Comparative ecophysiology and the effects of the mobilome on *Microcystis* physiology and genomic architecture

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Gwendolyn Fern Stark
December 2024

ACKNOWLEDGEMENTS

I'd like to start my acknowledgments by thanking the Microbiology department for allowing me to come to UTK, I feel incredibly grateful for being given this opportunity. I would also like to thank my advisor, Dr. Wilhelm for allowing me to join the lab, and having a lot of patience as I developed into a better scientist under his guidance. My committee has also been great in providing me with scientific feedback that has helped to shape this dissertation, especially my second chapter which was a long time coming. There have also been various Wilhelm lab members who have provided valuable feedback on this work over the years. Notably, I need to thank Dr. Martin, as someone who also works with *Microcystis*, we have had many insightful scientific discussions over the years, which have led to numerous ideas for experiments we will never have the time for. I have also had a great experience with REU students the past two summers, which I need to thank Dr. Gary LeClair for. My time here would have probably been much more difficult if it wasn't for the cohort I started this journey with. Jordan, Liz and Jake have been there for the hard times and the fun times, and I need to thank and acknowledge them for that. I also need to thank the people behind the scenes, my family and friends.

ABSTRACT

Harmful cyanobacterial blooms (HCABs) are an increasing threat to the preservation of freshwater systems. One of the top offenders of HCAB formation in freshwater systems is *Microcystis aeruginosa*, a toxic cyanobacterium with a global distribution. The toxin produced by *Microcystis*, called microcystin, is a hepatotoxic secondary metabolite. While microcystin poses an ecological threat, why the compound is produced by *Microcystis* is still unknown. In Chapter 2, we performed an ecophysiological comparison of a toxic wildtype *Microcystis* strain, and a non-toxic *Microcystis* mutant to elucidate microcystin's role in the cell. This was done during chemostat growth under optimal and suboptimal growth temperatures, and involved ROS assays, photo-physiological measurements, microcystin quota, and a transcriptomic time series. In Chapter 3, we investigated transposon movement between the toxic *Microcystis* wildtype and non-toxic mutant, which were thought to be genetically identical. In some instances, natural genomic rearrangements that occurred during long-term batch culture growth altered phenotype, which may give insight into microcystin's cellular role. These investigations were done using comparative genomics approaches after genome resequencing, growth assays, PCR validation, and transcriptomics. In Chapter 4, we found a yet undescribed identical plasmid in our toxic *Microcystis* wildtype and non-toxic *Microcystis* mutant, despite being a well-used system to study microcystin production for >20 years, which is how long ago the mutant was created. In this chapter we used protein models to help elucidate the plasmid gene functions, investigated gene flow between the plasmid and genome, and looked at plasmid activity in vitro and in situ using transcriptomics and meta-transcriptomics. As little is known about *Microcystis* plasmids in general, we extended this work looking at plasmids in other *Microcystis* strains, and the role they play in host physiology, cyanophage infection, and their ecological relevance.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION 1
Cyanobacterial blooms overview 2
Microcystis- dominated blooms 2
Differences between wildtype *Microcystis aeruginosa PCC 7806* and the non-toxigenic Microcystis aeruginosa PCC 7806 ΔmcyB 5
High-light studies 6
Temperature 6
Low dissolved inorganic carbon (DIC) studies 7
Oxidative Stress 8
Microcystis Genomics 8
Microcystis plasmids 9
Dissertation Overview 10
References 11
CHAPTER TWO: MICROCYSTIN AIDS IN COLD TEMPERATURE ACCLIMATION: DIFFERENCES BETWEEN A TOXIC *MICROCYSTIS* WILDTYPE AND NON-TOXIC MUTANT 19
Publication Note 20
Abstract 21
Introduction 22
Methods 24
Continuous cultures 24
Cell abundance and size 25
Measurement of Microcystin 25
Measurements of photosynthetic physiology 25
Oxidative stress assessment 26
RNA extraction 26
RNA sequencing and sequence analyses 27
Data visualization and statistical analyses 28
Data availability 28
Results 28
Cell concentration in chemostats 28
Cellular microcystin content 29
Photosynthetic physiology assessment 29
Reactive oxygen species damage measurements via the CM-H2DCFDA Assay 32
Transcriptomic differences between the mutant and wildtype in warm conditions (T1 and T10). 34
Transcript differences between the mutant and wildtype in cold conditions (T2-T8). 34
Discussion 38
Microcystin production alters cold stress management strategies 41
Acknowledgements 43
References 44
Appendix 49
CHAPTER THREE: PHYSIOLOGICAL REPROGRAMMING BY THE *MICROCYSTIS* MOBILOME 60

Publication Note 61
Abstract 62
Background 63
Methods 64
Strains, DNA Extraction & genome sequencing, assembly, and annotation 64
RNA sequencing 65
Genome alignments and analyses 66
PCR validation 67
Reverse Transcriptase PCR 67
Sulfate Growth Assays 68
Statistical analysis 68
Results 69
Whole genome alignments of *Microcystis aeruginosa* PCC 7806 and the discovery of a large chromosome inversion 69
IS200 family transposases 71
IS1-like element IMSae3 family transposases 71
IS1634 Family Transposase with DUF4277 75
Phage-acquired genes affiliated with transposable elements 75
A miniature inverted transposable element (MITE) occurs multiple times in both genomes 78
Insertion of the 187-nt long MITE in a sulfate binding protein in the Δ*mcyB* genome 81
Discussion 83
Chromosome inversion 84
Gene activation and silencing by transposases 85
MITES are abundant across *Microcystis* genomes 86
Can MITEs alter gene expression in *Microcystis aeruginosa* PCC 7806? 87
Are there other roles for transposable elements in *Microcystis* genomes? 88
Conclusions 89
Appendix 98
CHAPTER FOUR: THE PHYSIOLOGICAL AND ECOLOGICAL RELEVANCE OF CONSERVED *MICROCYSTIS* PLASMIDS 116
Publication Note 117
Abstract 118
Introduction 119
Methods 121
Plasmid assembly and annotation 121
Plasmid copy number determination 121
RNA-seq analysis of the plasmid 121
RNA-seq analysis of the PCC 7806 and plasmid for Lake Erie metatranscriptomes 122
RNA-seq analysis of the NIES-298 P1 and P2 Plasmids 122
Finding Viral and Plasmid Contigs in the Lake Erie Co-Assembly & RNA seq 123
Presence and Gene flow of the PCC 7806 plasmid in *Microcystis* 124
Determining if conserved plasmids are in other *Microcystis spp.* 124
Finding plasmid-genome interactions 125
Finding Phage-Plasmids, and Phage-Plasmid-Genome interactions 125

Microcystis aeruginosa PCC7806 has a conserved plasmid that hasn't changed in over 20 years 126
Plasmid structure and similarity to the chromosome 127
PCC 7806 plasmid copy number 129
The PCC 7806 plasmid is found in other scaffold-level genome assemblies 129
The PCC 7806 plasmid is expressed in lab cultures and in situ 129
Proportion of total plasmid versus virus genes and transcripts in Lake Erie 131
Other *Microcystis* may also harbor conserved plasmids 131
What is the mechanism of conservation for *Microcystis* plasmids ? 133
PCC 7806 plasmid sequences are found in other *Microcystis* CRISPR spacer sequences. 134
Other *Microcystis spp.* plasmid sequences are widespread in CRISPR spacer regions 134
Microcystis putative phage-plasmids 134
Plasmid-Genome & Plasmid-Genome-Phage interactions 136
Discussion 140
Ecological relevance of *Microcystis spp.* plasmids and their ties to phage 141
Physiological relevance of *Microcystis* plasmids 143
Conclusions 144
Acknowledgments 145
References 146
Appendix 151
CHAPTER FIVE: CONCLUSIONS 163
References 169
APPENDIX 170
APPENDIX A: Sequencing the mutant genome 170
Publication note 171
Abstract 172
Announcement 172
References 174
APPENDIX B: Experimental Development of Co-culture Chemostat Competition Assays 177
Publication Note 178
Background 179
Methods 180
Continuous Cultures 180
Flow Cytometry 180
Initial mixing the Chemostats & Sampling Schedule 180
Nitrogen depletion treatments 181
qPCR Primers 181
Microcystin Toxin Measurements 183
Photophysiology Readings 183
References 184
VITA 185

LIST OF TABLES

Table A2.1 Primers used for qPCR of co-cultures 182

LIST OF FIGURES

Figure 2.1 PCC7806 wild-type and mutant responses to temperature change in chemostats 30
Figure 2.2 Non-photochemical quenching data for mutant and wildtype chemostats at T1 and T5 33
Figure 2.3 NMDS of Square Root Transformed Transcriptomic Libraries for Mutant and Wildtype Chemostats 35
Figure 2.4 Gene Expression Profile of a Glutathione-Dependent Peroxiredoxin in the Mutant and Wildtype 36
Figure 2.5 Transcriptomic Profiles of *coxBAC* for the mutant and wildtype 37
Figure 2.6 Average TPM of *mcyA* and *mcyD* for the mutant and wildtype 49
Figure 2.7 Non-photochemical measurements quenching at all PAR for the mutant and wildtype 51
Figure 2.8 Excitation pressure measurements for the mutant and wildtype 52
Figure 2.9 CM-H_2DCFDA Assay for the Mutant and Wildtype 53
Figure 2.10 Putative Gene Cluster Differentially Expressed between the Mutant and Wildtype 55
Figure 2.11 Average TPM for a Thioredoxin-Dependent Peroxiredoxin 56
Figure 2.12 Gene expression of *pntAB* for the mutant and wildtype 57
Figure 2.13 Transcriptional profiles of electron transporters *petJ* and *petE* 59
Figure 3.1 Chromosome inversion region, gene arrangement and average TPM of genes 70
Figure 3.2 IS1-family transposase insertion in the wildtype and expression of effected genes ... 72
Figure 3.3 Insertion of a IS1-family transposase in the wildtype upstream of an Efflux RND transporter 74
Figure 3.4 Insertion of a IS1634 transposase in Δ*mcyB* 76
Figure 3.5 MITE alignments, insertion sites and structure 79
Figure 3.6 Gene expression of the sulfate transporter gene cluster in the mutant and wildtype .. 80
Figure 3.7 Conserved residues of the sulfate binding gene, reads mapping to *sbp* in the mutant, sulfate limitation growth assays 82
Figure 3.8 Location of primer design for *sbp* gene 98
Figure 3.9 Locations chosen for gene expression analysis of transposases 99
Figure 3.10 Duplicated region in *M.aeruginosa* PCC 7806SL genome 100
Figure 3.11 Mauve alignment of wildtype and Δ*mcyB* genomes 101
Figure 3.12 SLC permease interruption in Δ*mcyB* by *tnpA* 102
Figure 3.13 *trkH* interruption in the wildtype by *tnpA* 103
Figure 3.14 Glycosyltransferase interrupted by *tnpA* in the wildtype 104
Figure 3.15 Insertion of *tnpA* in the mutant 105
Figure 3.16 IS1-family transposase insertion in an uncharacterized gene in the mutant 106
Figure 3.17 Gene expression profile of IS1634 family transposase and a gene encoding a diflavin flavoprotein 107
Figure 3.18 Microcystin-related protein, gene arrangement and expression profile in the mutant and wildtype 108
Figure 3.19 Phage protein D in *M.aeruginosa* NIES-843 109
Figure 3.20 Phage tail lysozyme in *M.aeruginosa* LE-3 and NIES-298 110
Figure 3.21 Phage Tail lysozyme in LE-3 and ITR's of IS1-transposase 111
Figure 3.22 Inverted terminal repeats for IS1-family transposases in LE-3 112

Figure 3.23 Insertion of a MITE in *livG* in the mutant and expression of the gene 113
Figure 3.24 Shortened alignment of an ISL3-family transposase with the MITE..... 114
Figure 3.25 Reverse transcriptase PCR of the MITE insertion in *sbp*..... 115
Figure 4.1 Plasmid in *M. aeruginosa* PCC 7806 128
Figure 4.2 Expression of the plasmid *in vitro* and *in situ* 130
Figure 4.3 Average TPM/gene for viral and plasmid contigs from Lake Erie microcosm metatranscriptomes 132
Figure 4.4 *M.aeruginosa* NIES-298 P1 Plasmid, protein structure homologs, BLAST output against Bacteriophage_8960 contig 135
Figure 4.5 Alignments of the *M.aeruginosa* NIES-298 genome and *Microcystis Phage* LMM01 against the NIES-298 P1 plasmid 137
Figure 4.6 NIES-298 P2 Plasmid gene expression during Ma-LMM01 phage infection..... 138
Figure 4.7 Protein structure alignment of gene 5 on the plasmid with a chromsome-atpase like gene in the chromosome 151
Figure 4.8 Nucleotide BLAST of all NIES-298 P1 plasmids and contigs from scaffold level genomes 152
Figure 4.9 Protein structure alignment of tyrosine recombinases from NIES-298 P1 plasmid and PCC 7806..... 153
Figure 4.10 BLAST of *Microcystis* phage genomes against the NIES-298 P1 plasmid 154
Figure 4.11 BLAST of NIES-298 genome and NIES-298 P1 plasmid against *Microcystis phage* Ma-LMMO1 repeat region 155
Figure 4.12 Recombinase and intergenic region from NIES-298 P1 plasmid that alignments with genome CRISPR region..... 156
Figure 4.13 IS5 transposase that changed on NIES-298 P2 plasmid..... 157
Figure 4.14 Gene expression of NIES-298 P1 plasmid during *Microcystis phage* Ma-LMM01 infection 158
Figure 4.15 BLAST of NIES-298 genome against NIES-298 P2 plasmid..... 159
Figure 4.16 BLAST of *M.aeruginosa* FD4 genome against FD4 plasmid..... 160
Figure 4.17 BLAST of *M.aeruginosa* NIES-2481 genome against NIES-2481 P1 plasmid 161
Figure 4.18 TnpB and IS607 recombinase that are 99% identical between NIES-2481 P1 plasmid, the NIES-2481 genome and *Microcystis phage* Ma-LMM01..... 162

LIST OF ATTACHMENTS

Appendix Sheet 2.1

- Sheet 2.1A. List of RNA-Seq Time Points and Experimental Relevance
- Sheet 2.1B Statistics of Transcriptomic Libraries generated
- Sheet 2.1C Differentially Expressed Genes Time Point 1
- Sheet 2.1D Differentially Expressed Genes Time Point 2
- Sheet 2.1E Differentially Expressed Genes Time Point 3
- Sheet 2.1F Differentially Expressed Genes Time Point 4
- Sheet 2.1G Differentially Expressed Genes Time Point 5
- Sheet 2.1H Differentially Expressed Genes Time Point 6
- Sheet 2.1I Differentially Expressed Genes Time Point 7
- Sheet 2.1J Differentially Expressed Genes Time Point 8
- Sheet 2.1K Differentially Expressed Genes Time Point 9
- Sheet 2.1L Differentially Expressed Genes Time Point 10
- Sheet 2.1M Locus Tags for Gene Cluster 1

Appendix Sheet 3.1

- Sheet 3.1A. Mapping *tnpA* reads
- Sheet 3.1B Chromosome Inversion genes
- Sheet 3.1C Locations of *tnpA* in the *ΔmcyB* genome
- Sheet 3.1D Locations of *tnpA* in the wildtype genome
- Sheet 3.1E Coordinates of 737-nt long IS1-family transposases in *ΔmcyB* genome
- Sheet 3.1F Coordinates of 737-nt long IS1-family transposases in wildtype genome
- Sheet 3.1G Coordinates of IS1634 transposases in *ΔmcyB*
- Sheet 3.1H Coordinates of IS1634 transposases in the wildtype
- Sheet 3.1I geNomad identified phage genes
- Sheet 3.1J Phage holin homologs in other *Microcystis*
- Sheet 3.1K Phage gene BLASTN identity in LE3 and NIES-298
- Sheet 3.1L JGI BLAST
- Sheet 3.1M JGI virus blast output
- Sheet 3.1N MITE locations in *ΔmcyB* genome
- Sheet 3.1O MITE locations in wildtype genome
- Sheet 3.1P MITE homologs in other *Microcystis*
- Sheet 3.1Q MITE homologs in plasmids
- Sheet 3.1R MITE homologs in phage

Appendix Sheet 4.1

- Sheet 4.1A *Microcystis* plasmids on NCBI
- Sheet 4.1B Nucleotide BLAST output of the PCC 7806 plasmid against the nr database
- Sheet 4.1C Plasmid genes, PGAP annotations and PFAM domains
- Sheet 4.1D DALI scores of plasmid proteins structure homologs
- Sheet 4.1E DALI score for chromosome gene (pgaptmp_004807)
- Sheet 4.1F Whole genome shotgun sequencing BLASTN outputs from PCC 7806 plasmid

Sheet 4.1G geNomad scores of contigs with significant hits to the PCC 7806 plasmid
Sheet 4.1H BLAST results for closed plasmid contigs from Lake Erie
Sheet 4.1I *Microcystis* WGS nucleotide BLAST search of all *Microcystis* plasmids
Sheet 4.1J JGI CRISPR spacer BLAST of all *Microcystis* plasmids
Sheet 4.1K geNomad scores of all *Microcystis* plasmids
Sheet 4.1L BLAST of NIES-298 P1 plasmid
Sheet 4.1M DALI output for protein model structures of NIES-298 plasmid
Sheet 4.1N BLAST outputs of NIES-2549 genes
Sheet 4.1O TPM values for P1 plasmid 3hr control
Sheet 4.1P TPM values for P1 plasmid 3hr infected
Sheet 4.1Q TPM values for P2 plasmid 3hr control
Sheet 4.1R TPM values for P2 plasmid 3hr infected
Sheet 4.1S BLAST outputs for NIES-298 P2 plasmid and FD4 plasmid
Sheet 4.1T *tnpB* and IS607 blast outputs from NIES-2481 P1 Plasmid

CHAPTER ONE:
INTRODUCTION

Cyanobacterial blooms overview

Cyanobacterial blooms have been a natural occurrence for millions of years [1]. Paleoecologists have attributed mass mortality events to blooms dating back to the Pleistocene (2.6 million years ago) [1, 2] , and the first microorganisms observed under a microscope by Antonie von Leewenhoek in 1674 may have been cyanobacteria from a cyanobacterial bloom [1]. While history shows cyanobacterial blooms are not a uniquely modern-day inconvenience, increasing eutrophication due to anthropogenic inputs and climate-driven temperature changes are increasing the frequency and lifespan of these blooms [1, 3-5]. As temperatures rise, cyanobacteria will have a better chance at outcompeting other phytoplankton, due to their higher growth rates at higher temperatures, which eukaryotic phytoplankton do not tolerate as well [3].

Cyanobacterial blooms can alter ecosystem functions by various mechanisms. High levels of biomass in blooms can increase water turbidity, which can have negative effects on aquatic life and impact food webs [3, 6]. Hypoxia is another consequence of blooms, where oxygen is consumed from the freshwater environment during the decomposition of cyanobacterial biomass [7]. Hypoxic conditions can have trophic level effects, negatively impacting the health of aquatic plants, invertebrates and fish [8]. By altering ecosystem dynamics, blooms can lead to the loss of biodiversity in aquatic habitats [9]. Loss of biodiversity is not only bad ecologically, but can also have economic consequences when it comes to fisheries, and tourism [1].

Microcystis- dominated blooms

Microcystis spp. are globally distributed potentially-toxic cyanobacteria [10, 11]. Toxin-producing *Microcystis spp*. can form a cyclic hepatoxic secondary metabolite, known as microcystin [12]. Toxic *Microcystis* blooms can exacerbate harmful effects on aquatic life through the release of microcystin in the water when cells lyse [11]. Microcystin's deadly effects come from its ability to bind to eukaryotic protein phosphatases 1 and 2A, making it damaging to plants and animals [12-14]. As of 2016, *Microcystis* blooms were shown to have been reported in 108 countries [11]. Of those 108 countries that reported blooms, the presence of microcystin was confirmed in 79

countries [11]. As climate change advances, *Microcystis* blooms are projected to increase in frequency and time scale [3]. This is bad from an ecological, human health, and economic standpoint.

Microcystis blooms are usually composed of some composition of toxic and non-toxic *Microcystis* genotypes [10]. Understanding what drives the genotypic composition or the toxicity of *Microcystis* blooms is still an ongoing process, as many environmental models are oversimplified [15]. Additionally, there are many inconsistencies in *Microcystis* literature when it comes to abiotic factors and their influence on toxin production [16]. Such inconsistencies may be due to strain variability or the microbiome that is present with cultures [10, 17-20]. Research on the "phycosphere", defined as the microenvironment that surrounds phytoplankton cells, is beginning to unravel the importance of considering all members of a bloom- beyond the dominant *Microcystis* taxa [20]. In some instances, additions of heterotrophic bacteria into axenic *Microcystis* cultures have promoted *Microcystis* colony formation [21]. It has also recently been discovered that the heterotrophic bacteria present in blooms can aid *Microcystis* nitrogen uptake by modifying nitrogen sources [22]. As time goes on, it may be that inconsistencies found in the literature are driven by difference heterotopic communities that co-habitat with lab strains. It also means that all community members need to be accounted for in *Microcystis* ecology/eco-physiology work.

Microcystis blooms are constrained by numerous factors. Abiotic factors, such as nitrogen and phosphorus can promote bloom formation when in excess or, constrain blooms if in limiting supply [11]. There are also various biotic constraints on blooms. Parasitic fungi, known as chytrids, can infect *Microcystis* and promote the decline of cell populations [23]. Cyanophage have also been implicated in the decline of *Microcystis* populations [23-25]. However, it appears that many *Microcystis* phage are strain-specific, which may instead result in compositional shifts in *Microcystis* genotypes in a bloom [23, 26, 27].

Microcystin

Microcystin is a cyclic hepatoxic compound that is produced through non-

ribosomal peptide synthesis [12]. The genes encoding for the full length-cyclic microcystin metabolite are transcribed from a 55kb *mcy* gene cluster consisting of 10 open-reading frames (*mcy*A-J) [12]. However, partial active microcystin operons have been reported, which only consist of *mcyA-C* [28]. Since a bi-directional promotor is found between *mcyA* and *mcyD*, these two genes are often used when assessing toxin production, however, some of the other *mcy* tailoring genes may have their own individual promotors as well [29]. Currently, over 240 congener variants of microcystin have been discovered, all which vary in toxicity [30]. The most common congener, and most toxic congener found to date is the microcystin-LR (MC-LR) variant [30]. The World Health Organization guidelines are based around MC-LR, which should not exceed 1μg/L in drinking water sources [11]. However, chronic exposure of low levels of microcystins in the water have still been shown to be correlated/suspected in cases of liver disease and liver cancer [14]. More recently, research has found that microcystin congener composition can be influenced by environmental factors such as nitrogen [31]. We have also seen this with some of our temperature work (published and unpublished) [32]. In the future, as we learn more about microcystin congener variants, reporting microcystin congener composition may be more informationally useful then total microcystins alone.

Microcystin is a stable compound when isolated, with the ability to withstand high temperatures (>100°C), and extreme pH [33, 34]. It has also been shown that sunlight alone cannot breakdown microcystin, and photodegradation of the toxin is dependent on photosensitizers in the environment [35]. Heterotrophic bacteria have been also shown to influence the breakdown of the toxin. This has been seen in lab-based studies, and in situ, with the use of ^{15}N-Microcystin-LR isotope labeling [36, 37]. The enzymatic breakdown of microcystin-LR has been shown to be mediated by the *mlrABCD* pathway in heterotrophs, which has been well characterized [38, 39]. However, this pathway and homologs of it are often not present in metagenomic datasets, therefore it seems likely other biotic microcystin degradation pathways exist and are still elusive [36, 37].

Proposed functions of Microcystin

Microcystin acts on higher order eukaryotic organisms by irreversibly binding to cysteine residues on protein phosphatases 1 (PP1) and 2A (PP2A), inhibiting their phosphatase activity [40]. Due to the harmful effects of microcystin on eukaryotic organisms, some theorize that microcystin is produced as a deterrent to grazers, as negative effects on zooplankton have been reported [41]. However, this is likely a side benefit of and not the sole ecological purpose of microcystin production, as it has been shown the *mcy* gene cluster evolved before metazoans [42]. Additionally, research on microcystin's impact on zooplankton is mixed, where some *in situ* work has shown that zooplankton biomass increased with a *Microcystis*-dominated bloom [6]. Other studies have also shown contradictory results when it comes to microcystins and grazers [43]. These contradictions could be due to zooplankton adaptions to improve grazing [1]. Also, many other abiotic factors, such as light, temperature, and nutrient availability have been shown to independently effect microcystin cellular quota [44-47]. These reasons along with biochemical assays showing microcystin aggregating near the thylakoid membranes [48], and its ability to bind to intracellular proteins makes it likely microcystin serves an intracellular function [49].

In 2011 it was found that microcystin can covalently bind to cysteine residues on redox sensitive-proteins in *Microcystis* during oxidative stress conditions [49, 50]. The proteins that microcystin can intracellularly bind to have diverse functions in the cell, serving in both light and dark reactions. From this study, it was thought the intracellular binding of microcystin protects photosynthetic proteins from oxidative stress[49], however, *in vivo* stabilization of proteins was not, and has not yet been shown. Additionally, some research has shown conflicting results when it comes to the oxidative stress hypothesis [51], and there could be more than one role microcystin plays in the cell [52].

Differences between wildtype *Microcystis aeruginosa* PCC 7806 and the non-toxigenic *Microcystis aeruginosa* PCC 7806 Δ*mcyB*

Most of this dissertation work was performed using the toxic type-strain

Microcystis aeruginosa PCC 7806 (will be referred to as "wildtype"), and it's non-toxic mutant *Microcystis aeruginosa* PCC 7806 Δ*mcyB* (will be referred to as "mutant"). *Microcystis spp.* are difficult to genetically manipulate, as their genomes code for numerous restriction modification systems [53, 54]. For this reason, many *Microcystis* researchers who study toxin production use the non-toxic mutant strain, *Microcystis aeruginosa* PCC 7806 Δ*mcyB*, created by Elke Dittmann in 1997, and which is available to order from the Pasteur Culture Collection [54]. This mutant is unable to produce the cyclic microcystin compound, due to the insertion of a chloramphenicol cassette in the *mcyB* gene [54]. Since its creation, this mutant strain has been used alongside its toxic wildtype counterpart, *Microcystis aeruginosa* PCC 7806, to conduct comparative physiology studies in efforts to elucidate microcystin's functions by various researchers around the world [41, 45, 51, 55-64].

High-light studies

Many of the first studies done comparing the mutant and wildtype investigated high-light effects on the fitness of the cells. High-light can increase the generation of reactive oxygen species in photosynthetic organisms, damaging the D1 protein in photosystem II [65][54]. High-light was initially investigated because it increased the transcripts of *mcyB* and *mcyD* in the wildtype strain [44]. Since then, phenotypic differences have been observed in PCC 7806 and **Δ***mcy*B after exposure to high light and limiting dissolved inorganic carbon [55, 58]. Reported phenotypic changes between 7806 and **Δ***mcy*B result from altered abundances of chlorophyll-a molecules, which denotes a difference in photosynthetic acclimation between the strains when exposed to environmental stressors. The wildtype strain has also been shown to have a growth advantage under high-light conditions compared to the mutant, whereas there was no growth advantage for the wildtype at low or optimal light conditions [58].

Temperature

We still have gaps in knowledge when it comes to temperature and microcystin production, due in part to inconsistencies frequently seen between lab observed

temperature effects, and in situ temperature effects [10]. Additionally, aside from a handful of studies, there has been limited research into temperature effects on physiology using direct comparisons of the wildtype *Microcystis aeruginosa* PCC 7806 and the Δ*mcyB* mutant [46, 63, 66]. However, it has been shown that blooms dynamics often follow a trend where toxic genotypes are observed earlier in the growth season when waters are cooler, and non-toxigenic genotypes are more populous in later in the growth season [32]. Studies that utilized *Microcystis* strains other than PCC 7806 and the mutant, have shown that strains usually increase microcystin production or *mcy* gene transcription at suboptimal growth temperatures, i.e. 6.3°C below optimum growth, and growth above the optimal temperature decreases microcystin production [15, 32]. However, in some cases microcystin production was not affected by temperature [32].

Low dissolved inorganic carbon (DIC) studies

Previous studies involving wild type PCC7806 and **Δ***mcy*B showed that low DIC levels caused microcystin production to increase in the wildtype strain [55]. The wildtype also displayed increased chlorophyll-a quotas compared to the mutant, indicating differences in photosynthetic acclimation to limited inorganic carbon availability. The authors of the study speculated that microcystin may aid in this acclimation [55]. Based on other research, we can speculate that microcystin plays a role in the photosynthetic response to low DIC, either by regulation through carbon/nitrogen balance in the cell [67], and/or by affecting the activity of RubisCO, to which microcystin can bind [49]. While evidence for microcystin affecting enzymatic activity of RubisCO is limited, a study by Barchewitz *et al*. [61] discovered that under 4-hours of high light exposure, the wild type PCC7806 displayed increased carboxylase activity compared to the **Δ***mcy*B mutant, indicating RubisCO activity was altered in the wildtype strain. RubisCO is known for being an inefficient enzyme in cyanobacteria. It has been reported that RubisCO activity can be modulated through covalent binding of proteins to its cysteine residues [68]. Zilliges *et al*. [49] showed that microcystin interacts with both subunits of RubisCO, achieving this interaction by covalently binding to cysteine residues. Therefore, if microcystin plays a role in carbon metabolism it could be through the

modulation of RubisCO activity, and the resulting photosynthetic acclimation previously seen between the wildtype and mutant may be from redox-regulated processes associated with the electron transport system, which RubisCO is a part of [69-71].

Oxidative Stress

One of the leading hypotheses in microcystin research is that microcystin plays a role in oxidative stress, due to its ability to bind to redox-sensitive thiol groups on proteins [49]. One of the redox sensitive proteins that has been shown to bind to microcystin is glutathione reductase, which is needed for glutathione metabolism, an important antioxidant defense system [49][72]. However, when it comes to studies were H_2O_2 is added to growth media, results for whether the wildtype or mutant withstand external H_2O_2 better have been mixed, and dependent on other factors. At naturally occurring concentrations of H_2O_2 (34μg/L), the wildtype was shown to be less effected compared to the non-toxic mutant [49]. However, when exposed to high levels of H_2O_2 (10mg/mL), an amount usually used in bloom lake treatments, the microcystin-free mutant fared better than the wildtype, which was completely unable to grow [51]. This was also seen by [73]. Therefore, it may be the wildtype is better at tolerating a certain threshold of lower concentrations of H_2O_2, but is unable to tolerate unnaturally high amounts of H_2O_2.

Microcystis Genomics

The first *Microcystis* genome successfully closed was by Kaneko et al. in 2007, when they sequenced *Microcystis aeruginosa* NIES-843 [74]. Since this time, many draft genome *Microcystis spp.* assemblies have become available (361 draft genomes as of 2024), however only 16 of them have been closed. Sequencing of the *Microcystis aeruginosa* NIES-843 genome [74], and shortly after the *Microcystis aeruginosa* PCC 7806 genome [9] revealed the "plastic" nature of *Microcystis* genomes, which consist of numerous repetitive regions, transposable elements, and restriction enzymes, making up ~12% of the genome [9]. Despite a large fraction of the genome being made up of transposable elements, there have been very few studies that have investigated the rate or regulation of transposition events in *Microcystis* [75].

Microcystis genomes can range considerably in size, with the smallest complete

genome currently on NCBI being ~4.3Mb, and the largest being 5.8Mb. As of 2023, the pan genome of *Microcystis* contained 21,880 nonredundant genes, 1639 of which make up the core genome [76]. This means less than half of each *Microcystis* genome consists of shared core genes [76]. Most of the pangenome is made up of genes in the accessory genome (14,562 genes), showing that *Microcystis* have a great deal of genetic diversity [76]. Many accessory and unique genes in *Microcystis* genomes may be a result of cross-species horizontal gene transfer events [77]. Some population genomics studies have found that geography and local adaptation may be important factors in controlling the rates of horizontal gene transfer, as frequent homologous recombination was observed more within monophyletic clades [78]. However, more recent investigations using core-genes and non-core genes have determined geographic dispersal is not a major driver of *Microcystis* phylogenic relationships.

Microcystis plasmids

Microcystis plasmids have been an overlooked part of *Microcystis* physiology and ecology studies. The first manuscripts of *Microcystis* plasmids were published in the early 1980's, when researchers were still trying to figure out the genetics of microcystin production by *Microcystis* species. Some researchers speculated that plasmids may have carried the toxin [79]. One of the first papers investigating this showed that a toxic *Microcystis sp. WR70* that was cured of its plasmid by multiple chemical means was no-longer toxic after curing [79]. This led into additional investigations into plasmid-borne toxicity by other researchers, with varying, but doubtful results that plasmids were responsible for microcystin production [80-82]. Interesting observations did emerge from some of these studies, and in 1988, researchers noted that different strains of *Microcystis*, which were isolated from geographically distant areas, may harbor the same or very similar plasmids, and that *Microcystis* often harbored more than one plasmid [81]. However, since the confirmation of the microcystin biosynthesis PKS/PRKS gene cluster on the chromosome in 1997, there has not been much research on *Microcystis* plasmids, or attempts to understand how they influence the physiology of the host organism. We also have very little understanding of *Microcystis* plasmids and how they participate on

an ecological/evolutionary level, or the role they play in horizontal gene transfer among *Microcystis*, or other cyanobacterial species. Therefore, *Microcystis* plasmids and their cellular functions are still cryptic.

Dissertation Overview

This dissertation highlights key physiological differences between the toxic wildtype *Microcystis aeruginosa* PCC 7806 and it's non-toxic mutant *Microcystis aeruginosa* PCC 7806 Δ*mcyB* during growth at suboptimal temperatures in controlled continuous cultures. As these conditions cause a two-fold increase in microcystin in the wildtype, the mechanisms employed by the mutant to acclimate to these conditions provide more insight into elucidating microcystin's intracellular role and shows that ROS are higher in the mutant during these conditions. We also bring attention to changes in genome structure in the mutant and wildtype after long term batch cultivation in the lab. These two strains are often used in comparative studies in labs across the world to study the role of microcystin and are assumed to be genetically identical, aside from the chloramphenicol cassette insertion in the *mcyB* gene. We show that genetic rearrangements have occurred between these two strains in our cultures. We found five active families of transposable elements/insertion sequences that caused genetic rearrangements which lead to changes in gene expression, some of which had clear phenotypic consequences. With this, we demonstrate how important genome re-sequencing is when doing comparative physiology studies with *Microcystis*. As a continuation of investigating mobile elements in *Microcystis*, we investigated a conserved plasmid found in both the mutant and wildtype. This plasmid had identical nucleotide sequences in both strains. Although these two strains have been used in numerous research articles over the last 20 years, no one has reported this plasmid before. As we were interested in the plasmids role in the cell, we investigated gene functions, the plasmids transcriptional activity, and extended our search to all *Microcystis* plasmids, to see if plasmid conservation was a common phenomenon. Other findings include the genetic overlap between *Microcystis* plasmids and phage, and the importance of including plasmids in *in situ* bioinformatics investigations.

References

1. Huisman, J., et al., *Cyanobacterial blooms.* Nature Reviews Microbiology, 2018. **16**(8): p. 471-483.

2. Braun, A. and T. Pfeiffer, *Cyanobacterial blooms as the cause of a Pleistocene large mammal assemblage.* Paleobiology, 2002. **28**(1): p. 139-154.

3. *Paerl, H.* and *J. Huisman*, *Blooms like it hot.* Science, 2008. **320**: p. 57-58.

4. Paerl, H.W. and T.G. Otten, *Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls.* Microbial Ecology, 2013. **65**(4): p. 995-1010.

5. Paerl, H., *Mitigating Harmful Cyanobacterial Blooms in a Human- and Climatically-Impacted World.* Life, 2014. **4**(4): p. 988-1012.

6. Briland, R.D., et al., *Cyanobacterial blooms modify food web structure and interactions in western Lake Erie.* Harmful Algae, 2020. **92**: p. 101586.

7. Paerl, H., *Nutrient and other environmental controls of harmful cyanobacterial blooms along the freshwater–marine continuum*. 2008, Springer New York. p. 217-237.

8. Watson, S.B., et al., *The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia.* Harmful Algae, 2016. **56**: p. 44-66.

9. Frangeul, L., et al., *Highly plastic genome of Microcystis aeruginosa PCC 7806, a ubiquitous toxic freshwater cyanobacterium.* BMC Genomics, 2008. **9**(1): p. 274.

10. Dick, G.J., et al., *The genetic and ecophysiological diversity of Microcystis.* Environmental Microbiology, 2021. **23**(12): p. 7278-7313.

11. Harke, M.J., et al., *A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp.* Harmful Algae, 2016. **54**: p. 4-20.

12. Tillett, D., et al., *Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide–polyketide synthetase system.* Chemistry & Biology, 2000. **7**(10): p. 753-764.

13. Robert W. MacKintosh, K.N.D., David G. Campbell, Patricia T.W. Cohen, Philip Cohen, Carol ManKintosh, *The cyanobacterial toxin microcystin binds covalently to cysteine-273 on protein phosphatase 1.* FEBS Letters, 1995. **371**(3): p. 236-240.

14. Melaram, R., A.R. Newton, and J. Chafin, *Microcystin Contamination and Toxicity: Implications for Agriculture and Public Health.* Toxins, 2022. **14**(5): p. 350.

15. Hellweger, F.L., Martin, Robbie M. , Eigemann, Falk , Smith, Derek J., Dick, Gregory J. , Wilhelm, Steven W. , *Models predict planned phosphorus load reduction will make Lake Erie more toxic. Science*, 2022. **376**: p. 1001-1005.

16. Wilhelm, S.W., G.S. Bullerjahn, and R.M.L. McKay, *The Complicated and Confusing Ecology of Microcystis Blooms.* mBio, 2020. **11**(3).

17. Mikalsen, B., et al., *Natural Variation in the Microcystin Synthetase Operon mcyABC and Impact on Microcystin Production in Microcystis Strains.* Journal of Bacteriology, 2003. **185**(9): p. 2774-2785.

18. Bãnares-Espãna, E., et al., *Genetic variability associated with photosynthetic pigment concentration, and photochemical and nonphotochemical quenching, in strains of the cyanobacterium Microcystis aeruginosa.* FEMS Microbiology Ecology, 2007. **60**(3): p. 449-455.

19. Meyer, K.A., et al., *Genome sequences of lower Great Lakes Microcystis sp. reveal strain-specific genes that are present and expressed in western Lake Erie blooms.* PLOS ONE, 2017. **12**(10): p. e0183859.

20. Hoke, A.K., et al., *Genomic signatures of Lake Erie bacteria suggest interaction in the Microcystis phycosphere.* PLOS ONE, 2021. **16**(9): p. e0257017.

21. Shen, H., et al., *Morphological and physiological changes in Microcystis aeruginosa as a result of interactions with heterotrophic bacteria.* Freshwater Biology, 2011. **56**(6): p. 1065-1080.

22. Li, W., et al., *Microbiome processing of organic nitrogen input supports growth and cyanotoxin production of Microcystis aeruginosa cultures.* ISME J, 2024. **18**(1).

23. Van Wichelen, J., et al., *The common bloom-forming cyanobacterium Microcystis is prone to a wide array of microbial antagonists.* Harmful Algae, 2016. **55**: p. 97-111.

24. Steffen, M.M., et al., *Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH.* Environmental Science & Technology, 2017. **51**(12): p. 6745-6755.

25. McKindles, K.M., et al., *Dissolved microcystin release coincident with lysis of a bloom dominated by Microcystis spp. in Western Lake Erie attributed to novel cyanophage.* Applied and Environmental Microbiology, 2020. **86**(22).

26. Yoshida, T., et al., *Isolation and Characterization of a Cyanophage Infecting the Toxic Cyanobacterium Microcystis aeruginosa.* Applied and Environmental Microbiology, 2006. **72**(2): p. 1239-1247.

27. Wang, X.-Q., et al., *Profiling the interplay and coevolution of Microcystis aeruginosa and cyanosiphophage Mic1.* Microbiology Spectrum, 2024. **12**(6).

28. Yancey, C.E., et al., *Metagenomic and metatranscriptomic insights into population diversity of Microcystis blooms: Spatial and temperal dynamics of mcy genotypes, including a partial operon that can be abundant and expressed.* Applied and Environmental Microbiology, 2022. **88**(9).

29. Kaebernick, M., et al., *Multiple Alternate Transcripts Direct the Biosynthesis of Microcystin, a Cyanobacterial.* Applied and Environmental Microbiology, 2002. **68**(2): p. 449-455.

30. Diez-Quijada, L., et al., *Occurrence and toxicity of microcystin congeners other than MC-LR and MC-RR: A review.* Food Chem Toxicol, 2019. **125**: p. 106-132.

31. Chaffin, J.D., et al., *Microcystin congeners in Lake Erie follow the seasonal pattern of nitrogen availability. Harmful Algae*, 2023. **127**: p. 102466.

32. Peng, G., et al., *Seasonally Relevant Cool Temperatures Interact with N Chemistry to Increase Microcystins Produced in Lab Cultures of Microcystis aeruginosa NIES-843.* Environmental Science & Technology, 2018. **52**(7): p. 4127-4136.

33. Metcalf, J.S. and G.A. Codd, *Microwave oven and boiling waterbath extraction of hepatotoxins from cyanobacterial cells.* FEMS Microbiology Letters, 2000. **184**(2): p. 241-246.

34. Massey, I.Y. and F. Yang, *A Mini Review on Microcystins and Bacterial Degradation.* Toxins, 2020. **12**(4): p. 268.

35. Sun, Q., et al., *Ultraviolet photosensitized transformation mechanism of microcystin-LR by natural organic matter in raw water.* Chemosphere, 2018. **209**: p. 96-103.

36. Mou, X., et al., *Metagenomic Identification of Bacterioplankton Taxa and Pathways Involved in Microcystin Degradation in Lake Erie.* PLoS ONE, 2013. **8**(4): p. e61890.

37. Chaffin, J.D., et al., *Quantification of microcystin production and biodegradation rates in the western basin of Lake Erie.* Limnology and Oceanography, 2022. **67**(7): p. 1470-1483.

38. Bourne, D.G., et al., *Enzymatic pathway for the bacterial degradation of the cyanobacterial cyclic peptide toxin microcystin LR.* Applied and Environmental Microbiology, 1996. **62**(11): p. 4086-4094.

39. Shimizu, K., et al., *How microcystin-degrading bacteria express microcystin degradation activity.* Lakes & Reservoirs: Science, Policy and Management for Sustainable Use, 2011. **16**(3): p. 169-178.

40. Pereira, S.R., V.M. Vasconcelos, and A. Antunes, *Computational study of the covalent bonding of microcystins to cysteine residues - a reaction involved in the inhibition of the PPP family of protein phosphatases.* FEBS Journal, 2013. **280**(2): p. 674-680.

41. Rohrlack, T., et al., *Effects of cell-bound microcystins on survival and feeding of Daphnia spp.* Appl Environ Microbiol, 2001. **67**(8): p. 3523-9.

42. Rantala, A., et al., *Phylogenetic evidence for the early evolution of microcystin synthesis.* Proceedings of the National Academy of Sciences, 2004. **101**(2): p. 568-573.

43. Omidi, A., M. Esterhuizen-Londt, and S. Pflugmacher, *Still challenging: the ecological function of the cyanobacterial toxin microcystin – What we know so far.* Toxin Reviews, 2018. **37**(2): p. 87-105.

44. Kaebernick, M., et al., *Light and the Transcriptional Response of the Microcystin Biosynthesis Gene Cluster.* Applied and Environmental Microbiology, 2000. **66**(8): p. 3387-3392.

45. Hesse, K., E. Dittmann, and T. Börner, *Consequences of impaired microcystin production for light-dependent growth and pigmentation of Microcystis aeruginosa PCC 7806.* FEMS Microbiology Ecology, 2001. **37**(1): p. 39-43.

46. Martin, R.M., et al., *Episodic Decrease in Temperature Increases mcy Gene Transcription and Cellular Microcystin in Continuous Cultures of Microcystis aeruginosa PCC 7806.* Frontiers in Microbiology, 2020. **11**.

47. Jähnichen, S., B.M. Long, and T. Petzoldt, *Microcystin production by Microcystis aeruginosa: Direct regulation by multiple environmental factors.* Harmful Algae, 2011. **12**: p. 95-104.

48. Young, F.M., et al., *Quantification and localization of microcystins in colonies of a laboratory strain of Microcystis (Cyanobacteria) using immunological methods.* 2008. **43**(2): p. 217-225.

49. Zilliges, Y., et al., *The Cyanobacterial Hepatotoxin Microcystin Binds to Proteins and Increases the Fitness of Microcystis under Oxidative Stress Conditions.* PLoS ONE, 2011. **6**(3): p. e17615.

50. Wei, N., et al., *Microcystin-Bound Protein Patterns in Different Cultures of Microcystis aeruginosa and Field Samples.* Toxins, 2016. **8**(10): p. 293.

51. Schuurmans, J.M., Brinkmann, B.W., Makower, A.K., Dittmann, E., Huisman, J., Matthijs, H.C.P.,, *Microcystin interferes with defense against high oxidative stress in harmful cyanobacteria. Harmful Algae*, 2018. **78**.

52. Wilhelm, S.W. and G.L. Boyer, *Healthy competition.* Nature Climate Change, 2011. **1**(6): p. 300-301.

53. Papoulis, S.E., et al., *Nutrient Loading and Viral Memory Drive Accumulation of Restriction Modification Systems in Bloom-Forming Cyanobacteria.* mBio, 2021. **12**(3): p. e0087321.

54. Dittmann, E., et al., *Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium Microcystis aeruginosa PCC 7806.* Molecular Microbiology, 1997. **26**(4): p. 779-787.

55. Jähnichen, S., et al., *Impact of Inorganic Carbon Availability on Microcystin Production by Microcystis aeruginosa PCC 7806.* Applied and Environmental Microbiology, 2007. **73**(21): p. 6994-7002.

56. Dittmann, E., et al., *Altered expression of two light-dependent genes in a microcystin-lacking mutant of Microcystis aeruginosa PCC 7806.* Microbiology, 2001. **147**(11): p. 3113-3119.

57. Jüttner, F. and H. Lüthi, *Topology and enhanced toxicity of bound microcystins in Microcystis PCC 7806.* Toxicon, 2008. **51**(3): p. 388-397.

58. Phelan, R.R. and T.G. Downing, *A growth advantage for microcystin production by Microcystis PCC7806 under high light. Journal of Phycology*, 2011. **47**(6): p. 1241-1246.

59. Makower, A.K., et al., *Transcriptomics-Aided Dissection of the Intracellular and Extracellular Roles of Microcystin in Microcystis aeruginosa PCC 7806.* 2015. **81**(2): p. 544-554.

60. Yeung, A.C., et al., *Physiological and Proteomic Responses of Continuous Cultures of Microcystis aeruginosa PCC 7806 to Changes in Iron Bioavailability and Growth Rate.* Appl Environ Microbiol, 2016. **82**(19): p. 5918-29.

61. Barchewitz, T., et al., *Non-canonical localization of RubisCO under high-light conditions in the toxic cyanobacterium Microcystis aeruginosa PCC7806.* Environmental Microbiology, 2019. **21**(12): p. 4836-4851.

62. Sandrini, G., et al., *Sensitivity to hydrogen peroxide of the bloom-forming cyanobacterium Microcystis PCC 7806 depends on nutrient availability.* Harmful Algae, 2020. **99**: p. 101916.

63. Roy, S., A. Guljamow, and E. Dittmann, *Impact of temperature on the temporal dynamics of microcystin in Microcystis aeruginosa PCC7806.* Frontiers in Microbiology, 2023. **14**.

64. Alexova, R., et al., *Specific global responses to N and Fe nutrition in toxic and non-toxic Microcystis aeruginosa.* Environmental Microbiology, 2016. **18**(2): p. 401-413.

65. Latifi, A., M. Ruiz, and C.-C. Zhang, *Oxidative stress in cyanobacteria.* FEMS Microbiology Reviews, 2009. **33**(2): p. 258-278.

66. Dziallas, C. and H.-P. Grossart, *Increasing Oxygen Radicals and Water Temperature Select for Toxic Microcystis sp.* PLoS ONE, 2011. **6**(9): p. e25569.

67. Beversdorf, L.J., T.R. Miller, and K.D. McMahon, *Long-term monitoring reveals carbonâ€“nitrogen metabolism key to microcystin production in eutrophic lakes.* Frontiers in Microbiology, 2015. **6**.

