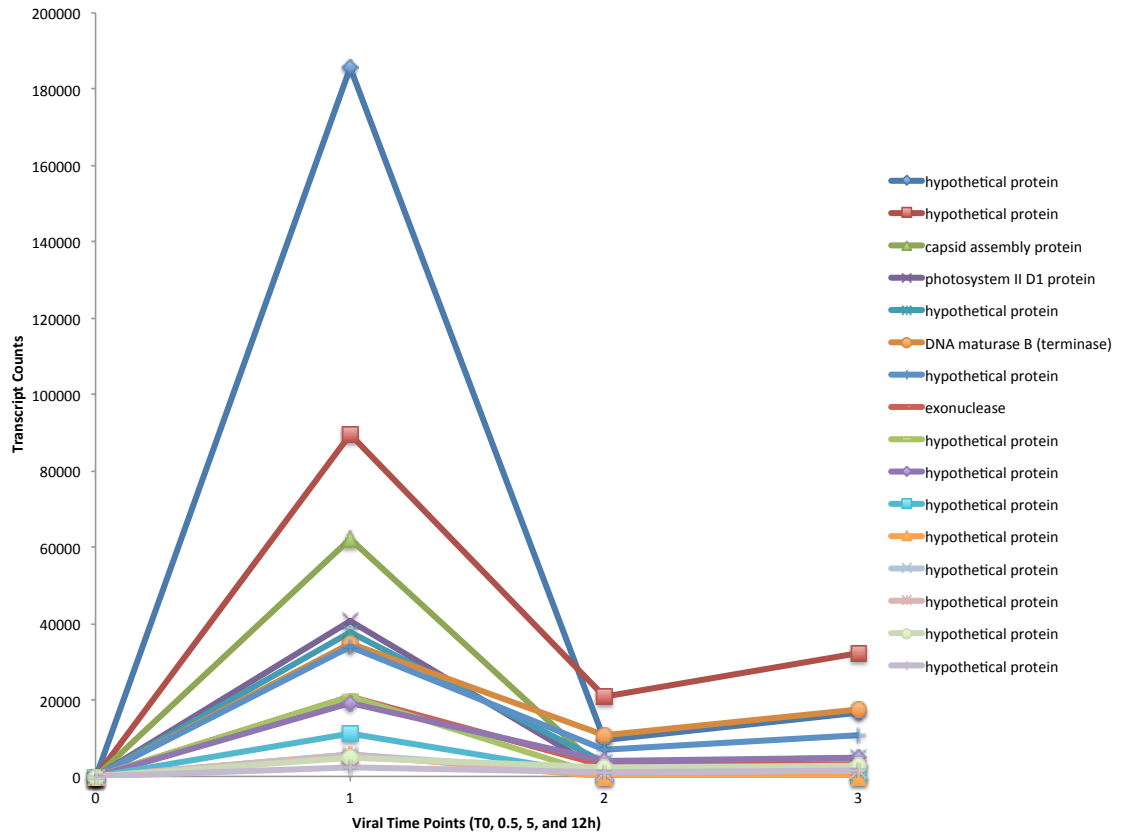


Figure 5.S6 Viral transcript cluster 2 reads mapping back to viral genes over the infection time series (0 min, 30 min, 5 h, and 12 h).

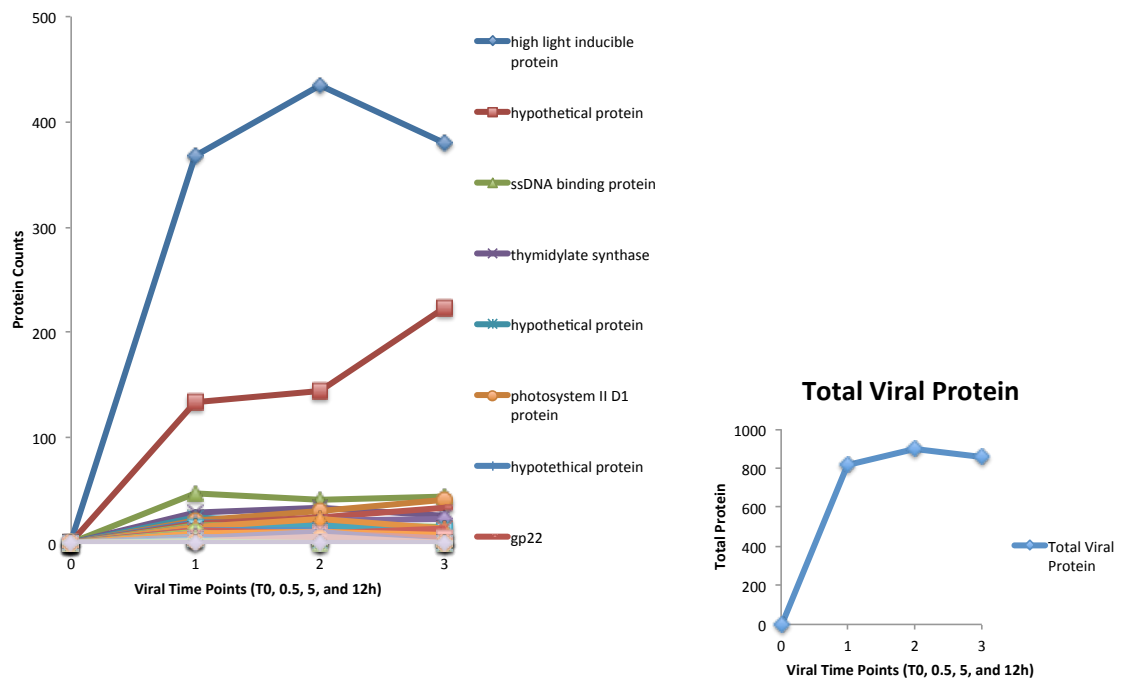


Figure 5.S7 Total viral protein broken down into clusters based on amount of counts mapping back to specific genes. Total counts mapping back to viral proteins was relatively constant over the infection time series (0 min, 30 min, 5 h, and 12 h).