68. Moreno, J., M.J. Garcia-Murria, and J. Marin-Navarro, *Redox modulation of Rubisco conformation and activity through its cysteine residues.* 2007. **59**(7): p. 1605-1614.

69. Irihimovitch, V. and M. Shapira, *Glutathione Redox Potential Modulated by Reactive Oxygen Species Regulates Translation of Rubisco Large Subunit in the Chloroplast.* Journal of Biological Chemistry, 2000. **275**(21): p. 16289-16295.

70. Schuller, J.M., et al., *Redox-coupled proton pumping drives carbon concentration in the photosynthetic complex I.* Nature Communications, 2020. **11**(1).

71. Sun, Y., et al., *Light Modulates the Biosynthesis and Organization of Cyanobacterial Carbon Fixation Machinery through Photosynthetic Electron Flow.* Plant Physiology, 2016. **171**(1): p. 530-541.

72. Mullineaux, P.M. and T. Rausch, *Glutathione, photosynthesis and the redox regulation of stress-responsive gene expression. Photosynthesis Research*, 2005. **86**(3): p. 459-474.

73. Latour, D., F. Perriere, and D. Purdie, *Higher sensitivity to hydrogen peroxide and light stress conditions of the microcystin producer Microcystis aeruginosa sp PCC7806 compared to non-producer strains.* Harmful Algae, 2022. **114**: p. 102219.

74. Kaneko, T., et al., *Complete Genomic Structure of the Bloom-forming Toxic Cyanobacterium Microcystis aeruginosa NIES-843.* DNA Research, 2007. **14**(6): p. 247-256.

75. Steffen, M.M., et al., *Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis.* The ISME Journal, 2014. **8**(10): p. 2080-2092.

76. Cai, H., et al., *Microcystis pangenome reveals cryptic diversity within and across morphospecies.* Science Advances, 2023. **9**(2).

77. Zhang, X., et al., *Trade-off in genome turnover events leading to adaptive evolution of Microcystis aeruginosa species complex.* BMC Genomics, 2023. **24**(1).

78. Perez-Carrascal, O.M., et al., *Coherence of Microcystis species revealed through population genomics.* ISME J, 2019. **13**(12): p. 2887-2900.

79. Hauman, J.H., *Is a Plasmid(s) Involved in the Toxicity of Microcystis Aeruginosa?* 1981, Springer US. p. 97-102.

80. Vakeria, D., et al., *Toxicity and extrachromosomal DNA in strains of the cyanobacterium Microcystis aeruginosa.* FEMS Microbiology Letters, 1985. **29**: p. 69-72.

81. Schwabe, W., et al., *Plasmids in toxic and nontoxic strains of the cyanobacteriumMicrocystis aeruginosa.* Current Microbiology, 1988. **17**(3): p. 133-137.

82. Bose, S.G. and W.W. Carmichael, *Plasmid distribution among filamentous toxic cyanobacteria.* Journal of Applied Phycology, 1990. **2**: p. 131-136.

CHAPTER TWO:
MICROCYSTIN AIDS IN COLD TEMPERATURE ACCLIMATION: DIFFERENCES BETWEEN A TOXIC *MICROCYSTIS* WILDTYPE AND NON-TOXIC MUTANT

Publication Note

This chapter was published in Harmful Algae in October 2023 and has been reformatted for this dissertation.

Citation:
Gwendolyn F. Stark, Robbie M. Martin, Laura E. Smith, Bofan Wei, Ferdi L. Hellweger, George S. Bullerjahn, R.Michael L. McKay, Gregory L. Boyer, Steven W. Wilhelm, Microcystin aids in cold temperature acclimation: Differences between a toxic Microcystis wildtype and non-toxic mutant. Harmful Algae, Volume 129, 2023, 102531, ISSN 1568-9883, https://doi.org/10.1016/j.hal.2023.102531.

GFS, RMM and SWW designed the experiments. GFS and LES did chemostat maintenance and daily sampling of RNA, DNA and metabolites. GFS did Phyto-PAM measurements, cell counts, ROS assays and RNA extractions. GFS with help of RMM prepared the transcriptomic libraries for RNA-seq analyses. BW and GB did LC-MS for microcystin samples. Data analysis was done by GFS. All authors contributed to writing and editing of the manuscript.

Abstract

For *Microcystis aeruginosa* PCC 7806*,* temperature decreases from 26° C to 19° C double the microcystin quota per cell during growth in continuous culture. Here we tested whether this increase in microcystin provided *M. aeruginosa* PCC 7806 with a fitness advantage during colder-temperature growth by comparing cell concentration, cellular physiology, reactive oxygen species damage, and the transcriptomics-inferred metabolism to a non-toxigenic mutant strain *M. aeruginosa* PCC 7806 Δ*mcyB*. Photo-physiological data combined with transcriptomic data revealed metabolic changes in the mutant strain during growth at 19° C, which included increased electron sinks and non-photochemical quenching. Increased gene expression was observed for a glutathione-dependent peroxiredoxin during cold treatment, suggesting compensatory mechanisms to defend against reactive oxygen species are employed in the absence of microcystin in the mutant. Our observations highlight the potential selective advantages of a longer-term defensive strategy in management of oxidative stress (*i.e.,* making microcystin) *vs* the shorter-term proactive strategy of producing cellular components to actively dissipate or degrade oxidative stress agents.

Introduction

Microcystis-dominated harmful algal blooms (HABs) occur throughout the summer season. For HABs that occur in temperate North American climates, such as Lake Erie, there are seasonal trends between temperature conditions experienced in early June, when temperatures can average 18° C, compared to those recorded later in the bloom season (August), when surface water temperatures can reach 26° C [1, 2]. As the season progresses and temperatures warm, harmful algal blooms in Lake Erie have shown a trend of becoming less toxic in conjunction with a shift from toxigenic- to non-toxigenic genotypes [1, 3, 4]. The drivers of this genotypic change remain unknown [5].

We previously observed that short-term temperature decreases (within the span of a week) can induce a "cool-temperature phenotype" in *Microcystis aeruginosa*, whereby microcystin production increased [2, 6]. Past research on HAB responses to temperature have largely focused on temperature increases due to concerns for a changing global climate [7]. However, when considering cold-to-warm seasonal bloom dynamics, understanding microcystin production across seasonal growth temperatures is relevant. Additionally, lakes subject to cooler stream discharge or which receive inputs from glacial or snow water melt can experience reductions in their thermal structure, even in warmer months [8]. Changes in microcystin production during growth at cooler temperatures remains an understudied component in understanding *Microcystis* bloom ecology [2, 6].

While regulation of the cold stress response in cyanobacteria is complex, both the physical properties of the thylakoid membranes (fluid or rigid) and the redox state of the plastoquinone pool (PQ pool) are involved in either sensing or responding to cold stress [9-12]. In cyanobacteria, the thylakoid membranes house both the photosynthetic and respiratory electron pathways, which converge at the PQ pool, making its redox status important for regulating electron transfer and energy dissipation processes in the cell [13-16]. Microcystins have been implicated to have a redox-related function. Research has shown differences in redox-related protein abundances between toxigenic and mutant strains of *Microcystis aeruginosa* during iron-limitation [17]. Furthermore, thiol-groups, important in maintaining redox balance through glutathione and thioredoxin networks,

bind to microcystins *in vivo* [18, 19]. The redox state of the of the photosynthetic apparatus in *Microcystis* grown under different light intensities was also negatively correlated with the microcystin content per cell [20]. Conclusions drawn from these studies suggest microcystin could play a role in redox-mediated metabolism.

Thiol-groups play a role in redox balance, where the presence of a reactive cysteine residue (thiolate) subjects them to post-translational modifications from reactive oxygen (ROS) and nitrogen species (RNS) [21, 22]. This property makes thiol-groups important for redox signaling and enzymatic redox regulatory pathways [23]. The original hypothesis offered to explain the binding of microcystin to redox-sensitive thiol groups was that it protected proteins and stabilized them from protease degradation/oxidation [18]. However, the idea of *in vivo* stabilization was not directly substantiated. While binding of microcystin to thiols could play a stabilizing role, another possibility is that binding of microcystin to proteins may interfere with repair or regulatory functions in the cell [24, 25]. Microcystins are known to form a covalent linkage with a key thiol in protein phosphatase 1A in eukaryotes, thereby permanently inactivating these regulatory proteins [26]. Similarly, microcystin has been shown to bind to phycobilisome subunits [18]. It was hypothesized this binding might serve to stabilize the phycobilisomes from protease degradation. However, in nutrient-limiting conditions, phycobilisome degradation *via* proteases can be beneficial, therefore microcystin binding might instead interfere in these processes [27, 28]. In these situations, binding of microcystin to proteins might alter metabolic activity that is reliant on protease-mediated degradation or on regulation mediated by the reduction/oxidation of thiol-groups. These interactions alone would mean that microcystin could have a multi-functional regulatory role in the cell, which could help explain why various environmental factors can influence microcystin production [29].

In continuous cultures of *M. aeruginosa* PCC 7806, a decrease in temperature from 26° C to 19° C induces a two-fold increase in microcystin quota [6]. Mechanistically, we hypothesized this response was linked to excitation pressure and the production of oxidative stress as a result of constant light at colder temperatures [30]. In our previous study, peak microcystin quotas coincided with cell recovery and the

establishment of a new steady-state cell concentration at 19° C. It was unclear if increased microcystin quota was a key acclimation mechanism, and to what extent other mechanisms were involved. Here we tested the effects of cool temperature on *M. aeruginosa* PCC 7806 *vs.* the microcystin-free mutant *M. aeruginosa* PCC7806 **Δ***mcyB* [31]. Through comparisons of transcriptional and physiological responses, we identified alternative mechanisms employed by the mutant to acclimate to 19º C. Our observations support the idea that in the absence of microcystin as a presumed physiological defense against oxidative stress, cells actively work to dissipate energy and degrade oxidative stress compounds that are generated as byproducts of excess light adsorption during cold stress. In *Microcystis*, the trade-offs between defensive *vs.* active approaches in mediating oxidative stress establish a potential mechanistic foundation, as well as a new hypothesis, for the regularly observed seasonal transition from toxigenic to non-toxigenic strains during bloom succession.

Methods

Continuous cultures

M. aeruginosa PCC 7806 was obtained from the Pasteur Culture Collection and has been maintained for the last decade as both active cultures and cryopreserved stocks. *Microcystis aeruginosa* PCC7806 *ΔmcyB* was generated by Dittmann and colleagues [31] and kindly provided by Professor E. Dittmann in 2016 (University of Potsdam, Germany). To assess gene mutations or transposon insertion/disruption [32], we re-sequenced and closed the genomes of both strains. Any differentially expressed genes discussed in this paper have been examined (along with adjacent intergenic spacer regions) to ensure phenotypic differences were not due to gene mutations occurring during this experiment [33]. Going forward we refer to these isolates as "wildtype" and "mutant". Chemostat design and construction followed our previous efforts [6]. Duplicate 1-L chemostats were established for each strain using modified CT medium [34], with nitrogen (N) concentrations at 323 μM supplied as a mix of $Ca(NO_3)_2$ and KNO_3 (N molar ratio of 1.37:1) [6]. Phosphorus (P) was supplied as 16.3 μM by K_2HPO_4, resulting in N:P molar ratios of ~20. Chemostats were maintained in the same

incubator at a measured light intensity of ~50 µmol photon m^{-2} s^{-1}, with a dilution rate of 0.25 d^{-1} controlled by independent peristaltic pumps. Chemostats were acclimated to 26° C, (referred to as the "warm" treatment), and 50 µmol photon m^{-2} s^{-1} for ~6 weeks, to ensure steady-state was achieved. After chemostat cell concentrations remained stable (coefficient of variation for each individual chemostat were all ≤ 5.55% over a six-day period), incubator temperature was lowered to 19° C, (referred to as the "cold" treatment) and maintained at 19° C for 11 days before the temperature was returned to 26° C. Re-acclimation to 26° C (referred to herein as "recovery") was monitored for one week. Sample collection (details below) occurred around 12:00 PM daily. Light was kept constant throughout the experiment. Temperature was monitored every 15 min with a Hobo Tidbit TempLogger (Onset Computer Corporation) to account for potential temperature fluctuations. At the end of the experiment, chemostat cultures were assessed for contamination by transferring aliquots to LB medium [35] followed by incubation at 26° C in the dark.

Cell abundance and size

Culture (1 ml) was collected from each chemostat through a sampling port using a sterile syringe. The samples were diluted ten-fold in fresh CT medium and assayed with a Guava EasyCyteHT flow cytometer (Millipore). Cell concentrations were determined by gating on red fluorescence, a proxy for chlorophyll *a*, and forward scatter, a proxy for size [6].

Measurement of Microcystin

Microcystin concentrations were estimated from 50 mL of culture collected daily. Samples were filtered onto 47-mm glass fiber filters (GF-75, Advantec) *via* vacuum filtration, immediately flash-frozen in liquid nitrogen, then stored at -80° C until processing. Quantification of microcystin and congeners was completed by LC-MS as outlined in detail at *protocols.io* [36].

Measurements of photosynthetic physiology

To assess changes in photo-physiology, 3-mL of chemostat culture was collected

daily. Rapid light curves (RLC) using a Pulse-Amplitude Fluorometry (Phyto-PAM II, Heinz Walz GmbH) were generated after 60 s of low-light far-red adaptation. Saturation pulse kinetics were checked to ensure a fluorescence peak was achieved before sample processing. For rapid light curves, saturating pulses (5,000 µmol photon m^{-1} s^{-2}) were applied every 30 s after actinic light exposure. The RLC consisted of 12 actinic light steps, starting at a PAR (µmol photon m^{-1} s^{-2}) of 1 and ending at 1275 PAR (µmol photon m^{-1} s^{-2}). From these rapid light curves, maximum potential quantum efficiency (F_v/F_m), nonphotochemical chlorophyll fluorescence quenching (NPQ), and photosystem II excitation pressure (1-qP) estimates were calculated by the Phyto-PAM II software.

Oxidative stress assessment

For the oxidative stress assessment, the green fluoroprobe, 5-(6)-Chloromethyl-2',7'-dichlorodihydrofluorescein diacetate, acetyl ester, (CM-H_2DCFDA), was used. A working concentration of approximately 1.15 x 10^7 cells was used for the oxidative stress assays. Cells were first pulled from the chemostats and pelleted by centrifugation at 3,700 xg for 15 min at 19° C. Cell pellets were washed and re-suspended in PBS buffer free of KCl and incubated in the dark with 25-µM of 2',7'-dichlorodihydrofluorescein diacetate (CM-H_2DCFDA) suspended in dimethylsulfoxide (DMSO) for 60 min at room temperature (Molecular Probes, Invitrogen). After incubation, cells were washed again and resuspended in PBS buffer. Changes in green fluorescence were measured *via* flow cytometry using a Guava easyCyteHT flow cytometer (ex_λ = 485 nm, em_λ = 535 nm). Each sample was run for 120 s, or until a maximum of 5000 cells were counted. Control culture green fluorescence (cells without CM-H_2DCFDA) was subtracted from cultures incubated with CM-H_2DCFDA, for net fluorescent gains. Green fluorescence emission is indicative of the thiol-reactive chloromethyl group of the CM-H_2DCFDA probe reacting with glutathione and other thiols (Invitrogen).

RNA extraction

Samples (25 mL) were collected daily on 47-mm diameter, 0.2 µm nominal pore-size polycarbonate filters and immediately flash-frozen in liquid nitrogen before storage at -80° C. We selected ten points across the time course (2 wildtype chemostats and 2

mutant chemostats x 10 time points, n = 40) for RNA sequencing based on cell concentration observations (Appendix sheet 2.1A). To extract RNA, we used the acid-phenol chloroform bead-beating method described in [37] and Turbo DNA-free kit to eliminate contaminating DNA (Invitrogen). Checks for contaminating DNA were completed by PCR (using 515F and 806R primers targeting the 16s rRNA gene) and visualization on an agarose gel. Where DNA contamination was present (i.e., PCR products observed in the gel), the sample was processed with another round of DNAse and rechecked until no band was present. Final RNA concentrations were determined using a high-sensitivity Qubit RNA assay per manufacturer's instructions (Invitrogen).

RNA sequencing and sequence analyses

RNA samples were sent to Discovery Life Sciences (previously Hudson Alpha, Huntsville, AL) where sample library prep and Illumina Novaseq 6000 platform sequencing (50 million reads, 150 bp, paired end) were performed after ribosomal RNA reduction. A TruSeq rRNA reduction was performed on all samples ex-silico, according to manufacturer protocol. Sequence data was loaded to CLC genomics Workbench (v. 21.0.4) where it was interleaved, filtered, and trimmed using default program settings. Due to the axenic nature of the chemostats, transcriptome RNA-sequencing reads were mapped (length fraction 0.8, similarity fraction 0.9) to the rRNA genes (PCC7806 has two sets of 5S, 16S and 23S genes) from the Microcystis aeruginosa PCC7806SL genome to remove ribosomal reads [38]. Reads that did not map to rRNA genes were then stringently mapped (length fraction 0.9, similarity fraction 0.9) to the annotated Microcystis aeruginosa PCC 7806SL genome (GenBank accession NZ_CP020771.1, annotated Jan-28-2020) [38], and normalized to transcripts per million (TPM). Differential expression analysis within each time point (calculated as expression of wildtype vs. expression of mutant at each time point) was performed using CLC's differential expression algorithm. Genes were considered differentially expressed if FDR-corrected p-values were ≤ 0.05 and fold change was ≥ 2. In our supplemental materials, we considered genes with fold changes ≥ 1.5 and FDR-p values < 0.05 for further investigation into genes of interest that had fold changes ≥ 2 and FDR p-values < 0.05 in

the main text. For downstream analysis of gene function, coding sequences (CDS) from Microcystis aeruginosa PCC7806SL (NZ_CP020771.1, annotated Jan-28-2020) were supplemented/merged with annotations from KEGG BlastKOALA. Genes of interest (displayed fold change ≥ 2) that were annotated as "hypothetical genes" were run through NCBI's conserved domain protein search and HMMER for protein homology. Reads mapping to genes that were not annotated and that failed to match with a homologous protein search were omitted from downstream analyses. Tables listing differentially expressed genes by RNA-seq time points (T1-T10) are provided in appendix attachment sheets 2.1C-2.1L. For this study, we discuss results in terms of transcript abundance: we define that here as "relative abundance" within the normalized transcriptomes from each time point.

Data visualization and statistical analyses

All statistical analyses were performed using R studio (4.2.1), PRIMER (v.7), or GraphPad Prism (v. 8.0.2). To check sphericity of the data for repeated measures ANOVA Phyto-PAM measurements were analyzed using the Mauchly test in R. If the data did not fit assumptions (p-value from Mauchly test was < 0.05), the Geisser Greenhouse correction was used [39]. Repeated measures and two-way ANOVA were run in GraphPad Prism (v. 8.0.2). GraphPad Prism was used for time point comparisons of the physiological measurements (cell counts, Fv/Fm and NPQ) within or between the strains. PRIMER (v.7) was used for non-metric multidimensional scaling (nMDS). Square-root-transformed TPM expression of all genes in the PCC7806SL genome (4,889) was used in calculating the similarity matrix used in the nMDS.

Data availability

All raw paired-read data used in this study is publicly available at NCBI under the BioProjectID PRJNA1008692.

Results

Cell concentration in chemostats

At 26° C, wildtype steady-state cell concentration averaged 3.07 x 106 mL-1 (Fig.

2.1A). When temperature was lowered to 19° C, growth rate dropped, causing cell concentration to decrease to an average of 1.76 x 106 mL-1. By T5 (168 h), wildtype cell concentration stabilized around 2.04 x 106 mL-1 for the remainder of the cold treatment (T6 through T8). In the mutant strain, a similar trend in growth rate decline and cell concentration was observed (Fig. 2.1A). At 26° C, the mutant steady-state concentration averaged 2.46 x106 mL-1. Temperature change to 19° C resulted in an average concentration decline to 1.27 x106 mL-1. By T6 (~192 h), cell concentrations stabilized around 1.39 x 106 mL-1 for the remainder of the cold treatment. When temperature returned to 26° C at T9, wildtype cell concentration increased to an average of 2.12 x 106 mL-1 over 24 h, while the mutant cell concentration increased to an average of 1.88 x 106 mL-1 at T9 (336 h).

Cellular microcystin content

Microcystin concentrations in wild-type cells were highly reproducible between the paired chemostats (largest difference ± 17%) and reached steady-state concentrations of ~10 fg cell-1 at 26° C (Fig. 2.1B). After temperature decreased to 19°C, toxin concentration peaked at T5, reaching 23.4 fg cell-1. Microcystin remained elevated (>20 fg cell-1) in the wildtype for the duration of the cold period (T6-T8). When the temperature returned to 26° C, toxin declined to pre-cold-treatment levels. No microcystins were detected in ΔmcyB cultures (Fig. 2.1B). Though no toxin was produced in the mutant, transcription of mcyD and mcyA followed the expression pattern of the wildtype strain throughout our RNA-seq time series (Appendix Fig. 2.6A-B).

Photosynthetic physiology assessment

In the wildtype, Fv/Fm values stayed consistent throughout the cold treatment, until T6 (~192 h) (Fig. 2.1C). At T6, Fv/Fm values decreased, becoming significantly lower at T7 (~240 h) ($p = 0.038$) and T8 (~263 h) ($p < 0.0001$) compared to T1 (0 h). In the mutant, F_v/F_m was consistently higher than the wildtype, regardless of treatment (Fig. 2.1C). During the cold treatment, F_v/F_m in the mutant was higher at T6 ($p= 0.016$) and T8 ($p= 0.046$) than at T1 (Fig. 2.1C).

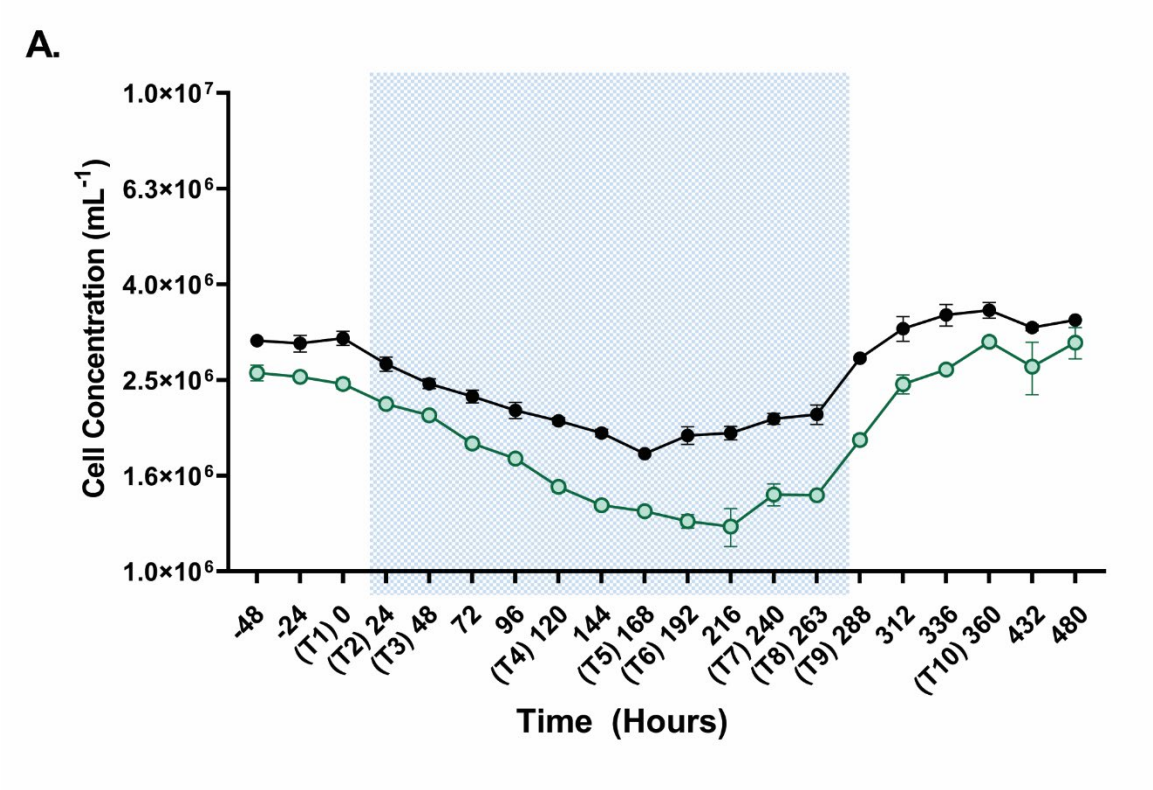

Figure 2.1 PCC7806 wild-type and mutant responses to temperature change in chemostats

(A) Average cell concentrations from duplicate wildtype and mutant mono-culture chemostats. Mutant strain is represented by the green open circles, wildtype is represented by black closed circles. Blue shading indicates the 19ºC cold period, white is indicative of the 26º C warm period. The x-axis is in units of hours, with each of the ten RNA-seq time points. (B) Average microcystin quota in fg/cell from the duplicate wildtype and mutant monoculture chemostats. The mutant strain had no detectable microcystins. The wildtype microcystin quota roughly doubled after 168h in cold treatment, which corresponds to RNA seq time point T5. The increased microcystin quota persisted in the wildtype for the rest of the cold treatments (T5-T8). (C) Average maximum quantum yield of PSII (Fv/Fm) from duplicate wildtype and mutant monoculture chemostats. The mutant shows a steady significant increase in Fv/Fm during cold treatment by T8 in comparison to T1 ($p= 0.0461$). The wildtype strain Fv/Fm shows an inverse trend from the mutant, significantly decreasing by T8 compared to T1 ($p < 0.0001$). In response to the temperature change back to 26º C, the mutant Fv/Fm significantly declined at 312h ($p = 0.044$) and 336h ($p = 0.004$), compared to Fv/Fm taken at T9.

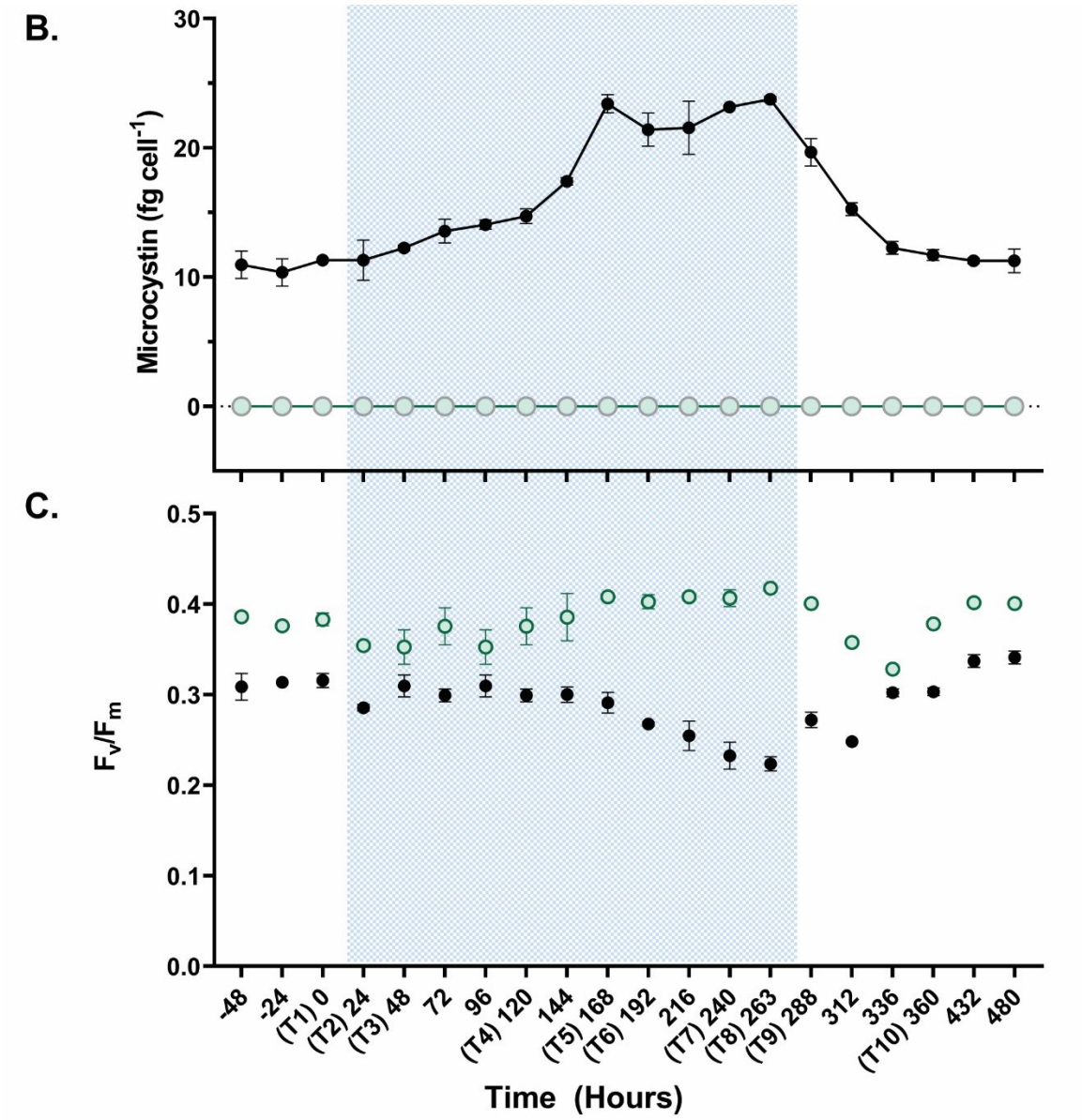

Figure 2.1 Continued.

The mutant and wildtype responded differently during re-acclimation to 26° C. F_v/F_m declined significantly in the mutant in the 48-h period after the temperature was returned to 26° C (T9 vs. 312 h, p = 0.044; T9 vs. 336 h, p = 0.004) (Fig. 2.1C). In contrast, in the wildtype, F_v/F_m increased in the same 48-h period, but the increase was not significant compared to T9.

Non-photochemical quenching (NPQ) measurements were derived from the rapid light curves (Fig. 2.2). In the wildtype, there were no significant differences in NPQ measurements during the cold treatment (T5) compared the control (T1) (Fig. 2.2). In the mutant, there was a significant increase in NPQ measurements taken during the cold treatment (T5) compared to the control (T1). These increases occurred at all PAR (μmol photon m-2 s-1) values >345 (Holm-Sidak padj all <0.05) (Fig. 2.2, Appendix Fig. 2.7). At T1, the mutant had higher NPQ than the WT across all PAR intensities measured, but the differences were not significant. However, NPQ was significantly higher in the mutant at T5 at all seven measured PAR values (p= 0.024, 0.005, 0.001, 0.006, 0.008, 0.005, 0.003). These data suggest differences in acclimation strategies between the two strains.

To assess stress on photosystem II at T5 versus control timepoints (T1 and T10), excitation pressure measurements were derived from the rapid light curves (Appendix Fig. 2.8). The wildtype had significantly higher excitation pressure at T5, compared to T1 and T10 (p= 0.0161, 0.0120). The mutant had significantly higher excitation pressure at T5 versus T10 (p= 0.0087).

Reactive oxygen species damage measurements via the CM-H2DCFDA Assay

The CM-H2DCFDA assay revealed significantly higher net green fluorescence ($p < 0.0001$) in the mutant compared to the wildtype strain during the eleven-day cold temperature treatment period (Appendix Fig. 2.9A-B). There was no significant difference in net green fluorescence between the strains during growth at 26°C.

Transcriptional profiles for Control, cold treatment and cold recovery

The transcriptional profiles of chemostat populations from the ten time points are

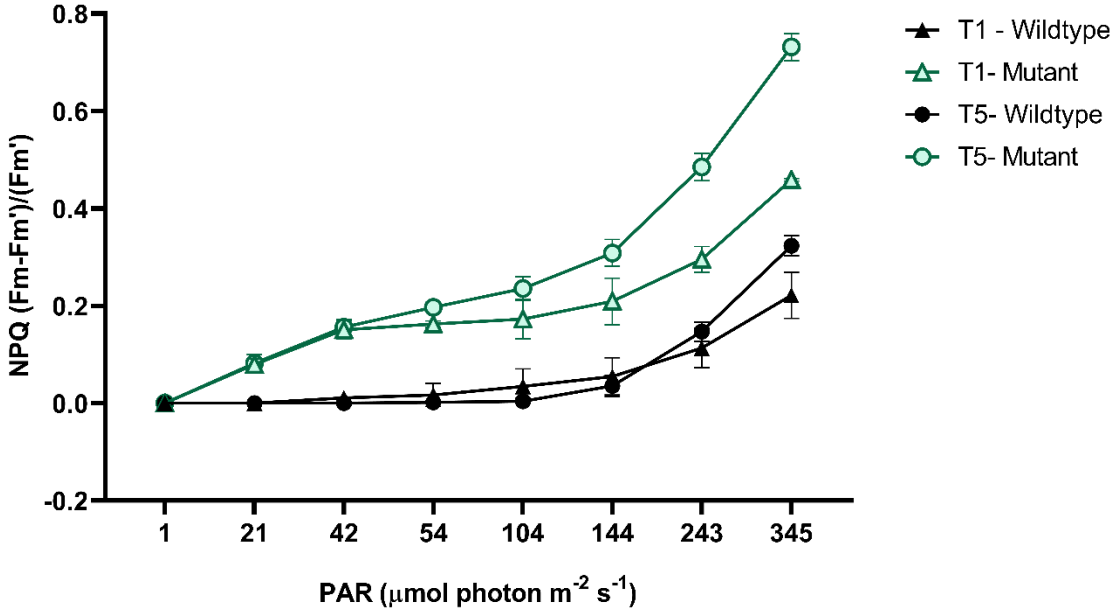

Figure 2.2 Non-photochemical quenching data for mutant and wildtype chemostats at T1 and T5

Average non-photochemical quenching measurements (NPQ) collected from the rapid light curves (RLC) for each chemostat, at T1 (0 h) and T5 (~168 h). NPQ significantly differs between the mutant and wildtype for measurements taken at the T5 time point across all actinic PAR (μmol photons m^{-2} s^{-1}) measured (Holm sidak adj. p-values = 0.024, 0.005, 0.001, 0.006, 0.008, 0.005, 0.003).

provided in Appendix sheet 2.1A. We sequenced 40 mRNA libraries with total reads ranging from 25,573,750 to 78,079,140 per library (Appendix sheet 2.1B). Broad transcriptional patterns were determined by projecting samples in two-dimensional space (nMDS) based on the expression of all genes in each genome (Fig. 2.3). This nMDS representation resulted in a circular progression across measured time points (T1-T10). In both strains, T2-T9 scale away similarly from T1 in a counter-clockwise pattern, showing an increasing divergence in gene expression during cold acclimation. However, the mutant transcriptomes diverge from T1 more than the wildtype. Re-acclimation to 26° C (T10) resulted in the transcriptional profile returning to a state similar to T1 in both strains.

Transcriptomic differences between the mutant and wildtype in warm conditions (T1 and T10).

There were 61 differentially expressed genes (fold change ≥2) at T1 and 94 at T10, many of which (42 genes) overlapped. A significant finding was differential expression of genes located in a putative gene cluster with homology to transmembrane proteins, and genes with functions related to redox metabolism/ membrane potential. At T1, all genes from gene cluster 1 were upregulated 4.6 to 22.8-fold (all $p < 0.001$) in the mutant strain (Appendix Fig. 2.10). At T10, transcripts from seven of the genes in putative gene cluster 1 were upregulated 2.3 to 9.1-fold in the mutant ($p < 0.001$). All other differentially expressed genes at these time points (T1 and T10), and the rest of the RNA seq time points (T2-T9) can be found in Appendix Attachment sheets 2.1C-M.

Transcript differences between the mutant and wildtype in cold conditions (T2-T8).

Generally, differences in gene expression during cold growth involved changes in a glutathione-dependent peroxiredoxin, and cytochrome-based electron transport pathways (Fig. 2.4, Fig. 2.5A-D). Cold treatment produced differential expression of a glutathione-dependent peroxiredoxin (*pgdx*), which increased in the mutant relative to the wildtype starting at T4. By T5, *pgdx* showed a 2.1-fold-change increase in the mutant ($p < 0.001$) (Fig. 2.4). There was no differential expression (fold change ≥2) observed for

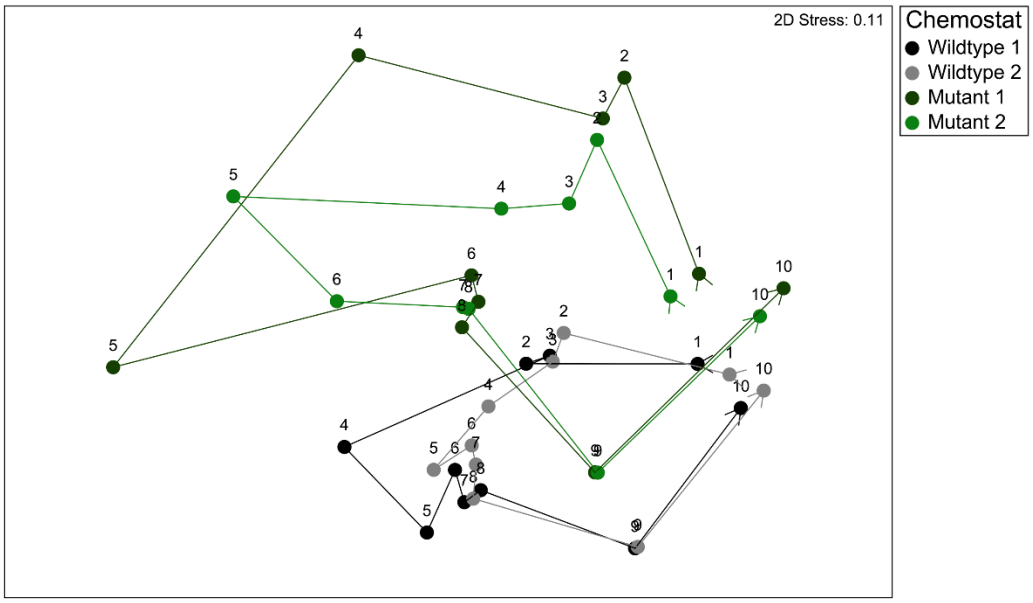

Figure 2.3 NMDS of Square Root Transformed Transcriptomic Libraries for Mutant and Wildtype Chemostats

NMDS of square root transformed transcriptomic data, normalized by transcripts per million (TPM), encompassing the whole transcriptome for wildtype (black and gray circles) and mutant (green and dark green circles) monoculture chemostats, n = 40. The nMDS was constructed to see how the transcriptomes changed over time throughout the cold temperature treatment at each RNA- seq timepoint (T1 through T10).

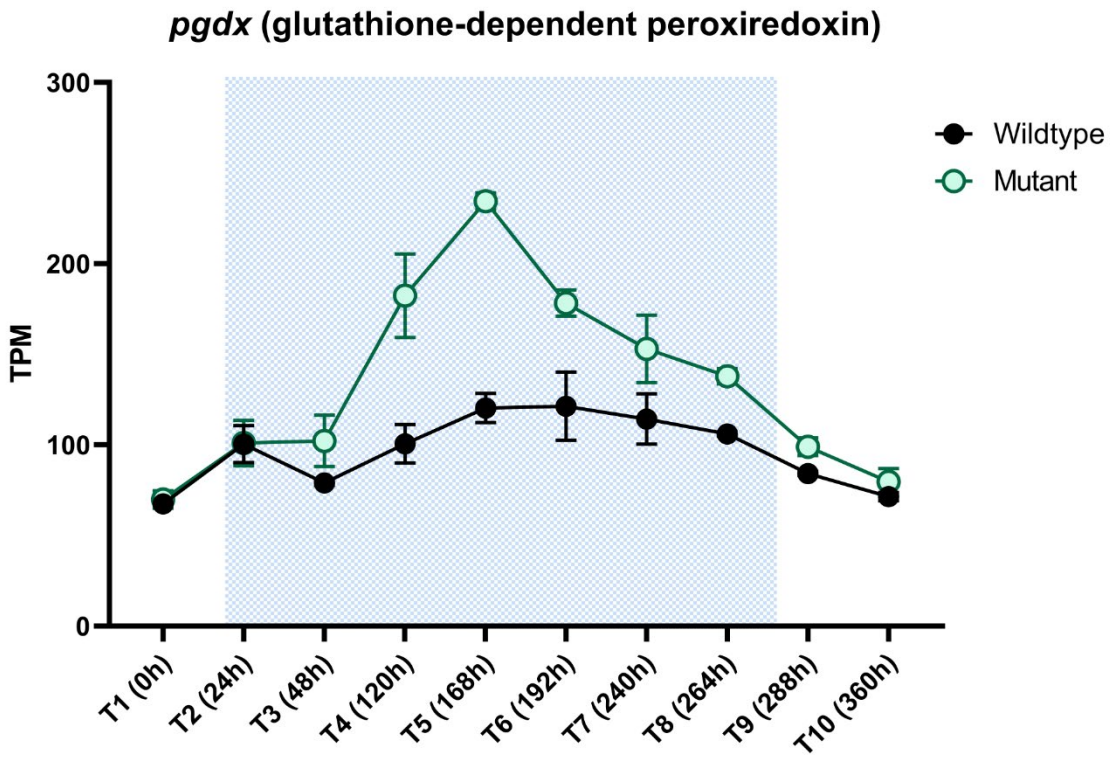

Figure 2.4 Gene Expression Profile of a Glutathione-Dependent Peroxiredoxin in the Mutant and Wildtype

Average transcriptomic data normalized by transcript per million (TPM), for the mutant (green circles) and wildtype (black circles) for the glutathione-dependent peroxiredoxin, *pgdx*. At T5 the mutant shows a 2.14 fold change increase ($p < 0.0001$) in transcript abundance of *pgdx*.

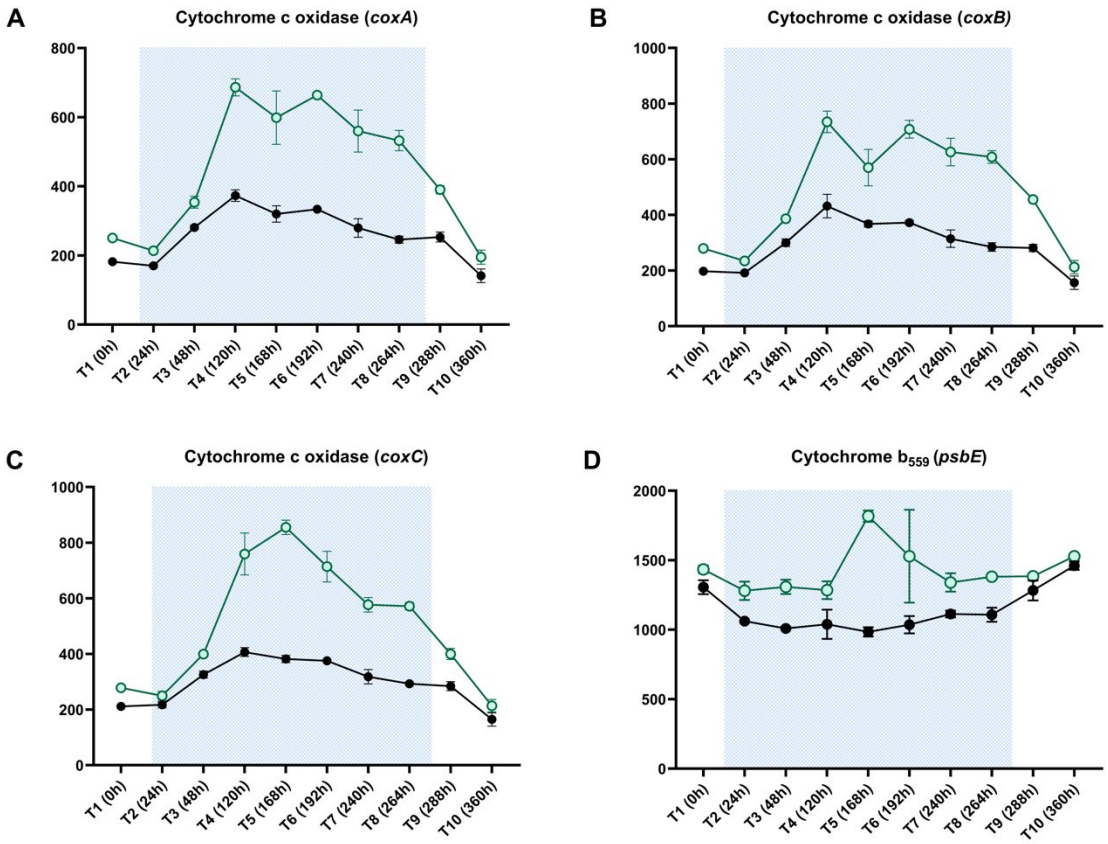

Figure 2.5 Transcriptomic Profiles of *coxBAC* for the mutant and wildtype

Average transcriptomic data, normalized by transcript per million (TPM), for the mutant (green circles) and wildtype (black circles) for cytochrome c oxidase, *coxBAC*. The x-axis shows each RNA-seq time point in order, from time 0 to 360 hours. (A) Starting at T4, *coxA* shows a fold change increase of 2.1 ($p < 0.0001$). This fold change increase >2 of *coxA* is seen through the rest of the cold treatment period (T5-T8) (Appendix sheets 2.1C-M). (B) Starting at T4, *coxC* shows a fold change increase of 2.1($p<0.001$) in the mutant. This fold change increase >2 of *coxC* is seen through the rest of the cold treatment period (T5-T8). (C) At T6, *coxB* shows fold change increase of 2.15 ($p < 0.0001$) in the mutant, fold change >2 of this gene is maintained through T7 and T8 in the mutant. (D) In the mutant, cytochrome b_{559} (*psbE*) saw a fold change increase of 2.04, with $p < 0.0001$ at T5.

other antioxidant genes (*sod*, *prx*) between the strains, however there was a thioredoxin-dependent 2-cys peroxiredoxin, (locus tag BH695_RS02415), which showed fold change increases of ≥1.5 but < 2 (FDR p <0.0001) from T6-T8 in the mutant (Appendix Fig. 2.11).

In the mutant, a subunit of the photoprotective cytochrome b_{559} complex in PSII (*psbE)* and the respiratory pathway based on cytochrome c oxidase (*coxBAC*) showed differential expression during cold treatment (Fig. 2.5A-D). From T4 to T8, cytochrome c oxidase expression (*coxBAC*) increased in the mutant (Fig. 2.5A-C, Appendix attachment sheets 2.1F-L), while cytochrome b_{559} alpha subunit (*psbE*), an electron sink, increased ~2-fold, (p <0.001) at T5 (Fig. 2.5D). Based on *psbE* and *coxBAC* expression, the mutant had increased expression of thylakoid-membrane-associated electron sinks during cold acclimation.

We also saw changes in expression for genes *pntA* (2 copies), *pntB, petJ*, and *petE*, all of which are associated with proton/electron transporting processes. The genes *pntA* (both copies), *pntB* and *petJ* did not meet our criteria of fold change ≥2, instead they showed fold changes ≥1.5 but <2, with FDR-p values <0.0001 from T6-T8 (Appendix Figs. 2.12, and 2.1.3A). The electron transporter, plastocyanin (*petE*), showed fold change decreases >2 (FDR-p < 0.0001) from T7-T9 in the mutant (Fig. 2.13B).

Discussion

The purpose of this study was to determine if changes in microcystin production provided an advantage to *Microcystis aeruginosa* PCC 7806 in its response to cold stress. Here we used cold acclimation under constant light – conditions known to generate oxidative stress in phototrophs [30, 40, 41]. We compared the response of non-toxigenic PCC 7806 Δ*mcyB* [31] with its corresponding toxic wild type. These strains were chosen to allow comparison within the most genetically similar backgrounds [e.g., 42]. We note that over time, genetic changes have accumulated between the two strains [33]. As part of this study, we validated that genes individually reported herein as differentially expressed were consistent in terms of genomic structure between strains. Genes that were identified as mutated were not considered in our gene expression analysis within the main

text or supplemental figures. The physiology of the wildtype and mutant strains were monitored with various methods during growth in chemostats shifted from 26° C to 19° C to induce cold stress [vis a vis 6], a condition known to induce microcystin production. By pairing the temporal RNA-seq analyses with photo-physiology and ROS measurements, we have identified differences in how a toxigenic *vs.* non-toxigenic strain responds and recovers to cool temperature exposure. We present these observations couched within a framework that describes how microcystin influences cell physiology and distinguishes between a defensive role microcystin may play in redox-homeostasis vs. the mutant's active metabolic restructuring to increase electron sink capacity and degrade ROS.

Microcystin has been implicated as playing a role in the oxidative stress response, due to its thiol-binding abilities [18, 25]. Oxidative stress can be generated by photosynthetic processes in cyanobacteria, where disruptions in the redox state of the thylakoid membranes can increase ROS generation if electron source exceeds electron sink capacity. During the cold temperature treatment, we observed a general trend of more reactive thiol groups in the mutant via our CM-H_2DCFDA assay compared to the wildtype, indicating more ROS damage in that strain (Fig. S3). Interestingly, we also observed an increase in expression of a glutathione-dependent peroxiredoxin (*pgdx*) in the mutant a week (T5) into cold treatment, which coincides with microcystin quota peaking and cell concentration recovery in the wildtype (mutant cold recovery was slightly delayed and started at T6). Glutathione-dependent peroxiredoxins are important for H_2O_2 detoxification and have also been implicated as acting as redox sensors [43, 44]. Differential expression of *pdgx* in the mutant during cold treatment is indicative of intracellular stress conditions absent in the wildtype. While other ROS scavengers, such as superoxide dismutase and other peroxiredoxins, behaved similarly in the mutant and wildtype (no fold change >2), the behavior of *pgdx* might indicate more specific regulatory activity than just ROS detoxification alone. Thus, while ROS protection may be one role of microcystin in the cell, the specifics as to the type or location of ROS-inducing damage (H_2O_2 v. O^-) needs more attention.

Glutathione-dependent peroxiredoxins (or glutaredoxins) are reliant on

glutathione reductase activity, whose disulfide reductase activity is powered by NADPH [22, 45]. Microcystin is capable of binding to glutathione reductase *in vivo*, and it has been hypothesized this binding might either aid or interfere in the oxidative stress response [18, 25]. While it is possible the presence of microcystin could be altering the activity of *pgdx* in the wildtype, broader implications of microcystin interference with the glutathione/thioredoxin networks might mean that less NADPH is required for cold metabolism in the wildtype, a scenario consistent with the decreased F_v/F_m beginning at T7. This reduction in F_v/F_m came with no impediment to wildtype growth. Conversely, the mutant not only showed increased F_v/F_m later into the cold treatment, but there was also an increase in transcription of an NADPH transhydrogenase (*pntAB*) in the mutant relative to the wildtype from T5 through T8 (Fig. S7). This may indicate the mutant takes more action to properly maintain NAD(P)+/NAD(P)H pools during cold growth, which could signify differences in redox coupling of these cofactors and intracellular redox homeostasis in the mutant [46].

We saw additional transcriptional and photo-physiological evidence of metabolic "re-wiring" in the mutant to acclimate to the cold, around the same time *pgdx* gene expression peaked. The mutant upregulated cytochrome b_{559} (*psbE*) and cytochrome c oxidase (*coxBAC*) in what we interpret as an effort to increase electron sink capacity to help reduce ROS generation. Cytochrome b_{559} has been shown to aid in photoprotection, by acting as an alternate electron pathway [47]. Cytochrome c oxidase (*coxBAC*) is a respiratory terminal oxidase that generates ATP by consuming oxygen and NAD(P)H reducing equivalents [48] . However, cytochrome c oxidase serves a dual role in the cell, both increasing ATP generation and having a photoprotective role, where it acts as an additional electron sink by diverting electrons away from the donor side of photosystem I via plastocyanin or cytochrome c_6 [48, 49]. Because of this, it is thought that activity of this complex in the light may be regulated by the redox state of the electron carriers plastocyanin and cytochrome c_6 [50]. Differential expression of these two electron transporters was observed in the mutant during the cold treatment (Fig. S8A-B). In terms of photoprotective abilities, there was an inverse trend of the *pgdx* expression with cytochrome c oxidase and cytochrome b_{559} expression, giving more credibility to the

hypothesis that they are upregulated to help modulate thylakoid redox state in the mutant. Additionally, cytochrome c oxidase relies on NAD(P)H for proton translocation, which further supports an increased need (or may function as a sink) for NAD(P)H in the mutant during cold temperature growth.

Based on transcriptional data presented, we conclude that the mutant and wildtype allocate excitation energy differently. Our NPQ data support this. The mutant had higher NPQ at all incident PAR at T5 compared to the wildtype. This means that during cold growth, the mutant would be more reliant on photo-protective energy dissipation through non-photochemical quenching, which can be achieved either by the orange carotenoid protein in photosystem II or state transitions [51, 52]. Increasing electron sinks and induction of NPQ processes in the mutant would decrease generation of ROS, which can result from an over-reduced thylakoid membrane [53, 54]. This is interesting, as the excitation pressure of PSII significantly increases in the wildtype at T5 (compared to controls T1 and T10) (Fig. S3). However, the wildtype manages this without increasing alternate electron pathways, or in the short term without increasing NPQ, and still recovers during cold growth despite reduced F_v/F_m after this time. When taken together, these findings show that the mutant and wildtype have different energy dissipation capabilities during cold growth, which may affect NAD(P)+/NAD(P)H pools or the thylakoid membranes. The absence of microcystin in the mutant increases the need for electron dissipating pathways, indicating microcystin may play a role in processes related to the photosynthetic electron transport chain. Complementary to this idea, if the mutant employs these strategies to mitigate or reduce ROS production, the absence of "turning on" these strategies in the wildtype could mean microcystin is in fact playing a role in the oxidative stress response, as has been previously hypothesized by other researchers.

Microcystin production alters cold stress management strategies

In response to cold treatment (11 days), the non-toxigenic mutant upregulated a glutathione-dependent peroxiredoxin and a cytochrome c oxidase complex, both of which help mitigate or reduce ROS generation in cyanobacteria. These responses temporally coincided with an increase of microcystin in the wildtype. Coincident in time, photo-

acclimation strategies differed, in which the strains showed opposite response in F_v/F_m, and NPQ measurements increased in the mutant. This implies cold acclimation alters electron transport capabilities differently in the mutant and wildtype strains. Our NPQ measurements also indicate inherent differences in the redox state of the thylakoid membranes between the strains during cool temperatures [14]. Indeed, differences in redox-related proteins have been shown to differ between these strains, and conditions that can cause over reduction of the thylakoid membranes, such as high-light growth, provide the toxigenic strain an advantage [17, 55, 56]. Collectively, our observations lead to the idea that increased microcystin per-cell during the cold treatment may provide a longer-term benefit (days versus hours) during conditions which can stress photosynthetic metabolism resulting in more ROS generation (here brought on by cold temperature growth). This leads us to our hypothesis that in locations that experience seasonal temperature trends like Lake Erie, microcystin quota may be higher under conditions where solar days are longer, and the water is clearer and colder (*e.g.,* June). This phenomenon may be overlooked by total microcystin measurements (μg/L) alone, as overall biomass is lower earlier in the growing season, and may be better observed with per-cell toxin quotas. The replacement of microcystin producers by non-producers in August (where temperatures are warmer, solar days shorter and water less clear) occurs during periods that may favor photo acclimation strategies that differ from those seen during high light and low temperature growth periods. This would include the use of enzymes that directly degrade reactive oxygen (*e.g.,* like peroxiredoxin), noting that specific strategies may be tailored to certain reactive oxygen species. Thus, there could be a connection between photo acclimation strategies for *Microcystis*, the nature of the photosynthetic stressors, and succession of toxigenic genotypes of *Microcystis*. We hypothesize that trade-offs in the differing strategies of managing photosynthetic stressors may therefore play an important role in regulating/selecting the toxigenic composition of *Microcystis* blooms (although other factors are also likely playing a role) [57]. Agent-based modeling has suggested that nitrogen availability may contribute to the succession of toxigenic to non-toxigenic strains in Lake Erie [58]. Data from our experiment can be integrated into models such as this, allowing us to quantify how

nutrients and temperature contribute to genotypic succession. Doing so may help improve management and mitigation strategies of *Microcystis*-dominated harmful algal blooms. Yet it is important to acknowledge that *Microcystis* blooms are complicated: the influence(s) of other environmental factors as well as competition with biology needs to be appropriately accounted for in our stewardship of these systems.

Acknowledgements

This work was supported by funds from the *National Science Foundation* (OCE-1840715) and the *National Institute of Environmental Health Sciences* (1P01ES028939–01) through funds provided to the *Great Lakes Center for Fresh Waters and Human Health* at Bowling Green State University, and the National Oceanographic and Atmospheric Administration (NA18NOS4780175). We also thank the generous support of the *Kenneth & Blaire Mossman Endowment* to the *University of Tennessee.*

References

1. Dick, G.J., et al., *The genetic and ecophysiological diversity of Microcystis.* Environmental Microbiology, 2021. **23**(12): p. 7278-7313.

2. Peng, G., et al., *Seasonally Relevant Cool Temperatures Interact with N Chemistry to Increase Microcystins Produced in Lab Cultures of Microcystis aeruginosa NIES-843.* Environmental Science & Technology, 2018. **52**(7): p. 4127-4136.

3. Rinta-Kanto, J.M., et al., *Lake Erie Microcystis: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. Harmful Algae*, 2009. **8**(5): p. 665-673.

4. Briand, E., et al., *Spatiotemporal changes in the genetic diversity of a bloom-forming Microcystis aeruginosa (cyanobacteria) population. The ISME Journal*, 2009. **3**(4): p. 419-429.

5. Wilhelm, S.W., G.S. Bullerjahn, and R.M.L. McKay, *The complicated and confusing ecology of Microcystis blooms. mBio*, 2020. **11**(3).

6. Martin, R.M., et al., *Episodic decrease in temperature increases mcy gene transcription and cellular microcystin in continuous cultures of Microcystis aeruginosa PCC 7806.* Frontiers in Microbiology, 2020. **11**.

7. Paerl, H.W. and J. Huisman, *Blooms like it hot.* Science, 2008. **320**: p. 57-58.

8. Wetzel, R.G., *Limnology: Lake and River Ecosystems*. Third Edition. 2001: Elsevier Science. 1005.

9. Suzuki, I., et al., *Cold-regulated genes under control of the cold sensor Hik33 in Synechocystis. Molecular Microbiology*, 2001. **40**(1): p. 235-244.

10. Mikami, K., et al., *The histidine kinase Hik33 perceives osmotic stress and cold stress in Synechocystis sp. PCC 6803. Molecular Microbiology*, 2002. **46**(4): p. 905-915.

11. Ritter, S.P.A., et al., *Evidence for convergent sensing of multiple abiotic stresses in cyanobacteria. Biochimica et Biophysica Acta (BBA) - General Subjects*, 2020. **1864**(1): p. 129462.

12. Maksimov, E.G., *Membrane fluidity controls redox-regulated cold stress responses in cyanobacteria.* Photosynthesis Research, 2017.

13. Williams, P.W. and D.J. Peter, *Control of excitation energy distribution in*

cyanobacteria: sensitivity to uncouplers and ATP synthase inhibitors. Biochimica et Biophysica Acta (BBA) 1989. **1015**: p. 121-130.

14. Misumi, M., et al., *Relationship between photochemical quenching and non-photochemical quenching in six species of cyanobacteria reveals species difference in redox state and species commonality in energy dissipation. Plant and Cell Physiology*, 2015: p. pcv185.

15. Ma, W., Y. Deng, and H. Mi, *Redox of plastoquinone pool regulates the expression and activity of NADPH dehydrogenase supercomplex in Synechocystis sp. PCC6803. Current Microbiology*, 2008. **56**(2): p. 189-193.

16. Liu, L.-N., et al., *Control of electron transport routes through redox-regulated redistribution of respiratory complexes. Proceedings of the National Academy of Sciences*, 2012. **109**(28): p. 11431-11436.

17. Alexova, R., et al., *Specific global responses to N and Fe nutrition in toxic and non-toxic Microcystis aeruginosa. Environmental Microbiology*, 2016. **18**(2): p. 401-413.

18. Zilliges, Y., et al., *The cyanobacterial hepatotoxin microcystin binds to proteins and increases the fitness of Microcystis under oxidative stress conditions. PLoS ONE*, 2011. **6**(3): p. e17615.

19. Mullineaux, P.M. and T. Rausch, *Glutathione, photosynthesis and the redox regulation of stress-responsive gene expression. Photosynthesis Research*, 2005. **86**(3): p. 459-474.

20. Deblois, C.P. and P. Juneau, *Relationship between photosynthetic processes and microcystin in Microcystis aeruginosa grown under different photon irradiances. Harmful Algae*, 2010. **9**(1): p. 18-24.

21. Sevilla, E., et al., *Redox-Based Transcriptional Regulation in Prokaryotes: Revisiting Model Mechanisms.* Antioxidants & Redox Signaling, 2019. **30**(13): p. 1651-1696.

22. Lu, J. and A. Holmgren, *The thioredoxin antioxidant system.* Free Radic Biol Med, 2014. **66**: p. 75-87.

23. Lindahl, M. and T. Kieselbach, *Disulphide proteomes and interactions with thioredoxin on the track towards understanding redox regulation in chloroplasts and cyanobacteria. Journal of Proteomics*, 2009. **72**(3): p. 416-438.

24. Nishiyama, Y., S.I. Allakhverdiev, and N. Murata, *Inhibition of the repair of Photosystem II by oxidative stress in cyanobacteria. Photosynthesis Research*, 2005. **84**(1-3): p. 1-7.

25. Schuurmans, J.M., Brinkmann, B.W., Makower, A.K., Dittmann, E., Huisman, J., Matthijs, H.C.P.,, *Microcystin interferes with defense against high oxidative stress in harmful cyanobacteria. Harmful Algae*, 2018. **78**.

26. MacKintosh, R.W., Dalby, K.N., Campbell, D.G., Cohen, P.T.W, Cohen, P, MacKintosh, C. , *The cyanobacterial toxin microcystin binds covalently to cysteine-273 on protein phosphatase 1. FEBS Letters*, 1995. **371**(3): p. 236-240.

27. Grossman, A.R., et al., *Phycobilisome degradation and responses of cyanobacteria to nutrient limitation and high light*, in *Photosynthesis: Mechanisms and Effects*. 1998, Springer Netherlands. p. 2853-2858.

28. Zhang, Z., et al., *Gene expression patterns of sulfur starvation in Synechocystis sp. PCC 6803.* BMC Genomics, 2008. **9**: p. 344.

29. Wilhelm, S.W. and G.L. Boyer, *Healthy competition.* Nature Climate Change, 2011. **1**(6): p. 300-301.

30. Maxwell, D.P., et al., *Growth at low temperature mimics high-light acclimation in Chlorella vulgaris. Plant Physiology*, 1994. **105**(2): p. 535-543.

31. Dittmann, E., et al., *Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium Microcystis aeruginosa PCC 7806.* Molecular Microbiology, 1997. **26**: p. 779-787.

32. Steffen, M.M., et al., *Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis.* The ISME Journal, 2014. **8**(10): p. 2080-2092.

33. Stark, G.F., et al., *Closed circular genome sequence of a Microcystis aeruginosa PCC7806 -mcyB (UTK) nontoxic mutant.* Microbiology Resource Announcements, 2023: p. in press.

34. Watanabe, M.M. and T. Ichimura, *Fresh-and salt-water forms of Spirulina platensis in axenic cultures.* Bulletin of the Japanesse Society for Phycology, 1977: p. 371-377.

35. Bertani, G., *Studies on Lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli.* Journal of Bacteriology, 1952. **62**: p. 293-300.

36. Boyer, G.L. *LCMS-SOP Determination of Microcystins in Water Samples by High Performance Liquid Chromatography (HPLC) with Single Quadrupole Mass Spectrometry (MS)*. protocols.io, 2020. DOI: 10.17504/protocols.io.bck2iuye.

37. Martin, R.M. and S.W. Wilhelm, *Phenol based RNA extraction from*

polycarbonate filters. protocols.io, 2020.

38. Zhao, L., et al., *The highly heterogeneous methylated genomes and diverse restriction-modification systems of bloom-forming Microcystis.* Harmful Algae, 2018. **75**: p. 87-93.

39. Greenhouse, S.W. and S. Geisser, *On methods in the analysis of profile data. Psychometrika*, 1959. **24**(2): p. 95-112.

40. Gray, G.R., et al., *Photosystem II excitation pressure and development of resistance to photoinhibition (II. Adjustment of photosynthetic capacity in Winter Wheat and Winter Rye). Plant Physiology*, 1996. **110**(1): p. 61-71.

41. Huner, N.P.A., G. Öquist, and F. Sarhan, *Energy balance and acclimation to light and cold. Trends in Plant Science*, 1998. **3**(6): p. 224-230.

42. Van de Waal, D.B., et al., *Reversal in competitive dominance of a toxic versus non-toxic cyanobacterium in response to rising CO_2.* The ISME Journal, 2011. **5**(9): p. 1438-1450.

43. Dietz, K.-J., et al., *The role of peroxiredoxins in oxygenic photosynthesis of cyanobacteria and higher plants: peroxide detoxification or redox sensing?*, in *Advances in Photosynthesis and Respiration* G.Garab, Editor. 2008, Springer Netherlands. p. 303-319.

44. Chae, H.Z., et al., *Protein Glutathionylation in the Regulation of Peroxiredoxins: A Family of Thiol-Specific Peroxidases That Function As Antioxidants, Molecular Chaperones, and Signal Modulators.* Antioxidants & Redox Signaling, 2012. **16**(6): p. 506-523.

45. Deponte, M., *Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes.* Biochimica et Biophysica Acta, 2013. **1830**(5): p. 3217-66.

46. Kämäräinen, J., et al., *Pyridine nucleotide transhydrogenase PntAB is essential for optimal growth and photosynthetic integrity under low-light mixotrophic conditions in Synechocystis sp. PCC6803. New Phytologist*, 2017. **214**(1): p. 194-204.

47. Hamilton, M.L., et al., *Investigating the photoprotective role of cytochrome b-559 in photosystem II in a mutant with altered ligation of the haem. Plant Cell Physiology*, 2014. **55**(7): p. 1276-85.

48. Lea-Smith, D.J., et al., *Photosynthetic, respiratory and extracellular electron transport pathways in cyanobacteria.* Biochimica et Biophysica Acta, 2016. **1857**(3): p. 247-55.

49. Nikkanen, L., et al., *Regulatory electron transport pathways of photosynthesis in cyanobacteria and microalgae: Recent advances and biotechnological prospects. Physiologia Plantarum*, 2021. **173**(2): p. 514-525.

50. Ermakova, M., et al., *Distinguishing the roles of thylakoid respiratory terminal oxidases in the cyanobacterium Synechocystis sp. PCC 6803.* Plant physiology, 2016: p. pp.00479.2016.

51. Campbell, D. and G. Oquist, *Predicting light acclimation in cyanobacteria from nonphotochemical quenching of photosystem II fluorescence, which reflects state transitions in these organisms. Plant Physiology*, 1996. **111**(4): p. 1293-1298.

52. Boulay, C., et al., *Occurrence and function of the orange carotenoid protein in photoprotective mechanisms in various cyanobacteria. Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2008. **1777**(10): p. 1344-1354.

53. Kanervo, E., M. Suorsa, and E.-M. Aro, *Functional flexibility and acclimation of the thylakoid membrane. Photochemical & Photobiological Sciences*, 2005. **4**(12): p. 1072-1080.

54. Gao, Y. and D.L. Erdner, *Cell death responses to acute high light mediated by non-photochemical quenching in the dinoflagellate Karenia brevis. Scientific Reports*, 2022. **12**(1): p. 14081.

55. Phelan, R.R. and T.G. Downing, *A growth advantage for microcystin production by Microcystis PCC7806 under high light. Journal of Phycology*, 2011. **47**(6): p. 1241-1246.

56. Alexova, R., et al., *Comparative Protein Expression in Different Strains of the Bloom-forming Cyanobacterium Microcystis aeruginosa. Molecular & Cellular Proteomics*, 2011. **10**(9): p. M110.003749.

57. Chaffin, J.D., et al., *Microcystin congeners in Lake Erie follow the seasonal pattern of nitrogen availability.* Harmful algae, 2023. **127**: p. 102466.

58. Hellweger, F.L., Martin, Robbie M. , Eigemann, Falk , Smith, Derek J., Dick, Gregory J. , Wilhelm, Steven W. , *Models predict planned phosphorus load reduction will make Lake Erie more toxic. Science*, 2022. **376**: p. 1001-1005.

Appendix

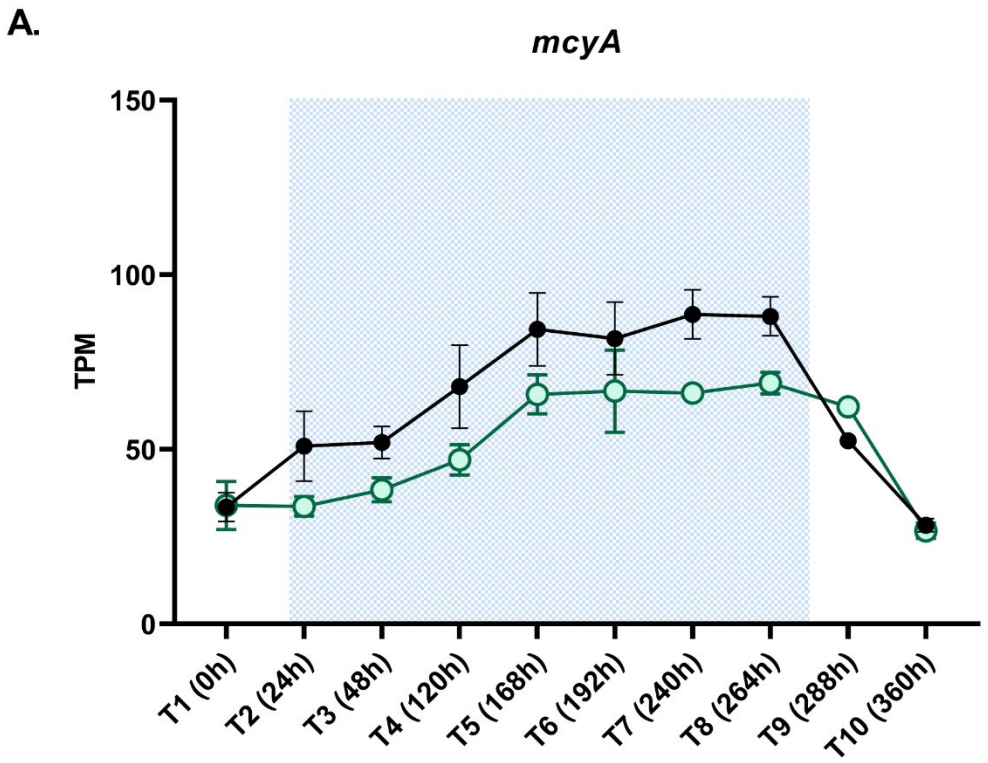

Figure 2.6 Average TPM of *mcyA* and *mcyD* for the mutant and wildtype

Average TPM for *mcyA* and *mcyB,* the two genes in the *mcy* operon induced by the bidirectional promotor. Blue shading is indicative of the cold treatment period. Both the mutant (green circles) and wildtype (black circles) show similar expression patterns (all fold changes < 1.5).

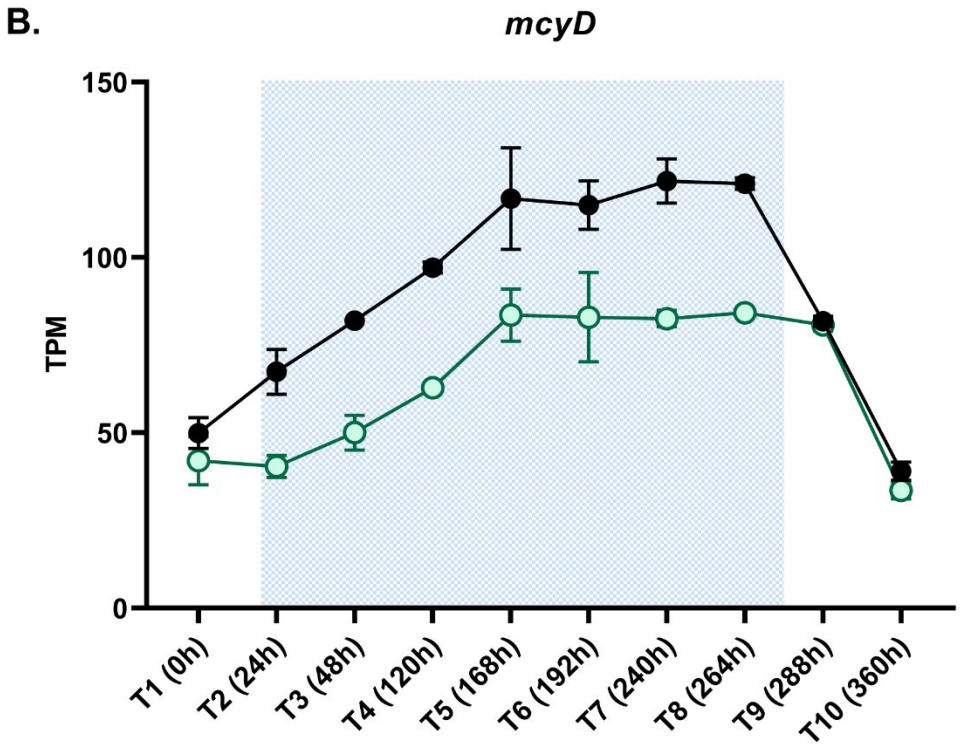

Figure 2.6 Continued

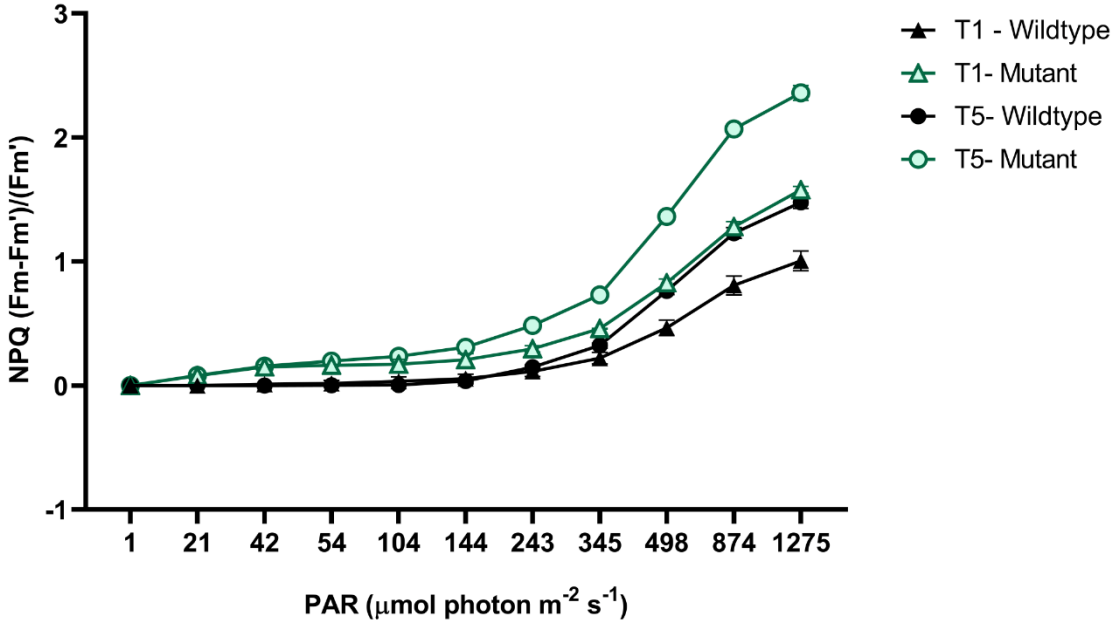

Figure 2.7 Non-photochemical measurements quenching at all PAR for the mutant and wildtype

NPQ at all measured PAR levels (µmol photons m^{-2} s^{-1}), for the wildtype and mutant strain. For the mutant, NPQ for PAR levels >345 are all significantly higher at T5 then T1 (Holm-Sidak adj. p= 0.005, 0.002, 0.002, 0.002, 0.002, 0.003, 0.003). For the wildtype, there is no significant difference in NPQ between T1 and T5 for any measured PAR.

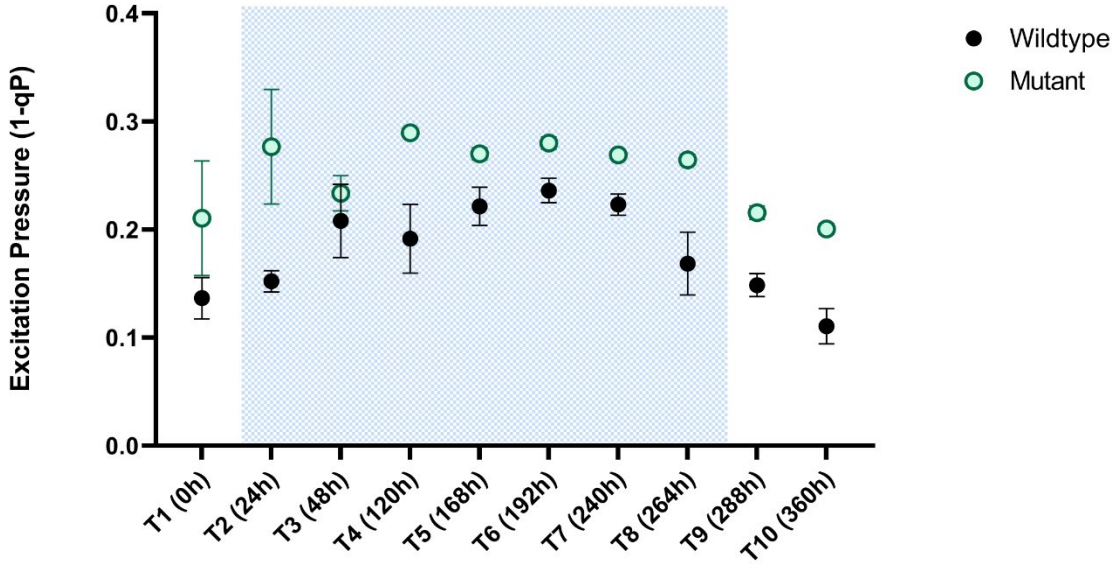

Figure 2.8 Excitation pressure measurements for the mutant and wildtype

Excitation pressure at 54 µmol photons m^{-2} s^{-1} PAR actinic exposure (30 seconds) for the mutant and wildtype extracted from the RLC's. This is around the same growth light intensity the strains experienced in chemostats (~ 50 µmol photons m^{-2} s^{-1}). Repeated measures ANOVA corrected for multiple comparisons with Dunnett's statistical hypothesis testing (only looking at T5 comparisons) was done to obtain p-values. The wildtype had significantly higher excitation pressure at T5, compared to T1 and T10 (p= 0.0161, 0.0120). The mutant had significantly higher excitation pressure at T5 versus T10 (p= 0.0087).

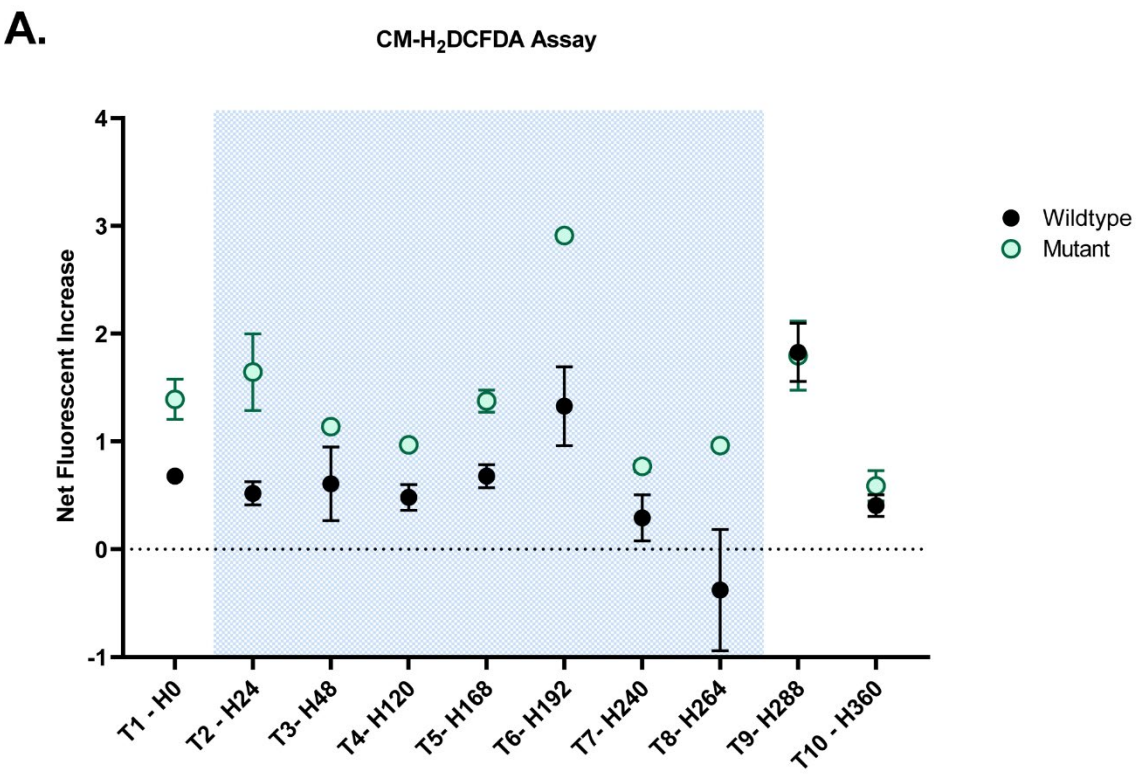

Figure 2.9 CM-H_2DCFDA Assay for the Mutant and Wildtype

Average net fluorescent increase in green fluorescence (CM-H2DCFDA signal) between the mutant (green circles) and wildtype, (black circles) that corresponds to each RNA-seq timepoint, starting at T1 (0 hrs) and ending at T10 (360 hrs). Blue shading is indicative of the cold treatment period. Within each time point, the mutant always showed a non-significantly higher net green florescence then the wildtype. (B) Combined net fluorescence increase taken from all time points from each chemostat at 26°C (7 days total) and 19°C (11 days). A t-test was conducted using FDR corrected p-value with the Benhamini, Kreiger and Yekutieli method to compare between strains within the given temperature treatment. The mutant showed a significantly higher increase in green fluorescence ($p < 0.0001$) after control green fluorescence was subtracted out compared to the wildtype.

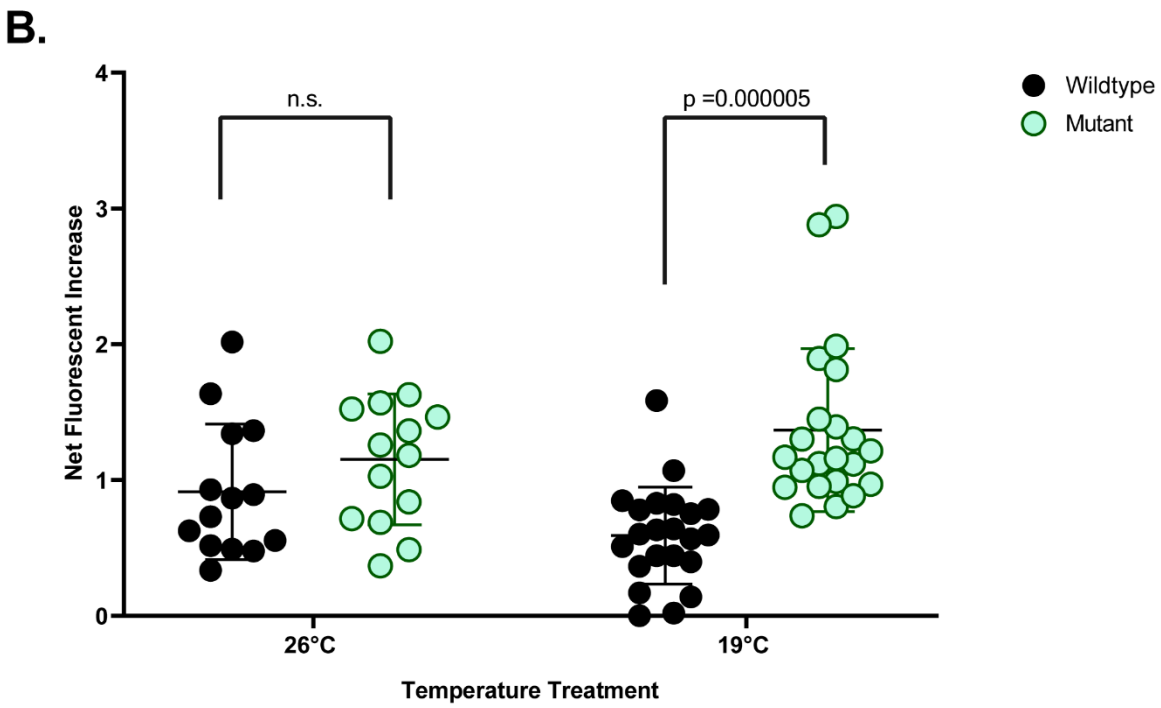

Figure 2.9 Continued.

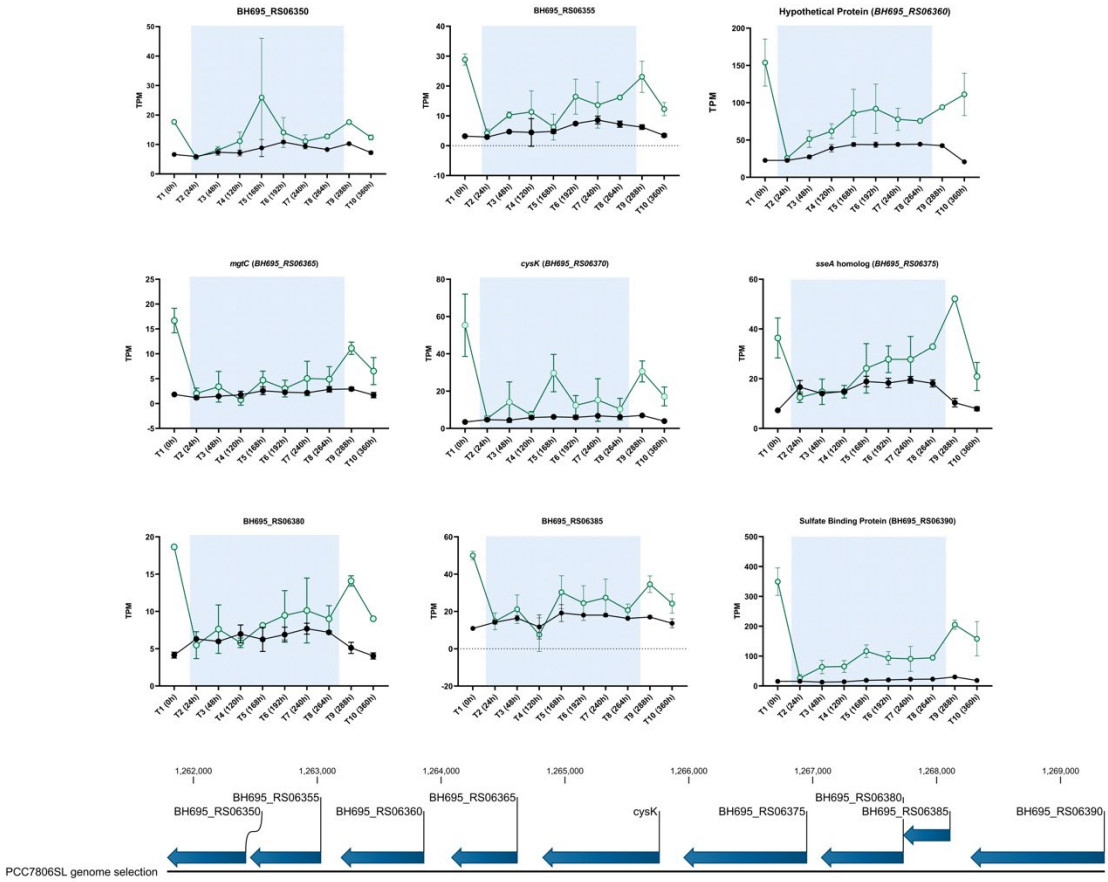

Figure 2.10 Putative Gene Cluster Differentially Expressed between the Mutant and Wildtype

Putative gene cluster 1 TPM values (BH695_RS06350, BH695_RS06355, BH695_RS06360_BH695_RS06365, cysK, sseA homolog, BH695_RS06380, BH695_RS06385, BH695_RS06390). Mutant TPM is represented by open green circles and the wildtype TPM is represented by black circles. Blue shading is indicative of the cold treatment period. All these genes show a fold change increase >2 in the mutant at T1 and nine of these genes show fold change >2 at T9, and are located next to one another in the genome. Annotation information and protein homology matches, along with fold change and FDR-corrected p-values for each locus tag can be found in appendix attachment sheets 2.1C-L.

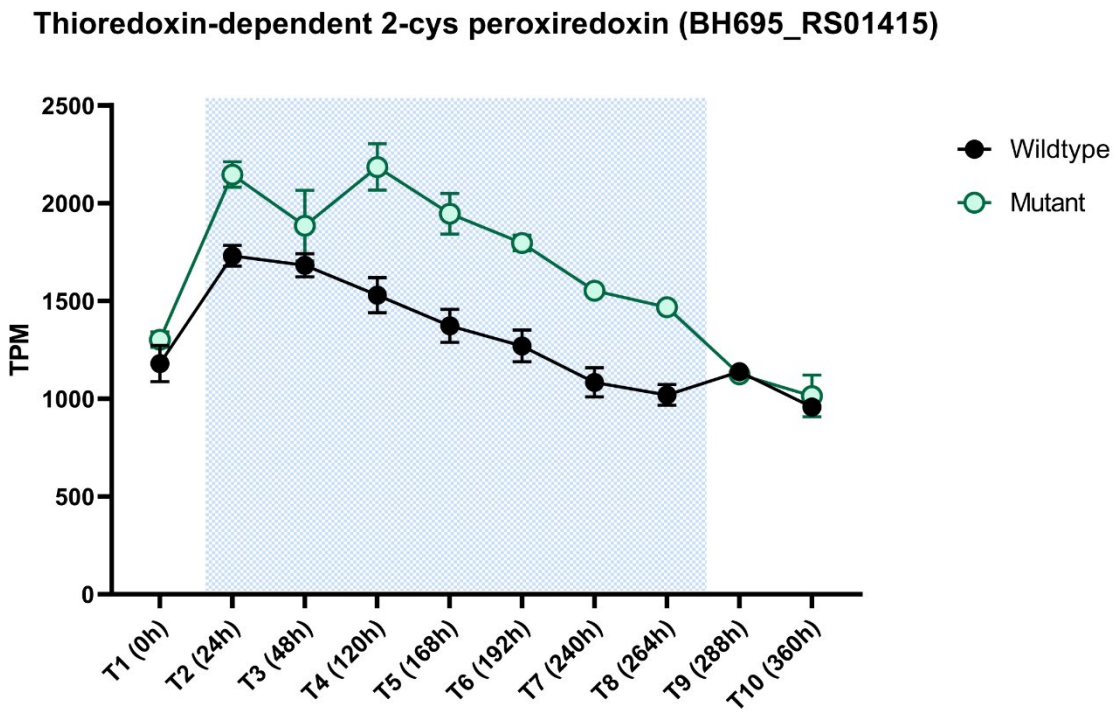

Figure 2.11 Average TPM for a Thioredoxin-Dependent Peroxiredoxin

Average TPM values for mutant (green circles) and wildtype (closed black circles) for a thioredoxin-dependent 2-cys peroxiredoxin. Blue shading is indicative of the cold treatment period. At T2, T4, and T6-T8, the mutant had fold change increases ≥ 1.5 but <2 all with FDR $p \leq 0.0001$, relative to the wildtype.

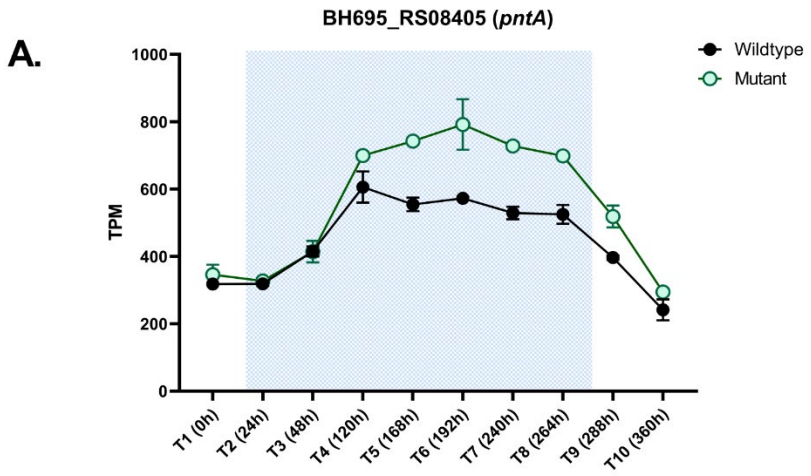

Figure 2.12 Gene expression of *pntAB* for the mutant and wildtype

Average TPM for mutant (green circles) and wildtype (black circles) chemostats at each of the 10 RNA-seq time points T1 (0 h) through T10 (360 h). Blue shading is indicative of the cold treatment period. (A) The mutant showed differential expression of *pntA* (locus tag BH695_RS08405) starting at T6 (fold change = 1.53, FDR p= 1.91E-5), from T7-T8 fold changes were 1.56 and 1.62 respectively, with p <0.0001. (B) Differential expression of *pntA* (locus tag BH695_RS08400) occurred at T5 and T6, with the mutant showing a fold change increase of 1.83 at T5 and 1.82 at T6, both with FDR p <0.0001. (C) The mutant showed differential expression of *pntB* from T5-T8, with fold change increases of 1.83, 1.62, 1.67 and 1.72, respectively, all with FDR-p <0.0001.

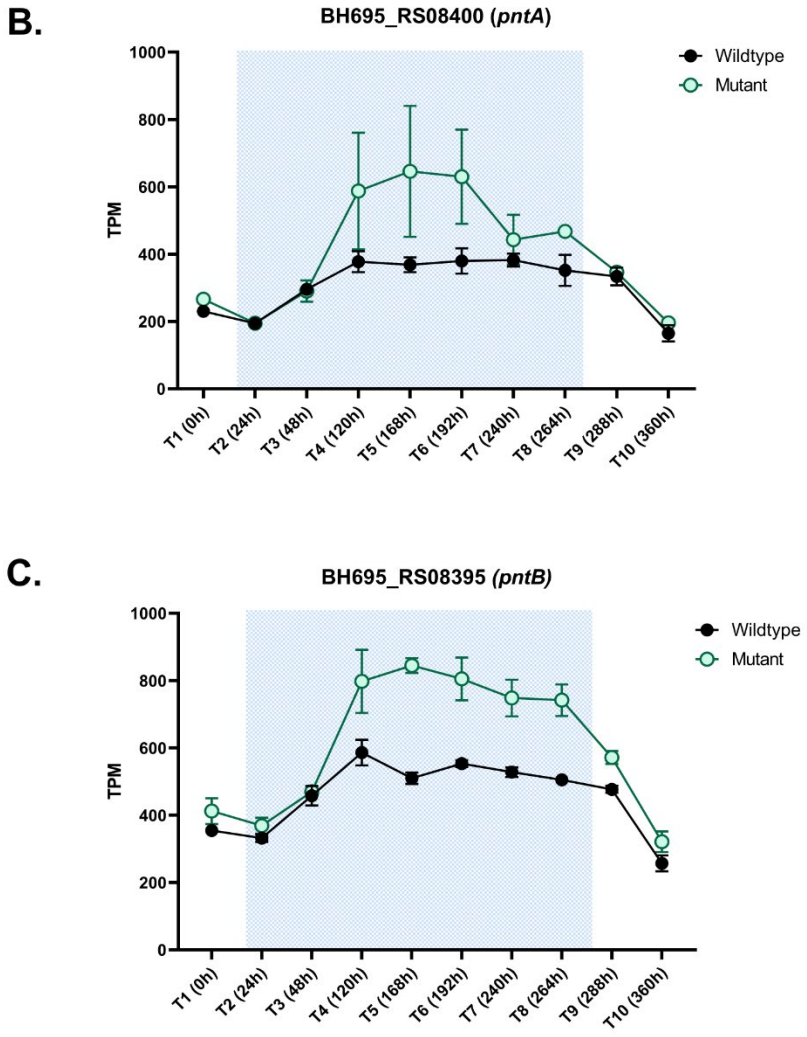

Figure 2.12 Continued.