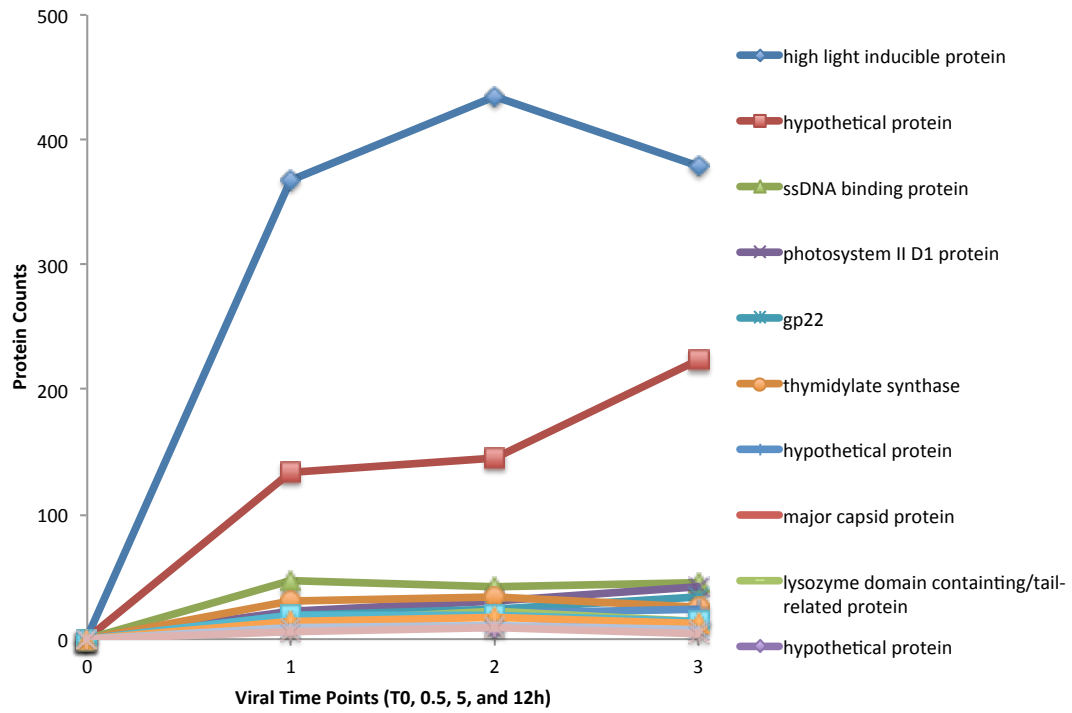


Figure 5.S8 Viral protein cluster 1 counts mapping back to viral proteins over the infection time series (0 min, 30 min, 5 h, and 12 h).

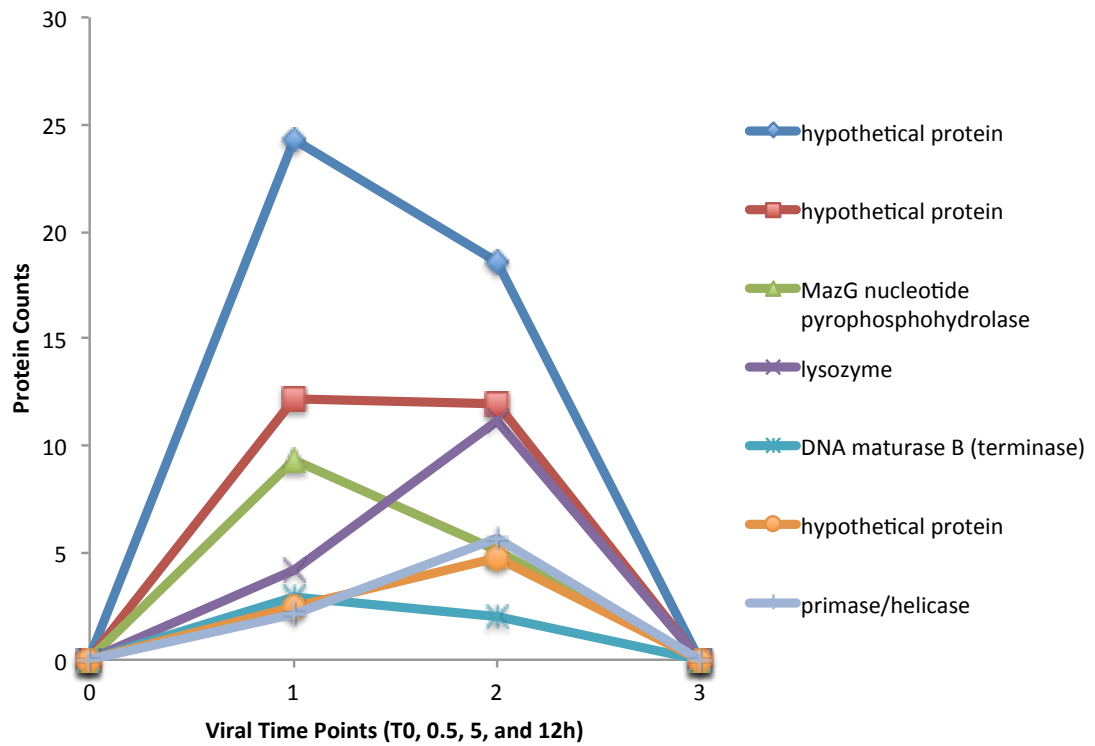


Figure 5.S9 Viral protein cluster 2 counts mapping back to viral proteins over the infection time series (0 min, 30 min, 5 h, and 12 h).