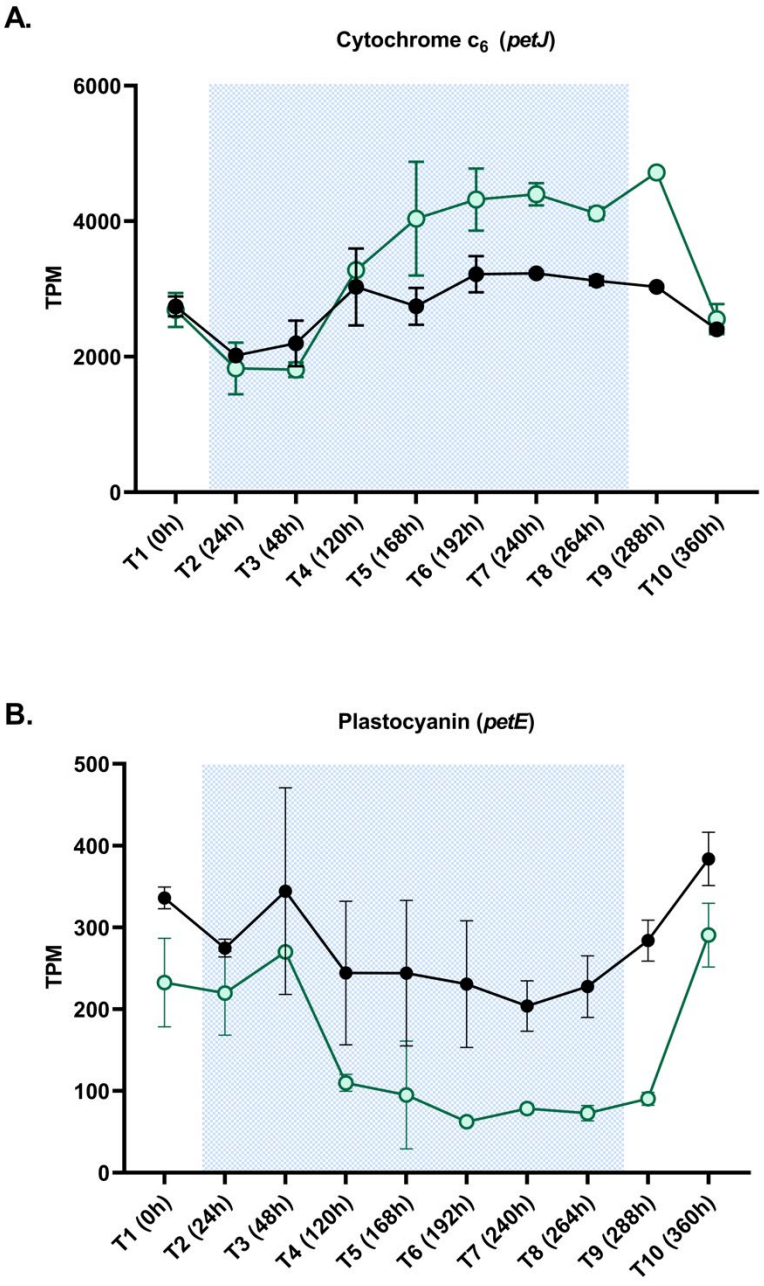

Figure 2.13 Transcriptional profiles of electron transporters *petJ* and *petE*

(A) Mutant petJ transcripts (green open circles) differentiate from the wildtype (black circles) from T6-T9 with fold change increases of 1.5, 1.6, 1.54, and 1.6 respectively, all FDR-corrected $p < 0.0001$. (B) Mutant petE transcripts showed fold change decreases compared to the wildtype from T6-T9 (fold changes 3.31, 2.22, 2.66, 3.06) all with FDR-corrected $p < 0.0001$.

CHAPTER THREE:
PHYSIOLOGICAL REPROGRAMING BY THE *MICROCYSTIS* MOBILOME

Publication Note

This chapter contains content adapted from a manuscript currently in review at BMC Genomics. The authors of the paper include Gwendolyn F. Stark, Alexander R. Truchon, and Steven W. Wilhelm.

GFS and SWW designed the experiments. GFS did all extractions, genomic annotations & analyses, transcriptomic analyses, growth assays and PCR validation of mutations. ART ran the minION long-read sequencer and did the genome assembly. All authors contributed to editing and review of the manuscript.

Abstract

The *Microcystis* mobilome is a well-known but understudied component of harmful algal bloom research. Through genomic and transcriptomic comparisons, we found five families of transposases that altered the expression of genes in the well-studied toxigenic type-strain, *Microcystis aeruginosa* PCC7086, and a non-toxigenic genetic mutant, *Microcystis aeruginosa* PCC 7806 Δ*mcyB*. These isolates were genetically separated when the *ΔmcyB* clone was created in 1997. Differences in gene expression between what were thought to be otherwise genetically identical strains have appeared due to insertion events in both intra- and intergenic regions. In one instance, an insertion event by a miniature inverted repeat transposable element (MITE) into a sulfate-binding gene (*sbp*) may have provided a phenotypic advantage for the Δ*mcyB* mutant under sulfate-limited conditions. This paper highlights how *Microcystis* strains continue to "evolve" in lab conditions, and the importance of insertion sequences / transposable elements in shaping genomic and physiological differences between *Microcystis* strains that we assume are genetically identical. In the current era, our observations force the necessity of knowing the genetic background of isolates, even in physiological experiments, to facilitate the correct conclusions from experiments.

Background

Microcystis spp. are potentially toxic bloom-forming cyanobacteria responsible for harmful algal blooms (HABs) around the globe [1]. These cyanobacteria can affect human health and disrupt ecosystems by producing excessive biomass and potentially toxic secondary metabolites, the best known being microcystin [2, 3]. Due to these negative ecological effects, significant research has examined *Microcystis spp.* physiology and ecology to provide insight that may prevent or mitigate HABs [1]. Much of this research has benefited from the global collection of *Microcystis* isolates in labs around the world, as well as model strains that are broadly shared and employed by researchers.

The tools of molecular biology have increased accessibility to sequenced *Microcystis spp.* genomes and made it apparent that *Microcystis* genomes exhibit some level of variability due to abundant transposable elements and repeat regions [4, 5]. From an ecological perspective, this is an important consideration when investigating *Microcystis* isolates in the laboratory. Mobile elements have been responsible for the loss of the microcystin (*mcy*) gene cluster in non-toxigenic strains [6], and the loss of buoyancy in *Microcystis aeruginosa* PCC 7806 has been shown to be mediated by insertion sequences [6, 7]. Transposable elements may therefore complicate comparisons between *Microcystis* species/ecotypes, as genes may be present but can change functionality if a transposition event occurs near a gene without disrupting the gene's architecture.

While previous studies have reported on the "plastic" genome of *Microcystis*, studies investigating transposition in *Microcystis* are limited [5, 8]. In culture-based studies, it has been shown that differing nitrogen chemistries and phosphorous concentrations affect the expression of transposases in *Microcystis aeruginosa* NIES 843 [9]. *In situ* metatranscriptome samples collected during a *Microcystis* bloom in western Lake Erie (July - October 2014) revealed expression of many strain-specific genes associated with gene plasticity [10]. These studies suggest transposases are transcriptionally active and can play an important role in the adaptive ecophysiology of *Microcystis* both *in vitro* and *in situ*.

A recent genome resequencing of the type-strain for *Microcystis aeruginosa* (PCC 7806, the "wildtype" isolate) and a non-toxic mutant thought to have a single insertion of a chloramphenicol resistance cassette in the *mcyB* gene (*Microcystis aeruginosa* PCC 7806 Δ*mcyB*) led to the discovery of chromosome re-arrangements in these strains relative to a publicly available genome (7806SL) and each other [11, 12]. To examine the effects of these changes, we used transcriptome libraries generated from a recent chemostat experiment [3] to examine gene expression under comparable growth conditions. We found multiple genes and gene clusters were differentially expressed between *ΔmcyB* and the PCC 7086 wildtype under identical conditions and concluded some differences in expression were due to endogenous mobile elements. Additionally, we observed putative phage-acquired genes situated near mobile elements and searched for associations between phage-acquired genes and mobile elements within these and other *Microcystis* genomes. Our findings are important caveats for comparative studies of *Microcystis* evolution and ecology, particularly with respect to extrapolation of observations between different labs and from the lab to the field. Our observations further confirm that multiple families of mobile elements are active in *Microcystis spp.* and can alter genomic architecture and gene regulation when inserted into intergenic and intragenic regions. These observations are particularly salient for efforts that compare wildtype PCC 7806 and Δ*mcyB* for effects of microcystin production, as we demonstrate that there are other relevant differences between these cyanobacterial genomes.

Methods

Strains, DNA Extraction & genome sequencing, assembly, and annotation

M. aeruginosa strains used in this study were obtained from either the Pasteur Culture Collection (*M. aeruginosa* PCC 7806) or directly from Professor E. Dittmann (University of Potsdam, Germany in 2016), who created the original *Microcystis aeruginosa* PCC 7806 Δ*mcyB* mutant in 1997 [14].We will refer to our *Microcystis aeruginosa* PCC 7806 Δ*mcyB* isolate as Δ*mcyB*. The wildtype isolate, *Microcystis aeruginosa* PCC 7806, which we will refer to as PCC 7806 wildtype, has been in our lab for the last decade and has been maintained as both active cultures and cryopreserved

stocks at -150˚ C. For both the *ΔmcyB* (GCA_030553035.1) and wildtype PCC 7806 (accession CP155078) strain growth conditions (25 mL batch cultures in modified CT [11], ~26° C, ~50 μmol photons m^{-2} s^{-1}), high-molecular weight DNA extraction, genome sequencing and assembly methods are as reported [11]. Genome annotation for the sequenced Δ*mcyB* and our wildtype PCC 7806 was done using NCBI's prokaryotic genome annotation software (PGAP) [15]. All other genomes used in this paper were downloaded from publicly available datasets available on NCBI's GenBank [16].

RNA sequencing

Transcriptional analyses of strains grown in chemostats during warm (26° C) and cold (19° C) temperature treatments was previously reported in [13]: these treatments have been shown to alter cellular physiology and biochemistry, including increasing the transcription and biosynthesis of microcystin [13, 17]. The RNA-seq dataset from [13] consisted of 10 time points (two control 26° C, eight 19° C time points). Transcriptomes of two PCC 7806 wildtype and two *ΔmcyB* chemostats were generated, resulting in a total of 40 libraries. In the current study, we sequenced the genome of *Microcystis aeruginosa* PCC 7806 *ΔmcyB* and re-sequenced the wildtype isolate of *M. aeruginosa* PCC 7806. This provided improved references for transcriptomic analysis to account for expression of transposases that were in different positions (or absent) in each genome.

Mapping parameters for all transcriptome libraries were as described in the methods of [13]. Briefly, CLC genomics workbench (v. 23.0.4) was used to map transcriptomic libraries to the *ΔmcyB* (GCA_030553035.1) and the PCC 7806 wildtype (CP155078) genomes. Stringent read mappings were performed using 0.9 length and 0.9 similarity coverage for mapped reads. All reads were normalized by transcripts per million (TPM). We note that the "maximum hits per read" was set to the default of 10, which would make the TPM for some highly-repetitive transposases with >10 copies (IS200 family and ISNYC- family) under-representative of their true expression values. Since increasing the "maximum hits per read" to >10 resulted in non-specific read-mapping, we decided that keeping the mapping parameters stringent and specific would be preferential. An example of this for IS200 family transposases can be seen in the

supplemental materials (Appendix sheet 3.1A). Normalized expression values for transposases that fall in this category were not shown as they were considered inaccurate. Differential expression analysis was done as described in [13], using our PCC 7806 wildtype genome. TPM values reported for the sulfate transporter gene cluster (*sbpA-cysTWA*) were from reads mapped to the wildtype PCC 7806 genome (since *sbp* is annotated as two genes in Δ*mcyB)*, all other figures have TPM values that came from mappings to the respective strains (Δ*mcyB* and PCC 7806 wildtype) genome.

Genome alignments and analyses

Unless noted, default parameters were used for all software. All genomes were aligned using Progressive Mauve (v. 2.4.0) [18]. Alignments of transposase sequences to intergenic regions of interest was done using Clustal Omega [19]. For genes of interest that were annotated as hypothetical or uncharacterized, yet were differentially expressed due to transposon re-arrangement, we ran the translated sequences through PFAM (v. 36.0), HMMER (v. 3.4), and NCBI conserved domain searches to characterize protein domains and predict putative functions [20-23]. BLASTP, BLASTN, and JGI/IMG BLAST (BLASTN – genome isolate and virus databases) were also used to find homologous proteins/sequences among *Microcystis* strains [23, 24]. To find putative open reading frames for the miniature inverted transposable element (MITE), NCBI ORFfinder was used [23]. The secondary structure of the 187-nt MITE was produced *in silico*, using mFOLD (v. 3.0) [25].

Phage gene(s) and MGE Mobile genetic elements (MGE) and phage-gene discovery

Transposable elements that altered gene expression were discovered by looking at annotated transposases that had a fold change in gene expression ≥ 2 from the differential expression analysis comparing the wildtype and Δ*mcyB* transcriptome libraries [13]. Manual inspection of gene clusters that were differentially expressed at all 10 RNA-seq time points [13] was also done to see if intra- or intergenic mutations were present that were not caused by annotated insertion sequences.

To look for genes of putative phage-origin, geNomad [26] was utilized, along

with manual searches (*e.g.,* scanning for genes annotated as "phage"), using publicly available complete *Microcystis spp.* genomes from NCBI GenBank (accessed 01/2024) [23, 26]. Default parameters were used for all geNomad runs. Genes with viral hallmark scores of "1" were considered for further analysis using PFAM, nucleotide BLAST, and conserved domain searches to confirm relevant domains/distribution of genes in bacteria.

PCR validation

To confirm the presence of the MITE in the *sbp* gene that we observed during sequencing, primers were constructed flanking the MITE insertion sequence (forward primer 5'- ACCAAAAAGGTGAGAAGTTAGC-3', reverse primer 5'- CGAAACCCCACCTTAGCA -3') (Appendix Figure 3.8). PCR reaction mix consisted of 12.5 µL Taq Polymerase, 1.25 µL of 10 µM forward and reverse primers, 9 µL of molecular grade water, and 1 µl of whole-cell template. The thermocycler program for PCR amplification was set to 95º C for a 5-min initial denaturing cycle. Then, 35 cycles were run with a 95º C / 30s denaturing, 53º C / 30 s annealing, and 72º C /60 s elongation. A final elongation step was set at 72º C and run for 5 min.

Reverse Transcriptase PCR

Due to the repetitive nature of the MITE in the genome, and the placement of a MITE in the 23S rRNA gene (reads were removed during *in silico* rRNA reduction), we completed a reverse transcriptase PCR to determine if the MITE in the *sbp* gene was transcribed with the gene. Reverse transcriptase PCR was accomplished on RNA samples extracted from chemostat-grown cultures as previously reported [13]. In the present study, we used time points denoted T4 and T10 (Appendix Figure 3.9, 3.25). All RNA was previously DNAsed with a Turbo DNA-free kit (Invitrogen) and confirmed DNA-free *via* PCR [13]. A Thermoscript Plus RT-PCR kit (Invitrogen) was used to convert isolated RNA to cDNA, using manufacturers protocols. Briefly, each 50 µL reaction had 1 µL of Thermoscript plus/platinum TAQ enzyme, 25 µL of 2X Thermoscript reaction mix, 1 µL of 10 µM forward and reverse primers, 20 µL of DEPC treated water, and 2 µL of ~15 ng/µL of template RNA. The thermocycle program for cDNA synthesis was set at

58° C and run for 25 min. Immediately after, a PCR cycle was run, consisting of a 5-minute 95° C incubation period, and 35 cycles of 95° C denaturing for 15 s, 55° C annealing for 30 s, and 72° C elongation for 30 s.

Sulfate Growth Assays

To address specific observations from the genomic sequencing and transcriptomics [11, 13], we examined the effect of sulfate availability on cell growth. At the onset of growth assays, we validated that both the *mcyB* chloramphenicol insertion [14] and an observed insertion of a miniature inverted transposable element (MITE) into a *sbp* gene were present in the Δ*mcyB* strain. For sulfate-limited culture conditions, CT medium [11] was reduced to 0.975 μM $MgSO_4$, whereas sulfate replete-medium contained 195 μM of $MgSO_4$. Cultures were acclimated to ~20° C (± 0.4° C) and ~50 μmol photons m^{-2} s^{-1} for six days prior to experimentation. Experimental growth temperatures were 19.5° C (± 0.3 °C) and ~50 μmol photons m^{-2} s^{-1}. Prior to inoculation for the respective sulfate treatments, all cultures were pelleted and washed twice with sulfate-limited medium. Cells were then inoculated in 25 mL batch cultures of either sulfate-replete or sulfate-limited media. Culture position in incubators underwent daily random re-shuffling to minimize any variance in irradiance. All cultures were grown in triplicate. Cell concentrations were estimated with a Cytoflex Flow Cytometer (Beckman Coulter), with populations gated by red fluorescence (chlorophyll a proxy) and forward scatter (size proxy).

Statistical analysis

Normalized RNA-seq libraries (n = 40) from a previous chemostat experiment [13] were analyzed with GraphPad Prism (v.8.0.2). For genes of interest, t-tests were done on the normalized TPM values between the *ΔmcyB* and PCC7806 wildtype strains for two acclimated control time points at 26° C (T1 and T10) and two 19° C acclimated time points (T7 and T8), see Figure S2 for more information. FDR-corrected p-values ≤ 0.05 were considered "significant". Both temperature treatments were considered, as we hypothesized genes altered by transposases should not be affected by either temperature

treatment.

Results

Whole genome alignments of Microcystis aeruginosa *PCC 7806 and the discovery of a large chromosome inversion*

The complete *M. aeruginosa ΔmcyB* genome was 5,103,923 bp long, whereas our wildtype PCC 7806 genome was 5,096,229 bp. These contrasted with the available *Microcystis aeruginosa* PCC 7806SL genome on NCBI which is 5,139,339 bp. However, a whole genome alignment of our *ΔmcyB* and PCC 7806 wildtype genome with the published *Microcystis aeruginosa* 7806SL (will be referred to as 7806SL) genome from NCBI lead to the discovery of a duplicated region of 44,534 bp in PCC 7806SL (Figure S3). It is possible the duplicated portion of the 7806SL genome may be due to an assembly error, which only used PacBio RS II long reads for genome assembly [12]. To this end, we decided to not further pursue comparisons to 7806SL for this study.

Whole genome alignment of the *ΔmcyB* and wildtype genome uncovered a large chromosome inversion (Figure 3.1, Appendix Figure 3.11). The inversion was ~2.5 Mbp and flanked on both ends by identical nucleotide sequences ~10 kb in length. The ~10 kb identical nucleotide sequences encoded a set of 14 genes that existed in 3 copies within both the *ΔmcyB* and PCC7806 wildtype genome (Figure 3.1A, Appendix sheet 2.1B). Of note, these regions all contained a gene encoding a protein with a cro/C1-type HTH domain, and another gene with tyrosine-type site-specific recombinase/phage integrase domains [20].

We examined the normalized gene expression (TPM) for the 14 genes in the regions flanking the chromosome inversion for control (26° C, n = 2) and cool temperature acclimated (19° C, n = 2) chemostat transcriptomes (Figure 3.1B-1C, Appendix Figure 3.9). Most genes (12 out of 14) had TPM values < 30 for both the warm and cold treatments across all three gene clusters (Figure 3.1C). The two most highly expressed genes in these regions (TPM >30 in some instances) in both *ΔmcyB* and the PCC 7806 wildtype were “gene 7”, which a PFAM search showed has no homology to any known protein domains, and “gene 8”, which is a tyrosine site-specific recombinase/

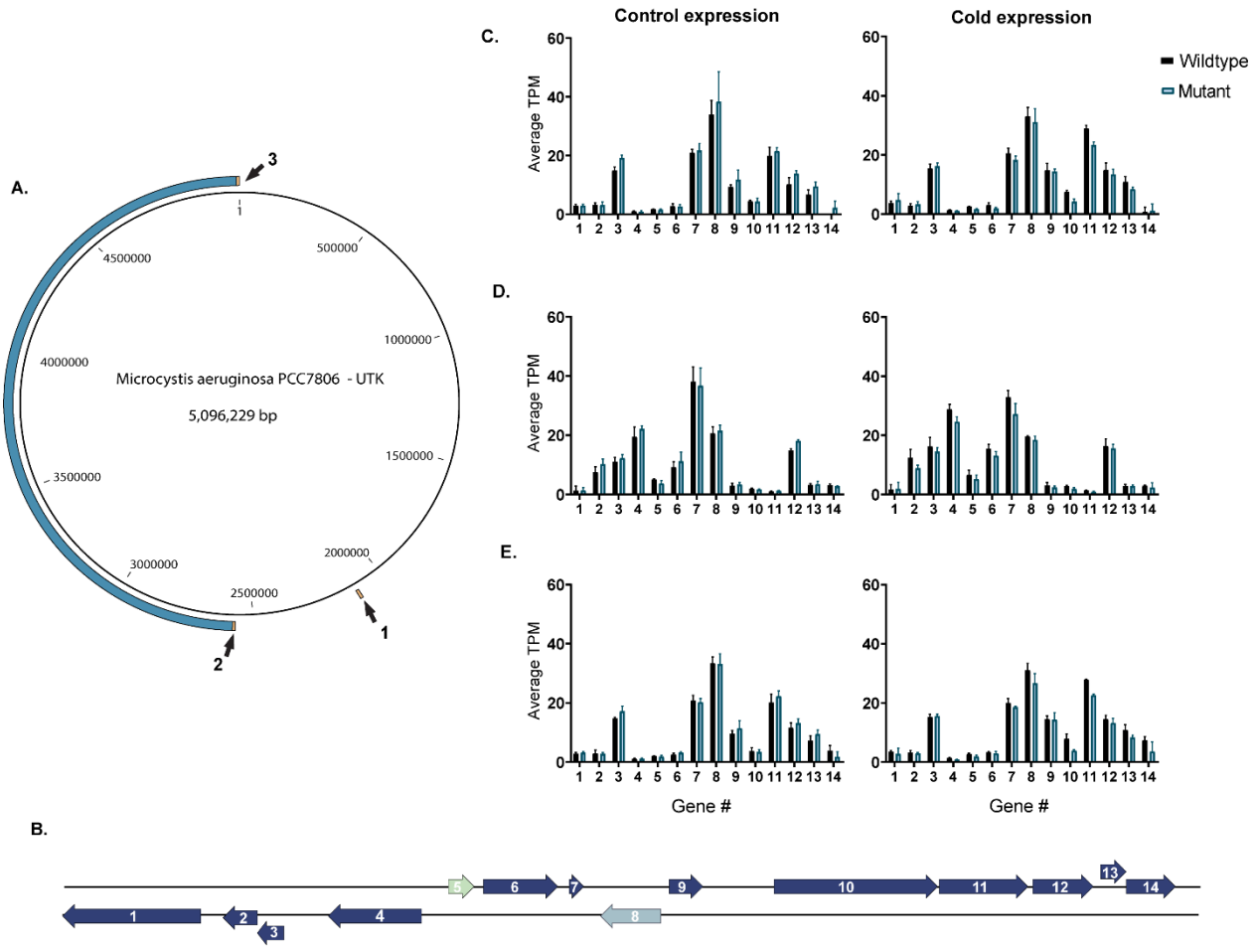

Figure 3.1 Chromosome inversion region, gene arrangement and average TPM of genes

(A) Chromosome map showing the inversion region, in blue of the *Microcystis aeruginosa* PCC7806 genome from our lab relative to alignments against the mutant *Microcystis aeruginosa* PCC7806 Δ*mcyB* mutant genome. Orange regions with arrows indicate the locations of a cluster of 14 identical genes, which occur in 3 locations in the genome. (B) Gene arrangement of the repetitive gene cluster that flanks the chromosome repeat regions in the genomes. Dark blue genes are all uncharacterized proteins. Gene 5 has a Cro/C1 HTH-like protein domain. Gene 8 has phage site-specific integrase/tyrosine recombinase protein domains. (C) Gene expression of chromosome repeat region 1 in control (26°C) (left) and cold (19°C) (right) growth conditions. (D) Gene expression of chromosome repeat region 2 in control (26°C) and cold (19°C) (right) growth conditions. (E) Gene expression of chromosome repeat region 3 in control (26°C) (left) and cold (19°C) (right) growth conditions.

phage-integrase (Figure 1B). In the wildtype, "gene 7" had average TPM values ranging from 20.49 to 38.13 across the warm and cold treatments. In *ΔmcyB*, average TPM values for "gene 7" ranged from 18.27 to 36.74 across treatments. Gene 8 had average TPM values that ranged from 19.60 to 33.91 for the wildtype and 18.41 to 38.36 for the *ΔmcyB*, across treatments.

IS200 family transposases

We found multiple copies of a 444-bp IS200/605 family transposase gene in both the wildtype and *ΔmcyB* genomes. In *ΔmcyB*, we found fourteen, where as our wildtype strain had sixteen of these transposases (Appendix sheets 3.1C,D). Most of the IS200 family genes have the same genome location between our two strains. In *ΔmcyB*, there was one IS200 family gene whose genomic location differed from the wildtype: it interrupted a solute carrier superfamily (SLC) permease gene (Appendix Figure 3.12). Our wildtype strain had three additional IS200 family transposase genes not found (location wise) in the *ΔmcyB:* one interrupted a potassium uptake gene (*trkH*), another interrupted a glycosyltransferase gene, and the third was inserted between a 3-mercaptopyruvate sulfurtransferase and PIN-domain containing protein gene (Appendix Figures 3.13-3.15). In *ΔmcyB*, the glycosyltransferase gene was interrupted by an ISNCY-like ISMae2 family transposase instead of a IS200 family transposase (Appendix Figure 3.14).

IS1-like element IMSae3 family transposases

Identical 737-bp IS1-like ISMae3 family transposases were found throughout the *ΔmcyB* and wildtype genomes. The *ΔmcyB* strain had five of such while our wildtype had six (Appendix sheets 3.1E, F). Notably, an IS1-like family transposase situated before an uncharacterized gene and a gene encoding a metallo-hydrolase with a metallo-beta-lactamase (MBL)-fold in the wildtype strain is missing in *ΔmcyB* (Figure 3.2A). In the wildtype, the IS1-family transposase is transcriptionally active, showing average TPM values of 17.63 in warm conditions, and 35.53 in cold conditions. Downstream of the transcriptionally active IS1-family transposase in the wildtype we saw significantly

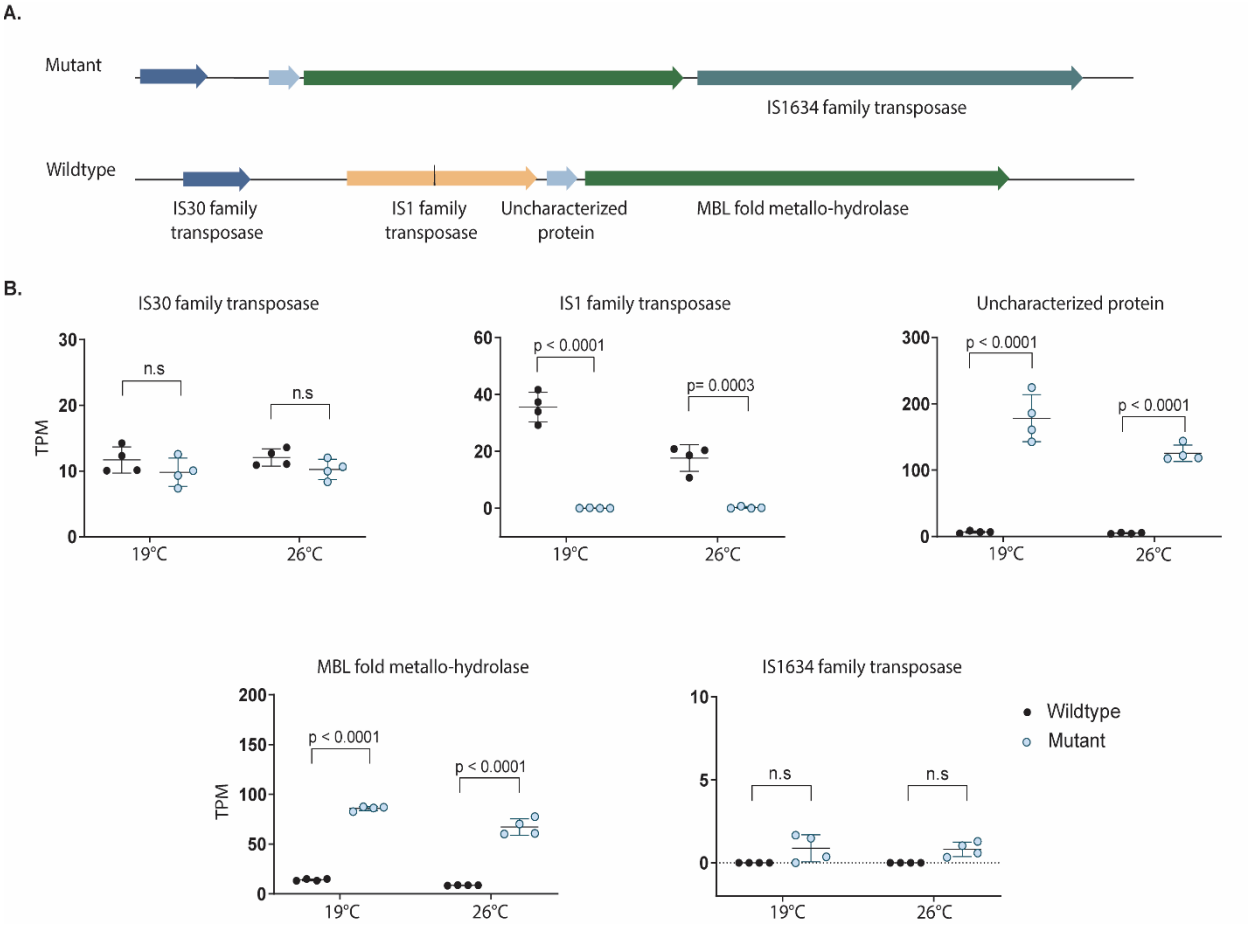

Figure 3.2 IS1-family transposase insertion in the wildtype and expression of effected genes

(A) Gene arrangement in Δ*mcyB* versus the PCC 7806 wildtype(bottom) genome. In the wildtype (bottom arrangement), an IS1-family transposase is inserted before genes encoding an uncharacterized protein and a MBL-fold metallo-hydrolase protein. In Δ*mcyB* (top) the IS1-family transposase has hopped out of the region before the genes encoding an uncharacterized protein and MBL-fold metallohydrolase. The placement of the IS1-family transposase is the same in our PCC 7806 wildtype strain as the 7806SL genome currently available on NCBI, making us believe it has "hopped out" of this region in the mutant. (B) Transcription of the genes in the order shown in Figure 2A. Removal of the IS1-family transposase in the mutant leads to significantly increased expression ($p < 0.001$) of the uncharacterized protein and MBL-fold metallo-hydrolase compared to the expression of these genes in the wildtype, which has the IS1 transposase insertion.

reduced expression ($p < 0.0001$, avg. TPM in cold 6.67, in warm 4.7) of a gene encoding an uncharacterized protein, and an MBL-fold metallo-hydrolase ($p < 0.0001$, average TPM 8.55 in warm and 14.06 in cold acclimated) (Figure 3.2B). An opposite effect was seen in *ΔmcyB*, where the absence of the IS1-like family transposase results in transcription of the uncharacterized protein (average TPM 125.4 in warm and 178.4 in cold acclimated) and the MBL fold metallo-hydrolase (average TPM 67.12 in warm and 85.78 in cold acclimated). In *ΔmcyB*, there was also an IS1634-family transposase inserted after the MBL-fold hydrolase gene which is not present in the wildtype, however, expression of this transposase was minimal (Figure 3.2B).

Another gene interrupted by the IS1-transposase was a gene encoding an uncharacterized protein. The gene encoding the uncharacterized protein was situated in-between genes that encoded for a MoaD/ThiS family protein and an Acyl-CoA dehydrogenase in *ΔmcyB* (Appendix Figure 3.16). A protein BLAST search of the translated sequence for the uncharacterized protein revealed it had 94% query cover and 32.25% amino acid sequence identity to an Acyl-CoA thioesterase in *Chloroflexota* (TMF61956.1) [23]. Insertion of the IS1 transposase in the gene encoding the uncharacterized protein (putative thioesterase) significantly ($p < 0.0001$) interrupted transcription of the uncharacterized protein in *ΔmcyB* compared to the wildtype (Appendix Fig. 3.16). The IS1-transposase was transcriptionally active in *ΔmcyB* , with average TPM values of 70.1 in warm and 118.8 in 19°C acclimated conditions. Next to the uncharacterized protein, and present in both the wildtype and *ΔmcyB* genome, was an ISL3-family transposase. The ISL3-family transposase showed minimal expression in *ΔmcyB* compared to the wildtype (all *ΔmcyB* TPM < 5, all wildtype TPM ≥ 13 and ≤ 50) (Appendix Figure 3.16). A gene encoding an acyl-CoA dehydrogenase downstream of the interrupted gene for the uncharacterized protein in *ΔmcyB* shows significantly reduced expression during cold treatment ($p < 0.001$) but not a significant decrease during warm temperature treatment.

In our wildtype type strain, there was an IS1-like ISMae3 family transposase situated next to a gene that encodes a protein with five tetratricopeptide repeats (TPR) (Figure 3.3A) [20, 21]. The presence of the IS1-like ISMae3 transposase led to very low

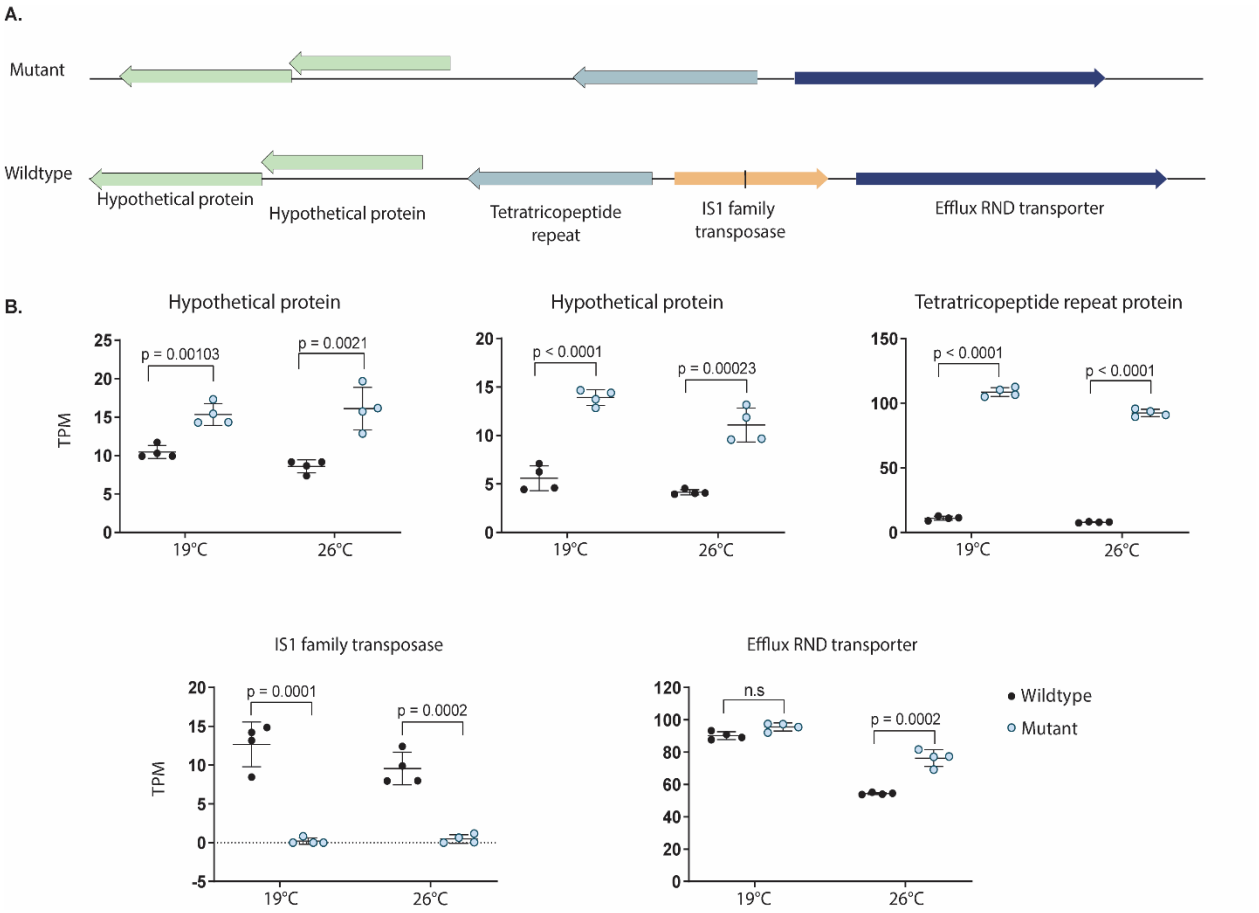

Figure 3.3 Insertion of a IS1-family transposase in the wildtype upstream of an Efflux RND transporter

A) Gene arrangement in Δ*mcyB* (top) versus the PCC 7806 wildtype (bottom) genome. In the PCC7806 wildtype, an IS1-family transposase is inserted in the region between genes encoding a tetratricopeptide repeat protein and an efflux RND transporter. (B) Transcription of the genes in the order shown in Figure 3A. In the PCC 7806 wildtype, insertion of the IS1-family transposase in the region between genes encoding a tetratricopeptide repeat protein and an efflux RND transporter leads to significantly decreased expression ($p < 0.001$) of the gene encoding a tetratricopeptide repeat protein and a hypothetical protein. Expression of the efflux RND transporter is significantly different between the Δ*mcyB* and PCC 7806 wildtype at control growth temperature (26°C) ($p < 0.001$) but not cold (19°C). Differences in expression for the RND transporter are probably not due to the IS1-family transposase, since the two strains have similar expression profiles in the cold.

low gene expression of the TPR-protein in the wildtype (avg TPM 11.01 at 19° C, 7.90 at 26°C) (Figure 3.3B). However, in *ΔmcyB*, the absence of the IS1-like family transposase was accompanied by significantly increased gene expression of the TPR protein ($p < 0.0001$, avg TPM 108.6 at 19° C, 92.46 at 26° C). The two hypothetical proteins located downstream of the TPR-protein also showed significantly higher expression in *ΔmcyB* compared to the wildtype (all $p \leq 0.0021$), which may mean they were also affected by the position of the IS1- like ISMae3 family transposase. The efflux RND transporter, downstream of the IS1-family transposase had similar expression at 19° C in the *ΔmcyB* and wildtype but showed higher expression in *ΔmcyB* ($p = 0.0002$) at 26° C.

IS1634 Family Transposase with DUF4277

We found movement/duplication of a 1746-bp long IS1634 family transposase in the *ΔmcyB* genome. In *ΔmcyB*, ten copies of this specific 1746-bp long IS1634 family transposase exist, whereas only eight exist in our wildtype strain (Appendix sheets 3.1G,H). In *ΔmcyB*, the IS1634-family transposase is inserted in a putative gene cluster encoding genes with protein homology to a diflavin flavoprotein, a putative lipoprotein, a DUF1995-domain containing protein, and a SAM-dependent adenine specific methyltransferase (Figure 3.4A). In *ΔmcyB*, the insertion of the IS1634-family transposase between the diflavin flavoprotein and the lipoprotein resulted in expression of the IS1634-family transposase, which mimics the expression profile of the diflavin flavoprotein (Appendix Figure 3.17). Transcription of the IS1634 family transposase subsequently decreased ($p < 0.001$) expression of the putative lipoprotein, a DUF1995 containing protein, and an adenine-specific SAM- methyltransferase gene in *ΔmcyB* (Figure 3.4B).

Phage-acquired genes affiliated with transposable elements

While investigating differences between the Δ*mcyB* and PCC7806 wildtype genome, we noticed genes annotated as “phage” and located adjacent to transposable elements. This led us to explore the congruity of *Microcystis spp.* genomes retaining

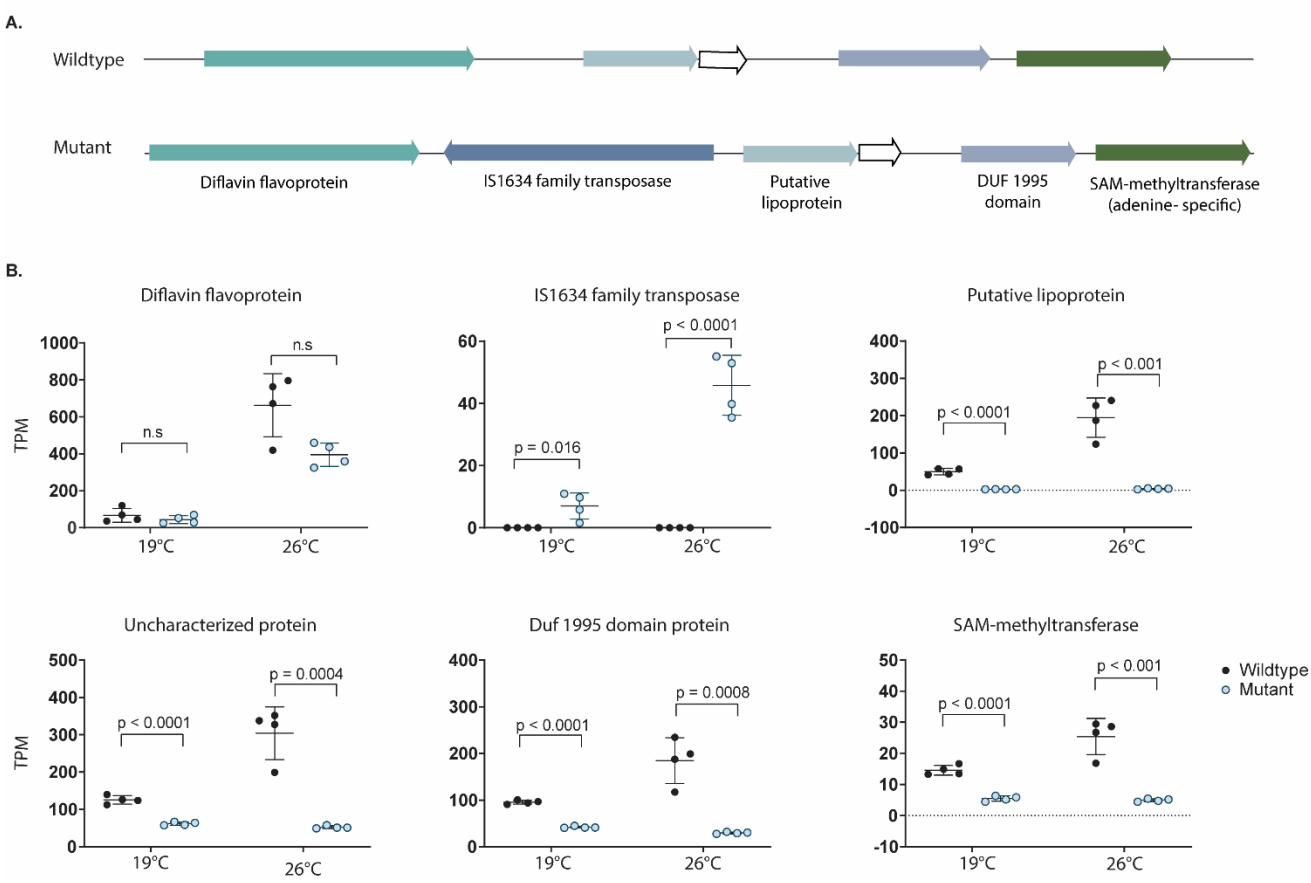

Figure 3.4 Insertion of a IS1634 transposase in Δ*mcyB*

A) Gene arrangement in PCC 7806 wildtype genome (top) versus the Δ*mcyB* genome (bottom). In Δ*mcyB*, an IS1634 family transposase is inserted between genes encoding a diflavin flavoprotein and a putative lipoprotein. In the PCC7806 wildtype, this region is absent of the transposase.(B) Transcription of the genes in the order shown in Figure 4A. In Δ*mcyB*, insertion of the IS1634 family transposase significantly reduced ($p < 0.001$) the expression of four genes, which include a putative lipoprotein, an uncharacterized protein (white arrow), a DUF1995- domain containing protein, and a SAM-methyltransferase.

phage genes and the associations of these acquired phage genes with transposable elements. We investigated this with manual searches (genes annotated as “phage”), and with the aid of GeNomad [26]. From the manual searches we found that most *Microcystis spp.* genomes had retained four common “phage” genes: a phage holin, a phage integrase, a putative phage lysozyme and DUF5131 (phage gp37/gp68) genes (Appendix sheet 3.1I). Many of the phage holin, integrase and DUF5131 gene arrangements appear to be conserved across *Microcystis spp.* genomes in both nucleotide sequence and location (an example for phage holin can be found in Appendix sheet 3.1J. For these genes, further work would need to be done to determine if they are truly “phage” or phage-derived, as they were not labeled as “phage” by geNomad. In some instances, we found putative phage-acquired genes that were not conserved (Appendix sheet 3.1I). Some examples include a phage tail gene (previously annotated as microcystin related protein B, *mrpB*) that is co-transcribed with a "microcystin related protein A” (*mrpA*) in *Microcystis aeruginosa* PCC 7806 (Appendix Fig 3.18), a phage tail lysozyme and associated peptidoglycan binding/recognition proteins in *Microcystis aeruginosa* strains LE3 and NIES298, and a “phage protein D”, which has peptidoglycan-binding properties, found in *Microcystis aeruginosa* NIES843 (Appendix Figures 3.18-3.20).

A gene annotated as a phage-tail lysozyme was found in *Microcystis aeruginosa* LE3 (LE3) and *Microcystis aeruginosa* NIES 298 (NIES 298), next to three other genes which are 99% identical between the two strains (Appendix Figure 3.20, Appendix sheet 3.1K). A BLASTN search showed these four genes were only found in the complete genomes of LE3 and NIES 298. However, a JGI IMG BLAST (genome & virus) search showed that these four genes and the flanking intergenic regions are present in 11 other scaffold-level *Microcystis* genomes at matching nucleotide identities >98%, and in five putative (low completeness) viral scaffolds with nucleotide identities >95% (Appendix sheets 3.1L,M). Flanking these four genes are intergenic sequences that have inverted terminal left and right repeats which align with IS1-family transposase inverted terminal repeat (ITR) regions (Appendix Figs 3.20-3.22). Interestingly, the IS1-family transposases that had similar ITRs as these regions were either 756 or 758 bp in length.

The 756 bp long IS1 transposes did not have DTR's, while the 758 bp ones did (Appendix Figure 3.22).

A miniature inverted transposable element (MITE) occurs multiple times in both genomes

We discovered a 187-bp nucleotide sequence that existed in various regions of the *ΔmcyB* and wildtype genome (Appendix sheets 3.1N,O). The insertion was flanked by direct terminal repeats that were eight nucleotides in length and duplicated from the target sequence, a characteristic of transposable elements [27, 28]. The sequence also has imperfect ITRs nine nucleotides in length (Figure 3.5). As the insertion sequence did not encode a transposase, we concluded it was a small, non-autonomous transposable element, which are known as miniature inverted repeat transposable elements (MITEs) [29, 30]. While the 187-nt MITEs were located mainly in intergenic regions, often between toxin anti-toxin (TA) systems and transposable elements, in some instances they inserted in protein coding genes. Notably, the MITE is present in one of the two 23S rRNA genes in *ΔmcyB*, our wildtype, and the 7806SL genomes. In the *ΔmcyB* genome, there were thirty-three identical 187-bp long MITEs and thirty-one in our wildtype (Appendix sheets 3.1N,O). The *ΔmcyB* strain had two additional mutations due to this sequence, one located in a gene encoding a sulfate-binding protein (*sbp*) associated with the sulfate transporter gene cluster *cysTWA* (Figure 3.6A), and another in a gene encoding the protein LivG (Appendix Figure 3.23).

To assess whether the 187-nt MITE had coding potential, it was run through NCBI ORFfinder: two ORF's less than 30 amino acids were found [23]. ORF 1 encodes a putative peptide twenty-one amino acids in length (aa sequence MSFSIRVPSLITPVYCIRQTEE), which starts at residue 120 and ends at residue 185 in Frame 3. ORF 2 encoded a peptide eleven amino acids in length (aa sequence MPPYRTITSNEE), starts at residue 38 and ends at residue 3 in Frame 3 (Figure 3.5A). *In silico* modeling (mFOLD) of the RNA structure of the 187-nt MITE revealed it has a stem-loop structure, which is typical of MITEs (Figure 3.5C) [31, 32].

MITEs are non-autonomous insertion sequences, relying on a transposase to mediate their transposition [29]. To find the family of transposase responsible for MITE

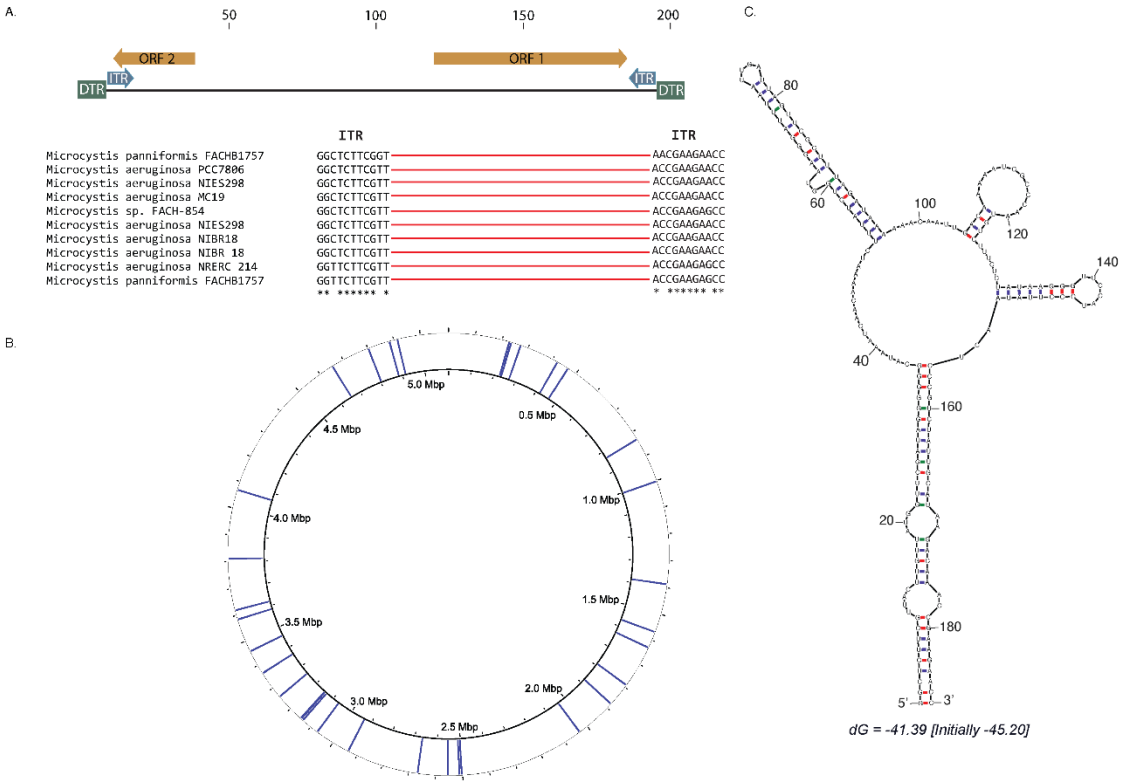

Figure 3.5 MITE alignments, insertion sites and structure

A) Simplified overview of the 187-nt insertion sequence in the *sbp* gene in Δ*mcyB*. The insertion sequence is flanked by 8-nt long direct terminal repeats, duplicated from the target site. There are also imperfect 11-nt long inverted terminal repeats (ITR), and two putative open reading frames. Homologous insertion sequences from other *Microcystis spp*. show similar ITR's. (B) Locations of identical 187-nt MITEs in the Δ*mcyB* genome. (C) *In-silico* Mfold [25] prediction of the RNA secondary structure of the 187-nt MITE sequence (Predicted ΔG= -41.39).

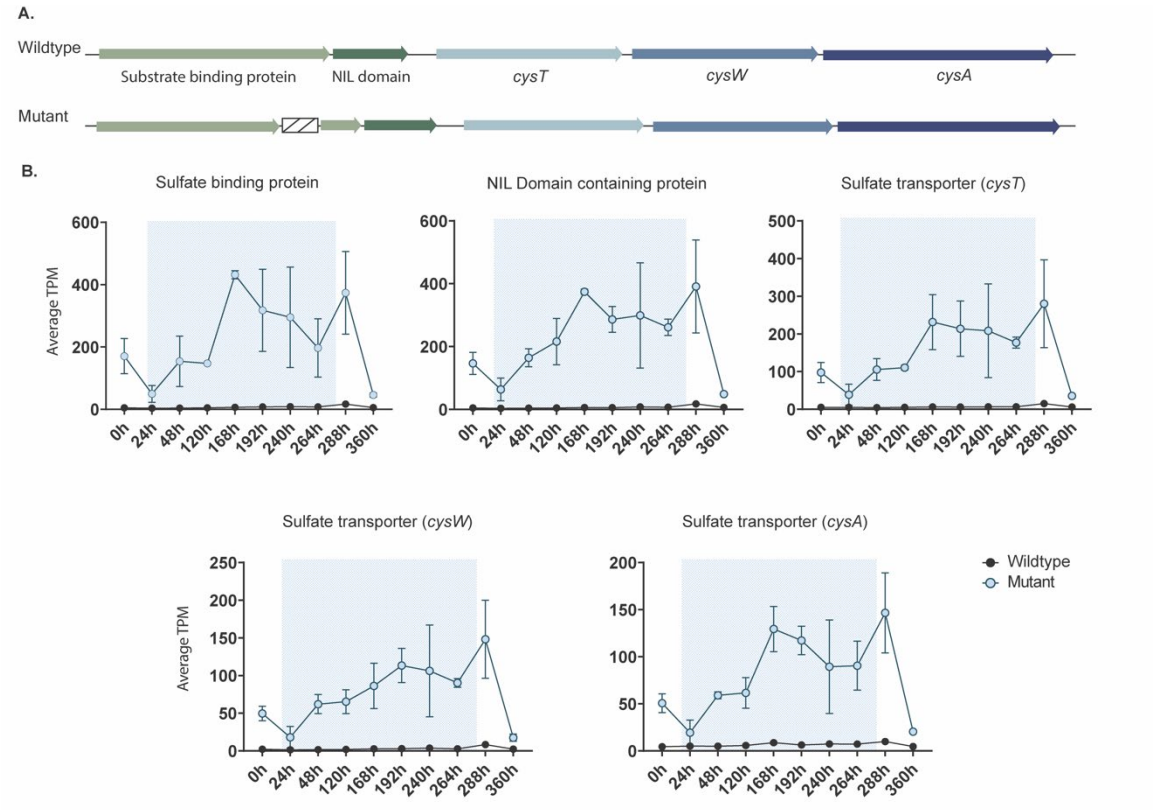

Figure 3.6 Gene expression of the sulfate transporter gene cluster in the mutant and wildtype

A) Gene arrangement of the high-affinity sulfate transport gene cluster in *Microcystis aeruginosa* PCC7806 wildtype (top) and Δ*mcyB* (bottom). Genes are labeled, homologous genes in the mutant are the same colors as those seen in the wildtype arrangement. In the mutant, the sulfate/substrate binding protein gene has a 187-nt long insertion, which is denoted by the striped box interrupting the gene. (B) Average normalized gene expression (normalized by TPM) of the sulfate transporter genes, in order of the gene arrangement seen in figure 6A. The open blue circles represent average TPM of Δ*mcyB*, and black circles are the average TPM of the PCC 7806 wildtype. The mutant has a marked differential expression of all five genes, which show fold changes >2. The wildtype shows very little transcriptional activity of any of these genes (all TPM <20).

transposition, we searched for sequences in the genome that matched the conserved 9-nt long inverted terminal repeat (ITR) of the MITE. Doing so led to the observation that some 1,215-bp long ISL3-family transposases are flanked at both ends by imperfect inverted terminal repeats starting with the sequence 5'-"GGCTCTTCG"-3', which matches the MITE's ITRs (Appendix Figure 3.24). Furthermore, the ISL3-family transposases with homologous ITRs are flanked by 8-nt long direct terminal repeats, the same length as the DTRs for the MITE. These characteristics make it likely IS3-family transposases mediate the transposition of the MITE [29].

A nucleotide BLAST search of the MITE found homologs with > 90% nucleotide identity in twelve other *Microcystis* genomes (Appendix sheet 3.1P). We also found three *Microcystis* plasmids with homologs of the MITE with ≥ 85% nucleotide identity (Appendix sheet 3.1Q). Additionally, a JGI-IMG Blast search found three putative viral scaffolds with homologous hits to the MITE, with identities of ~86 to 90 % (Appendix sheet 3.1R) [33]. A nucleotide alignment of some of the homologous sequences found in *Microcystis* genomes, all of which had confirmed direct terminal repeats flanking the sequences in the genomes, suggests the 9-nt long imperfect inverted terminal repeat is a conserved feature of the homologous MITE's (Figure 3.5).

*Insertion of the 187-nt long MITE in a sulfate binding protein in the Δ*mcyB *genome*

In *ΔmcyB*, the insertion of a 187-bp MITE in the sulfate-binding protein gene resulted in > 2-fold change in expression of the high-affinity sulfate transporter *sbp-cysTWA* operon, with expression highest during cool temperature conditions (Figure 3.6B). In the wildtype, expression of these sulfate transporters was minimal (TPM < 20). This trend was not observed for the disrupted *livG* gene in *ΔmcyB*, where average TPM values for the gene regions flanking the MITE were like those of the wildtype (Appendix Figure 3.23). For the *sbp* nucleotide sequence, the conserved residues of the substrate binding site occur from residues 136 to 678, upstream of the MITE insertion sequence, which occurs from residues 859-1053 [23] (Figure 3.7A). To understand whether the MITE insertion interrupted transcription of the *sbp* gene, we analyzed the normalized (TPM) reads for the altered *sbp* gene (upstream and downstream of the insertion).

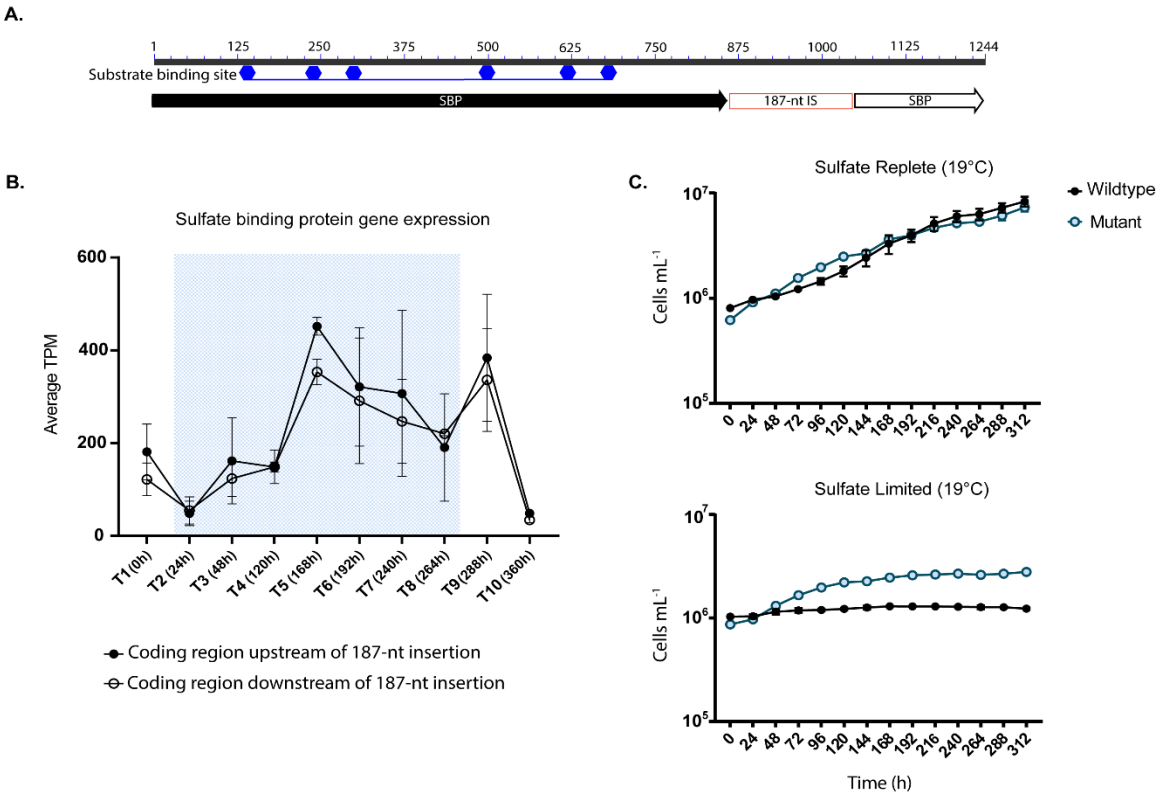

Figure 3.7 Conserved residues of the sulfate binding gene, reads mapping to *sbp* in the mutant, sulfate limitation growth assays

A) *In-silico* determination of the substrate binding site conserved residues of the sulfate binding protein gene in the Δ*mcyB* strain, based on and adapted from the output of NCBI's conserved domain search[23]. The 187-nt insertion is located downstream from the substrate binding domain. (B) Gene expression profile of the sulfate binding protein coding regions upstream and downstream of the MITE insertion sequence in the Δ*mcyB* strain. The 187-nt long MITE does not appear to affect the expression of the gene downstream of the insertion sequence in the Δ*mcyB* strain. (C) Growth assays of the PCC 7806 wildtype and Δ*mcyB* strain in sulfate replete, or sulfate limited media, grown at 19°C. The mutant Δ*mcyB* strain appears to have a growth advantage in sulfate limited media when grown at cold temperature compared to the PCC 7806 wildtype. The PCC7806 wildtype barely grew above the initial inoculum of (avg. starting ~1.03 x10^6, avg. ending ~1.23 x 10^6), whereas the Δ*mcyB* starting inoculum was (avg. ~8.65 x 10^5), and it reached maximum cell concentrations of (avg. ~2.8 x 10^6).

Reads consistently mapped to the region of the gene before the MITE insertion and the region of the gene after the MITE among all *ΔmcyB* transcriptome libraries (Figure 3. 7B).

Due to the repetitive nature of the MITE in the genome, we could not determine normalized gene expression of the insertion at specific sites, including in the *sbp* gene. To see if the MITE in the *sbp* gene was transcribed, we used reverse transcriptase PCR (RT-PCR) to amplify the region around the MITE. Primers were designed to amplify a 189-bp amplicon in the wildtype genome, and a 376-bp amplicon (larger due to the 187-bp MITE insertion sequence) in the *ΔmcyB* genome (Appendix Figure 3.25). For the wildtype, we observed similar sized transcripts (189-bp in length) in the RT-PCR (cDNA) reaction and the DNA-PCR reaction. For *ΔmcyB*, RT-PCR amplified faint bands 376-bp in length, signifying that the *sbp* gene was transcribed through the MITE in *ΔmcyB* (Appendix figure 3.25).

We assessed whether the increased gene expression of the sulfate-binding protein (*sbp*) and *cysTWA* gene cluster in the *ΔmcyB* allowed for improved growth in a sulfate-limited environment than the wildtype (Figure 3.7C). *ΔmcyB* and the wildtype showed similar growth when grown in sulfate-replete conditions at 19.5˚ C (Figure 3.7C). In sulfate-limited conditions, the *ΔmcyB* strain reached higher cell density than the wildtype (Figure 3.7C).

Discussion

In this study we observed mobile genetic features of *Microcystis aeruginosa* PCC 7806 *via* comparison to a Δ*mcyB* mutant strain generated ~ 26 years ago from the original cell line. Some changes in gene expression were caused by mobile elements, which include IS200/IS605, IS1-like, ISNCY and IS1634 family transposases. Genetic changes were also caused by a 187-nt long MITE, which is transposed by an ISL3-type transposase. Three of these families of transposases (IS1-like, IS1634, and ISL3) are DDE-type dsDNA transposons, whose catalytic triad (Asp, Asp, and Glu) use Mg^{+} or Mn^{+} as cofactors to facilitate transposition [34, 35]. IS200 transposases are part of the HuH- enzyme superfamily, and work on ssDNA through a “peel and paste” mechanism

[36]. In some instances, as is the case for IS200 transposases, transposons can produce regulatory sRNAs, having widespread effects on gene expression [37]. Some transposon-produced sRNAs can also function as dual regulators, effecting host and transposon gene expression [38]. The RNA-binding protein, Hfq, is also able to post-transcriptionally inhibit different families of transposases [39, 40]. Aside from the various regulatory mechanisms inhibiting transposition, it is also thought that transposition activity is maintained at a low-level in cells due to deleterious mutations [41-43]. However, in this paper, we confirm through gene expression data and genomic comparisons, that these five families of transposable elements are active in *Microcystis* and altered gene expression by insertion in both inter and intra-genic regions. Movement of transposable elements and the resulting phenotypic changes from these events may play a role in producing contradictory results found in the literature with respect to different environmental variables, *Microcystis* physiology, and microcystin production [44]. While natural strain variability already complicates *Microcystis* work, the effects of transposable elements on *Microcystis* genomes creates further confusion as to how to characterize *Microcystis* "strains" to make *Microcystis* research consistent and reproducible from lab-to-lab.

Chromosome inversion

A major difference between the *ΔmcyB* and wildtype genome was a chromosome inversion, roughly 2.5 Mbp in length. In *Staphylococcus aureus*, chromosome inversions are reversible, and have been shown to result in phenotype switching and prophage activation [45, 46]. In some *E. coli* strains, chromosome inversions have been shown to be flanked by prophages [47]. Interestingly, of the 14 genes flanking both ends of the inversion region, only two had identifiable protein domains, but they were similar to domains found in bacteriophage proteins. These included a gene that encodes a protein with a Cro/C1-type HTH domain and a tyrosine-type site-specific recombinase/phage integrase (which is one of the most highly expressed genes in the cluster). Cro/C1-type HTH domains are involved in site-specific DNA binding and are well characterized as lytic/lysogenic switches in bacteriophage, but have been shown to have roles in bacteria

as well [48, 49]. Tyrosine site-specific recombinases also have numerous roles, where they facilitate the rearrangement (which includes integration/excision and inversion) of DNA molecules [50]. While large chromosome inversions have been reported and shown to impact genes in cyanobacteria before, they have not yet been described in *Microcystis* [51, 52].

Gene activation and silencing by transposases

We found five families of transposases whose insertion into intragenic or intergenic regions resulted in the differential expression of nearby genes in Δ*mcyB* and the wildtype. These include the IS1-like Ismae3 family transposases, and IS1634 family transposases. In one instance in *ΔmcyB*, the loss of an IS1 family transposase appears to have "turned on" transcription of an MBL-fold metallo-hydrolase gene. We think this IS1 transposase may have "jumped out" of this location in the mutant, because it is present in this region in our wildtype, in PCC 7806SL (sequenced 2017) [12], and on a contig from PCC 7806 sequencing data from 2007 [5].

However, we were unable to find any transposase "footprints" in this region in Δ*mcyB* to definitively confirm this. This same IS1 transposase was also located by a gene encoding a putative TPR-protein in the wildtype, were we saw differential gene expression of the TPR-protein and three hypothetical/uncharacterized genes. The gene encoding the MBL-fold metallo-hydrolase has homology to RNA-processing exonucleases, which may indicate it has RNAse activity [53]. Tetratricopeptide repeat proteins mediate protein-protein interactions and can be involved in various functions including transcription and formation of multi-protein complexes [54, 55]. While it has previously been shown that IS1-like transposases insert in the gas vesicle gene *gvpV*, inactivating it in *Microcystis aeruginosa* PCC 7806 [7, 56], here we show that IS1-like elements can also inactivate genes by inserting in intergenic regions. These IS1-family transposases may also have the capability of "jumping out" of these regions, reactivating genes, making these differences in gene expression and any resulting phenotypes reversible in *Microcystis*.

IS1634 transposases also appeared transcriptionally active, and lead to gene

silencing/decreased expression of a putative gene cluster of unknown function in *ΔmcyB*. Insertion of a IS1634 transposase in the intergenic region between a diflavin flavoprotein and a putative lipoprotein led to silencing of the putative lipoprotein, and decreased expression of genes downstream of the lipoprotein, including a SAM-methyltransferase. Intragenic transposition events like this could complicate genomic comparisons, as these genes are present in both strains, yet are silenced/have significantly reduced expression due to the intergenic insertion of IS1634 and IS1 family transposases.

MITES are abundant across Microcystis *genomes*

While *Microcystis* genomes have been suggested to be "plastic" due to the abundant transposases, there has not been much focus on the role of MITE's in *Microcystis* [4, 5, 8]. MITE's are short (usually < 500 bp in length) non-coding sequences with high copy numbers that are mobilized in the genome *via* transposases [57]. We found that the 187-long MITE in the *sbp* gene had imperfect inverted terminal repeats that matched those of ISL3-family transposases. It also had a stem-loop secondary RNA structure, another characteristic of MITES [31]. One exception we found was that that *in silico* analysis of the 187-bp MITE suggested it has two putative open reading frames overlaying the inverted terminal repeats, which is uncommon for MITEs [58]. MITE homologs were also present in other *Microcystis* strains, plasmids, and putative *Microcystis* virus sequences. MITE's have previously been described in *Microcystis aeruginosa* NIES-843 and FACHB-843 and have been shown to be present in *Microcystis* CRISPR sequences [4, 59, 60]. While it's apparent MITEs are widespread in *Microcystis*, MITEs' effect on gene expression in *Microcystis* has not been explored. This is important, as MITEs have been shown to influence gene expression in eukaryotes and prokaryotes through production of miRNA's, regulatory motifs, and promotor propagation [30, 58, 61]. Additionally, transposition of the MITE into one of the 23S rRNA genes could have widespread phenotypic effects. The 23S rRNA gene product functions as a ribozyme that catalyzes the peptidyl transferase step of protein synthesis [62]. Mutations in the 23S rRNA gene have been shown to confer antibiotic resistance in bacteria [62]. What roles the MITEs play in the functionality of the mutated 23S rRNA

gene or its product needs further lab validation to see if it offers phenotypic advantages/disadvantages.

Can MITEs alter gene expression in Microcystis aeruginosa *PCC 7806?*

One striking difference between *ΔmcyB* and the wildtype strain transcriptomes was the differential expression of the sulfate transporter gene cluster, *sbp-cysTWA.* All genes in this cluster were identical between these two strains (at a sequence level) except for the presence of a MITE inserted in the *sbp* gene in the *ΔmcyB* isolate downstream of putative ligand-binding residues. Studies in other cyanobacteria have shown the activity of *sbpA-cysTWA* is regulated by sulfate availability [63, 64]. In *Synechococcus sp.* strain PCC 7942, it has been shown that *sbpA*-deficient mutants are able to grow on sulfate, and only show differences in phenotype from wildtype cells when grown on sulfate-limited medium [63]. Instead, it appeared that *cysT*, *cysW*, and *cysA* genes were essential for sulfate uptake, as *cysT*, *cysW* and *cysA* deficient mutants could not uptake sulfate [63]. Based on our growth assays in sulfate-replete and sulfate-limited media, it seems *ΔmcyB* may have a growth advantage over the wildtype in sulfate-limited conditions, as the wildtype was unable to grow beyond the initial inoculum. We concluded that the sulfate transporters were most likely functional in some capacity in *ΔmcyB*, despite the presence of the MITE. It remains unknown why the wildtype has such low transcription of *sbp-cysTWA* in standard and cold-temperature growth conditions and is unable to grow to the same density as Δ*mcyB* in sulfate-limited medium. Data from the sulfate growth assays and transcriptomes imply sulfate transport is either regulated differently in *Microcystis aeruginosa* PCC 7806, or the MITE is making this gene cluster functional in *ΔmcyB*. Unfortunately, to confirm the MITE's role in regulation, it would need to be removed from *sbp* in *ΔmcyB*, a task which is difficult to do in a non-model organism like *Microcystis*, which has abundant endonucleases that interfere with genetics [14, 65]. Regardless, our genomic comparisons suggest there is active transposition of MITE's in *Microcystis*, and future work should be done to understand how they may contribute to gene expression/regulation, especially because they may be easily overlooked if annotated genome features are only considered. Additionally, if regulation of this gene

cluster is the same in *Microcystis* as other model cyanobacteria, and expression is not influenced by the presence of the MITE, it would imply *ΔmcyB* requires more sulfur then the wildtype [63, 66].

Are there other roles for transposable elements in Microcystis *genomes?*

While we have illustrated how transposases have contributed to the lab-evolution of the commonly used *ΔmcyB* and wildtype PCC 7806 strains in batch culture, we also wondered if transposable elements may serve other roles. A specific question that arose was the role of these elements in the horizontal transfer of genetic material from phage to the host cell. While we have not addressed this conclusively, we did observe some putative phage-related genes in PCC 7806 situated near transposases. In the *Microcystis aeruginosa* PCC 7806 genome, there is a gene encoding a "microcystin related protein A" (*mrpA*) associated with a phage tail gene, which was previously annotated as *mrpB* (microcystin related protein B). These two "microcystin related proteins", one of phage origin, are transcribed together in both the wildtype and Δ*mcyB*, but have higher expression in *ΔmcyB* at one time point in our RNA-seq dataset (Appendix Fig. 3.18). Co-transcription of *mrpAB* has been reported previously, although transcripts were found to be much lower in microcystin-free mutants under varying light regimes versus the wildtype PCC 7806 [67]. The two genes, *mrpA* and *mrpB* (phage tail), are situated in-between IS4 and ISNCY-family transposases and a ribonuclease type II BrnT/BrnA toxin anti-toxin (TA) system [68]. While we could not determine the normalized gene expression of this ISNCY-family transposase due to repetitive copies of this gene in the genome, we found through a nucleotide BLAST search that it has previously been reported as "ORF3" in a study dating back to 2001 [67]. This indicates these three genes have likely been present in this gene array for >20 years.

To determine how widespread phage-acquired genes may be in other *Microcystis* spp., and any possible associations they had with transposable elements, we examined additional examples of putative phage-genes in other *Microcystis* genomes. One interesting finding in *Microcystis aeruginosa* LE3 (LE3) and *Microcystis aeruginosa* NIES 298 (NIES 298) was a putative composite transposon consisting of a phage-tail

lysozyme gene next to three other genes, one of which has peptidoglycan binding/recognition domains. The regions flanking these four genes aligned with the left and right inverted terminal repeats of an IS1-family transposase, which is why we refer to it as a putative composite transposon. Composite transposons have transposable elements, or sometimes remnants of transposable elements, flanking genes that do not function in transposition [69]. Phage tail lysozymes are involved in host entry, where they help degrade the peptidoglycan layer of the cell wall [70]. The role of these genes in these strains would be an interesting route to pursue, as they could have implications for HAB microbiome research. Other putative phage acquired genes seen in the *Microcystis spp.* genomes can self-propagate (group I intron endonucleases and group II intron reverse transcriptases). Which may be the case in the *Microcystis aeruginosa* NIES-2549 genome, which has two identical copies of a phage-derived group II intron endonuclease.

Although we found several examples of putative phage genes situated near transposable elements, it cannot be definitively determined that transposition events facilitated gene acquisition. As shown in several examples in this paper, transposable elements may be activating or inactivating these putative phage-acquired genes and could be there by chance. While we only looked at single genes, in the case of prophage regions in *Microcystis* genomes, transposition of insertion sequences in or near prophage regions may have potential consequences for lytic/lysogenic phage lifestyles. In other instances, phage-acquired genes may themselves be autonomous elements, contributing to genetic/physiological changes. While we can only provide observations, as there is still much to be learned about the mechanisms of transposition and acquired resistance to phage infection in *Microcystis*, we feel it is important to draw attention to putative phage-acquired genes and any possible associations they could have with MGE's.

Conclusions

We compared a *ΔmcyB* and wildtype PCC7806 genome, which were genetically separated for over 25 years and maintained in lab cultures. Our work has revealed the activation/silencing/differential expression of certain gene clusters mediated by IS1, ISL3, IS1634, ISNCY and IS200 type transposable elements. Transposition events are

thought to be rare as transposition has been shown to be tightly regulated [38, 41], however as we and others have shown, this may not be the case in *Microcystis* [7]. As transposases can be activated by specific stressors, future work should focus on the activators of various families of transposable elements in *Microcystis*, and how each might result in genomic re-arrangements [71]. This includes phage-infection, as it is evident that phage-related genes may be acquired by hosts in some capacity. Insertion of transposases both inter- and intragenically can complicate genomic and physiological comparisons of *Microcystis* strains and interfere with the reproducibility of experiments. Active transposition in *Microcystis* may explain contradictory studies which investigate abiotic effects on *Microcystis* physiology and microcystin production. Because of this, we suggest labs should routinely re-sequence strains to account for phenotypic variability, especially in the case of "omics"-based studies. As these mutations have been compounded over the last 25 years of strain maintenance in culture, the activation/silencing of some of these gene clusters may also provide additional insight into microcystin's intracellular role.

Data Availability

The *Microcystis aeruginosa* PCC 7806 wildtype genome is available on NCBI Genbank, under Bioproject PRJNA1104278, accession CP155078. The genome of the *ΔmcyB* isolate was previously published on NCBI Genbank under GCA_030553035.1.

Funding

This work was supported by funding from the Simons Foundation (735077). This work was also support by the National Science Foundation (OCE-1840715) and the National Institute of Environmental Health Sciences (1P01ES028939–01) through funds provided to the Great Lakes Center for Fresh Waters and Human Health.

References

1. Harke, M.J., et al., *A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp.* Harmful Algae, 2016. **54**: p. 4-20.

2. Dick, G.J., et al., *The genetic and ecophysiological diversity of Microcystis.* Environmental Microbiology, 2021. **23**(12): p. 7278-7313.

3. Steffen, M.M., et al., *Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH.* Environmental Science & Technology, 2017. **51**(12): p. 6745-6755.

4. Kaneko, T., et al., *Complete Genomic Structure of the Bloom-forming Toxic Cyanobacterium Microcystis aeruginosa NIES-843.* DNA Research, 2007. **14**(6): p. 247-256.

5. Frangeul, L., et al., *Highly plastic genome of Microcystis aeruginosa PCC 7806, a ubiquitous toxic freshwater cyanobacterium.* BMC Genomics, 2008. **9**(1): p. 274.

6. Christiansen, G., et al., *Nontoxic Strains of Cyanobacteria Are the Result of Major Gene Deletion Events Induced by a Transposable Element.* Molecular Biology and Evolution, 2008. **25**(8): p. 1695-1704.

7. Mlouka, A., et al., *The Gas Vesicle Gene Cluster from Microcystis aeruginosa and DNA Rearrangements That Lead to Loss of Cell Buoyancy.* Journal of Bacteriology, 2004. **186**(8): p. 2355-2365.

8. Humbert, J.F., et al., *A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium Microcystis aeruginosa.* PLoS One, 2013. **8**(8).

9. Steffen, M.M., et al., *Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis.* The ISME Journal, 2014. **8**(10): p. 2080-2092.

10. Meyer, K.A., et al., *Genome sequences of lower Great Lakes Microcystis sp. reveal strain-specific genes that are present and expressed in western Lake Erie blooms.* PLOS ONE, 2017. **12**(10): p. e0183859.

11. Stark, G.F., Truchon, A. R. , Dittmann, E. , Wilhelm, S. W. , *Closed circular genome sequence of a Microcystis aeruginosa PCC7806 -mcyB (UTK) nontoxic mutant.* Microbiology Resource Announcements, 2023.

12. Zhao, L., et al., *The highly heterogeneous methylated genomes and diverse restriction-modification systems of bloom-forming Microcystis.* Harmful Algae, 2018. **75**: p. 87-93.

13. Stark, G.F., et al., *Microcystin aids in cold temperature acclimation: Differences between a toxic Microcystis wildtype and non-toxic mutant.* Harmful Algae, 2023. **129**: p. 102531.

14. Dittmann, E., et al., *Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium Microcystis aeruginosa PCC 7806.* Molecular Microbiology, 1997. **26**(4): p. 779-787.

15. Tatusova, T., et al., *NCBI prokaryotic genome annotation pipeline.* Nucleic Acids Research, 2016. **44**(14): p. 6614-6624.

16. Benson, D.A., et al., *GenBank.* Nucleic Acids Research, 2012. **41**(Database issue): p. D36-42.

17. Martin, R.M., et al., *Episodic decrease in temperature increases mcy gene transcription and cellular microcystin in continuous cultures of Microcystis aeruginosa PCC 7806.* Frontiers in Microbiology, 2020. **11**.

18. Darling, A.C.E., et al., *Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements.* Genome Research, 2004. **14**(7): p. 1394-1403.

19. Madeira, F., et al., *Search and sequence analysis tools services from EMBL-EBI in 2022.* Nucleic Acids Research, 2022. **50**(W1): p. W276-W279.

20. Mistry, J., et al., *Pfam: The protein families database in 2021.* Nucleic Acids Research, 2021. **49**(D1): p. D412-D419.

21. Teufel, F., et al., *SignalP 6.0 predicts all five types of signal peptides using protein language models.* Nature Biotechnology, 2022. **40**(7): p. 1023-1025.

22. Finn, R.D., J. Clements, and S.R. Eddy, *HMMER web server: interactive sequence similarity searching.* Nucleic Acids Research, 2011. **39**(Web Server issue): p. W29-37.

23. Sayers, E.W., et al., *Database resources of the national center for biotechnology information.* Nucleic Acids Research, 2022. **50**(D1): p. D20-D26.

24. Nordberg, H., et al., *The genome portal of the Department of Energy Joint Genome Institute: 2014 updates.* Nucleic Acids Research, 2014. **42**(D1): p. D26-D31.

25. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Research, 2003. **31**(13): p. 3406-3415.

26. Camargo, A.P., et al., *Identification of mobile genetic elements with geNomad.* Nature Biotechnology, 2023.

27. Cunha, V.D., et al., *Integrative and Conjugative Elements Encoding DDE Transposases*, in *Bacterial Integrative Mobile Genetic Elements*. 2013. p. 250-260.

28. Guerillot, R., et al., *The diversity of prokaryotic DDE transposases of the mutator superfamily, insertion specificity, and association with conjugation machineries.* Genome Biol Evol, 2014. **6**(2): p. 260-72.

29. Feschotte, C., X. Zhang, and S.R. Wessler, *Miniature Inverted-Repeat Transposable Elements and Their Relationship to Established DNA Transposons*, C.e.a. N.L., Editor. 2002, American Society of Microbiology. p. 1147-1158.

30. Xin, Y., et al., *Amplification of miniature inverted-repeat transposable elements and the associated impact on gene regulation and alternative splicing in mulberry (Morus notabilis).* Mobile DNA, 2019. **10**(1).

31. Tu, Z., *Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito,Aedes aegypti.* Proceedings of the National Academy of Sciences, 1997. **94**(14): p. 7475-7480.

32. Delihas, N., *Impact of Small Repeat Sequences on Bacterial Genome Evolution.* Genome Biology and Evolution, 2011. **3**(0): p. 959-973.

33. Chen, I.M.A., et al., *The IMG/M data management and analysis system v.7: content updates and new features.* Nucleic Acids Research, 2023. **51**(D1): p. D723-D732.

34. Wu, S., P. Tian, and T. Tan, *Genomic landscapes of bacterial transposons and their applications in strain improvement.* Applied Microbiology and Biotechnology, 2022. **106**(19-20): p. 6383-6396.

35. Hickman, A.B., M. Chandler, and F. Dyda, *Integrating prokaryotes and eukaryotes: DNA transposases in light of structure.* Critical Reviews in Biochemistry and Molecular Biology, 2010. **45**(1): p. 50-69.

36. He, S., et al., *The IS200/IS605 Family and "Peel and Paste" Single-strand Transposition Mechanism.* Microbiology Spectrum, 2015. **3**(4).

37. Ellis, M.J., et al., *A transposon-derived small RNA regulates gene expression in Salmonella Typhimurium.* Nucleic Acids Research, 2017. **45**(9): p. 5470-5486.

38. Ellis, M.J. and D.B. Haniford, *Riboregulation of bacterial and archaeal transposition.* WIREs RNA, 2016. **7**(3): p. 382-398.

39. Ellis, M.J., R.S. Trussler, and D.B. Haniford, *Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation.* Molecular Microbiology, 2015. **96**(3): p. 633-650.

40. Ross, J.A., et al., *Tn5 transposition in Escherichia coli is repressed by Hfq and activated by over-expression of the small non-coding RNA SgrS.* Mobile DNA, 2014. **5**(1): p. 27.

41. Machida, C. and Y. Machida, *Regulation of IS1 transposition by the insA gene product* Journal of Molecular Biology, 1989. **208**: p. 567-574.

42. Mahillon, J. and M. Chandler, *Insertion Sequences.* Microbiology and Molecular Biology Reviews, 1998. **62**(3): p. 725-774.

43. Wu, Y., R.Z. Aandahl, and M.M. Tanaka, *Dynamics of bacterial insertion sequences: can transposition bursts help the elements persist?* BMC Evolutionary Biology, 2015. **15**(1).

44. Wilhelm, S.W., G.S. Bullerjahn, and R.M.L. McKay, *The complicated and confusing ecology of Microcystis blooms. mBio*, 2020. **11**(3).

45. Cui, L., et al., *Coordinated phenotype switching with large-scale chromosome flip-flop inversion observed in bacteria.* Proceedings of the National Academy of Sciences, 2012. **109**(25): p. E1647-E1656.

46. Guérillot, R., et al., *Unstable chromosome rearrangements in Staphylococcus aureus cause phenotype switching associated with persistent infections.*

Proceedings of the National Academy of Sciences, 2019. **116**(40): p. 20135-20140.

47. Kotewicz, M.L., et al., *Optical maps distinguish individual strains of Escherichia coli O157 : H7.* Microbiology 2007. **153**(Pt 6): p. 1720-1733.

48. Eckstein, S., et al., *Two novel XRE-like transcriptional regulators control phenotypic heterogeneity in Photorhabdus luminescens cell populations.* BMC Microbiology, 2021. **21**(1).

49. Lewis, R.J., et al., *An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti repressor complex.* Journal of Molecular Biology, 1998. **283**: p. 907-912.

50. Meinke, G., et al., *Cre Recombinase and Other Tyrosine Recombinases.* Chemical Reviews, 2016. **116**(20): p. 12785-12820.

51. Sugita, C., et al., *Complete nucleotide sequence of the freshwater unicellular cyanobacterium Synechococcus elongatus PCC 6301 chromosome: gene content and organization.* Photosynthesis Research, 2007. **93**(1-3): p. 55-67.

52. Entfellner, E., et al., *Toxic/Bioactive Peptide Synthesis Genes Rearranged by Insertion Sequence Elements Among the Bloom-Forming Cyanobacteria Planktothrix.* Frontiers in Microbiology, 2022. **13**.

53. Kato, Y., et al., *RNA-hydrolyzing activity of metallo-beta-lactamase IMP-1.* PLoS One, 2020. **15**(10): p. e0241557.

54. Cerveny, L., et al., *Tetratricopeptide Repeat Motifs in the World of Bacterial Pathogens: Role in Virulence Mechanisms.* Infection and Immunity, 2013. **81**(3): p. 629-635.

55. Blatch, G.L. and M. Lässle, *The tetratricopeptide repeat: a structural motif mediating protein-protein interactions.* BioEssays, 1999. **21**(11): p. 932-939.

56. Mlouka, A., K. Comte, and N. Tandeau De Marsac, *Mobile DNA elements in the gas vesicle gene cluster of the planktonic cyanobacteria Microcystis aeruginosa.* FEMS Microbiology Letters, 2004. **237**(1): p. 27-34.

57. Yang, G., et al., *Transposition of the rice miniature inverted repeat transposable element mPing in Arabidopsis thaliana.* Proceedings of the National Academy of Sciences, 2007. **104**(26): p. 10962-10967.

58. Klein, B.A., et al., *Identification and characterization of a minisatellite contained within a novel miniature inverted-repeat transposable element (MITE) of Porphyromonas gingivalis.* Mobile DNA, 2015. **6**(1).

59. Min, X., R. Kong, and X. Xu, *A miniature insertion element transposable in Microcystis sp. FACHB 854.* Progress in Natural Science, 2006. **16**(5).

60. Kuno, S., et al., *Intricate Interactions between the Bloom-Forming Cyanobacterium Microcystis aeruginosa and Foreign Genetic Elements, Revealed by Diversified Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Signatures.* Applied and Environmental Microbiology, 2012. **78**(15): p. 5353-5360.

61. Matus-Garcia, M., H. Nijveen, and M.W.J. Van Passel, *Promoter propagation in prokaryotes.* Nucleic Acids Research, 2012. **40**(20): p. 10032-10040.

62. Long, K.S., et al., *Mutations in 23S rRNA at the Peptidyl Transferase Center and Their Relationship to Linezolid Binding and Cross-Resistance.* Antimicrobial Agents and Chemotherapy, 2010. **54**(11): p. 4705-4713.

63. Laudenbach, E.D. and R.A. Grossman, *Characterization and mutagenesis of sulfur-regulated genes in a cyanobacterium: evidence for function in sulfate transport.* Journal of Bacteriology, 1991. **173**(9): p. 2739-2750.

64. Zhang, Z., et al., *Gene expression patterns of sulfur starvation in Synechocystis sp. PCC 6803.* BMC Genomics, 2008. **9**: p. 344.

65. Papoulis, S.E., et al., *Nutrient Loading and Viral Memory Drive Accumulation of Restriction Modification Systems in Bloom-Forming Cyanobacteria.* mBio, 2021. **12**(3): p. e0087321.

66. Long, B.M., *Evidence that sulfur metabolism plays a role in microcystin production by Microcystis aeruginosa.* Harmful Algae, 2010. **9**(1): p. 74-81.

67. Dittmann, E., et al., *Altered expression of two light-dependent genes in a microcystin-lacking mutant of Microcystis aeruginosa PCC 7806.* Microbiology, 2001. **147**(11): p. 3113-3119.

68. Heaton, B.E., et al., *Molecular Structure and Function of the Novel BrnT/BrnA Toxin-Antitoxin System of Brucella abortus.* Journal of Biological Chemistry, 2012. **287**(15): p. 12098-12110.

69. Wagner, A., *Cooperation is fleeting in the world of transposable elements.* PLoS Computational Bioliology, 2006. **2**(12): p. e162.

70. Xiang, Y., et al., *Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage φ29 tail.* Proceedings of the National Academy of Sciences, 2008. **105**(28): p. 9552-9557.

71. Casacuberta, E. and J. González, *The impact of transposable elements in environmental adaptation.* Molecular Ecology, 2013. **22**(6): p. 1503-1517.

Appendix

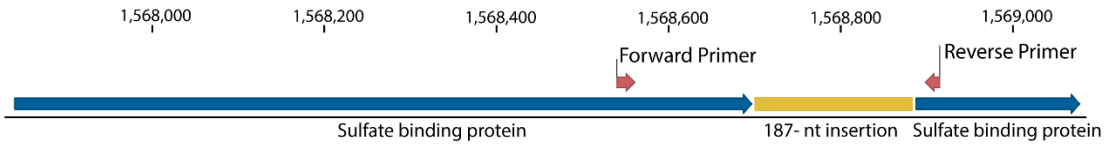

Figure 3.8 Location of primer design for *sbp* gene

Layout of the gene encoding a sulfate binding protein in the *Microcystis aeruginosa* PCC7806 Δ*mcyB* strain. Blue regions are coding sequences for the sulfate binding gene. In orange is the 187-nt MITE, located from 1568698 to 1568884 in the Δ*mcyB* genome. Primers were made to amplify a 181 bp region in the PCC 7806 wildtype (no insertion present), and a 376 bp region in Δ*mcyB*.

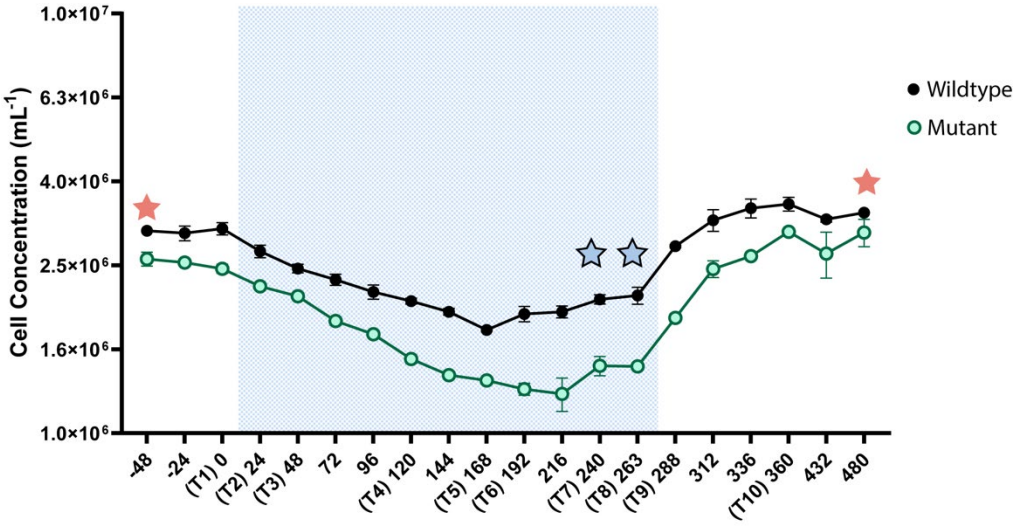

Figure 3.9 Locations chosen for gene expression analysis of transposases

Cell concentrations (y-axis) of a chemostat experiment done for a previous manuscript (Stark et al. 2023a), time in hours is on the x-axis. Normalized gene expression between the *Microcystis aeruginosa* PCC7806 Δ*mcyB* and *Microcystis aeruginosa* PCC7806 wildtype strains were compared for the cold (19°C) treatment (T7 and T8), and control "warm" (26°C) treatment (T1 and T10).

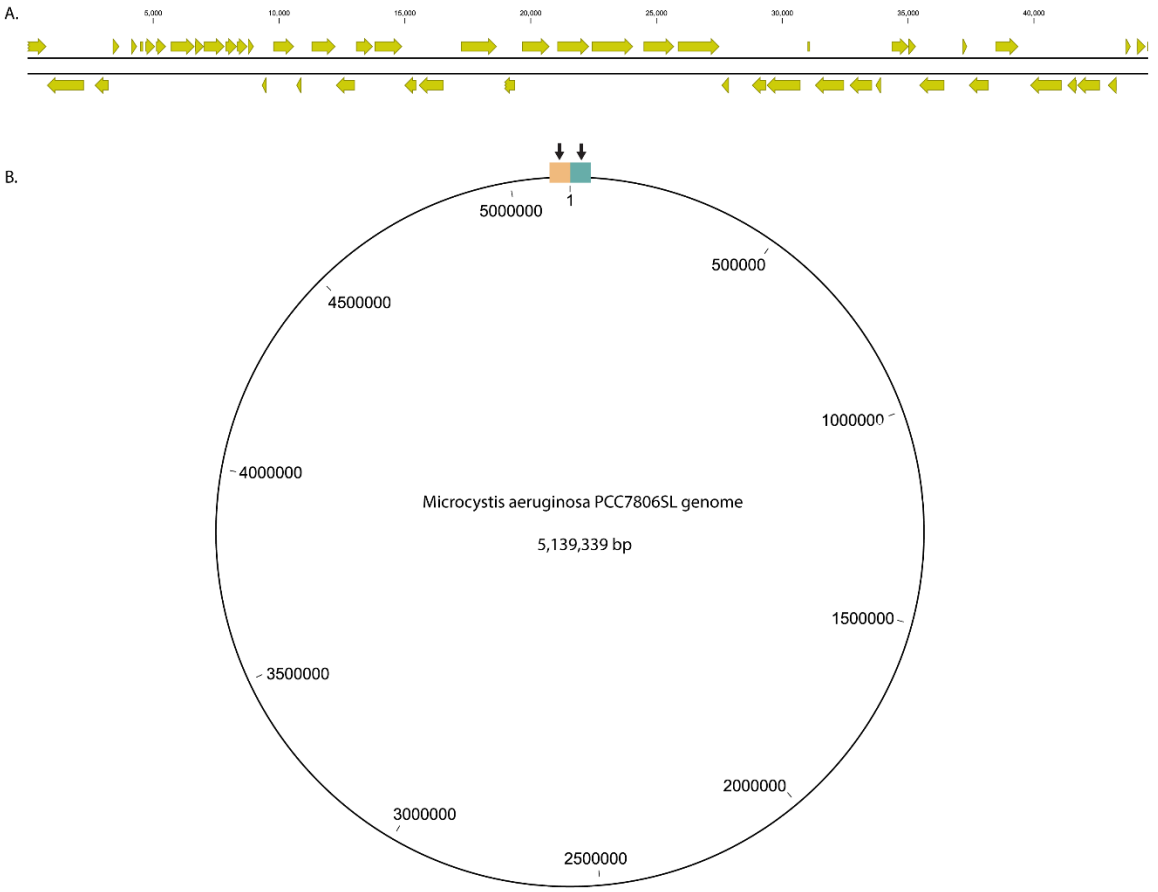

Figure 3.10 Duplicated region in *M.aeruginosa* PCC 7806SL genome

A) Duplicated region of the *Microcystis aeruginosa* PCC7806SL genome downloaded off NCBI's Genbank. The region is ~45kb in size, and codes ~49 genes. The region located from 5094806 to 5139339 codes a transposase which is not seen in the duplicated region from 1 to 43125. B) Genome map of *Microcystis aeruginosa* PCC7806SL. Duplicated regions of the genome (seen above) are denoted by the orange and the green box. The orange box represents the region that has an additional transposase in the region.