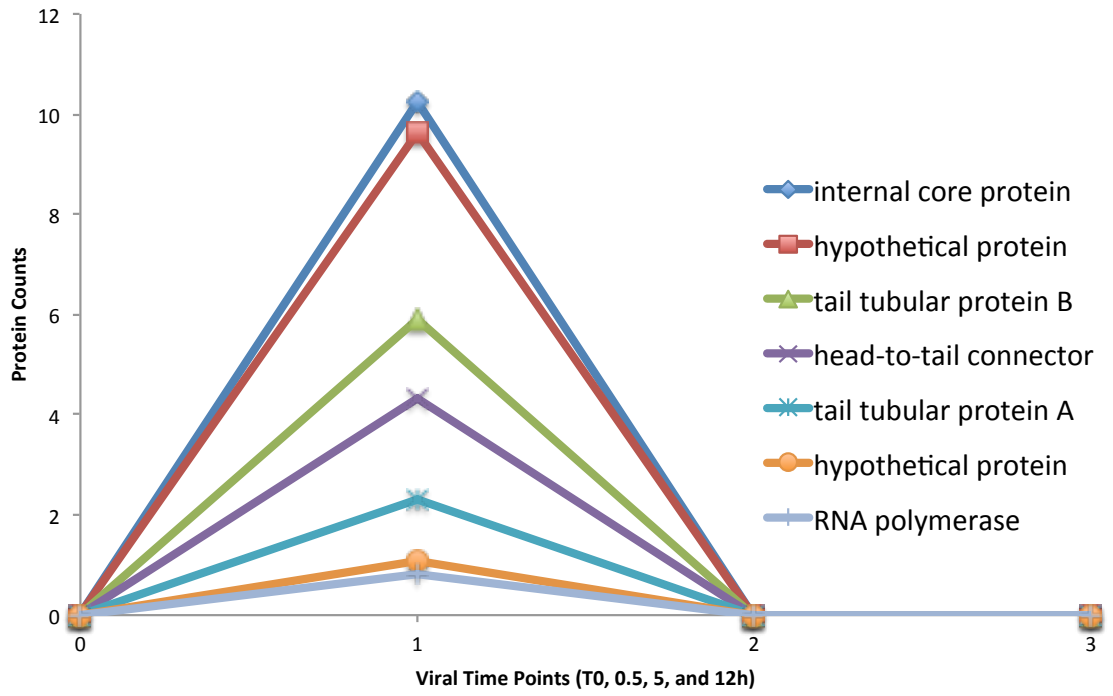
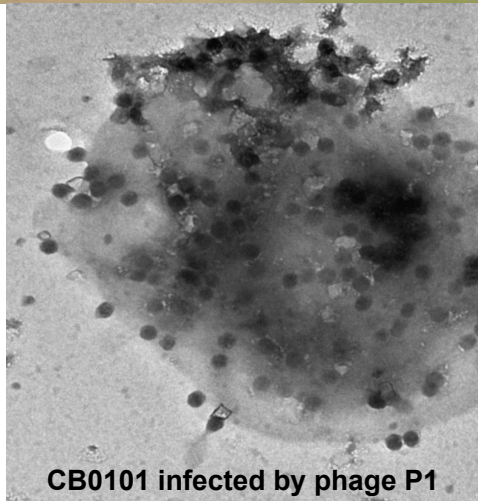
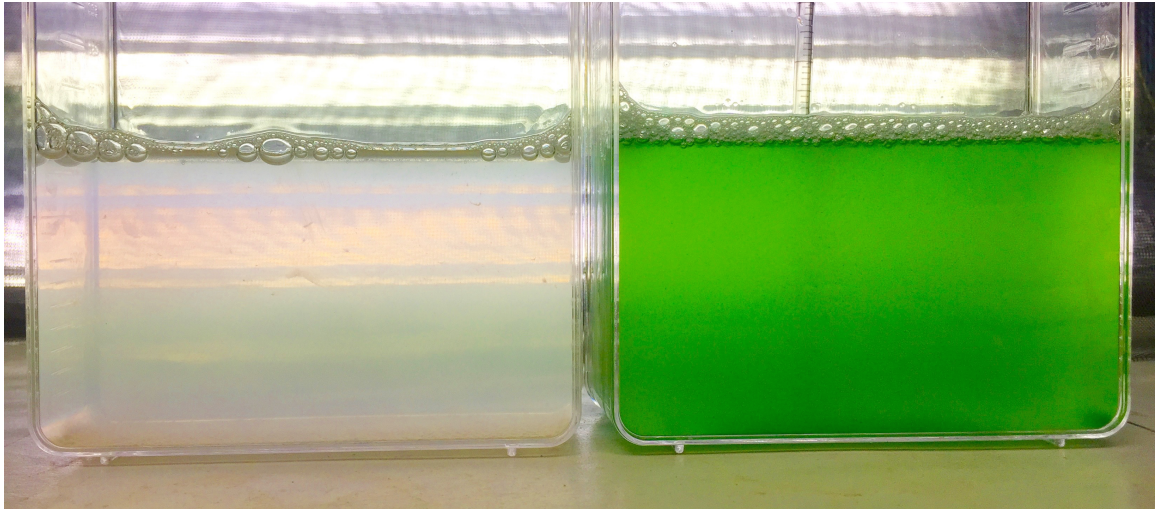


Figure 5.S10 Viral protein cluster 3 counts mapping back to viral proteins over the infection time series (0 min, 30 min, 5 h, and 12 h).



CB0101 infected by phage P1

Figure 5.S11 Culture of CB0101 before and after infection by cyanopodovirus S-CBP1. Top left after 48 h of infection. Top right before infection. Bottom, TEM image of S-CBP1 bursting from CB0101 cell.

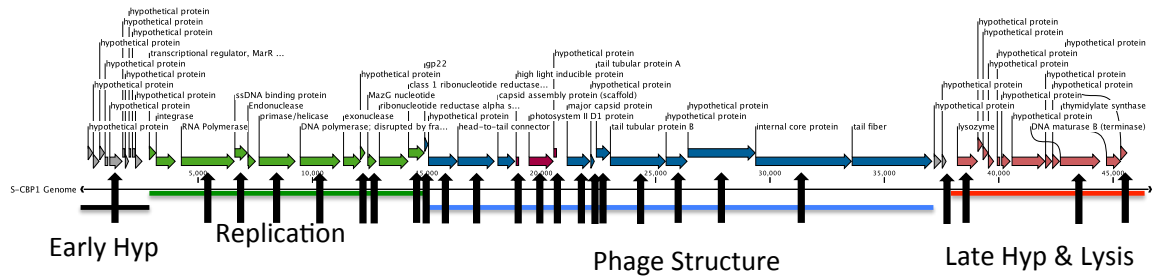


Figure 5.S12 Genome map of cyanopodovirus S-CBP1. Arrows note which proteins were identified through mass spec.

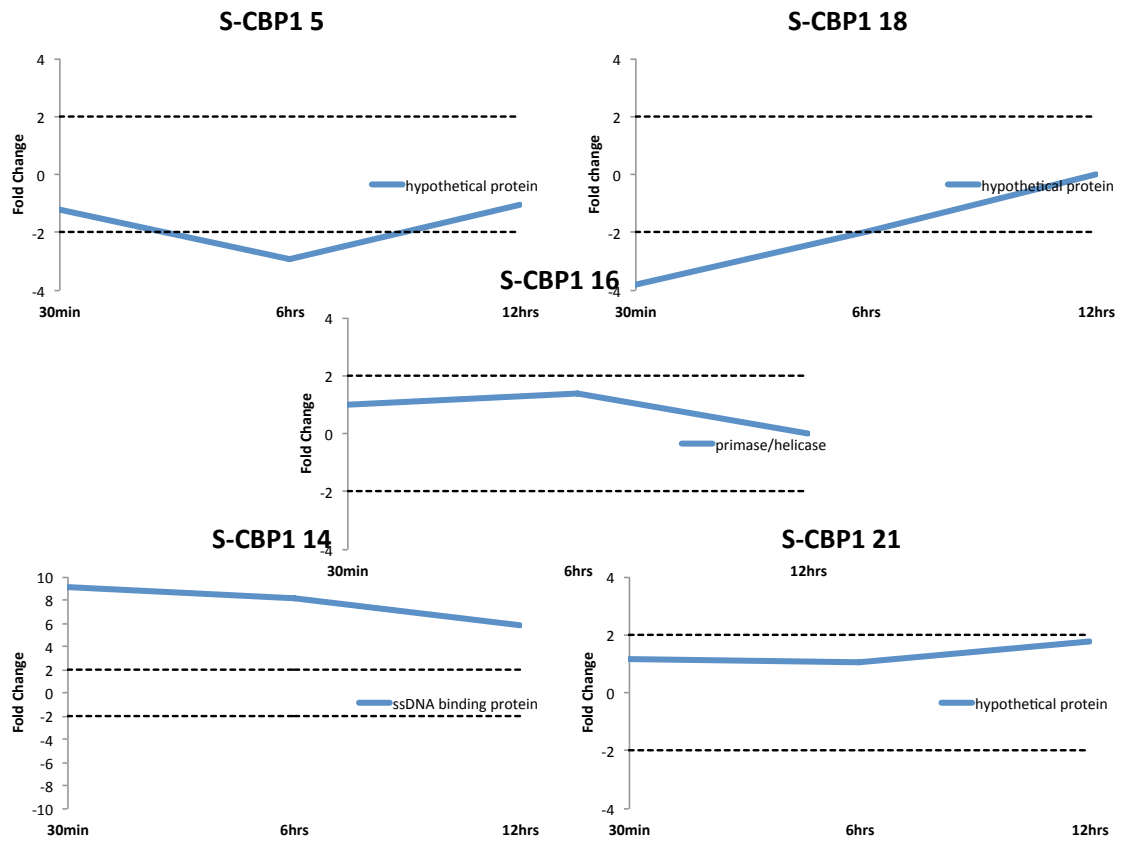


Figure 5.S13 S-CBP1 protein response at 30 min, 5 h, and 12 h post infection. Dashed lines delineate 2-fold increase in protein expression compared to control.

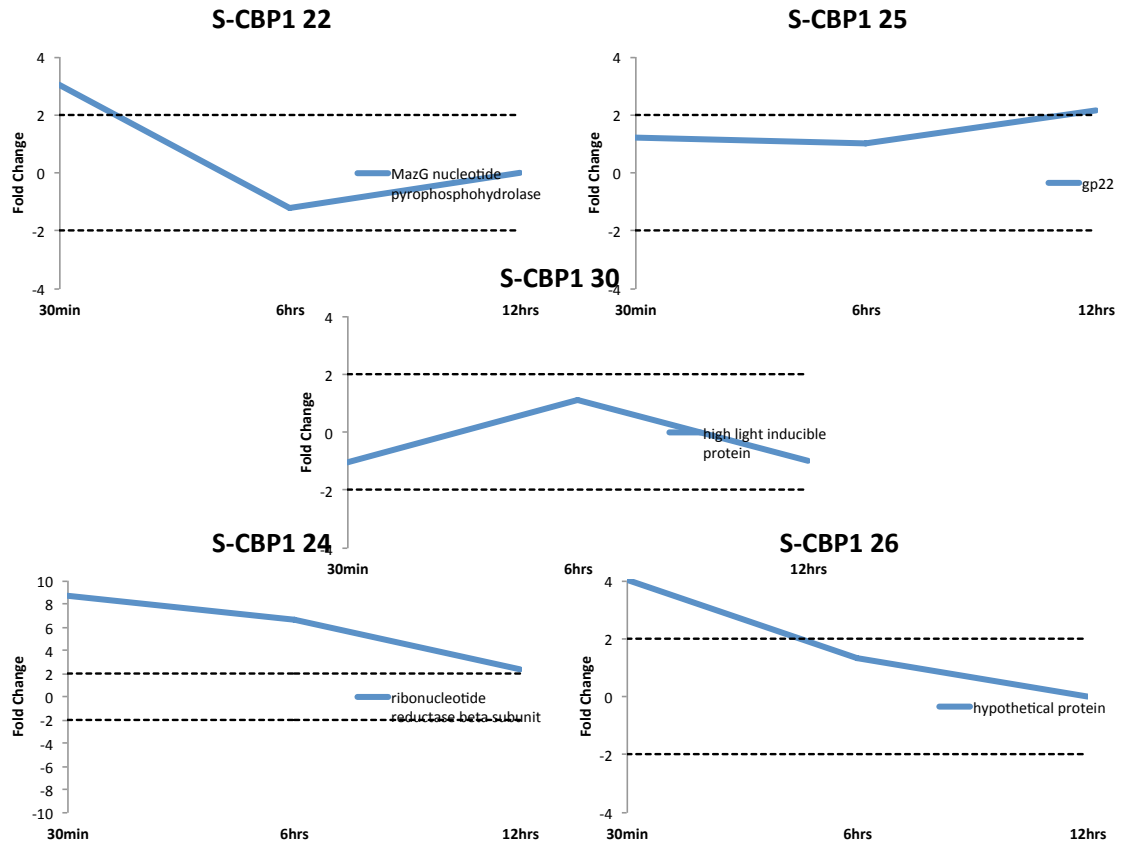


Figure 5.S14 S-CBP1 protein response at 30 min, 5 h, and 12 h post infection. Dashed lines delineate 2-fold increase in protein expression compared to control.