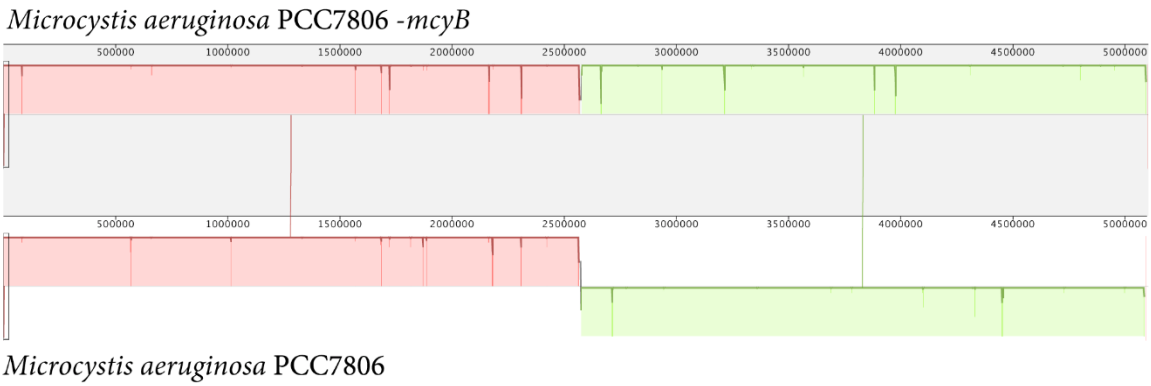

Figure 3.11 Mauve alignment of wildtype and Δ*mcyB* genomes

Mauve whole-genome alignment of our *Microcystis aeruginosa* PCC 7806 wildtype (bottom) and Δ*mcyB* (top) strains. Genome regions that are the same color correspond to one another. In green is the region of the genome that is inverted between the PCC 7806 wildtype and Δ*mcyB*.

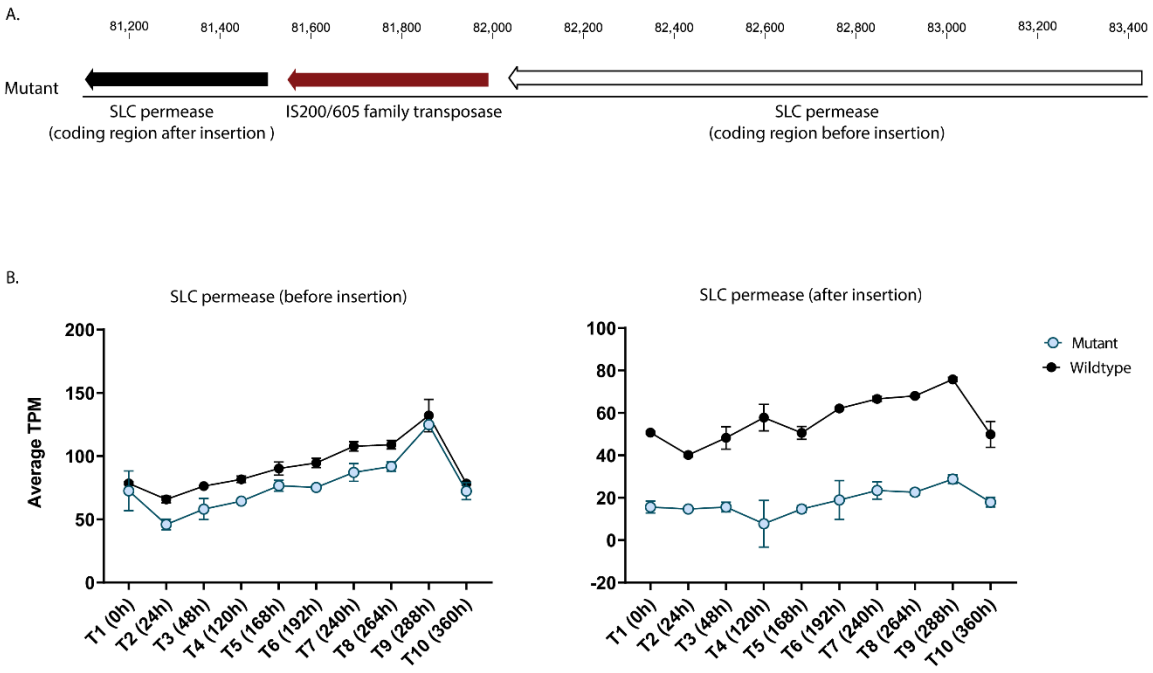

Figure 3.12 SLC permease interruption in Δ*mcyB* by *tnpA*

A) SLC permease gene was interrupted in the *Microcystis aeruginosa* PCC7806 Δ*mcyB* strain by a IS200/605 family transposase. B) The gene expression of the coding region of the SLC permease before the transposase is similar to the gene expression profile seen in the PCC7806 wildtype, however, transcripts for the coding region of the gene after the insertion are about half that of the wildtype.

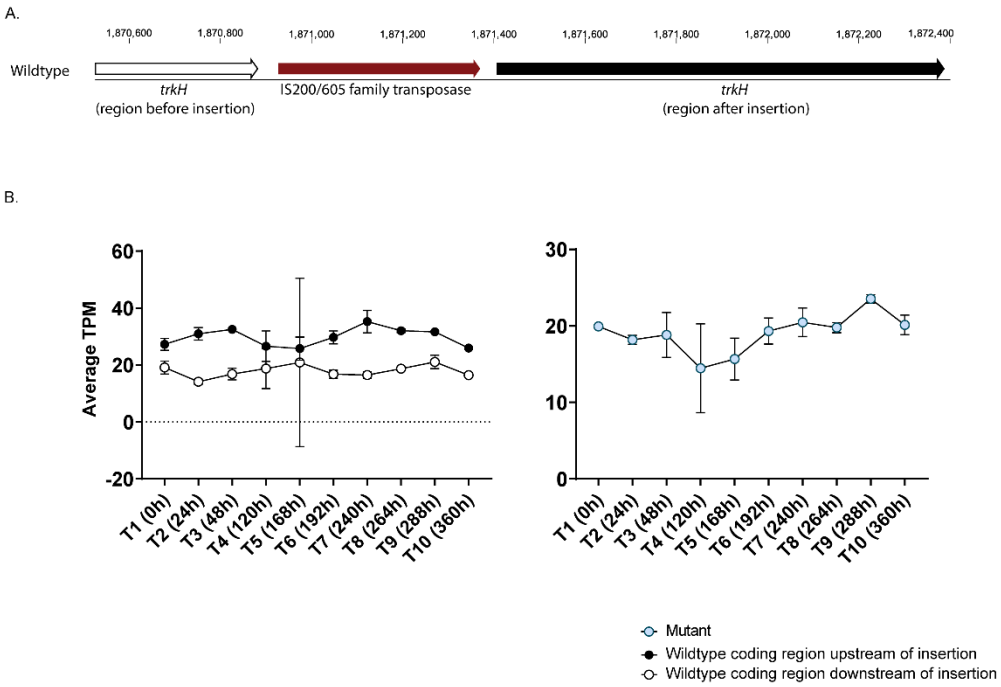

Figure 3.13 *trkH* interruption in the wildtype by *tnpA*

A) A IS1200/605 family transposase is present in the *Microcystis aeruginosa* PCC7806 wildtype genome, interrupting a *trkH* gene. The *trkH* gene is not interrupted in the Δ*mcyB* genome. The mutation does not appear to cause a significant difference in expression from the *Microcystis aeruginosa* Δ*mcyB* genome.

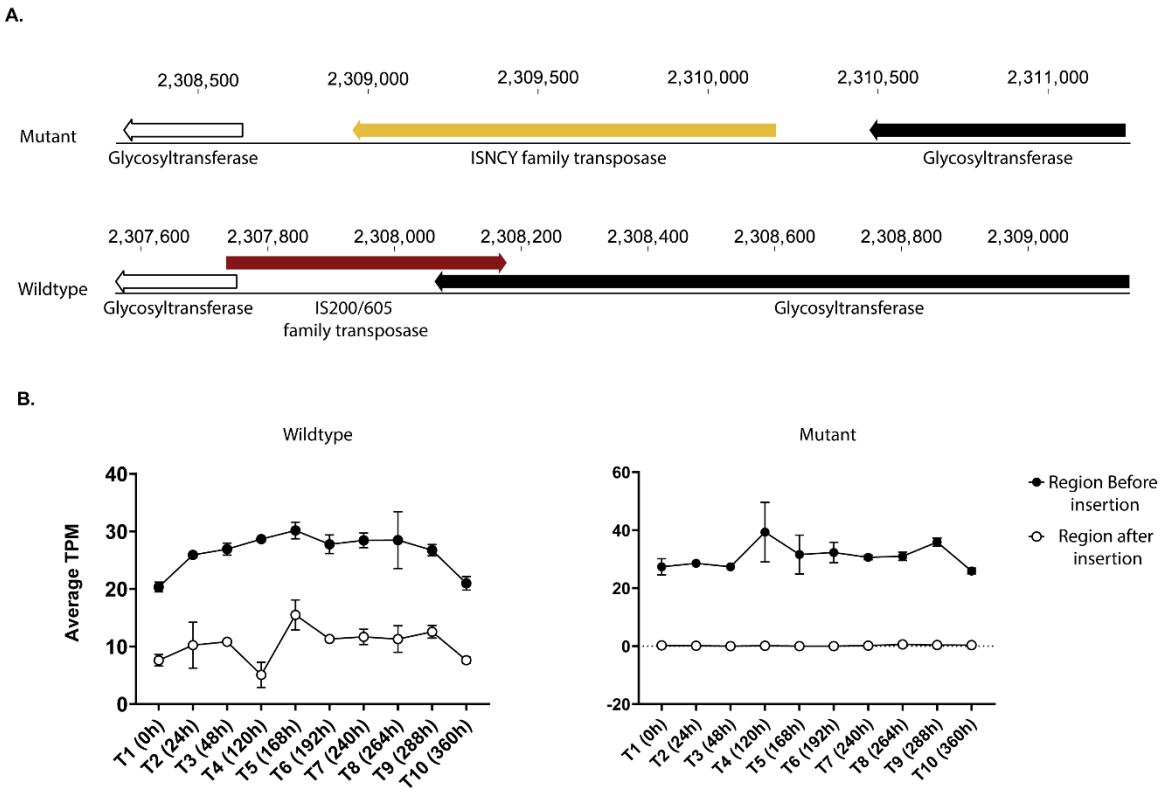

Figure 3.14 Glycosyltransferase interrupted by *tnpA* in the wildtype

A) A glycosyltransferase gene in the *Microcystis aeruginosa* PCC7806 Δ*mcyB* and PCC 7806 wildtype genome was interrupted in both strains, by two different types of transposases. In Δ*mcyB*, the glycosyltransferase gene was interrupted by a ISNCY (insertion sequence not characterized yet) transposase, whereas in the wildtype the glycosyltransferase gene was interrupted by a IS200/605 family transposase. (B) Gene expression of the interrupted glycosyltransferase gene in Δ*mcyB* and the PCC 7806 wildtype. In the wildtype, the IS200/605 insertion reduced transcript abundance by approximately half for the region of the gene after the insertion sequence. In the mutant, insertion of the ISNCY transposase in the middle of the gene resulted in no expression for the coding region after the insertion.

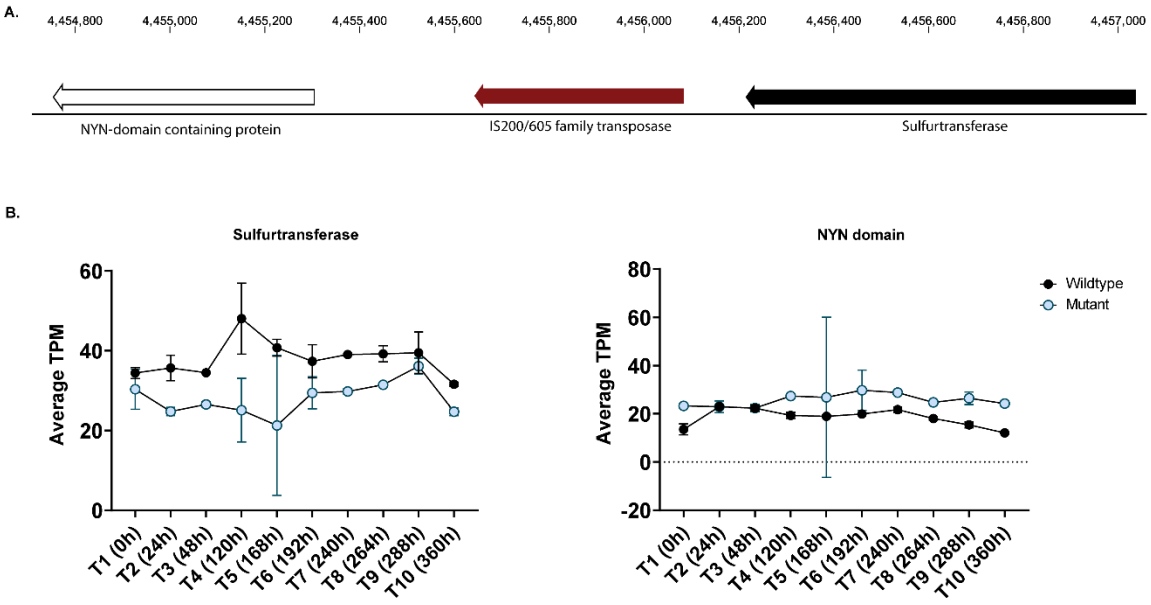

Figure 3.15 Insertion of *tnpA* in the mutant

A) A IS200/605 family transposase inserted between genes coding for a sulfurtransferase and a NYN-domain protein in the *Microcystis aeruginosa* PCC 7806 wildtype strain. B) Insertion of the transposase did not seem to effect gene expression of the gene encoding a NYN-domain containing protein in the PCC 7806 wildtype, as the expression profile was similar to that of the PCC7806 Δ*mcyB* strain.

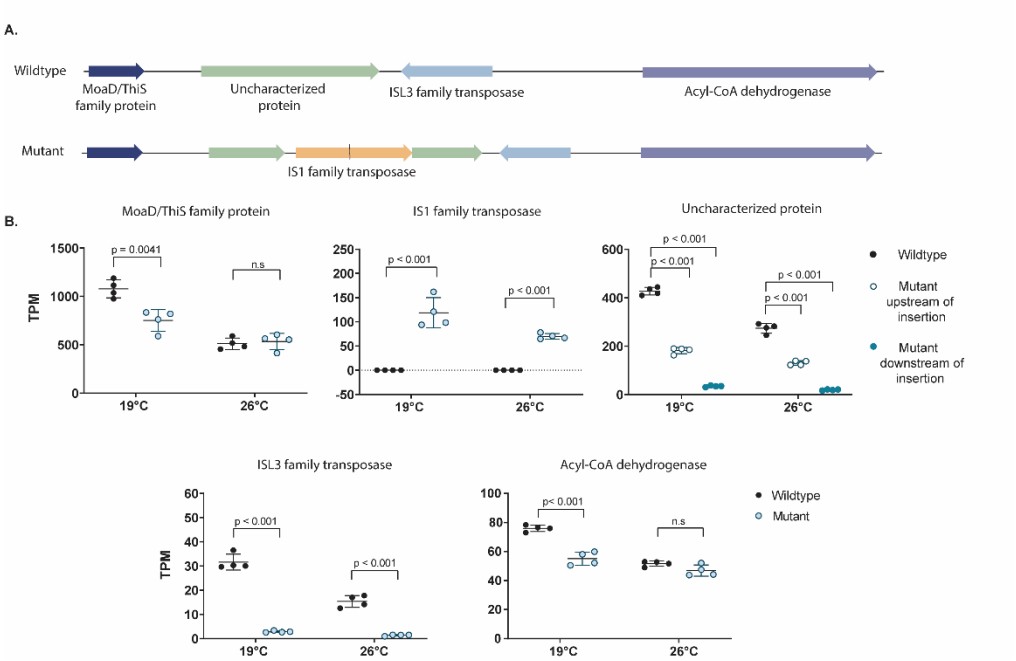

Figure 3.16 IS1-family transposase insertion in an uncharacterized gene in the mutant

A) Gene arrangement in *Microcystis aeruginosa* PCC7806 wildtype (top) versus the *Microcystis aeruginosa* PCC7806 Δ*mcyB* (bottom) genome. The gene arrangement shows where a IS1-family transposase is inserted in a gene encoding an uncharacterized protein (putative thioesterase) in Δ*mcyB*, and genes surrounding the mutated gene. The PCC7806 wildtype arrangement shows the same hypothetical protein and gene area, unmutated. B) Transcription of the genes in the order shown in 9A. Insertion of the IS1-family transposase in the gene encoding a hypothetical protein significantly reduces its expression in the mutant ($p<0.0001$), while the IS1-family transposase is transcribed instead (average TPM ~70 at 26°C and ~118 at 19°C). Consequently, this insertion causes a significant decrease in expression/little activity ($p <0.001$) of an ISL3-family transposase downstream of the hypothetical protein in Δ*mcyB* compared to the PCC7806 wildtype.

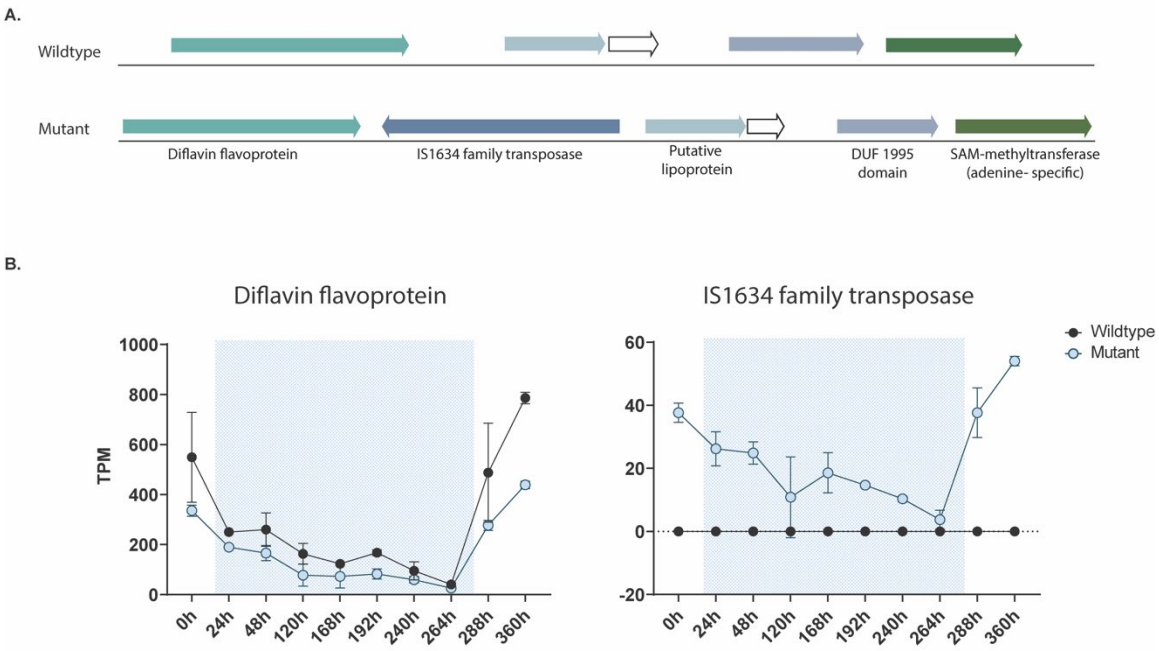

Figure 3.17 Gene expression profile of IS1634 family transposase and a gene encoding a diflavin flavoprotein

(A) In *Microcystis aeruginosa* PCC7806 Δ*mcyB*, an IS1634 transposase inserted between a diflavin flavoprotein and a putative lipoprotein. B) The IS1634 transposase inserted in the Δ*mcyB* strain mimics gene expression of the diflavin flavoprotein.

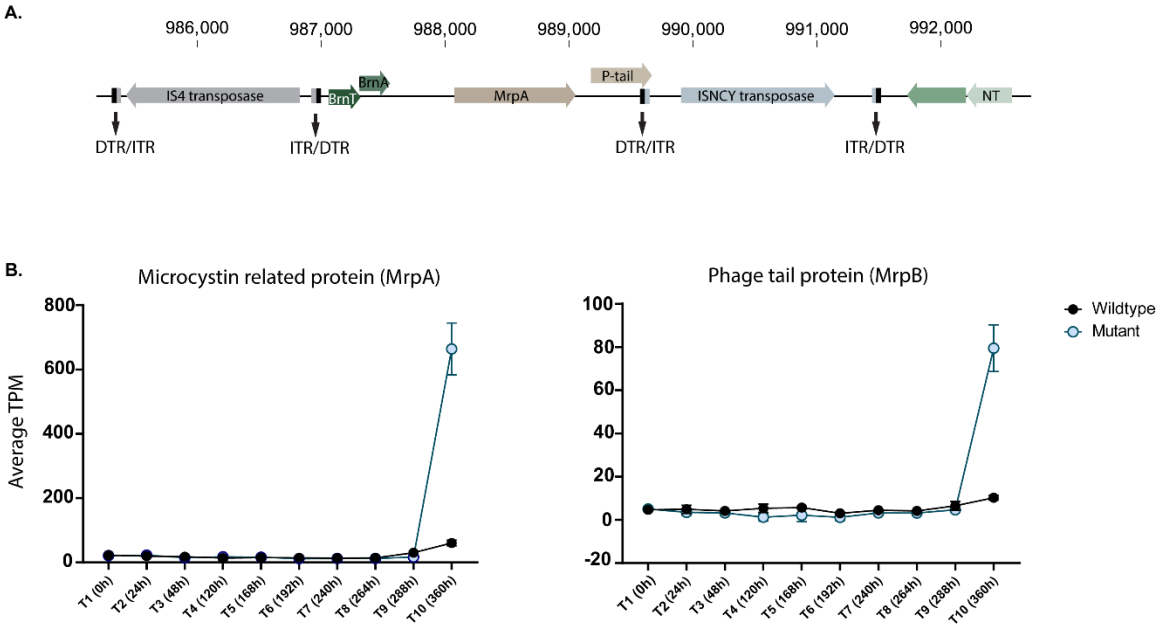

Figure 3.18 Microcystin-related protein, gene arrangement and expression profile in the mutant and wildtype

A) A gene encoding a phage tail protein, P-tail (previously annotated as mrpB) was found in *Microcystis aeruginosa* PCC7806, next to a microcystin-reliant protein (mrpA). These two genes were situated near a BrnTA toxin-antitoxin system (which is flanked by an IS4-transpoase), and an ISNCY (insertion element not characterized yet). B) Normalized gene expression (average TPM), for mrpA and the phage tail gene. These two genes appear to be active in the mutant at one time point, providing some sort of function to the cells.

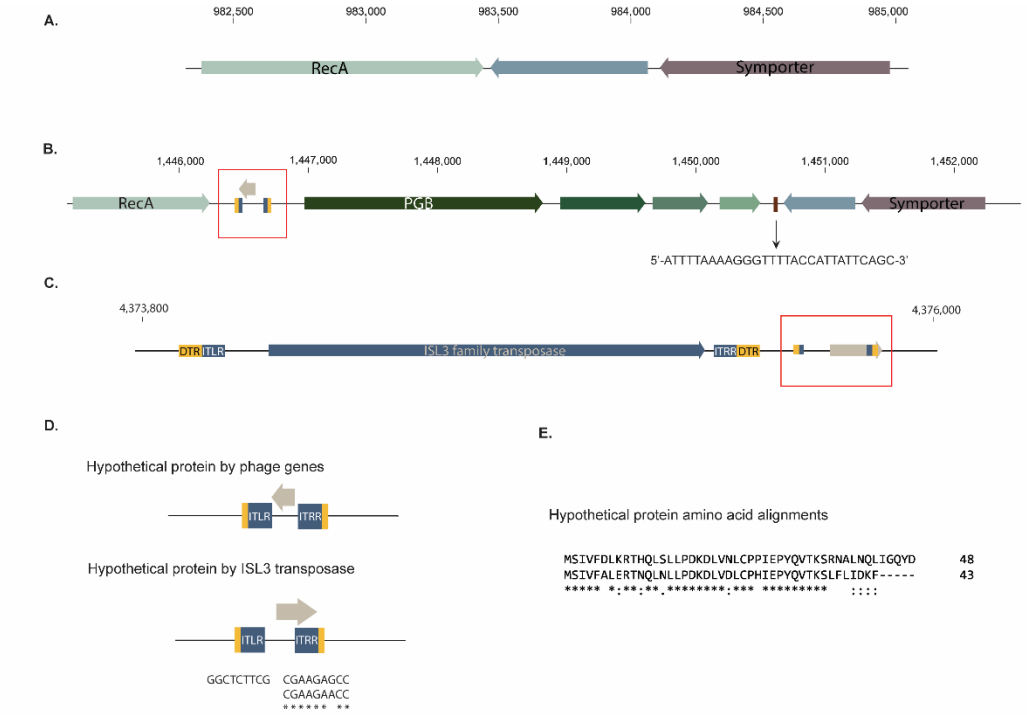

Figure 3.19 Phage protein D in *M.aeruginosa* NIES-843

A) phage protein gene with peptidoglycan binding domains (PGB- dark green) was identified in *Microcystis aeruginosa* NIES 843. Figure A is an unmutated gene arrangement in *Microcystis aeruginos*a LE3. Genes with similar colors are homologous to one another. B) In Microcystis aeruginosa NIES843, the phage gene (PGB) was found next to three other unique genes (nucleotide sequences not found in other complete *Microcystis* spp. genomes) downstream of PGB (all green) may have also been acquired with this gene. Flanking these four genes upstream is a gene encoding a hypothetical protein (marked by a red box) which appears to have been inserted by a ISL3 family transposase, as it has similar inverted terminal repeats to ISL3 and is flanked by direct terminal repeats, and a homolog of this gene is present in another area of the genome (seen in figure III, marked by a red box), next to an ISL3 transposase (B, C, D). Downstream of these four genes is a sequence which is repetitive in the genome (exists in 25 areas). This sequence is 5'-ATTTTAAAAGGGTTTTACCATTATTCAGC-3'. Of the 25 repeats, 10 are by transposases, and of those 10, four are by TnpB transposable elements. Similar gene arrangements without the four acquired genes are found in other *Microcystis* genomes (*Microcystis aeruginosa* NIES-88,LE3 and *Microcystis virdis* NIES-102 were checked).

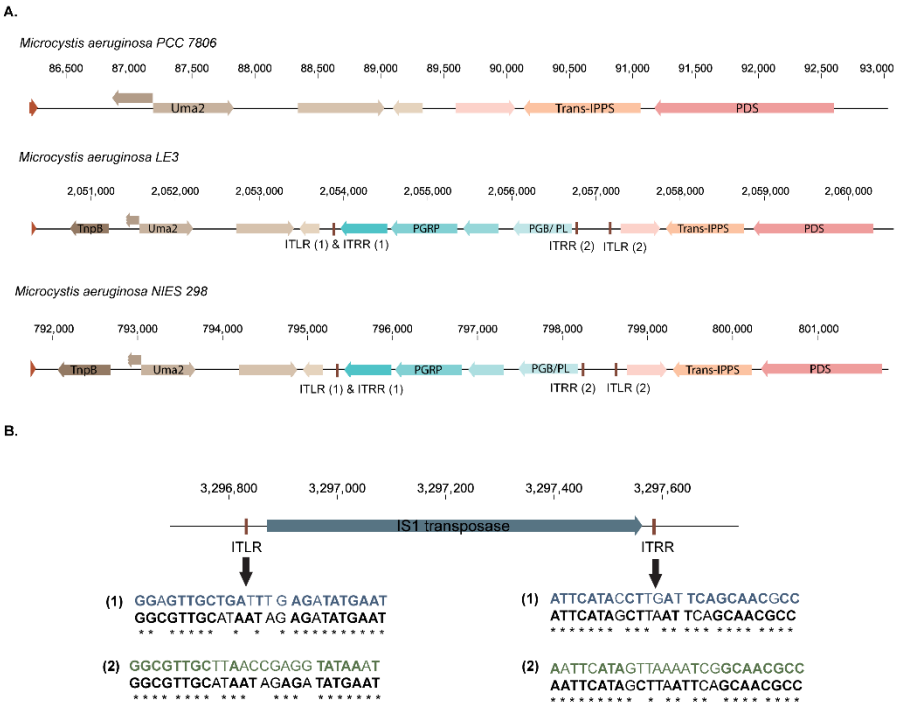

Figure 3.20 Phage tail lysozyme in *M.aeruginosa* LE-3 and NIES-298

A) *Microcystis aeruginosa* LE3 and *Microcystis aeruginosa* NIES298 both have a putative phage tail lysozyme (PGB/PL gene). Along with this gene are three other genes (in blue) that may have been acquired with the putative phage (PBG/PL) gene, one of them has a peptidoglycan recognition/binding protein domain (PGRP). In *Microcystis aeruginosa* LE3 and *Microcystis aeruginosa* NIES298, these four genes are flanked by mutated areas which have homology IS1-family transposases in the genome. These intergenic regions have areas that align with the inverted terminal repeat regions of an IS1-family transposase (ITLR – inverted left terminal repeat, ITRR- inverted right terminal repeat). ITLR (1) is an imperfect ITR of ITRR (1). In other *Microcystis* strains (NIES843, 7608), the gene arrangement is similar to LE3 and NIES298, where an uma2 endonuclease is located near a tRNA, and a phytoene synthase (Trans-IPPS) and a phytoene desaturase (PDS) are present. The unlabeled genes are uncharacterized genes. B) Alignment of the inverted terminal repeats for an IS1-family transposase. The colored sequences are those from the intergenic regions in LE3 and NIES298, and black are the sequences of the IS1-transposase ITR's. The numbers by the sequences correspond to the location of the ITLR and ITRR, above.

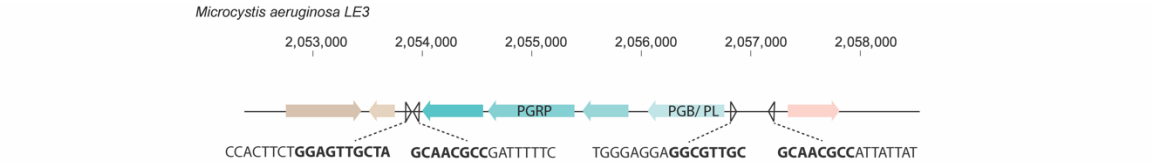

Figure 3.21 Phage Tail lysozyme in LE-3 and ITR's of IS1-transposase

A more detailed view of the region in Microcystis aeruginosa LE-3 which has a phage-tai lysozyme gene (PGB/PL), and three other genes associated with it, one of which has peptidoglycan recognition/binding domains (PGRP). Flanking these four genes in the intergenic regions are sequences which are homologous the inverted terminal repeats (ITR) of certain 756 & 758-bp long IS1 family transposases present in the LE3 genome. In bold are the ITR's, the sequences flanking the repeats are not bolded/part of the transposase unit. No apparent DTR's flank the ITR's in these regions, which is also seen in the 756-bp long IS1-family transposases.

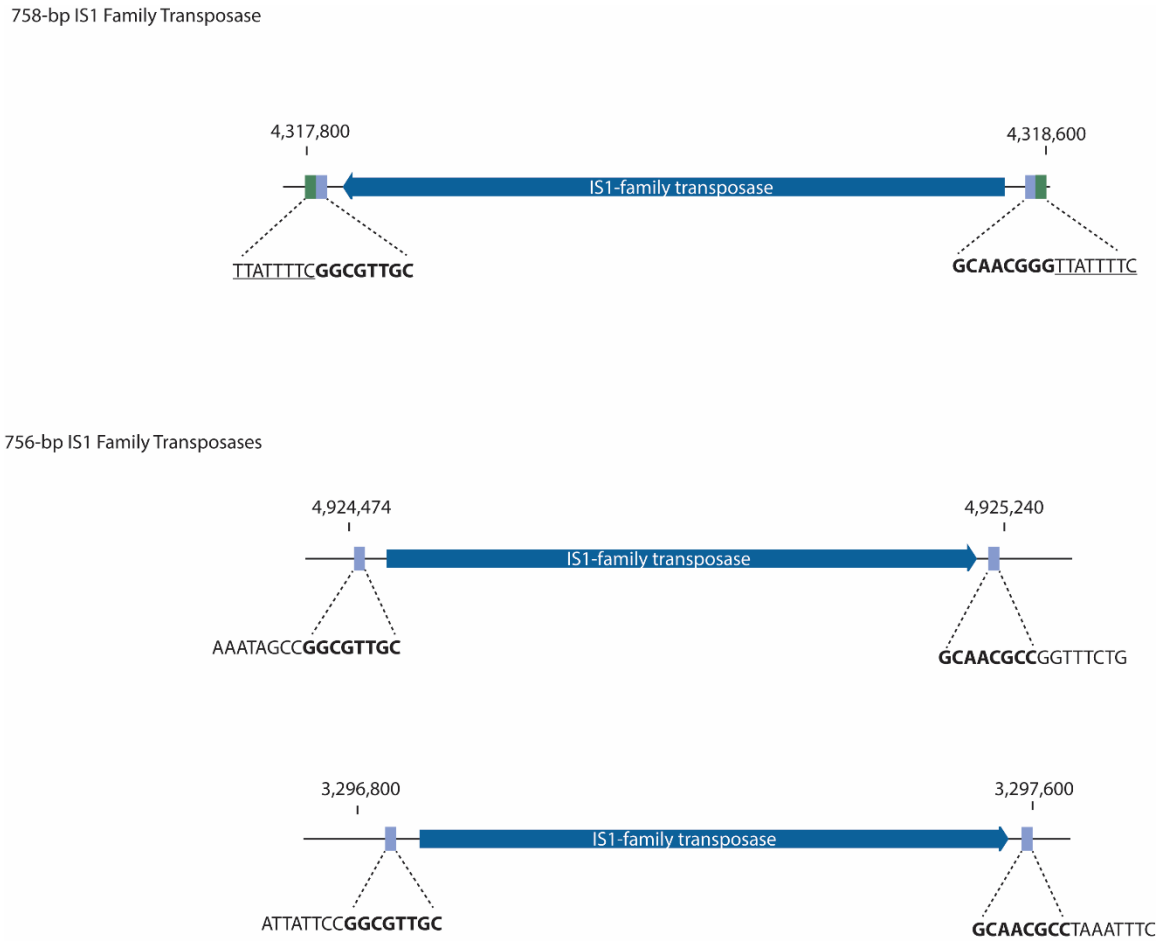

Figure 3.22 Inverted terminal repeats for IS1-family transposases in LE-3

Microcystis aeruginosa LE3 has three IS1-family transposases which have inverted terminal repeats (ITR's) that match those of the intergenic region flanking four genes which may have been horizontally acquired in LE3 & other *Microcystis* strains(Figs S13, S14). One of the IS1-family transposases is 758bp long and has direct terminal repeats flanking the ITR's. Two of the IS1 family transposases are 756bp long and do not appear to have any direct terminal repeats.

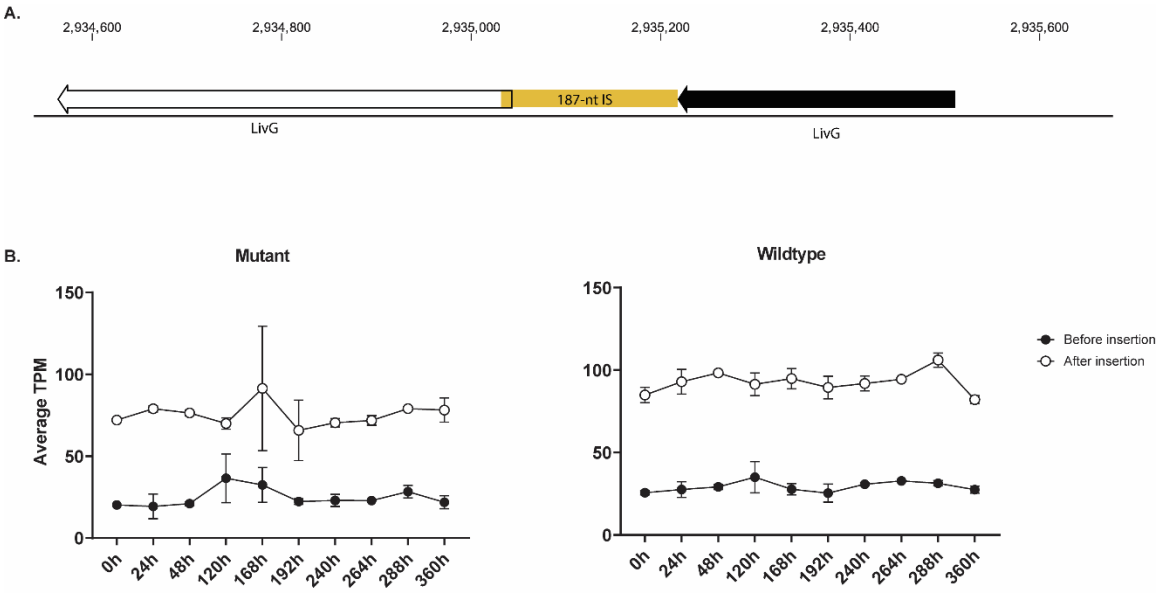

Figure 3.23 Insertion of a MITE in *livG* in the mutant and expression of the gene

A) 187-nt long insertion sequence (IS) was inserted in a gene encoding a branched chain amino acid transporter (LivG) in the mutant strain. B) The *livG* gene shows expression values roughly half that of the wildtype, before and after the insertion sequence. This is unlike the sulfate binding protein in the mutant, which has the same 187-nt IS, but has upregulated expression of sulfate transporter gene cluster (*sbp-cysTWA*).

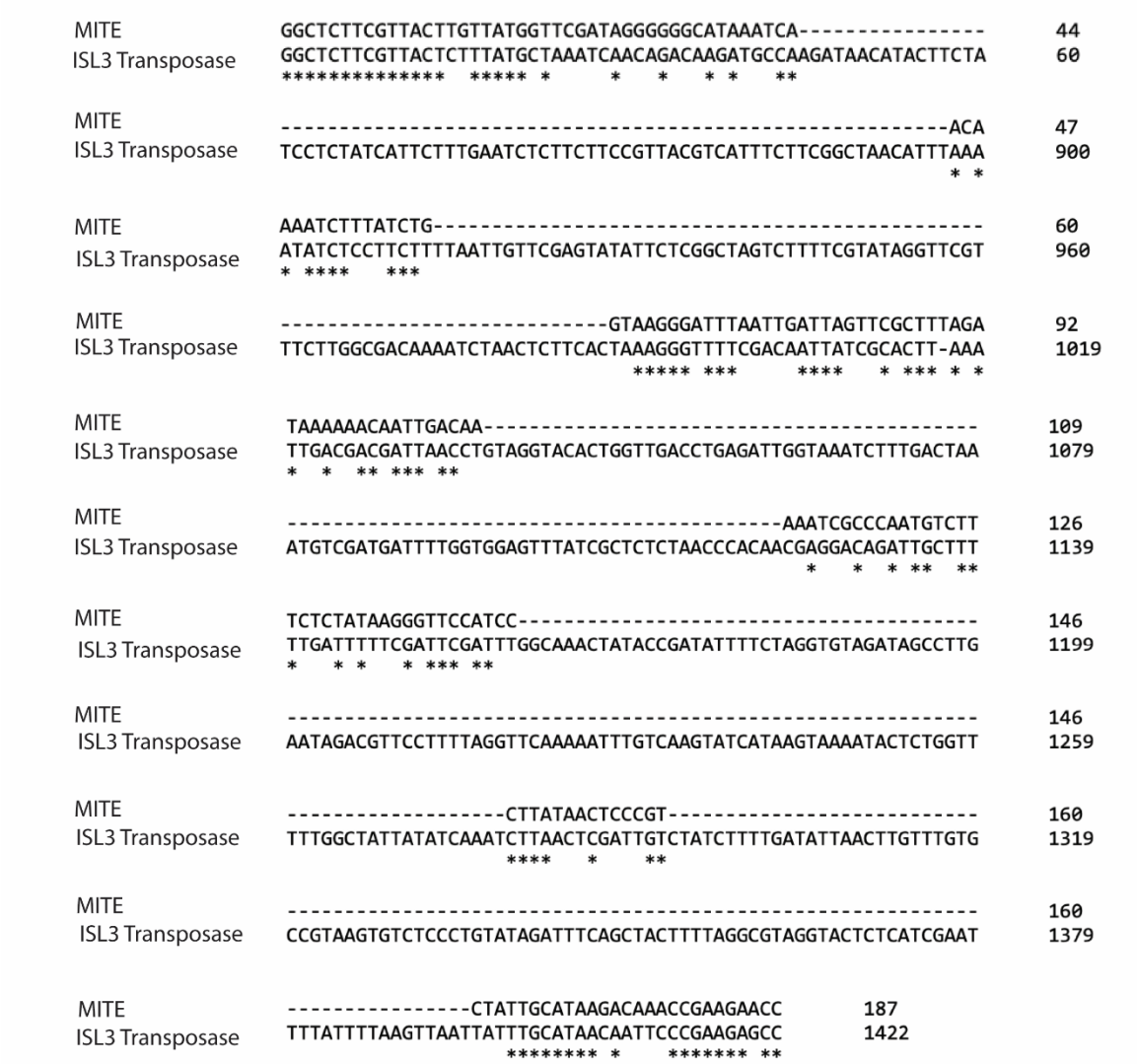

Figure 3.24 Shortened alignment of an ISL3-family transposase with the MITE

Shortened alignment (Clustal Omega output) of an ISL3-family transposase with inverted terminal repeats (ITR) that match the imperfect ITR's of the 187-nt miniature inverted repeat transposable element that is active in *Microcystis aeruginosa* PCC7806.

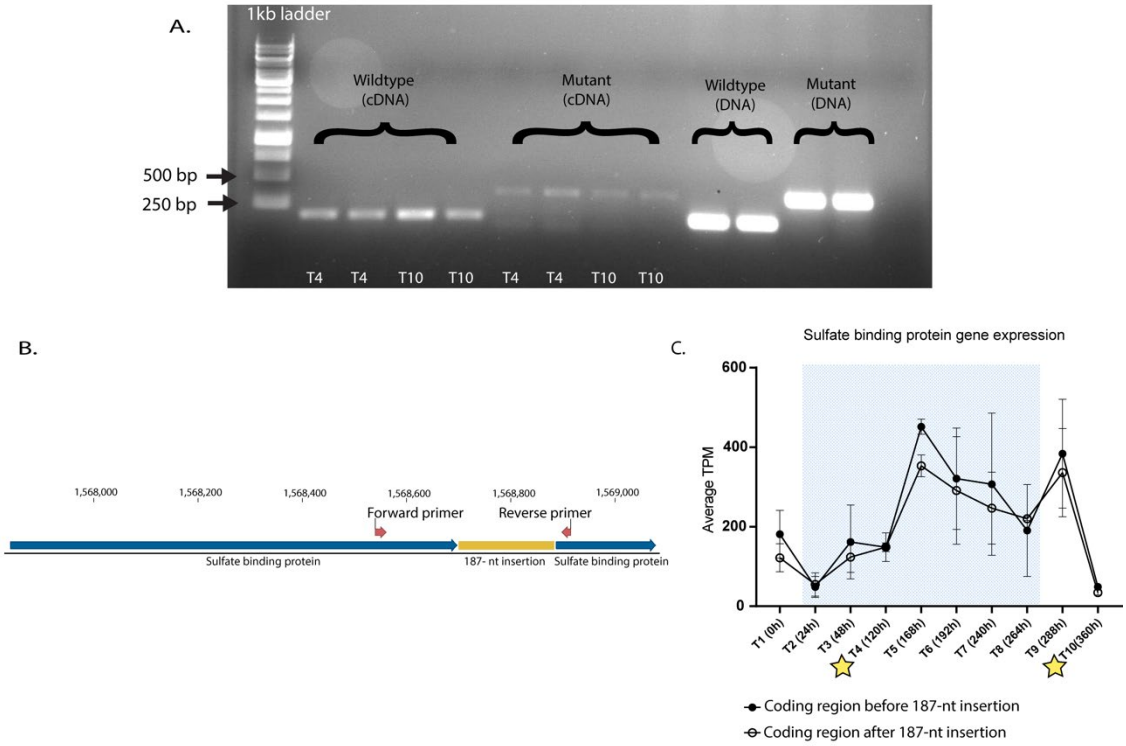

Figure 3.25 Reverse transcriptase PCR of the MITE insertion in *sbp*

A) Results from reverse transcriptase PCR on different timepoints taken from the original RNA-seq experiment (Stark et al. 2023a). Transcriptomic libraries from the original experiment were used in this manuscript. In the *Microcystis aeruginosa* PCC 7806 wildtype, a 181-bp amplicon was generated, consistent with the lack of the MITE in the gene encoding a sulfate binding protein. For the PCC 7806 Δ*mcyB* strain, faint bands around 376-bp in length were generated, consistent with the presence of the MITE insertion in the gene encoding the sulfate binding protein. B) Points that RT-PCR analysis was done are indicated with a star in the RNA-seq time series. This gel image has not been modified beyond the text added to the figure. All bands present were run on the same gel at the same time.

CHAPTER FOUR:
THE PHYSIOLOGICAL AND ECOLOGICAL RELEVANCE OF CONSERVED *MICROCYSTIS* PLASMIDS

Publication Note

This chapter is still in prep but will be submitted to a peer-reviewed journal. The contributors of this chapter are Gwendolyn F. Stark, Laura E. Smith, Alexander P. Truchon, Robbie M. Martin, Steven W. Wilhelm.

GFS and SWW designed the experiment. GFS did data/transcriptional analysis. LES collected, extracted and sequenced all *M. aeruginosa* NIES-298 genomes. APT ran the minION sequencer and assembled the genomes/plasmids. RMM made the Lake Erie co-assembly.

Abstract

Re-sequencing the genome of a commonly used harmful algal bloom (HAB) forming cyanobacterium, *Microcystis aeruginosa* PCC 7806, and it's non-toxic mutant, *Microcystis aeruginosa* PCC 7806 Δ*mcyB* lead to the discovery a of previously unreported 8.5kb plasmid present in both strains. Despite these two strains being separated for over 25 years, the plasmid was identical in both strains. Using RNA-seq datasets, we found this plasmid to be transcriptionally active during growth in chemostats, and *in situ* metatranscriptomes also revealed the plasmid or closely related genes to the plasmid were active in Lake Erie. Additionally, metatranscriptomes revealed plasmids and viruses show similar levels of transcripts *in situ*. We further investigated the phenomenon of "conserved" plasmids in *Microcystis spp*. through additional sequencing using *Microcystis aeruginosa* NIES-298, a strain often used in phage-infection research. Re-sequencing the NIES-298 uncovered a putatively conserved phage-like plasmid, and another plasmid that may contribute to chromosomal rearrangements, both of which are also transcriptionally active. These discoveries lead us to investigate the ecological importance of *Microcystis* plasmids, the potential mechanism(s) of plasmid conservation, and the similarities between plasmids and phage.

Introduction

Microcystis aeruginosa is a well-studied harmful bloom forming cyanobacteria, yet there are still many aspects of *Microcystis spp*. physiology and lifestyle we are still uncovering. One often overlooked aspect of *Microcystis spp.* are the plasmids they harbor, and what role these plasmids have in the cell. Usually, physiology-based studies and ecologically based studies ignore *Microcystis spp.* plasmids completely. Due to the lack of research and acknowledgment of *Microcystis* plasmids, we were initially surprised to find an uncharacterized plasmid in the well-used type-strain *Microcystis aeruginosa* PCC 7806. This wildtype strains and its non-toxic mutant, *Microcystis aeruginosa* PCC7806 Δ*mcyB,* have been used in comparative studies with one another since the mutant's creation in 1997, to elucidate the role of microcystin in the cell [1-12]. As these two strains are used in labs around the world, we wanted to investigate this seemingly conserved plasmid, and investigate other *Microcystis spp.* to see if conserved plasmids are common features in this genera.