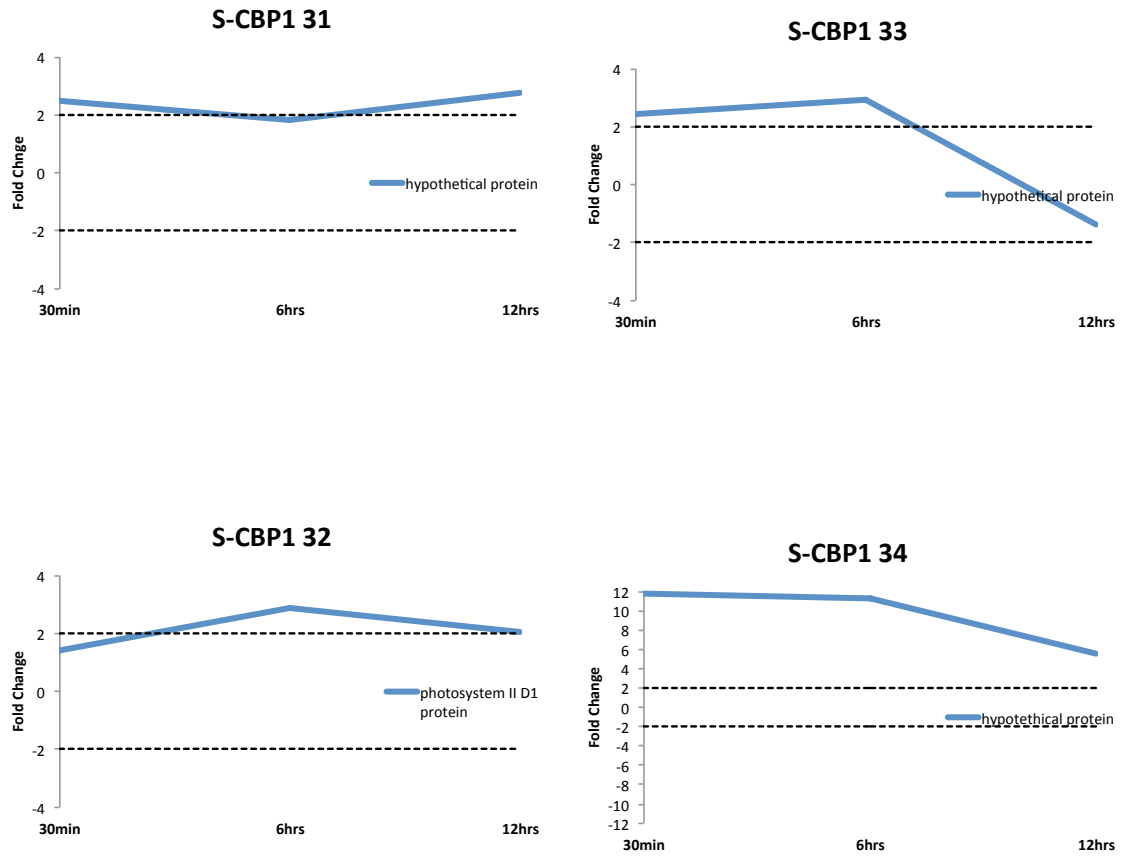


Figure 5.S15 S-CBP1 protein response at 30 min, 5 h, and 12 h post infection. Dashed lines delineate 2 fold increase in protein expression compared to control.

0.68574024 1.15307198 0.67331505 1.1706031 0.49760406 1.24040032 0.24791341 0.88088914 211 10.4666037 24.5 5.7222847 200 12.0286681 651 8.57227034

Gene	Protein	Accession	Length
Pepide methionine sulfoxidase 1	Peptide (glutathionylated) L16	AF274934	478
Lipid A-disaccharide synthase EC 2.4.1.1	Phosphatidylserine acetyltransferase	AF274935	494
Acy (beta-carrier) protein	Acy (beta-carrier) protein	AF274936	451
3-hydroxyacyl-beta-carrier-1-acyl-FAS	3-hydroxyacyl-beta-carrier-1-acyl-FAS	AF274937	472
UDP-3-O-(beta-D-mannosyl)-UDP-2-oxo-L-valine	UDP-3-O-(beta-D-mannosyl)-UDP-2-oxo-L-valine	AF274938	488
Outer membrane protein C	Outer membrane protein C	AF274939	445
Phosphotyrosine phosphatase 1 (Lyn)	Phosphotyrosine phosphatase 1 (Lyn)	AF274940	470
Oxidoreductase	Oxidoreductase	AF274941	460
Phosphorylase	Phosphorylase	AF274942	450
Phenylalanyl-tRNA synthetase	Phenylalanyl-tRNA synthetase	AF274943	475
Phosphonate kinase II, beta isoenzyme	Phosphonate kinase II, beta isoenzyme	AF274944	485
Citric acid cycle protein	Citric acid cycle protein	AF274945	475
Citric acid cycle protein	Citric acid cycle protein	AF274946	475
Citric acid cycle protein	Citric acid cycle protein	AF274947	475
Citric acid cycle protein	Citric acid cycle protein	AF274948	475
LSU ribosomal protein L21	LSU ribosomal protein L21	AF274949	475
LSU ribosomal protein L21 (c1)	LSU ribosomal protein L21 (c1)	AF274950	475
SAM dependent methyltransferase (Ac)	SAM dependent methyltransferase (Ac)	AF274951	475
Puative pre60S protein	Puative pre60S protein	AF274952	475
RNA pseudouridine synthase CBSS-138	RNA pseudouridine synthase CBSS-138	AF274953	475
Glucosamine 6-phosphate	Glucosamine 6-phosphate	AF274954	475
Type II secretory pathway	Type II secretory pathway	AF274955	475
FIG02685	FIG02685	AF274956	475
Specific sporulation protein SpoII	Specific sporulation protein SpoII	AF274957	475
Ribonuclease Z (EC 3.1.26)	Ribonuclease Z (EC 3.1.26)	AF274958	475
Phageoelmer protein	Phageoelmer protein	AF274959	475
FIG114972	FIG114972	AF274960	475
soluble (P2-25) ferredoxin id(1)	soluble (P2-25) ferredoxin id(1)	AF274961	475
Ribosomal protein L11	Ribosomal protein L11	AF274962	475
D-3-phosphoglycerate dehydrogenase	D-3-phosphoglycerate dehydrogenase	AF274963	475
FIG114967	FIG114967	AF274964	475
Thylakoid phosphatase	Thylakoid phosphatase	AF274965	475
UDP-N-acetylmuramylalanine-D-glucosamine	UDP-N-acetylmuramylalanine-D-glucosamine	AF274966	475
Protein of unknown function DUF5	Protein of unknown function DUF5	AF274967	475
hypothetical protein	hypothetical protein	AF274968	475
hypothetical protein	hypothetical protein	AF274969	475
hypothetical protein	hypothetical protein	AF274970	475
2-hydroxy-3-oxopropionate id(1)	2-hydroxy-3-oxopropionate id(1)	AF274971	475
DNA-directed RNA polymerase	DNA-directed RNA polymerase	AF274972	475
Possible Ribosomal protein	Possible Ribosomal protein	AF274973	475
Chaperone protein DnaK	Chaperone protein DnaK	AF274974	475
FIG115136	FIG115136	AF274975	475
Ferredoxin-nitrogen reductase	Ferredoxin-nitrogen reductase	AF274976	475
2,3-bisphosphoglycerate-independent	2,3-bisphosphoglycerate-independent	AF274977	475
Preprotein translocase subunit CBSS-31	Preprotein translocase subunit CBSS-31	AF274978	475
Pepide methionine sulfoxidase 1 (2)	Pepide methionine sulfoxidase 1 (2)	AF274979	478
Amino acid transport	Amino acid transport	AF274980	475
Heat shock protein 60 family id(1)	Heat shock protein 60 family id(1)	AF274981	475
Heat shock protein 60 family id(2)	Heat shock protein 60 family id(2)	AF274982	475
ATP synthase beta chain (Euo1)	ATP synthase beta chain (Euo1)	AF274983	475
ATP synthase epsilon chain (Euo1)	ATP synthase epsilon chain (Euo1)	AF274984	475
Multimeric transpeptidase id(2)	Multimeric transpeptidase id(2)	AF274985	475
hypothetical protein	hypothetical protein	AF274986	475
hypothetical protein	hypothetical protein	AF274987	475
hypothetical protein	hypothetical protein	AF274988	475
hypothetical protein	hypothetical protein	AF274989	475
hypothetical protein	hypothetical protein	AF274990	475
hypothetical protein	hypothetical protein	AF274991	475
hypothetical protein	hypothetical protein	AF274992	475
hypothetical protein	hypothetical protein	AF274993	475
hypothetical protein	hypothetical protein	AF274994	475
hypothetical protein	hypothetical protein	AF274995	475
hypothetical protein	hypothetical protein	AF274996	475
hypothetical protein	hypothetical protein	AF274997	475
hypothetical protein	hypothetical protein	AF274998	475
hypothetical protein	hypothetical protein	AF274999	475
hypothetical protein	hypothetical protein	AF275000	475

1.62788377 0.06170953 1.10229177 10 7.50000000 8 7.50000000 16 7.68007312 21 17.914088 70 81.9767124 608 205.932071 0.02796399 1.8825001 0.05914651 3.4412282

2.9339059 4.0956405 2.46052931 8 3.18704379 6 3.20240347 11 3.48213358 14 7.00239286 1079 205.201893 0.202015 1.7485791 0.5467347 1.5106134

Table with multiple columns containing numerical data, likely representing coordinates or indices in a grid. The table is organized in a regular grid format with varying column widths.

0.8415648 1.57979209 0.32960898 1.4206238 0.8171931 1.07042999 0.5050767 1.327152 21 1.29038126 11 1.80931856 11 1.07003031 24 0.40930831 18 6.71935932 14 7.3737875 27 1.6278005

30977185 1.08809E-01 3.757874... 41 5.42471617 42 5.88958133... 4067 73.7455813 3812 1.97797331... 0.13227843 0.1566845 0.0686485 1.14789747... 0.5079141 1.1792165 0.5168149 1.24157477... 0.2371786 0.1866426 0.2063339 1.0219693

0.38041815 1.3974904 0.05623283 2.04546312 0.07191153 2.8732753 0.17964654 1.16713301 34. 2.40341926 13. 1.88563311 12. 1.04832412 26. 4.07189105 56. 2.8337024

Table with multiple columns of numerical data, likely representing a coordinate grid or a data matrix. The data is organized in a regular grid pattern across the page.

Table with multiple columns containing numerical data, likely representing a coordinate grid or a data matrix. The table contains a dense grid of numbers with some missing values (zeros).

Cytochrome c oxidase polypeptide I1) Terminus ... 22999 ... 200169399 ...

hypothetical protein ... 23001 ...

Glyoxylate hydratase, family 2 ... 23002 ...

phosphotransferase I subunit VII (PtcVII) ... 23003 ...

Copper translocating P-type ATPase ... 23004 ...

conserved hypothetical membrane protein ... 23005 ...

FIG014441 hypothetical protein ... 23006 ...

FIG024026 hypothetical protein ... 23007 ...

Possible protein precursor ... 23008 ...

hypothetical protein ... 23009 ...

hypothetical protein ... 231 ...

Possible protein precursor ... 2311 ...

possible heat shock protein DnaJ ... 2312 ...

FIG014978 hypothetical protein ... 2313 ...

Nan(Y)H1) antiporter ... 2314 ...

Sr509 protein ... 2315 ...

Cell division inhibitor ... 2316 ...

FIG015026 hypothetical protein ... 2317 ...

TPR repeat protein, specific for cyanobact ... 2318 ...

FIG015032 hypothetical protein ... 2319 ...

Pro-zeaxanthin desaturase, pro-zeaxanthin ... 2320 ...

Oligopeptide cyclase/peptidyl transferase ... 2322 ...

Uroerythrin/lipo II synthase/iron home ... 2323 ...

hypothetical protein ... 2324 ...

Possible glycoyltransferase 2 fused to ... 2325 ...

Beta-hexosaminidase (EC: iCow1) ... 2326 ...

Beta-hexosaminidase (EC: iCow1) ... 2327 ...

Ribosome-binding factor A ... 2328 ...

Possible protein precursor ... 2329 ...

Possible DNA-binding response regulator ... 2331 ...

FIG015032e hypothetical protein ... 2330 ...

Glutathione S-transferase ... 2331 ...

Skimstone kinase (EC: 2.7.1) ... 2332 ...

FIG0149671 hypothetical protein ... 2333 ...

6-phospho-1-tryptophan hydroxylase ... 2334 ...

FIG0151977 hypothetical protein ... 2335 ...

6-aminocaproate decarboxylase ... 2336 ...

FIG015102e hypothetical protein ... 2337 ...

plein-binding family ... 2338 ...

Hypothetical with similarity to ... 2339 ...

FIG0240827 hypothetical protein ... 2340 ...

hypothetical protein ... 2341 ...

Multidrug efflux transporter ... 2342 ...

Ribosome protein S12a ... 2343 ...

L-asparatase oxidase (EC: 1.4.1) ... 2344 ...

FIG0150424 hypothetical protein ... 2345 ...

Phosphatase 112 ... 2346 ...

ATP-dependent DNA helicase I(1) ... 2347 ...

Udecaprenyl-phosphatase ... 2348 ...

Possible Fe-S oxidoreductase ... 2349 ...

FIG0150383 hypothetical protein ... 2350 ...

Outer membrane efflux protein ... 2351 ...

FIG0151508 hypothetical protein ... 2352 ...

Histidinol phosphate [histidine] ... 2353 ...

FIG0211150 ... 2354 ...

Hypothetical protein ... 2355 ...

Calcium/calmodulin-dependent protein ... 2356 ...

2,3-diketio-5-methylpyridoxal ... 2357 ...

Methylcrotonyl-CoA lyase ... 2358 ...

Formanoyltransferase ... 2359 ...

phosphotransferase I subunit VIII ... 2360 ...

FIG0150508 hypothetical protein ... 2361 ...

alginate regulatory protein AlgP ... 2362 ...

ATP-dependent DNA helicase I(2) ... 2363 ...

Aldehyde dehydrogenase ... 2364 ...

hypothetical protein ... 2365 ...

Adenylyl cyclase (EC: 4.6.1) ... 2366 ...

Two-component response regulator ... 2367 ...

3-dehydroquinate dehydratase ... 2368 ...