In the 1980's, before microcystin production was discovered to be chromosomally encoded by a PKRS/non-ribosomal peptide synthase complex, it was thought that plasmids may play a role in microcystin production [13]. For this reason, many researchers set out to elucidate *Microcystis* spp. plasmids. One study from that time showed *Microcystis* can harbor multiple plasmids, something that was often observed more in toxic strains [14]. While we know now that the microcystin toxin is not encoded by plasmids, research of this time did provide insights into *Microcystis* plasmid structure and conservation. For instance, it was found that there was sequence homology between co-existing plasmids and chromosomes of the host organism, and identical plasmids were found in strains isolated from different geographic regions through southern hybridization techniques [14]. However, since that time there has been limited research on *Microcystis* plasmids, aside from a small handful of studies [14-16]. Because of this, we are missing out on understanding how or if *Microcystis* plasmids effect physiology, and in turn bloom ecology. With more current sequencing techniques from those available in the 1980's, we were able to investigate these initial observations in more depth with -omics approaches, exploring homology between *Microcystis spp*. plasmids, their "conserved" nature, and

their activity in cells.

While plasmids are important to consider when conducting and investigating *Microcystis* physiology studies, ecologically they may also play relevant roles. *Microcystis* bloom ecology has also come a long way due to advancements in sequencing technology, however, plasmids and their role in *Microcystis* bloom research have not received as much attention. More recently, researchers are beginning to understand the importance of including the "phycosphere" or "interactome" in bioinformatic-based bloom research [17-19]. There is also a lot of focus on phage and their role in bloom dynamics [20-24]. As harbingers of horizontal gene transfer (HGT) between organisms, more attention should be paid to plasmids, as they could confer fitness impacts, such as antibiotic resistance, to susceptible host cells [25]. In fact, the co-occurrence of *Microcystis* and plasmid-mediated antibiotic resistance genes has been observed before [26]. Such things would impact *Microcystis* bloom dynamics and the microbiome composition of blooms. Additionally, with likeness between plasmid, hosts, and possibly phage genes, there could be inaccurate representations of *in situ* gene expression if plasmids are left unaccounted for.

In the last decade, genome sequencing has advanced to the point where it's easier to close genomes in-house with long-read MinION sequencing. Using a combination of long-read sequencing and short read Illumina sequencing, we generated an identical extra-chromosomal circular contig in *Microcystis aeruginosa* PCC 7806 and its mutant, *Microcystis aeruginosa* PCC 7806 *ΔmcyB* , despite their separation over two decades ago [1]. We assessed the functions of the plasmid using Alphafold2 models, looked at its transcriptional activity i*n vitro* and *in situ*, and assessed whether other *Microcystis* strains harbor the plasmid. Additionally, we looked at another one of our *Microcystis* strains that we re-sequenced, to see if it also harbored conserved plasmids. Doing so lead us to find that *Microcystis aeruginosa* NIES-298, a strain often used for *Microcystis* phage Ma-LMM01 studies [27], may harbor a conserved, phage-like plasmid, and another plasmid which may putatively contribute to chromosomal rearrangements. Altogether, we discuss the conserved nature of *Microcystis* plasmids, their functionality, and the overlap of host, plasmid and phage genomes.

Methods

Plasmid assembly and annotation

Originally, we re-sequenced the genomes for *Microcystis aeruginosa* PCC7806, and *Microcystis aeruginosa* PCC7806 Δ*mcyB* [28]. Re-sequencing these two strains led to the discovery of an additional circular contig beyond the chromsome (plasmids). In addition to *Microcystis aeruginosa* PCC7806, we also re-sequenced *Microcystis aeruginosa* NIES-298. Of the NIES-298 strains sequenced, one, we refer to as "host" was capable of being infected by the *Microcystis* phage Ma-LMM01, whereas the other three mutants "B", "C" and "D" were incapable of infection by Ma-LMM01. All genome/plasmid sequences generated from our lab were extracted, assembled with Unicycler (v.0.4.9b) [29], and annotated by NCBI's prokaryotic Genome Annotation Pipeline (PGAP) [30] according to the methods of [28].

Plasmid copy number determination

To determine plasmid copy number, we used Illumina short read sequencing coverage of the mutant and wildtype genome. Before mapping the sequencing reads to the genomes, the reads were paired and trimmed using CLC genomics workbench (v.21.0.4). For genome mapping, the paired trimmed reads from *Microcystis aeruginosa* PCC 7806 Δ*mcyB* (5,103,923 bp) and *Microcystis aeruginosa* PCC 7806 wildtype (5,096,229 bp) were mapped to their respective genomes. Reads from the PCC 7806 wildtype and Δ*mcyB* were then mapped to the plasmid (8505bp). Average coverage of the genome and the plasmid were calculated for each sequencing set. To get the copy number, the average coverage of the plasmid was divided by the average coverage of each respective genome (PCC 7806 wildtype and Δ*mcyB).*

RNA-seq analysis of the plasmid

To see if the plasmid is actively expressed in both Δ*mcyB* and the PCC 7806 wildtype, we used transcriptomic libraries generated from a previous chemostat temperature experiment (growth at 19°C or 26°C) done in our lab [8]. We used these transcriptomic libraries since they were generated from axenic monoculture chemostats

for Δ*mcyB* and the PCC 7806 wildtype. Both the Δ*mcyB* and PCC 7806 wildtype chemostats had an n of 2. All transcriptomic libraries were paired and trimmed for quality using CLC Genomics (v.21.0.4) according to the methods of [8]. To get an accurate portrait of plasmid gene expression in comparison to the PCC 7806 genome, prepared reads were mapped to a concatenated assembly of the *Microcystis aeruginosa* PCC 7806 wildtype genome (accession CP155078), and the plasmid using CLC Genomics workbench (v. 21). The genome and plasmid were both annotated using NCBI's prokaryotic genome annotation pipeline (PGAP) [30]. Mapping parameters and *in silico* ribosomal reduction for the transcriptomic reads are the same as those published in [8]. Briefly, we did stringent read mappings, using length fractions of 0.9, and similarity fractions of 0.9. Maximum number of hits allowed for a read was set to 10, and all reads were normalized by TPM, where paired reads were not counted as two. GraphPad Prism (v.10.2.2) was used to visualize and compare gene expression.

RNA-seq analysis of the PCC 7806 and plasmid for Lake Erie metatranscriptomes

Lake Erie microcosm metatranscriptomes were previously generated by Pound et al. [31]. In this study, we used metatranscriptomes from the control (no treatment), nitrogen amended (180 μM as KNO_3), and phosphorus amended (1μM as KH_2PO_4 : K_2HPO_4) treatments [31]. Each treatment had an n of 3. All libraries were trimmed, paired, and had *in silico* ribosomal reduction done according to the methods of Martin et al. [32]. As a reference genome to map the reads, we used the same concatenated assembly of the PCC 7806 wildtype genome (accession CP155078) and the plasmid as we did for mapping the chemostat-generated transcriptomes. Transcriptomic reads were mapped as described above, using stringent mapping parameters (0.9 length, 0.9 similarity fractions). Maximum reads number of reads for a hits for a read was set to 30, and all reads were normalized by TPM, and paired reads were not counted as two.

RNA-seq analysis of the NIES-298 P1 and P2 Plasmids

As *M.aeruginosa* NIES-298 is often used in phage infection research, we wanted to look at expression of the P1 and P2 plasmid in response to phage infection. To look at gene expression during phage infection, we used transcriptomes generated by Morimoto

et al. [33]. Transcriptomes were generated at hours T0, T1,T3 and T6. We chose the 3 hour infected and uninfected transcriptomes, since they served as the halfway point to when the viral latent period begins (6-8hr) [33]. Raw paired reads were downloaded off NCBI SRA database (SAMD00093588 3hr control, SAMD00093592 3hr infected). Transcriptomes for 3 hr control and 3 hr infected had an n of 1. Reads were uploaded into CLC Genomics workbench, and trimmed for quality. As the P1 and P2 plasmids have no rRNA, we did not do an i*n silico* ribosomal reduction on the reads. Reads were mapped to only the either the P1 or P2 plasmid gene sequences, as determined by PGAP annotation software [34]. Mapping parameters were set to 0.95 similarity score, and 0.9 length fraction score. A maximum of 30 hits per read was allowed. As differential expression analysis requires an n>1, we just looked for genes that had at least doubled the amount of TPM and reads mapped (or halved expression) between the infected transcriptome library and uninfected transcriptome library.

Finding Viral and Plasmid Contigs in the Lake Erie Co-Assembly & RNA seq

A co-assembly was created from the Lake Erie microcosm experiments, as described in [32]. Briefly, reads were trimmed for quality using CLC Genomics workbench (v.20.0.4), and an *in silico* ribosomal reduction was done using SortMeRNA (c. 4.2.0). All non-ribosomal reads were assembled into a co-assembly using MegaHit (v.1.2.9). To determine which contigs were viral or plasmid, we ran the co-assembly through geNomad [35]. From there, virus and plasmid contig sequence lists (had viral or plasmid scores > 0.70) were annotated using the .tsv files (coverted to .gff) from the geNomad output. Annotated plasmid and virus sequence lists were then concatenated into a single file, and all contigs were size selected to be ≥1000 bp. From there, the metatranscriptomic libraries we used for mapping to the PCC 7806 genome and plasmid (control, nitrate, and phosphate) were mapped to the viral and plasmid contigs using CLC genomic workbench. For read mapping, we used a length fraction of 0.9, a similarity fraction of 0.9, and maximum hits per read was set to 30. All reads were normalized by TPM, and paired reads were not counted as two. To get an understanding of the ratio of total gene expression for virus contigs versus plasmid contigs, total TPM was calculated

for all virus genes and all plasmid genes. Total TPM for virus genes and total TPM for plasmid contigs was then divided by the total amount of viral (8051) or plasmid (3651) genes, respectively. We did this for all genes in the assembly. As a more conservative approach, we also did separate calculations looking at contigs with direct terminal repeats (DTR's) or inverted terminal repeats (ITR's) at the ends (no contigs had ITR's), signifying a putatively closed viral/plasmid sequence.

Presence and Gene flow of the PCC 7806 plasmid in Microcystis

Unless noted, all search parameters were kept as default. A nucleotide BLAST search of the full PCC 7806 plasmid sequence (8.5kb), was conducted against the whole genome shotgun sequencing database (WGS), narrowed down to the *Microcystis* database (taxid:1125 *)* (341 scaffold-level genomes as of access data 04/2024). From there we recorded the top 4 hits, because all hits after that were <32% query cover, and <90% identity.

To see if any gene flow occurred between the plasmid and the PCC 7806 genome, we did a nucleotide blast search of the PCC 7806 plasmid against the genomes for *Microcystis aeruginosa* PCC 7806 wildtype and *Microcystis aeruginosa* PCC 7806 Δ*mcyB*.

Determining if conserved plasmids are in other Microcystis spp.

As our lab had recently re-sequenced *Microcystis aeruginosa* NIES 298 on four separate occasions, which also has a plasmid sequence on NCBI (accession CP046059, uploaded in 2020), we decided to compare the plasmids between our NIES-298 sequencing runs, and the NIES-298 P1 plasmid (accession CP046059) on Genbank. We also investigated scaffold-level *Microcystis aeruginosa* NIES-298 assemblies for contigs that matched the NIES-298 P1 plasmid. To find the NIES-298 P1 plasmid in scaffold-level *Microcystis* genomes, we used the NCBI *Microcystis* WGS database (taxid:1125). A BLASTN megablast (highly similar sequences) search was done. Contigs from genomes that had 100% query and 100% identity were considered to be identical.

We were also interested in seeing how common *Microcystis* plasmids are in

general, and if any of the other *Microcystis* plasmids (see appendix sheet 4.1A) on NCBI were conserved or widespread. To investigate this, we did a nucleotide search (all settings left default) for all 10 *Microcystis* plasmid nucleotide sequences on NCBI against the *Microcystis* WGS database (taxid:1125). We recorded any contigs that had high similarity (>80 % query, >90% identity) to any of the *Microcystis* plasmids in Table S1. BLASTN searches were visualized against the annotated plasmid sequences using Proksee [36].

Finding plasmid-genome interactions

We used all the *Microcystis* plasmids (Appendix sheet 4.1A) as search queries for JGI's CRISPR BLAST tool. Viral spacer sequences were searched. An e-value of E-05 was used, all other settings were kept as default. Only BLAST hits with 100% identity to a CRISPR spacer sequence were considered as a possible plasmid-host interaction.

Finding Phage-Plasmids, and Phage-Plasmid-Genome interactions

All *Microcystis* plasmids used in this study were run through geNomad [35] to get plasmid, virus and chromosome scores. Plasmids with plasmid scores <0.7 (NIES-298 P1 and NIES-2547) were run through NCBI nucleotide BLAST (slightly similar sequences, all other settings left as default) to look at gene distribution.

We were interested in any possible gene flow between the NIES-298 P1 and P2 plasmids with the NIES-298 host genome, as well as the infective *Microcystis* phage, Ma-LMM01 [27]. To look at this, we did a nucleotide blast (e-value set to E-05, all other settings left default) of the Ma-LMM01 genome (NC_008562) and NIES-298 genome (our re-sequenced genome was use) against the P1 or P2 plasmids. We also did this for another plasmid, from the *M.aeruginosa* NIES-2481 genome, (NIES-2481 P1plasmid), after noticing it had CRISPR-cas repeats and proteins on it.

Through manual inspection of plasmid annotations, we noticed three plasmids, *Microcystis aeruginosa* FD4 P1 plasmid, *Microcystis aeruginosa* NIES-2481 P1, and the NIES-298 P2 plasmid, all had multiple transposases present. As transposases are responsible for genomic rearrangements, we did a nucleotide BLAST search (e value set to E-05, all other settings left as default) for each of these plasmids against their

respective host genomes.

Protein structure models and protein structure homology

Colabfold, which combines MMseqs2 with Alphafold2 or RoseTTAFold [37], was used to generate protein models of all the genes for the *Microcystis aeruginosa* PCC 7806 plasmid. We also used this tool to model some genes from the *Microcystis aeruginosa* NIES-298 P1 plasmid that had multiple virus BLAST hits, and a gene located in the PCC 7806 genome (pgaptmp_004807) which had similarity to PCC 7806 plasmid gene 5. Protein input sequences were derived from PGAP annotations for all the genes. For all runs, PDB100 template was used. Number of recycles was set to 6, number of structures to relax was set to 5. All other settings were kept as default. To visualize protein structures, NCBI iCn3D 3D structure viewer was used [38, 39]. To find putative predicted functions for the PCC 7806 plasmid genes and NIES-298 P1 plasmid genes that had no homology to known protein domains on PFAM or HMMER [40], we ran our AlphaFold predicted structures through DALI [41]. The AlphaFold models run through DALI were all the relaxed fit, rank 1 structure against the whole PDB database. Structures with pLDDT values >70 were considered for protein structure predictions. The top result (highest Z score) was recorded for each protein model, and all Z-scores >2 were considered significant [41].

To align protein structures, we used TM-align for pairwise alignments [42] (version 20190822) . TM-scores for pairwise alignments were normalized by the length of the reference protein.

Results

Microcystis aeruginosa *PCC7806 has a conserved plasmid that hasn't changed in over 20 years*

Genome assembly uncovered an 8.5kb circular contig in both the *Microcystis aeruginosa* PCC7806 wildtype and *Microcystis aeruginosa* PCC7806 Δ*mcyB* genome, two strains that were separated in 1997 [1]. A BLASTN search of the 8.5kb contig showed that the plasmid was partially present (98% query, 100 % nucleotide identity) on

a 8,373-nt long contig (contig272, accession AM778902) from a scaffold-level genome of *Microcystis aeruginosa* PCC 7806 from 2007 (appendix sheet 4.1B) [43].

Plasmid structure and similarity to the chromosome

For the PCC 7806 plasmid, nine genes were annotated via PGAP (Figure 4.1). The original annotations of these genes can be seen in Table S3. To get a better idea of the protein products of all the nine genes, the amino acid sequences were run through PFAM. Genes 6-9 all had known protein domains, however, for genes 1-5 the PFAM search had no domain hits (Appendix sheet 4.1C). For a more thorough understanding of the plasmid's cellular functions, we generated protein models for all the plasmid genes. Using DALI [41] we were able to find putative functions/protein homologs for 4 of the 5 hypothetical plasmid genes (Appendix sheet 4.1C). Gene 2 had no protein structure homology matched to it. For gene 4, the best DALI hit (z=5.6, RMSD = 8.1) was to a conserved putative transcription factor (pdb id: 1y9b-A). For gene 5, the best DALI hit (z= 6.7, RMSD= 17) was to a structural cell protein (pdb id: 8ahl-A). All other predicted homologous protein structures can be seen in Table S4. Genes 6-9 already had known protein domains based on PFAM searches (Appendix sheet 4.1C), and the DALI outputs were similar to the PFAM domains (Appendix sheet 4.1C). One exception to this was gene 9, which came up as a RNA helicase, however, DnaG has been shown to be a species of RNA polymerase [44]. *In silico* prediction using OriFinder determined a putative origin of replication occurred from residues 8248 to 370 on the plasmid (Figure 1).

To see if any genetic transfer occurred between the PCC7806 plasmid, and PCC 7806 genome, we did a nucleotide BLAST search, finding gene 5 had some sequence similarity to a gene in the chromosome (locus tag pgaptmp_4807) (Appendix Fig. 4.7). The chromosomal gene was annotated as a hypothetical protein, however, PFAM showed it had a tropomyosin superfamily domain, and a HMMer homology search top hit was to a chromosomal ATPase (e-value: 2.6E-17). DALI search of the protein also showed closest structural homology to a centriole protein (z-score 7.9, RMSD=4) (Appendix Sheet 4.1E). After aligning the two protein structures, they had a TM-align score of

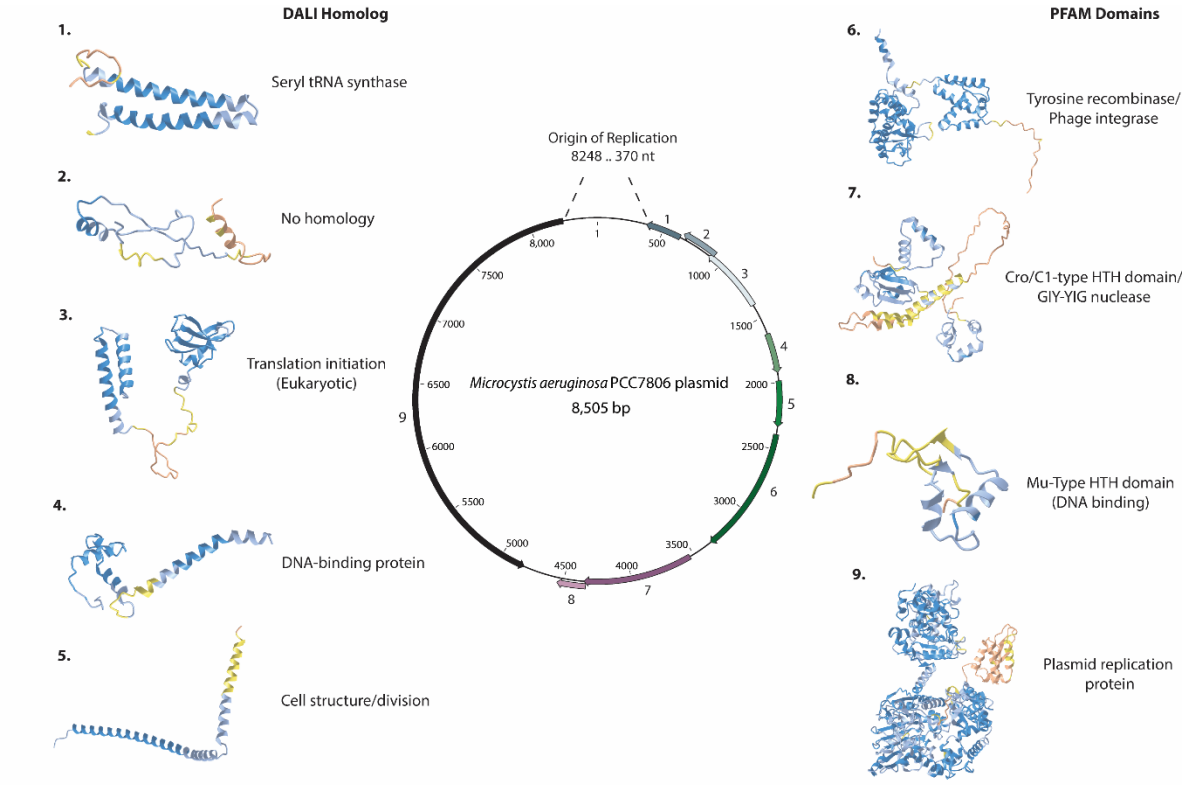

Figure 4.1 Plasmid in *M. aeruginosa* PCC 7806

Plasmid found in *Microcystis aeruginosa* PCC 7806 wildtype and Δ*mcyB*. The plasmid is 8,505bp in size, and has nine genes present on it. Genes 1-6 were annotated as hypothetical proteins. Through PFAM protein domain searches and protein structure homology searches (DALI), we generated putative generalized functions of the unknown proteins.

0.62225 with an RMSD of 1.65.

PCC 7806 plasmid copy number

Mapping the Illumina sequencing reads to the PCC7806 genomes and PCC7806 plasmid resulted in an approximate copy number of 14.77 per chromosome for the Δ*mcyB* genome and 15.08 per chromosome for the wildtype PCC 7806 genome.

The PCC 7806 plasmid is found in other scaffold-level genome assemblies

Searching the full PCC 7806 plasmid against the *Microcystis* WGS database revealed three contig hits from geographically distant *Microcystis* strains, which were determined as plasmid by geNomad (Appendix sheet 4.1E,F). The other contig with portion of the plasmid present was considered to be chromosomal (Appendix sheet 4.1G). One of the plasmid hits, in a scaffold-level genome from *Microcystis aeruginosa* Ma_AC_P_19900807_S300 (accession SFBP01000158) was 100% identical to the PCC 7806 plasmid, aside from the DTR's at the contig ends (signifies a putatively closed contig). The second closest hit was from *Microcystis aeruginosa* PMC 728.11 (JADCRC010000097), isolated from France, which covered 81% of the PCC7806 plasmid sequence, and was 99.8% identical.

The PCC 7806 plasmid is expressed in lab cultures and in situ

To assess how transcriptionally active the plasmid is, we looked at a transcriptomic dataset generated in our lab for *Microcystis aeruginosa* PCC7806 and Δ*mcyB* under warm or cold temperature regimes in chemostats [8]. We found that regardless of optimal or cold growth conditions, that the two most highly expressed plasmid genes in both strains were genes 4 and 5 (Figure 4.2A). For Δ*mcyB,* TPM for gene 4 was 84.33 at 19°C, and 137.94 at 26°C. The TPM for gene 5 was similar, being 86.67 at 19°C and 133.22 at 26°C. For the wildtype, gene 4 had an average TPM of 130.25 at 19°C and 113.74 at 26°C. Average TPM for gene 5 was similar, and was 134.35 at 19°C and 133.87 at 26°C. To get an idea of what a "normal" TPM value was, we looked at gene expression of *dnaA*, which encodes the chromosome replication protein, an essential component of cell division. Average TPM for *dnaA* was similar

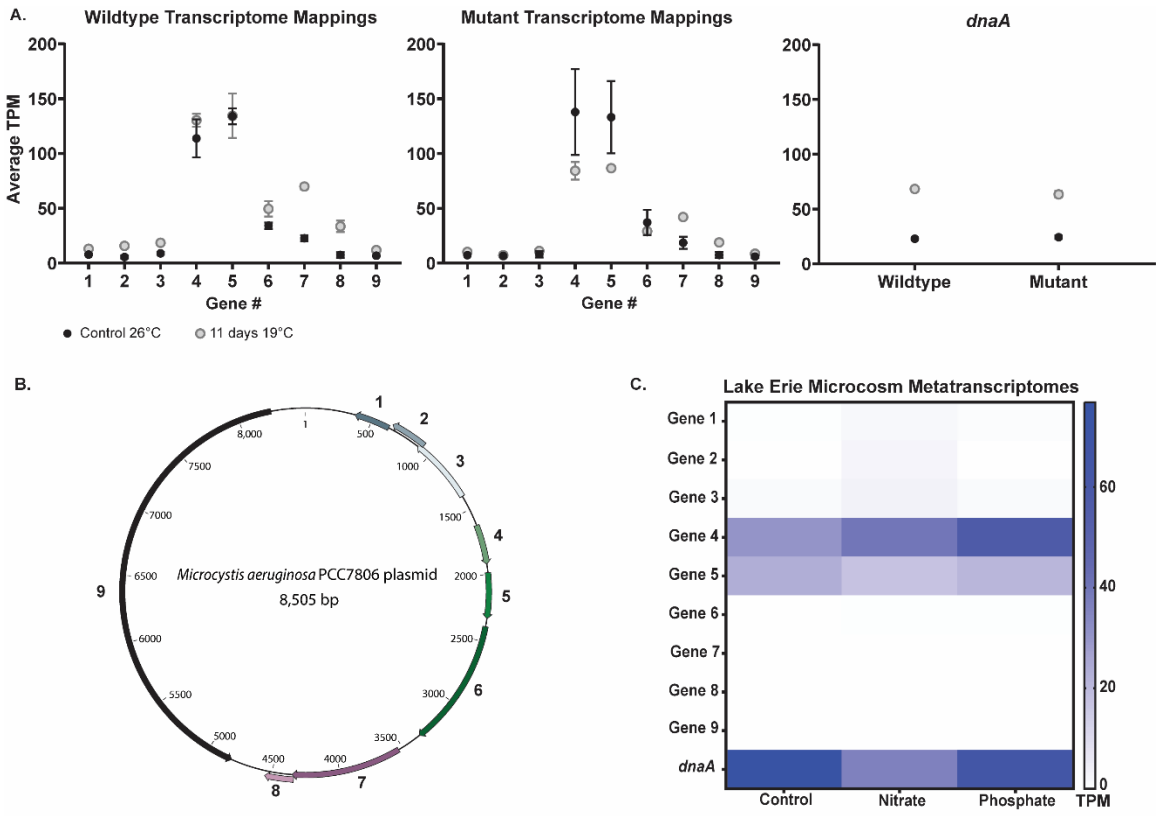

Figure 4.2 Expression of the plasmid *in vitro* and *in situ*

A) Average TPM of the nine plasmid genes from PCC 7806 wildtype transcriptome read mappings (far left), and average TPM of the nine plasmid genes from Δ*mcyB* transcriptome read mapping (middle. As a comparison, we are showing the average TPM of chromosome replication protein, *dnaA*, from the PCC 7806 wildtype genome from the same wildtype and mutant RNA-seq libraries. The transcriptomes were previously generated during a chemostat temperature experiment, during growth at 19ºC and 26ºC. The *Microcystis aeruginosa* PCC 7806 wildtype genome assembly was combined with the PCC 7806 plasmid, to which transcriptomic libraries were mapped. B) Image of the PCC 7806 plasmid with PGAP generated annotations. Genes are numbered 1-9, which correspond to the gene numbers in A and C. C) Heatmap showing the average TPM of the nine PCC 7806 plasmid genes based on mappings from Lake Erie microcosm metatranscriptomes generated by Pound et al. 2022. The average TPM of the plasmids genes was compared to the average TPM of *dnaA* in the *Microcystis aeruginosa* PCC 7806 genome.

between the wildtype and Δ*mcyB*, averaging around 22-24 at 26°C and 63-68 at 19°C (Figure 4.2A).

The Lake Erie microcosm metatranscriptomes also showed evidence of the plasmid, or genes very similar to the plasmid, being expressed in cyanobacteria sampled rom Lake Erie (Figure 4.2C). Again, it appeared that genes 4 and 5 were the most highly expressed genes *in situ*, regardless of the nutrient amended N or P treatments (Figure 4. 2C). For the nitrate and phosphate amendments, the average TPM of gene 4 was comparable to the average TPM of the PCC 7806 genome *dnaA.* For the nitrate treatment, the average TPM of gene 4 was 40.07, whereas *dnaA* was 36.31. For the phosphorus treatment, the average TPM of gene 4 was 57.25 and the average TPM of *dnaA* was 64.64. By comparison, for the non-nutrient amended control group, the expression of *dnaA* was 76.84, whereas gene 4 average TPM was 30.96 (Figure 4.2C).

Proportion of total plasmid versus virus genes and transcripts in Lake Erie

By calculating a ratio of total viral/plasmid TPM over total viral/plasmid genes, we found that for every plasmid gene, average TPM/gene was 176.875 for the control treatment, 180.806 for nitrogen treatments, and 194.858 for phosphate treatments (Figure 4.3B). By contrast, for every viral gene, average TPM/gene was 35.9351 for control, 35.4783 for nitrate, and 28.5824 for phosphate treatments. When looking at the putatively closed viral/plasmid sequences only, we saw much more comparable TPM's per viral/plasmid gene. For the closed plasmid contigs, average TPM/gene was 329.344 for control, 264.904 for nitrate, and 518.761 for phosphate treatments (Figure 4.3D). For the closed viral genomes, TPM/gene was 409.675 for control, 271.219 for nitrate, and 364.414 for phosphate treatments (Figure 4.3D). Nucleotide BLAST search results for the six closed plasmid contigs revealed that 3/6 contigs had significant hits to cyanobacterial genomes/*Microcystis* plasmids, the other 3 contigs had no significant hits to any organism (Appendix sheet 4.1H).

Other Microcystis *may also harbor conserved plasmids*

As the plasmid in *Microcystis aeruginosa* PCC7806 had not changed in >20 years we were curious if other *Microcystis* plasmids displayed similar sequence conservation in

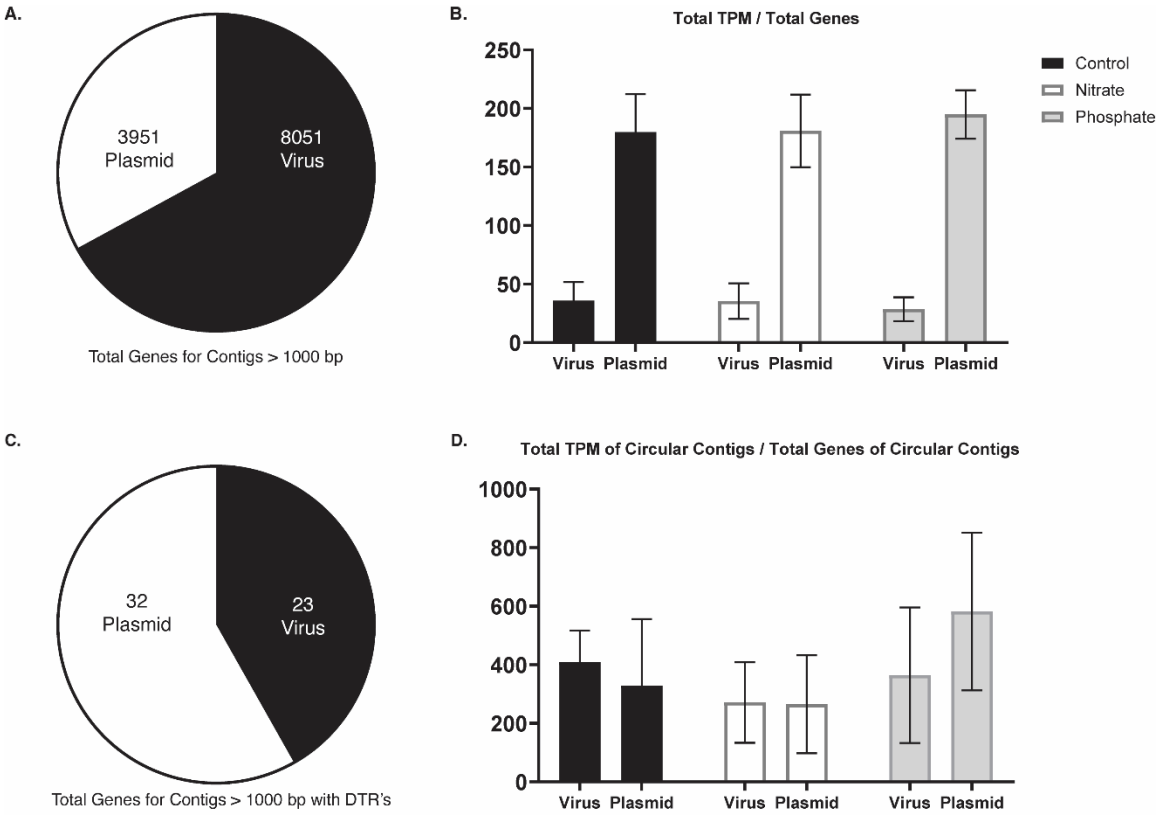

Figure 4.3 Average TPM/gene for viral and plasmid contigs from Lake Erie microcosm metatranscriptomes

A) Putative viral and plasmid contigs were determined with geNomad and pulled out of the Lake Erie microcosm co-assembly. This represents the amount of total genes present on ”viral” or “plasmid” contigs, as determined by genomad, after contigs were size selected to be 1000 bp or greater. B) The ratio of total TPM for all viral or all plasmid contig genes, divided by the total viral (8051) or plasmid (3951) genes for control, nitrate or phosphorus treatments. C) The total amount of genes for viral or plasmid contigs that were >1000 bp, and were also determined to be closed sequences based on the presence of direct terminal repeats at the end of the contigs. D) The ratio of total TPM for the closed viral or plasmid contigs, divided by the the total amount of genes present in all the closed contigs (23 viral genes and 32 plasmid genes).

their host strains. Based on four separate sequencing runs, and the scaffold level *Microcystis aeruginosa* NIES-298 genomes on NCBI, it appears that NIES-298 may also have a conserved plasmid 18,380 bp in size (Appendix Figure 4.8). Some of the sequencing runs in our lab were ~ a year apart, and the sequences for the P1 plasmid had 100% nucleotide identity to one another (though there were differences in whether genes were on + or – strand), despite the host genome undergoing genomic rearrangements during that time span (unpublished, Smith et al.). When comparing the P1 plasmid in our sequencing runs to contigs in NIES-298 draft assemblies, one of which was from 2017, sequenced in Japan, and one from 2019, sequenced in Korea, we found identical contigs in these assemblies (aside from DTR's present on the ends of the contigs-signifying a circular plasmid) (Appendix Figure 4.8).

To see how widespread plasmids may be in *Microcystis*, and if any of the other *Microcystis* plasmids were conserved, we searched the *Microcystis* WGS database (Appendix Sheet 4.1I). We did not get any results that had 100 % query coverage and 100% identity, which would signify identical plasmids. However, many of the *Microcystis* plasmids did have blast outputs with query coverage ≥80 % with % identity scores ≥80% (Appendix Sheet 4.1I). Contigs with high similarity scores such as these may be homologous plasmids, especially as many of the homologous contigs were similar in length.

What is the mechanism of conservation for Microcystis *plasmids ?*

Of our putatively conserved plasmids, aligning the full sequences from the PCC 7806 P1 plasmid and the NIES-298 P1 plasmid does not reveal much sequence similarity between the two plasmids. However, the two plasmids possess genes which encode for a tyrosine recombinase/phage integrase. While the tyrosine recombinases shared no sequence similarities, aligning the protein structures of the tyrosine recombinases between the three plasmids produced a TM-score of 0.62680 (out of 1) with a RMSD of 3.70 (Appendix Figure 4.9).

PCC 7806 plasmid sequences are found in other Microcystis *CRISPR spacer sequences*

To see if other *Microcystis* genomes encountered the PCC7806 plasmid, we decided to look at CRISPR spacer sequences (Appendix sheet 4.1J). In fully assembled genomes, two strains, *Microcystis aeruginosa* NIES-2481, and Microcystis aeruginosa NIES-2549 had CRISPR spacer sequences with a 35-nt long sequence 5'-GCCCAACTTCGAGGAAAAGATCGACGCGGCGATCGA-3', which matches a part of gene 4 in the *Microcystis aeruginosa* PCC7806 plasmid. Both NIES-2481 and NIES-2549 were isolated from Kasumigaura Lake, Japan, and Lake Kawaguchi, Japan, respectfully.

Other Microcystis spp. *plasmid sequences are widespread in CRISPR spacer regions*

Since we saw putative interactions of the PCC 7806 plasmid with two *Microcystis* CRISPR sequences from Japan, we decided to investigate this for all the available NCBI Genbank *Microcystis* plasmids (Appendix sheet 4.1J). Out of the 12 plasmids we searched, 10 had *Microcystis* CRISPR hits with 100% identity and lengths ≥30 nucleotides. Of these 10, the plasmid with the highest number of CRISPR spacer hits was the P1 plasmid from *Microcystis aeruginosa* NIES-88, which matched CRISPR spacers in 18 different strains of *Microcystis*. Second to that was the plasmid from *Microcystis aeruginosa* NIES-2481, which had hits to 14 different *Microcystis* CRISPR spacers. The P1 plasmid from NIES-298 had the third highest amount of hits, being found in 10 different *Microcystis* CRISPR spacers.

Microcystis *putative phage-plasmids*

One of the most unexpected findings we came across was the similarity of the NIES-298 P1 plasmid to bacteriophage. The geNomad score for this plasmid was 0.283 for plasmid and 0.6876 for phage (Appendix Sheet 4.1K). Our nucleotide BLAST results further confirmed this score, as the plasmid had 42% query cover and 87.91% identity to a bacteriophage contig (MT840190) (Figure 4.4, Appendix Sheet 4.1L). Many other genes on the plasmid also had additional BLAST hits to several other *Microcystis* phage

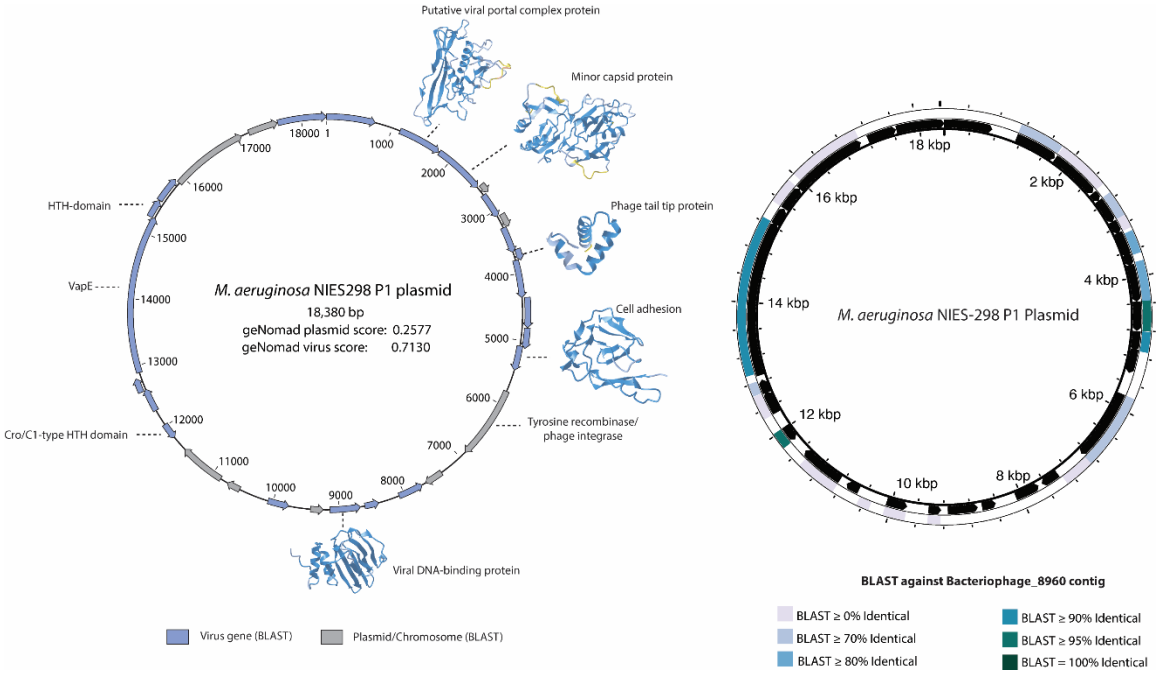

Figure 4.4 *M.aeruginosa* NIES-298 P1 Plasmid, protein structure homologs, BLAST output against Bacteriophage_8960 contig

On the left is a plasmid map of the NIES-298 plasmid, with genes color coded based on a nucleotide BLAST search result of each gene. If a bacteriophage was the top hit for the gene, it is colored blue. If the top hit for the gene was not bacteriophage, then the gene is colored grey. B) Many of the genes for the NIES-298 P1 plasmid had sequence similarity to a Bacteriophage_8960 contig. We did a BLASTx search of the NIES-298 P1 plasmid against the Bacteriphage_8960 contig. The CDS regions for the NIES-298 are in black in the inner ring. The outer ring shows the BLASTX hits against the Bacteriophage_8960 contig, ring colors correspond to the % identity of the region.

(Appendix Figure 4.10). Using protein models, we determined some of the phage-like hypothetical genes had homology to known viral protein structures, such as a phage tail tip protein, a minor capsid protein, a DNA-binding protein, and a portal protein (although this protein protein had the same z-score as a dehydrogenase protein, but second and third place hits were also to a viral portal protein/ a type IV secretion structural protein) (Figure 4.5, Sheet 4.1M). Most of the other *Microcystis* plasmids had geNomad determined plasmid scores > 0.75, aside from the NIES-2549 P1 plasmid, which had a plasmid score of ~0.65 and a virus score of ~0.3 (Sheet 4.1K). However, most of the nucleotide BLAST searches for the NIES-2549 plasmid genes did not have any significant hits beyond the plasmid (Sheet 4.1N).

Plasmid-Genome & Plasmid-Genome-Phage interactions

Through our BLAST searches of the NIES-298 genome and Ma-LMM01 genome against the NIES-298 P1-plasmid, we found an area of homology among the three (Figure 4.5,Appendix Fig. 4.11). For both the NIES-298 P1 plasmid and Ma-LMM01, this area was similar to a CRISPR repeat spacer region in the NIES-298 genome, where the repeats were similar to genomic CRISPR repeats (Figure 4.5, Appendix Fig. 4.11). Additionally, a tyrosine recombinase/phage integrase on the P1 plasmid also shared sequence similarity to an area adjacent to a CRISPR spacer repeat region in the NIES-298 genome (Appendix Figure 4.12).

While the P1 plasmid for NIES-298 remained the same between our sequencing runs, the NIES-298 P2 plasmid we found in our sequencing datasets did change slightly between our sequencing runs. This plasmid was not previously reported on Genbank, and is located in the assembled genome for the published *M.aeruginosa* NIES-298 genome.The NIES-298 P2 plasmid has abundant transposases, some of which varied slightly in nucleotide sequence between our four sequencing datasets (Figure 4.6, Appendix Figures 4.13, 4.14). In one instance, in one of our re-sequenced NIES-298 closed genomes (“Mutant B”), an identical transposon which was present on the P2 plasmid (and in one copy in the “host” genome), had duplicated in the genome of “mutant B”, where it inserted in a methyltransferase gene (Appendix Figure 4.13). Conversely, on

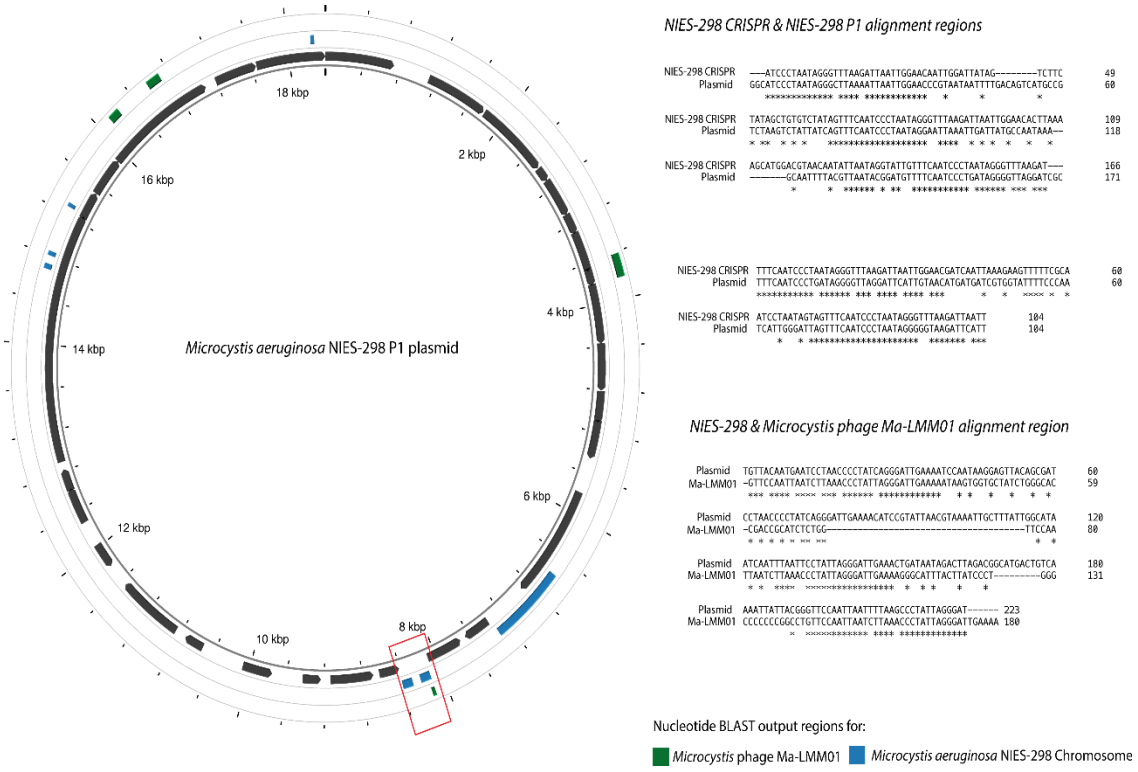

Figure 4.5 Alignments of the *M.aeruginosa* NIES-298 genome and *Microcystis Phage* LMM01 against the NIES-298 P1 plasmid

Nucleotide BLAST outputs against the NIES-298 P1 Plasmid for the Microcystis phage Ma-LMM01 (green ring regions) and the Microcystis aeruginosa NIES-298 genome (blue ring regions). The only region of similarity among the three sequences was and intergenic region on the NIES-298 P1 plasmid. This region had spacer sequences similar to the NIES-298 CRISPR repeats. This was also seen for Ma-LMM01 (Appendix Figure 4.11).

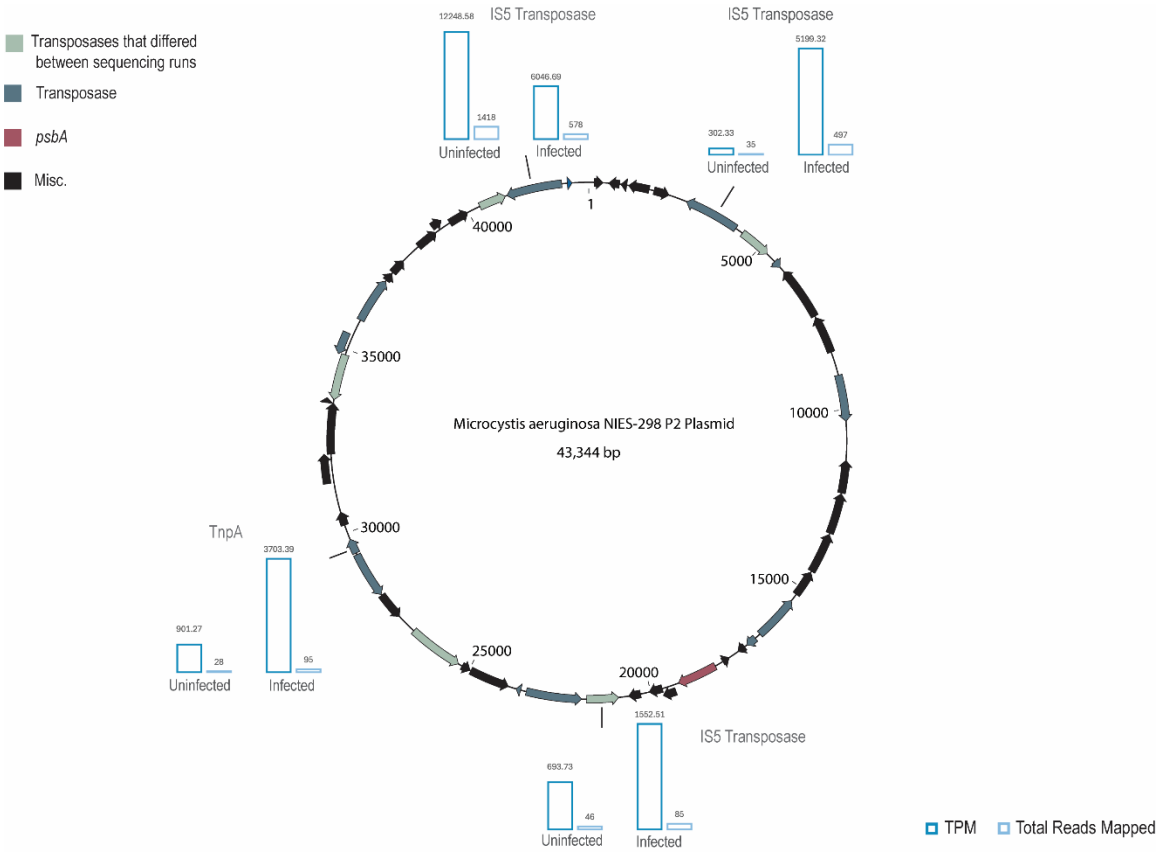

Figure 4.6 NIES-298 P2 Plasmid gene expression during Ma-LMM01 phage infection The P2 plasmid from *M.aeruginosa* NIES-298. Using transcriptomes generated by Yoshida et al. during a Microcystis phage Ma-LMM01 infection, at 3 hours post infection, we found many of the transposases present on the plasmid increase in transcript abundance in response to phage infection.

the “Mutant B” plasmid, this transposase amino acid sequence had changed from the host plasmid transposase (Appendix Figure 4.14).