RNA[met2(9)](6) hydrolysid(1) ... 2369 ...

hypothetical protein ... 2370 ...

hypothetical protein ... 2371 ...

hypothetical protein ... 2372 ...

possible Keim motif ... 2373 ...

hypothetical protein ... 2374 ...

hypothetical protein ... 2375 ...

hypothetical protein ... 2376 ...

putative membrane protein ... 2377 ...

hypothetical protein ... 2378 ...

hypothetical protein ... 2379 ...

Cobalt zinc-cadmium resistant protein ... 2380 ...

FIG0150686e hypothetical protein ... 2381 ...

Two-component system ... 2382 ...

hypothetical protein ... 2383 ...

hypothetical protein ... 2384 ...

hypothetical protein ... 2385 ...

hypothetical protein ... 2386 ...

CopG protein ... 2387 ...

FIG0151978 hypothetical protein ... 2388 ...

hypothetical protein ... 2389 ...

hypothetical protein ... 2390 ...

hypothetical protein ... 2391 ...

hypothetical protein ... 2392 ...

hypothetical protein ... 2393 ...

hypothetical protein ... 2394 ...

hypothetical protein ... 2395 ...

putative lysozyme (Lys) protein ... 2396 ...

hypothetical protein ... 2397 ...

AbrB family transcriptional regulator ... 2398 ...

hypothetical protein ... 2399 ...

Glucose-methanol choline (cwi-1) ... 2400 ...

4'-diphosphonorelate ... 2401 ...

hypothetical protein ... 2402 ...

hypothetical protein ... 2403 ...

hypothetical protein ... 2404 ...

Cation-transporting ATPase, E1-E2 ... 2405 ...

Mobile element protein ... 2406 ...

Mobile element protein ... 2407 ...

Cation-transporting ATPase ... 2408 ...

hypothetical protein ... 2409 ...

Beta-glucosidase (EC: 3.2.1) ... 2410 ...

Beta-hexosaminidase (EC: 3.2.1) ... 2411 ...

Large-conductance mechanosensitive ... 2412 ...

GTP pyrophosphatase (EC: 2.7.6.5) ... 2413 ...

hypothetical protein ... 2414 ...

hypothetical protein ... 2415 ...

hypothetical protein ... 2416 ...

hypothetical protein ... 2417 ...

hypothetical protein ... 2418 ...

hypothetical protein ... 2419 ...

Glyoxylate 4,4'-diphosphonorelate (cwi-1) ... 2420 ...

hypothetical protein ... 2421 ...

hypothetical protein ... 2422 ...

hypothetical protein ... 2423 ...

hypothetical protein ... 2424 ...

hypothetical protein ... 2425 ...

hypothetical protein ... 2426 ...

2-phosphoryl-6-methylthiohydrosulfonate ... 2427 ...

Mobile element protein ... 2428 ...

hypothetical protein ... 2429 ...

hypothetical protein ... 2430 ...

hypothetical protein ... 2431 ...

hypothetical protein ... 2432 ...

0.1792489 1.756214 1.10693889 0.49140017 1.00783306 0.40999946 1.80505683 30 8.38571995 19 4.121040395 41 1.71908012 8 2.10093986 63 4.44911744 41 1.0958993 1 0.4111185 28 4.2592494

Table with multiple columns of numerical data, likely representing a large dataset or a list of identifiers and values. The data is organized in a grid-like structure with varying column widths and some missing values.

0.38995 1.0919588 0.58818933 1.3047995 0.63239751 1.288922 0.2698061 1.6817459 26 5.21800271 16 1.84515577 111 1.25245157 13 1.5442845 10 4.78729986 20 1.8518972

Table with multiple columns of numerical data, likely representing a coordinate grid or a data matrix. The data is organized in a regular grid pattern across the page.

FIG0115447: hypothetical protein...
FIG11347: putative deacetylase subunit...
ATP-dependent Cys protease...
7-8-dimethyl-8-hydroxy-5-oxo-Coenzyme A...
FIG014311: hypothetical protein...
ATP-dependent Cys protease...
Rhodanese-related sulfufurilase...
Amidase...
FIG014323: hypothetical protein...
Cell division protein FtsH...
FIG014324: hypothetical protein...
Cell division protein FtsX...
FIG014328: hypothetical protein...
FIG015070: hypothetical protein...
Bltn carnitine...
Pyruvate kinase...
Xenopus...
Zn-dependent hydrolase...
Transcriptional regulator...
Bovine...
Phytochrome...
FIG015086: hypothetical protein...
FIG015087: hypothetical protein...
FIG015088: hypothetical protein...
FIG015089: hypothetical protein...
FIG015090: hypothetical protein...
FIG015091: hypothetical protein...
FIG015092: hypothetical protein...
FIG015093: hypothetical protein...
FIG015094: hypothetical protein...
FIG015095: hypothetical protein...
FIG015096: hypothetical protein...
FIG015097: hypothetical protein...
FIG015098: hypothetical protein...
FIG015099: hypothetical protein...
FIG015100: hypothetical protein...
FIG015101: hypothetical protein...
FIG015102: hypothetical protein...
FIG015103: hypothetical protein...
FIG015104: hypothetical protein...
FIG015105: hypothetical protein...
FIG015106: hypothetical protein...
FIG015107: hypothetical protein...
FIG015108: hypothetical protein...
FIG015109: hypothetical protein...
FIG015110: hypothetical protein...
FIG015111: hypothetical protein...
FIG015112: hypothetical protein...
FIG015113: hypothetical protein...
FIG015114: hypothetical protein...
FIG015115: hypothetical protein...
FIG015116: hypothetical protein...
FIG015117: hypothetical protein...
FIG015118: hypothetical protein...
FIG015119: hypothetical protein...
FIG015120: hypothetical protein...
FIG015121: hypothetical protein...
FIG015122: hypothetical protein...
FIG015123: hypothetical protein...
FIG015124: hypothetical protein...
FIG015125: hypothetical protein...
FIG015126: hypothetical protein...
FIG015127: hypothetical protein...
FIG015128: hypothetical protein...
FIG015129: hypothetical protein...
FIG015130: hypothetical protein...
FIG015131: hypothetical protein...
FIG015132: hypothetical protein...
FIG015133: hypothetical protein...
FIG015134: hypothetical protein...
FIG015135: hypothetical protein...
FIG015136: hypothetical protein...
FIG015137: hypothetical protein...
FIG015138: hypothetical protein...
FIG015139: hypothetical protein...
FIG015140: hypothetical protein...
FIG015141: hypothetical protein...
FIG015142: hypothetical protein...
FIG015143: hypothetical protein...
FIG015144: hypothetical protein...
FIG015145: hypothetical protein...
FIG015146: hypothetical protein...
FIG015147: hypothetical protein...
FIG015148: hypothetical protein...
FIG015149: hypothetical protein...
FIG015150: hypothetical protein...
FIG015151: hypothetical protein...
FIG015152: hypothetical protein...
FIG015153: hypothetical protein...
FIG015154: hypothetical protein...
FIG015155: hypothetical protein...
FIG015156: hypothetical protein...
FIG015157: hypothetical protein...
FIG015158: hypothetical protein...
FIG015159: hypothetical protein...
FIG015160: hypothetical protein...
FIG015161: hypothetical protein...
FIG015162: hypothetical protein...
FIG015163: hypothetical protein...
FIG015164: hypothetical protein...
FIG015165: hypothetical protein...
FIG015166: hypothetical protein...
FIG015167: hypothetical protein...
FIG015168: hypothetical protein...
FIG015169: hypothetical protein...
FIG015170: hypothetical protein...
FIG015171: hypothetical protein...
FIG015172: hypothetical protein...
FIG015173: hypothetical protein...
FIG015174: hypothetical protein...
FIG015175: hypothetical protein...
FIG015176: hypothetical protein...
FIG015177: hypothetical protein...
FIG015178: hypothetical protein...
FIG015179: hypothetical protein...
FIG015180: hypothetical protein...
FIG015181: hypothetical protein...
FIG015182: hypothetical protein...
FIG015183: hypothetical protein...
FIG015184: hypothetical protein...
FIG015185: hypothetical protein...
FIG015186: hypothetical protein...
FIG015187: hypothetical protein...
FIG015188: hypothetical protein...
FIG015189: hypothetical protein...
FIG015190: hypothetical protein...
FIG015191: hypothetical protein...
FIG015192: hypothetical protein...
FIG015193: hypothetical protein...
FIG015194: hypothetical protein...
FIG015195: hypothetical protein...
FIG015196: hypothetical protein...
FIG015197: hypothetical protein...
FIG015198: hypothetical protein...
FIG015199: hypothetical protein...
FIG015200: hypothetical protein...

Table with multiple columns of numerical data, organized in approximately 10 vertical columns per page.

gsyne_997	Synechococcus_assembled_mapped	conserved hypothetical protein	GO:0008150,GO:0003674,GCPFD4087.9	867235	868366	-1	156	fig11129.58.png.989
gsyne_998	Synechococcus_assembled_mapped	conserved hypothetical protein	GO:0008150,GO:0003674,GUniref100_B	868414	869269	-1	156	fig11129.58.png.990
gsyne_999	Synechococcus_assembled_mapped	sodium/hydrogen exchanger family protein	PF00999.10	869281	870616	-1	145	fig11129.58.png.991

hypothetical protein
 hypothetical protein
 Nav1H+ antiporter

atggcgaaacccMAEPTSLASLEAGFEREARLDEPFVLTVAAGLIATLGLLADSAVVGAMLIAPWILPLRSASFAMIDGRNLNLAGRALLTAVGVTTIALSAGLGLWLAQgyne_997 0.033427729, 1.39711347, 0.16484817, 1.24543471, 0.73042551, -1.0613432, 0.42177213, -1.1217878, 0.00143836
 atggccctdctogtaMPSPYATARIALATRHGKERALALPFRWGLGAEIALCPADTDQLGTFSGEIPRLADAFSTCKAKARLGEASGLQLGASEAFGPHAMPALAVQgyne_998 0.267613382, -1.4065184, 0.59551462, -1.21263751, 1.67341-07, 3.007531073, 0.63369216, 1.15988362, 3.50246-26
 atgctcccccggMLPPLALAEISPHGLEVAETLQVGRFLVFAAKAAEIVNRGLFTLGEIAGVILGVSGLIHVPDYGQISDALLSLLGLAEIRPEEVAEYNGgyne_999 0.0543117348, -1.3593088, 0.59637536, -1.01401091, 0.36839876, 1.02987119, 0.04915447, 1.24206367, 0.01038692

Gene	Nitrogen	Phosphate	Zinc	Viral 30min	Viral Shr	Viral 12hr
YefM protein (antitoxin to YoeB)	-1.6	-1.4	-1.3	-1.2	-1.1	-1.2
RetE/SSDE replicon stabilization toxin	2.2	1.5	2.1	-1.0	-2.0	-1.6
Death on curing protein, Doc toxin	1.8	1.2	2.7	-1.8	1.3	-1.3
VapC toxin protein	1.4	-1.2	1.7	-1.0	1.4	-1.1
RetE/SSD replicon stabilization protein (antitoxin to RetE/SSDE)	-1.0	-1.0	-2.6	1.0	-1.0	-2.3
RetE/SSDE replicon stabilization toxin	2.3	3.1	30.6	2.3	-1.3	1.7
Death on curing protein, Doc toxin	-1.5	-1.3	1.1	-1.4	-1.4	-1.3
Pleasant host death protein, Pho antitoxin	40.4	24.4	191.0	1.0	1.0	1.0

-1.48281669	0.02988919	-1.3218337	58	8.7701884	34	6.49824576	93	11.5511884	185	17.9154821	86	14.7892566	119	14.3129862	339	80.0321341	2843	193.610629	0.4349732	1.24171955	0.65472971	1.15698159
-2.23011921	2.7287121	3.6470204	14	2.80030609	9	2.27539194	15	2.46451502	20	2.56202437	17	3.86717036	17	2.7047598	263	82.1328586	2311	208.184682	0.19923997	-2.1664609	0.91635801	1.09894756
-1.40372450	0.76248667	1.04522834	53	6.78966018	56	9.06747948	67	7.95917443	106	8.69648947	78	11.3638060	117	11.9230378	410	82.0930202	3982	229.759108	0.18013844	-1.51925096	0.14731198	-1.5913801

0.04306582, 1.77833302, 0.80701492, -1.0732405, 0.22605998, 1.43215352, 0.14681343, 1.53704521,
0.62219660, -1.19874911, 0.16417487, 2.38082695, 0.42597742, 1.80726798, 0.54720613, -1.3173624,
0.03487772, -2.089865, 0.91127247, -1.04747691, 0.46814851, -1.3780364, 0.57726501, -1.3108835

58, 5.33780693, 36, 3.94236862, 36, 6.68936482, 47, 5.33662786, 30, 7.45631488, 44, 5.27717439, 33, 5.86579956, 115, 7.56342981,
14, 1.70432339, 9, 1.30974822, 1, 0.2457983, 7, 1.05138939, 8, 2.63020417, 9, 1.42786704, 1, 0.23513093, 18, 1.56599343,
53, -1.13240227, 57, 5.28823713, 17, 2.67616349, 38, 3.65592233, 13, 2.73733126, 37, 3.75951509, 15, 2.25844217, 28, 1.56012823