As NIES-298 is often used in phage infection research, we were curious about. whether the NIES-298 P1 or P2 plasmid play a relevant role during phage infection. Using transcriptome libraries previously generated from a Ma-LMM01 infection by Morimoto et al. [33], we looked at expression of the P1 and P2 plasmid after 3 hours of infection. For the P1 plasmid, only two genes had decreased expression by >1/2 in the infected group compared to the uninfected group (Figure 4.15, Appendix sheets 4.1O, P). For the P2 plasmid, most gene expression changes came from transposable elements, most of which increased in expression in response to phage infection at 3 hours (Figure 4.6, Appendix sheets 4.1Q, R). However, we want to note that some of the transposases which had a change in transcript abundance in response to phage infection are very similar (98-100%) to transposases present on the chromosome. Due to the repetitive nature of transposases, this makes it difficult to definitively conclude mapped reads were plasmid in origin. Regardless, it does still show that transcripts from IS5 transposases, or very similar IS5 transposases in the genome, have a response to phage infection at 3 hours.

While we found that many of the transposases on the NIES-298 P2 plasmid were identical or nearly identical to transposases in the NIES-298 genome, in general, the NIES-298 P2 plasmid is very similar to the NIES-298 chromsome (Figure 4.16, Appendix Sheet 4.1S). This is also the case for the *M.aeruginosa* FD4 P1 Plasmid, and the NIES-2481 plasmid, which also have abundant transposable elements on them and share a large amount of sequence similarity to their respective host genomes (Figures 4.17, 4.18, Appendix sheet 4.1Q). Of note, the NIES-2481 P1 plasmid has a CRISPR cas system, which only shares some sequence similarity to the NIES-2481 chromosome CRISPR spacer regions (Figure 4.18). For this reason, we decided to extend our BLAST search to *Microcystis* phage, to see if there was any putative phage-interaction with this plasmid. We found that the NIES-2481 P1 plasmid, the NIES-2481 genome, and the *Microcystis* phage Ma-LMM01 all had a *tnpB* and IS604 element, which were 98-99% similar to one another (Figure 4.19, Appendix Sheet 4.1T).

Discussion

While *Microcystis aeruginosa* PCC7806 and it's non-toxigenic mutant, *Microcystis aeruginosa* PCC7806 Δ*mcyB* have been the subject of numerous comparative studies over the last two decades, the presence of its plasmid, and the plasmid's role in cell physiology has been left unexplored. This paper highlights the conserved nature of the plasmid in PCC 7806 and shows that it is transcriptionally active throughout log-phase growth in chemostats. There was also some evidence of homologous genes being present and active in Lake Erie, based on microcosm metatranscriptome libraries [31]. Additionally, we found another geographically isolated *Microcystis* strain from Canada that harbored this identical plasmid, and some other *Microcystis* strains had plasmids very similar to the one found in PCC 7806. For these reasons, along with the plasmid being partially present on a contig from a PCC 7806 draft genome sequenced in 2007 (France) [43], we do not think the plasmid was picked up in our lab and was most likely present when the Δ*mcyB* strain was created and separated from the wildtype in 1997 by Dittmann et al. [1]. CRISPR spacer BLAST results also showed evidence of other geographically isolated *Microcystis* genomes from Japan encountering the plasmid, which gives more credence to the idea of the plasmid being present (and possibly conserved) in other *Microcystis* strains [45].

Due to the discovery of a conserved plasmid in *Microcystis aeruginosa* PCC 7806 we wanted to explore if conserved plasmids are a common feature in *Microcystis*. While we were limited by the availability of known *Microcystis* plasmid sequences and sequenced *Microcystis* genomes, we concluded based on our results that *Microcystis aeruginosa* NIES-298 may also have a putatively conserved plasmid. This is surprising, as *Microcystis* has been known to have a plastic genome with highly abundant transposases, which could make them prone to genetic rearrangements [43]. Additionally, through separate studies (currently unpublished), we have found that while these plasmids have remained unchanged in PCC 7806 and NIES-298, the respective host genomes have undergone chromosome rearrangements throughout our sequencing datasets. We also saw that many of the *Microcystis* plasmids on NCBI had BLAST hits to contigs of similar size with >80% coverage, >90% identity to them, which could be

homologous plasmids. As sequencing datasets increase, and *Microcystis* plasmids are investigated further, we may find many more plasmids that are conserved in nature, or have conserved backbones.

While the PCC 7806 and NIES-298 P1 plasmids did not have significant sequence similarity, they both possessed tyrosine recombinases with phage integrase domains. In general, it seems for *Microcystis* plasmids >5000 bp, that tyrosine recombinases are a common feature (6/8 *Microcystis* plasmids >5000 bp had tyrosine recombinases). Aligning the tyrosine recombinases from the PCC 7806 plasmid and NIES-298 P1 plasmids also showed that while they share very little sequence similarity, structurally, they have a 63% alignment overlap. Toxin-antitoxin systems and tyrosine recombinases have been shown to be involved in plasmid stability [46]. Since none of the putatively conserved plasmids had a characterized TA system present, it seems likely the tyrosine recombinases could play a role in their stability.

Ecological relevance of Microcystis spp. *plasmids and their ties to phage*

There are still many aspects of *Microcystis* bloom ecology to uncover. While the phycosphere and mobilome have been topics of interest that are more easily investigated now thanks to advancements in sequencing technologies, the role of plasmids, or phage plasmids in bloom ecology is less accounted for. Based on the Lake Erie microcosm metatranscriptomes used in this paper, it seems likely plasmids play some role in bloom ecology, as average plasmid gene expression may either exceed that of viruses or, if we are more conservative with our mappings and only look at putatively complete viral or plasmid contigs, is very similar to average viral TPM. While our metatranscriptomic analyses captured all viral and all plasmid genes regardless of taxonomy, 50% of the closed plasmid contigs were cyanobacterial in nature, whereas the other three closed plasmid contigs had no matches.

Viral impacts on *Microcystis* blooms have been well documented, and in some cases the presence of phage has been correlated with bloom collapse [23, 24, 47]. The NIES-298 P1-plasmid is very similar to a bacteriophage contig assembled from *in situ* sequencing data collected after a *Microcystis* bloom collapse in 2019 [23]. Based on our

findings with the NIES-298 P1 plasmid, there can be much overlap between "phage" genomes, and "plasmid" genomes, which calls into question if some of the "viral" impacts on blooms could be due to plasmids/ phage-plasmid hybrids. While many of the genes on the P1 plasmid are hypothetical, our protein models for some structural homology to viral tail/portal/ capsid, for these reasons we think it likely that the NIES-298 P1 plasmid is putatively a "phage-plasmid" [48]. While the transcriptomes mapped to the P1 plasmid during phage infection did not reveal many changes in gene expression for a useful interpretation, we did find that there is an intergenic region on the P1 plasmid which shares sequence similarity to the CRISPR repeat regions in the NIES-298 genome. This is also the case for the *Microcystis* phage MA-LMM01. It has recently been discovered that phage and plasmids can utilize solitary repeat units, which resemble CRISPR repeats, as an anti-CRISPR defense mechanism [49, 50]. It is also thought they may serve as a potential mechanism to integrate into CRISPR regions in the host genome [49, 51]. Currently, we cannot determine whether acquisition of CRISPR repeats and the similarity of the integrase to a CRISPR region has allowed the transition of the P1 plasmid into a stabile plasmid lifestyle. This, along with the possible interplay of the P1 plasmid and Ma-LMM01 anti-CRISPR defense systems are worthwhile as future follow up studies.

Defense mechanisms may be a shared trait among *Microcystis* phage, plasmids, and genomes, as we also saw this with the NIES-2481 plasmid, which had a CRISPR-cas defense system, as well as *tnpB* and IS607 genes which were >98% identical between the NIES-2481 plasmid, the NIES-2481 genome, and the *Microcystis* phage LMM01 (MA-LMM01). Previously, Kuno et al. found that IS607 was most likely transferred from *Microcystis aeruginosa* (strains NIES-112, NIES605, NIES90, RM6) to Ma-LMM01, using PCR amplification methods and phylogenetics [20]. As more sequencing data has become available since that time, we are adding to that story, showing the interplay between this specific plasmid, its host, and a phage genome. We are unsure if Ma-LMM01 (isolated in Japan) is capable of infecting NIES-2481 (also isolated in Japan) [52]. Ma-LMM01 is used for infection assays with *Microcystis aeruginosa* NIES-298, a strain also isolated in Japan [27]. However, in this case, a plasmid equipped with a phage-

fighting defense system brings up the question of whether acquiring this plasmid would provide added protection to strains for phage they have not previously encountered.

In the future, as more focus is put into elucidating *Microcystis* plasmids and their functions, it may be discovered that "phage-plasmids" are common, and possibly important components of *Microcystis* blooms. It also calls into question how *Microcystis* phage and plasmids interact, as horizontal gene transfer between the two, as well as host genomes, seems likely based on our data. The potential phage-plasmid was found in NIES-298, one of the few *Microcystis* strains that has an isolated infective phage, Ma-LMM01 [27]. As gene flow between plasmids and phage has been shown to be mediated by phage-plasmids, future research should begin accounting for the P1 and P2 plasmids and their role in NIES-298 [53]. Additionally, as many plasmid sequences were found in *Microcystis* CRISPR spacer regions, it shows not only how widespread plasmids are, but it also makes it hard to distinguish whether some spacer sequences are phage or plasmid derived.

Physiological relevance of Microcystis *plasmids*

While it seems evident plasmids may have *in situ* impacts on blooms, when we investigated gene expression of the PCC 7806 plasmid during growth in chemostats, it was clear that although it is a low copy-number plasmid, four of the plasmid genes (4,5,6, and 7) had comparable expression values to the chromosome replication protein, *dnaA*. For additional context, only ~12.4-14.8% of the 4875 genes in the PCC 7806 genome had TPM values >200. For gene 4, our protein model showed some structural homology to DNA-binding proteins and gene 5 had some structural homology to proteins related to the cell structure/organization/ and possibly division. Gene 5 also had some nucleotide sequence similarity to a gene with HMMer homology to a chromosome ATPase in the PCC 7806 genome. We also saw that in the Lake Erie metatranscriptomes, genes 4 and 5 were the two most highly expressed genes out of the 9. Between our transcriptomes generated from chemostat temperature treatments, and the microcosm nutrient amendments, it seems these two genes are likely co-transcribed. We also hypothesize, based on our protein models, and gene 5's similarity to a chromosome ATPase, that they

may play some role in cell division/partitioning (potentially plasmid segregation/ partitioning).

Re-sequencing the NIES-298 genome on four separate occasions also uncovered a second, previously unreported plasmid in this strain, which we refer to as "P2". This plasmid was interesting, because unlike the P1 plasmid, the P2 plasmid did not remain 100% identical in nucleotide sequence between our sequencing runs. Instead, we found that some of the transposases present on the plasmid sequences had changed. In some instances, these transposases had identical copies present in the chromosome. While we cannot be sure the transposition of the transposase came from the plasmid, as an identical transposase is also in the genome, it is an interesting observation and shows a potential HGT transfer event between plasmid and chromosome. Additionally, we also saw that transposases (*tnpA* and IS5-family) on the NIES-298 P2 plasmid, or very similar transposases in the NIES-298 genome to those on the plasmid, have different gene expression profiles in response to Ma-LMM01 phage infection at 3 hours. Therefore, it seems one NIES-298 plasmid remains identical (P1) and may have the ability to interact with the hosts CRISPR array, while the other plasmid, P2, may potentially contribute to chromosomal rearrangements, which could have impacts on gene function and host physiology (Stark et al., unpublished). Recently, it was shown that the genetic transfer of a transposable element between *Microcystis* FACHB1339 and *Microcystis* phage Mic-1 coincided with increased host resistance [54]. Therefore, it seems plausible the P2 plasmid could play a relevant role during phage infection if it is successfully contributing to transposition events, and more consideration should be paid to these plasmids and the effects they have on genomic structure and physiology.

Conclusions

This paper brings attention to *Microcystis* plasmids, and their physiological and ecological relevance. *Microcystis* physiology-based studies should begin considering the role of plasmids, if their strains harbor them. *Microcystis aeruginosa* PCC 7806 has been harboring a plasmid in culture for >25 years, which was also found in a geographically distant *Microcystis* isolate. Thus it seems likely that the PCC 7806 plasmid may serve

some sort of important role to the cell. Additionally, for in-situ based studies investigating bloom ecology, more attention should be paid to genes that are considered to be “viral”, “host”, or as we have seen from this paper “plasmid”, as there can be a great overlap between the three. This is especially evident for the putative phage-plasmid present in *M.aeruginosa* NIES-298, and similar defense systems seen in plasmid and phage (CRISPR repeats and *tnpB*). Additionally, as we saw with the NIES-298 P2 plasmid, plasmids may contribute to genomic rearrangements in their hosts, which could have physiological effects.

Acknowledgments

Thanks to Dr. Jordan Cannon, Liz Denison and Dr. Emily Chase.

References

1. Dittmann, E., et al., *Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium Microcystis aeruginosa PCC 7806.* Molecular Microbiology, 1997. **26**(4): p. 779-787.

2. Hesse, K., E. Dittmann, and T. Börner, *Consequences of impaired microcystin production for light-dependent growth and pigmentation of Microcystis aeruginosa PCC 7806.* FEMS Microbiology Ecology, 2001. **37**(1): p. 39-43.

3. Dittmann, E., et al., *Altered expression of two light-dependent genes in a microcystin-lacking mutant of Microcystis aeruginosa PCC 7806.* Microbiology, 2001. **147**(11): p. 3113-3119.

4. Zilliges, Y., et al., *An Extracellular Glycoprotein Is Implicated in Cell-Cell Contacts in the Toxic Cyanobacterium Microcystis aeruginosa PCC 7806.* Journal of Bacteriology, 2008. **190**(8): p. 2871-2879.

5. Phelan, R.R. and T.G. Downing, *A growth advantage for microcystin production by Microcystis PCC7806 under high light.* Journal of Phycology, 2011. **47**(6): p. 1241-1246.

6. Makower, A.K., et al., *Transcriptomics-Aided Dissection of the Intracellular and Extracellular Roles of Microcystin in Microcystis aeruginosa PCC 7806.* Applied and Environmental Microbiology, 2015. **81**(2): p. 544-554.

7. Barchewitz, T., et al., *Non-canonical localization of RubisCO under high-light conditions in the toxic cyanobacterium Microcystis aeruginosa PCC7806.* Environmental Microbiology, 2019. **21**(12): p. 4836-4851.

8. Stark, G.F., et al., *Microcystin aids in cold temperature acclimation: Differences between a toxic Microcystis wildtype and non-toxic mutant.* Harmful Algae, 2023. **129**: p. 102531.

9. Jähnichen, S., et al., *Impact of Inorganic Carbon Availability on Microcystin Production by Microcystis aeruginosa PCC 7806.* Applied and Environmental Microbiology, 2007. **73**(21): p. 6994-7002.

10. Rohrlack, T., et al., *Effects of cell-bound microcystins on survival and feeding of Daphnia spp.* Applied and Environmental Microbiology, 2001. **67**(8): p. 3523-9.

11. Schuurmans, J.M., Brinkmann, B.W., Makower, A.K., Dittmann, E., Huisman, J., Matthijs, H.C.P.,, *Microcystin interferes with defense against high oxidative stress in harmful cyanobacteria.* Harmful Algae, 2018. **78**. 47-55.

12. Zilliges, Y., et al., *The cyanobacterial hepatotoxin microcystin binds to proteins and increases the fitness of Microcystis under oxidative stress conditions.* PLoS ONE, 2011. **6**(3): p. e17615.

13. Hauman, J.H., *Is a Plasmid(s) Involved in the Toxicity of Microcystis Aeruginosa?* The Water Environment, 1981, Springer US. p. 97-102.

14. Schwabe, W., et al., *Plasmids in toxic and nontoxic strains of the cyanobacterium Microcystis aeruginosa.* Current Microbiology, 1988. **17**(3): p. 133-137.

15. Tominaga, H., et al., *Structure and Replication of Cryptic Plasmids, pMA1 and pMA2, from a Unicellular Cyanobacterium,Microcystis aeruginosa.* Bioscience, Biotechnology, and Biochemistry, 1995. **59**(7): p. 1217-1220.

16. Wallace, M.M., D.W. Miller, and S. Raps, *Characterization of pMa025, a plasmid from the cyanobacterium Microcystis aeruginosa UV025.* Archives of Microbiology, 2002. **177**(4): p. 332-338.

17. Hoke, A.K., et al., *Genomic signatures of Lake Erie bacteria suggest interaction in the Microcystis phycosphere.* PLOS ONE, 2021. **16**(9): p. e0257017.

18. Smith, D.J., et al., *Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time.* Environmental Microbiology, 2021. **23**(6): p. 3020-3036.

19. Cook, K.V., et al., *The global Microcystis interactome.* Limnology and Oceanography, 2020. **65**(S1).

20. Kuno, S., et al., *The Distribution of a Phage-Related Insertion Sequence Element in the Cyanobacterium, Microcystis aeruginosa.* Microbes and Environments, 2010. **25**(4): p. 295-301.

21. Yoshida-Takashima, Y., et al., *Cyanophage Infection in the Bloom-Forming Cyanobacteria Microcystis aeruginosa in Surface Freshwater.* Microbes and Environments, 2012. **27**(4): p. 350-355.

22. Pound, H.L. and S.W. Wilhelm, *Tracing the active genetic diversity of Microcystis and Microcystis phage through a temporal survey of Taihu.* PLOS ONE, 2020. **15**(12): p. e0244482.

23. McKindles, K.M., et al., *Dissolved microcystin release coincident with lysis of a bloom dominated by Microcystis spp. in Western Lake Erie attributed to novel cyanophage.* Applied and Environmental Microbiology, 2020. **86**(22).

24. Steffen, M.M., et al., *Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH.* Environmental Science & Technology, 2017. **51**(12): p. 6745-6755.

25. Stalder, T., et al., *Linking the resistome and plasmidome to the microbiome.* The ISME Journal, 2019. **13**(10): p. 2437-2446.

26. Li, W., et al., *Impacts of Microcystis on the Dissemination of the Antibiotic Resistome in Cyanobacterial Blooms.* ACS ES&T Water, 2021. **1**(5): p. 1263-1273.

27. Yoshida, T., et al., *Isolation and Characterization of a Cyanophage Infecting the Toxic Cyanobacterium Microcystis aeruginosa.* Applied and Environmental Microbiology, 2006. **72**(2): p. 1239-1247.

28. Stark, G.F., Truchon, A. R. , Dittmann, E. , Wilhelm, S. W. , *Closed circular genome sequence of a Microcystis aeruginosa PCC7806 -mcyB (UTK) nontoxic mutant.* Microbiology Resource Announcements, 2023. 12:e00700-23.

29. Wick, R.R., et al., *Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.* PLOS Computational Biology, 2017. **13**(6): p. e1005595.

30. Sayers, E.W., et al., *Database resources of the national center for biotechnology information.* Nucleic Acids Research, 2022. **50**(D1): p. D20-D26.

31. Pound, H.L., et al., *Changes in Microbiome Activity and Sporadic Viral Infection Help Explain Observed Variability in Microcosm Studies.* Frontiers in Microbiology, 2022. **13**.

32. Martin, R.M., et al., *Sulfolipid substitution ratios of Microcystis aeruginosa and planktonic communities as an indicator of phosphorus limitation in Lake Erie.* Limnology and Oceanography, 2023. **68**(5): p. 1117-1131.

33. Morimoto, D., et al., *Transcriptome Analysis of a Bloom-Forming Cyanobacterium Microcystis aeruginosa during Ma-LMM01 Phage Infection.* Frontiers in Microbiology, 2018. **9**.

34. Tatusova, T., et al., *NCBI prokaryotic genome annotation pipeline.* Nucleic Acids Research, 2016. **44**(14): p. 6614-6624.

35. Camargo, A.P., et al., *Identification of mobile genetic elements with geNomad.* Nature Biotechnology, 2023.

36. Grant, J.R., et al., *Proksee: in-depth characterization and visualization of bacterial genomes.* Nucleic Acids Res, 2023. **51**(W1): p. W484-W492.

37. Mirdita, M., et al., *ColabFold: making protein folding accessible to all.* Nature Methods, 2022. **19**(6): p. 679-682.

38. Wang, J., et al., *iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures.* Bioinformatics, 2020. **36**(1): p. 131-135.

39. Wang, J., et al., *iCn3D: From Web-Based 3D Viewer to Structural Analysis Tool in Batch Mode.* Frontiers in Molecular Biosciences, 2022. **9**.

40. Finn, R.D., J. Clements, and S.R. Eddy, *HMMER web server: interactive sequence similarity searching.* Nucleic Acids Research, 2011. **39**(Web Server issue): p. W29-37.

41. Holm, L., *Using Dali for Protein Structure Comparison.* Methods Mol Biol, 2020. **2112**: p. 29-42.

42. Zhang, Y., *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic Acids Research, 2005. **33**(7): p. 2302-2309.

43. Frangeul, L., et al., *Highly plastic genome of Microcystis aeruginosa PCC 7806, a ubiquitous toxic freshwater cyanobacterium.* BMC Genomics, 2008. **9**(1): p. 274.

44. Smiley, B.L., et al., *Sequences of the Escherichia coli dnaG primase gene and regulation of its expression.* Proceedings of the National Academy of Sciences, 1982. **79**(15): p. 4550-4554.

45. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA.* Nature, 2010. **468**(7320): p. 67-71.

46. Sengupta, M. and S. Austin, *Prevalence and Significance of Plasmid Maintenance Functions in the Virulence Plasmids of Pathogenic Bacteria.* Infection and Immunity, 2011. **79**(7): p. 2502-2509.

47. Harke, M.J., et al., *A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp.* Harmful Algae, 2016. **54**: p. 4-20.

48. Pfeifer, E., et al., *Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires.* Nucleic Acids Research, 2021. **49**(5): p. 2655-2673.

49. Faure, G., et al., *CRISPR–Cas in mobile genetic elements: counter-defence and beyond.* Nature Reviews Microbiology, 2019. **17**(8): p. 513-525.

50. Camara-Wilpert, S., et al., *Bacteriophages suppress CRISPR–Cas immunity using RNA-based anti-CRISPRs.* Nature, 2023. **623**(7987): p. 601-607.

51. Varble, A., et al., *Prophage integration into CRISPR loci enables evasion of antiviral immunity in Streptococcus pyogenes.* Nature Microbiology, 2021. **6**(12): p. 1516-1525.

52. Yamaguchi, H., et al., *Complete Genome Sequence of Microcystis aeruginosa NIES-2481 and Common Genomic Features of group G M. aeruginosa.* Journal of Genomics, 2018. **6**: p. 30-33.

53. Pfeifer, E. and E.P.C. Rocha, *Phage-plasmids promote recombination and emergence of phages and plasmids.* Nat Commun, 2024. **15**(1): p. 1545.

54. Wang, X.-Q., et al., *Profiling the interplay and coevolution of Microcystis aeruginosa and cyanosiphophage Mic1.* Microbiology Spectrum, 2024. **12**(6).

Appendix

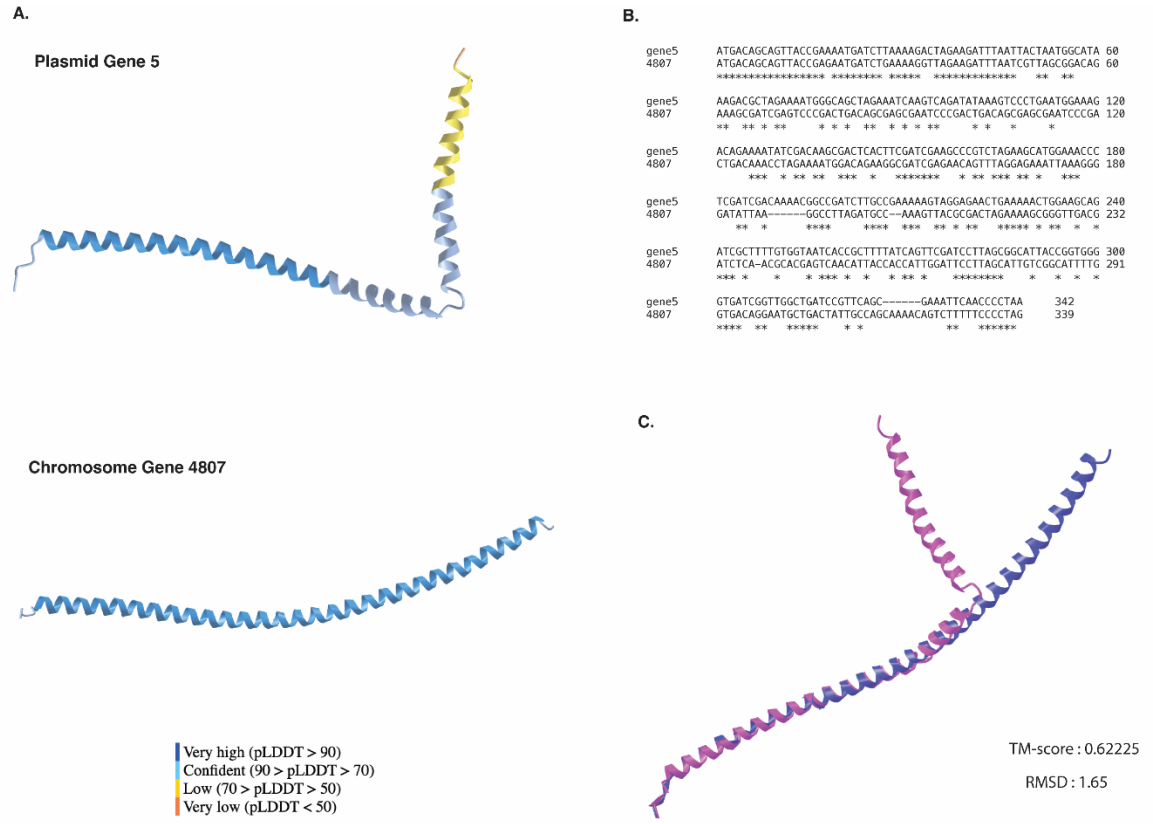

Figure 4.7 Protein structure alignment of gene 5 on the plasmid with a chromsome-atpase like gene in the chromosome

A) Protein model of gene 5 on the PCC 7806 plasmid (top), and a protein model for a gene on the PCC 7806 chromosome (locus tag pgaptmp_4807) (bottom). B) Clustal W nucleotide alignments of gene 5 on the PCC 7806 plasmid, and pgaptmp_4807. C) Structural alignments of gene 5 and pgaptmp_4807, using TM-align. The TM-score was ~0.62, indicating the structures share a similar fold (highest TM score is 1).

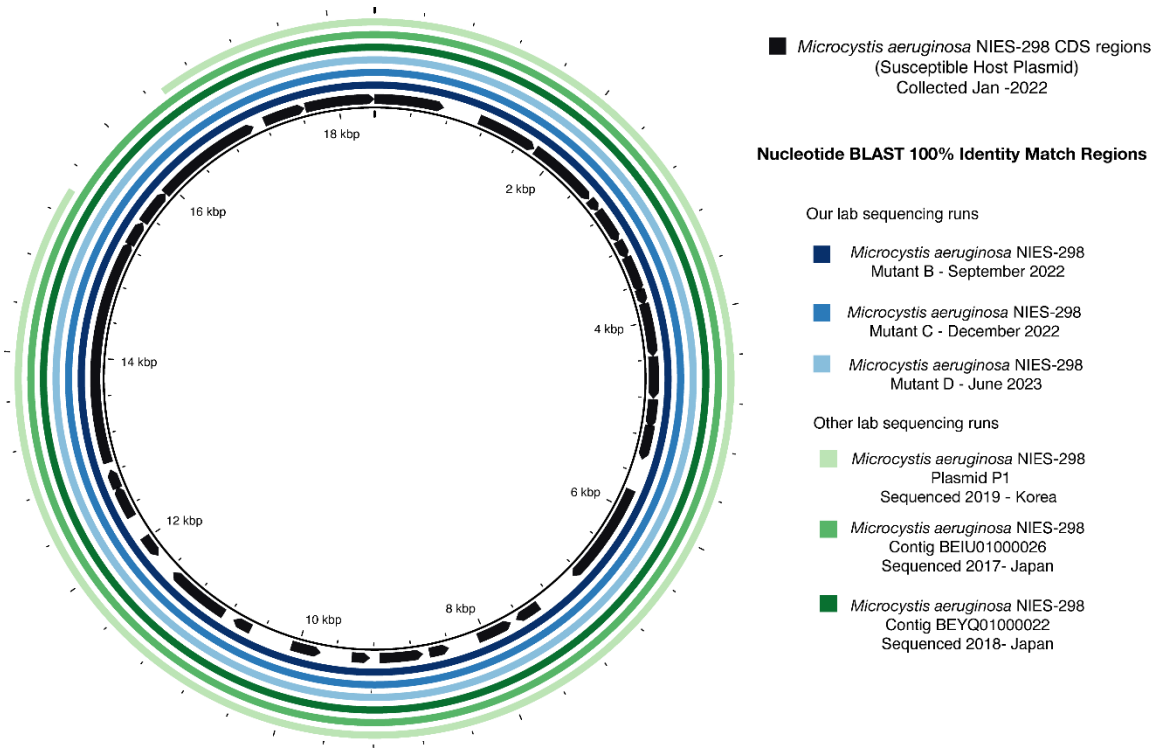

Figure 4.8 Nucleotide BLAST of all NIES-298 P1 plasmids and contigs from scaffold level genomes

Nucleotide BLAST searches of a plasmid present in our *Microcystis aeruginosa* NIES-298 strain. Between our four sequencing runs (backbone coding sequences (CDS), and blue rings), the plasmid nucleotide sequences remained identical, though gene orientation on the + or – strands did change between some runs. Our BLAST of the NIES-298 plasmid against scaffold-level NIES-298 genomes (green rings) on NCBI also revealed contigs with DTR's at the ends, which were identical to the NIES-298 plasmid found in our lab sequencing runs (outside of the duplicate DTR). The NIES-298 P1 plasmid on NCBI currently is identical, aside from missing two genes (incompete outer ring).

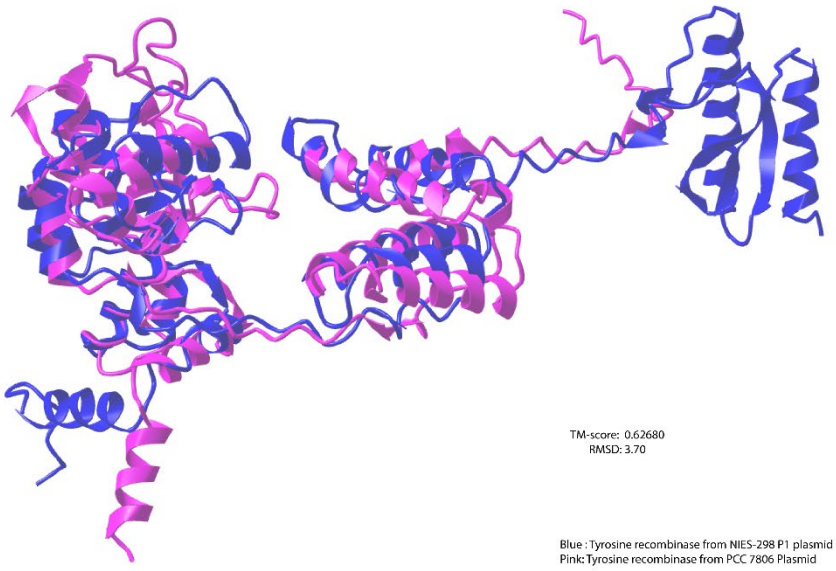

Figure 4.9 Protein structure alignment of tyrosine recombinases from NIES-298 P1 plasmid and PCC 7806

Pairwise alignment of the tyrosine recombinases from *Microcystis aeruginosa* NIES-298 P1, and PCC 7806 plasmids. Structures were aligned with TM-align, with a TM-score of 0.62680 (out of 1), and RMSD score of 3.70.

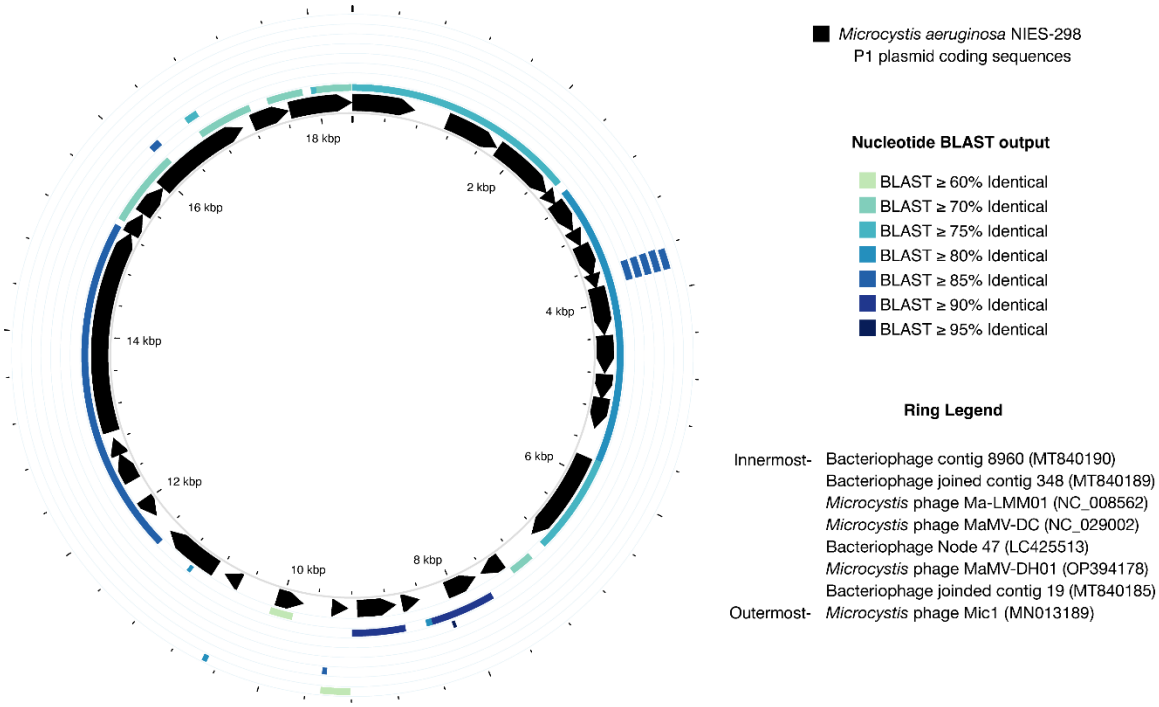

Figure 4.10 BLAST of *Microcystis* phage genomes against the NIES-298 P1 plasmid

Nucleotide BLAST outputs from *Microcystis* phage genome sequences. Whole phage genome sequences available on NCBI Genbank were used and BLAST against the NIES-298 P1 plasmid. The BLASTn outputs are color coded based on the percent identity to the regions of the NIES P1 plasmid.

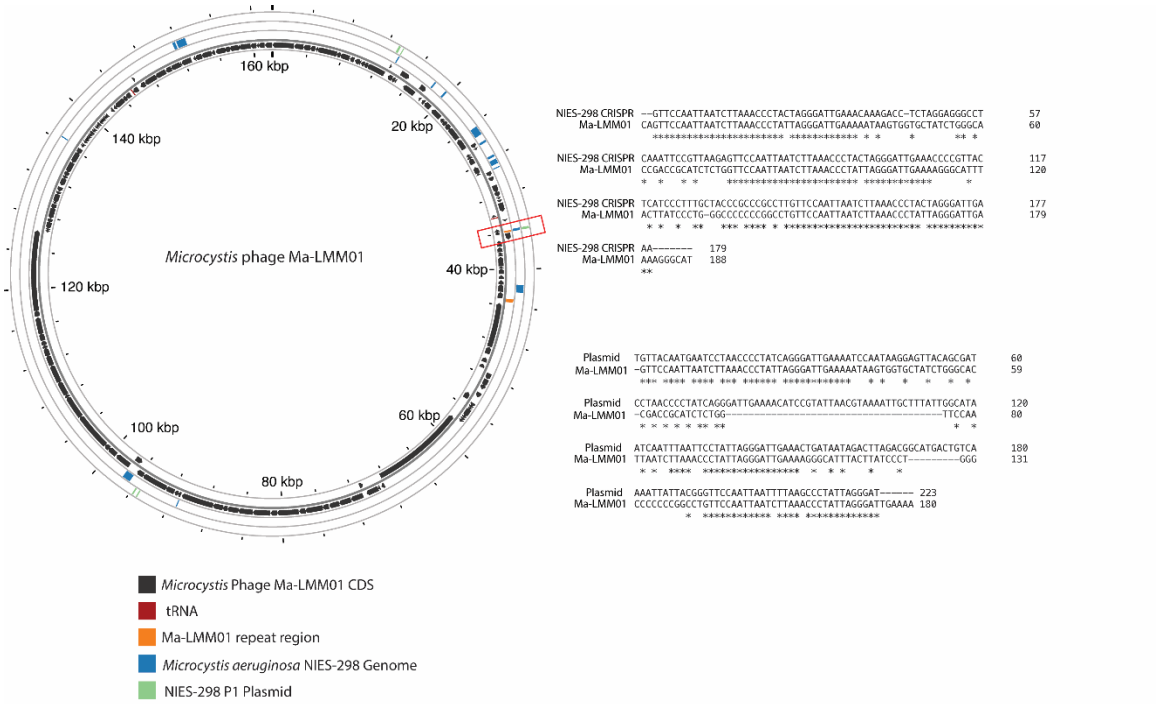

Figure 4.11 BLAST of NIES-298 genome and NIES-298 P1 plasmid against *Microcystis phage* Ma-LMMO1 repeat region

Nucleotide BLAST comparison of the NIES-298 genome (blue ring regions) and the NIES-298 P1 plasmid (green ring regions) against *Microcystis* phage Ma-LMM01. The three sequences share similairty to the CRISPR repeat region in the NIES-298 genome.

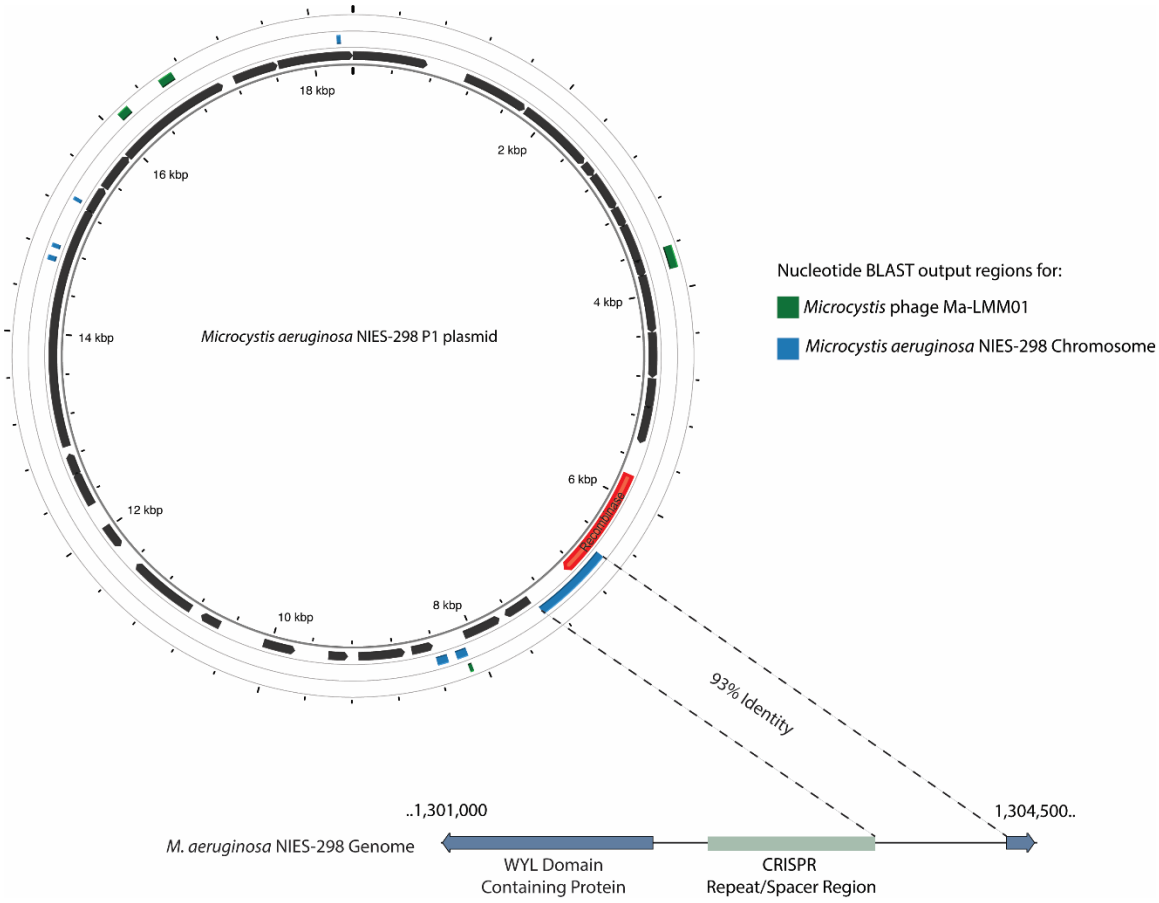

Figure 4.12 Recombinase and intergenic region from NIES-298 P1 plasmid that alignments with genome CRISPR region

Two separate regions of the NIES-298 P1 plasmid have sequence similairty to CRISPR regions in the NIES-298 genome. Shown here, part of the tyrosine recombinase/integrase along with the intergenic region downstream is 93% identical to a region directly following a CRISPR repeat in the NIES-298 genome.

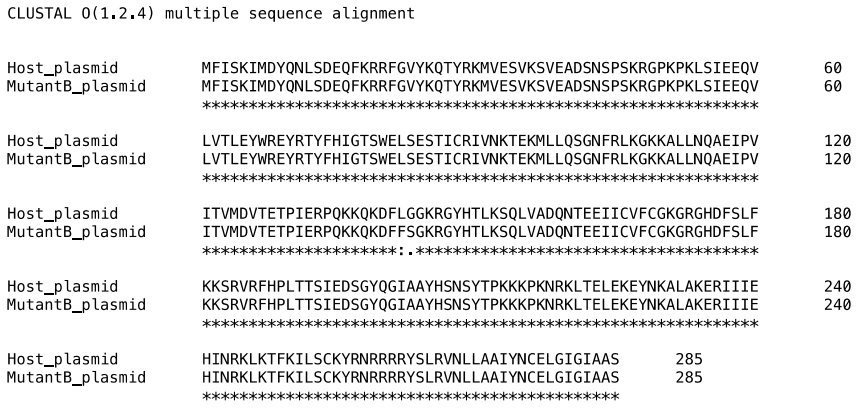

Figure 4.13 IS5 transposase that changed on NIES-298 P2 plasmid

Clustal W. amino acid sequence alignment of the IS5-family transposase that changed between the “host” plasmid, and “mutant b” plasmid, as seen in Appendix Figure 4.14.

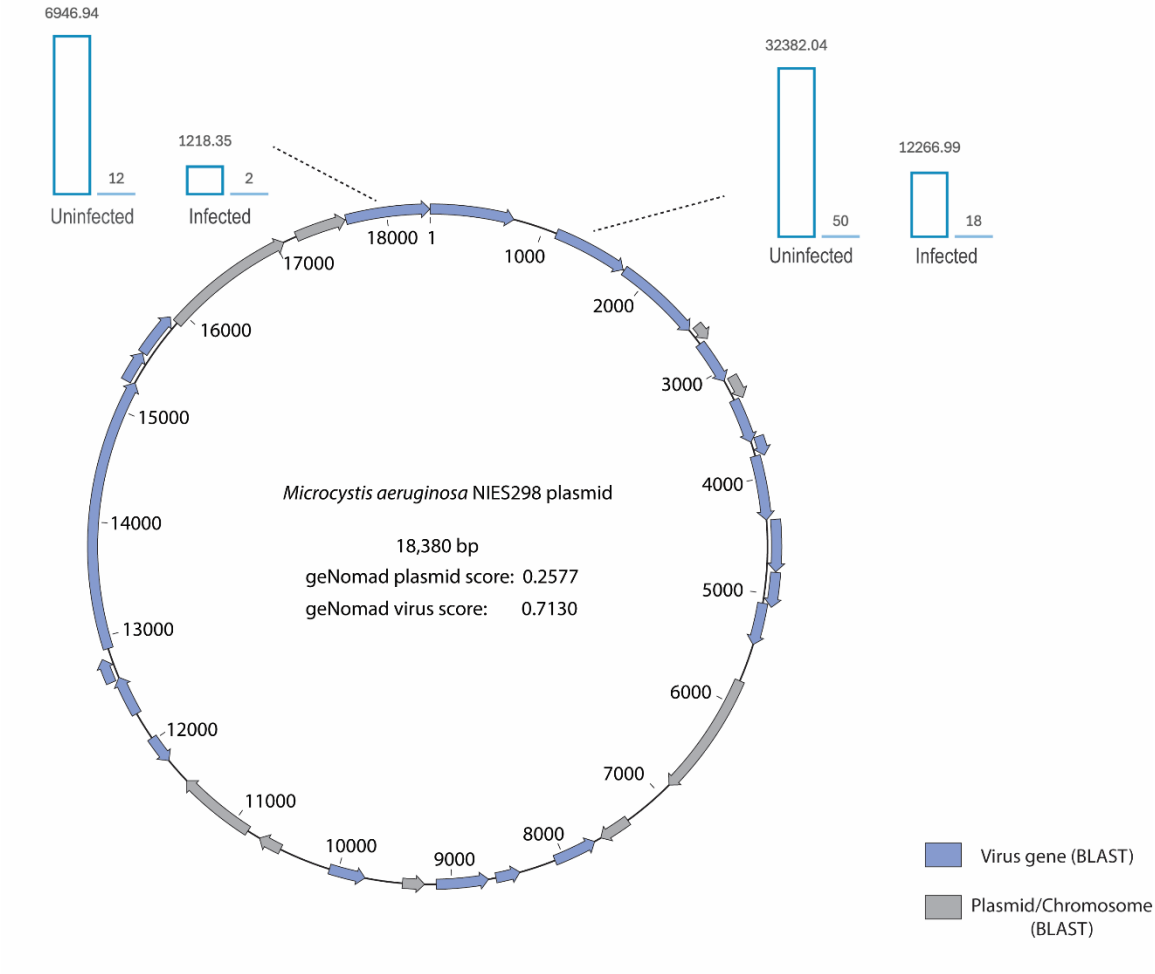

Figure 4.14 Gene expression of NIES-298 P1 plasmid during *Microcystis phage* Ma-LMM01 infection

NIES-298 P1 plasmid genes that had TPM's and read mappings that doubled between the infected and uninfected transcriptome libraries at 3 hours. TPM for all genes can be found in Tables S15 and S16.

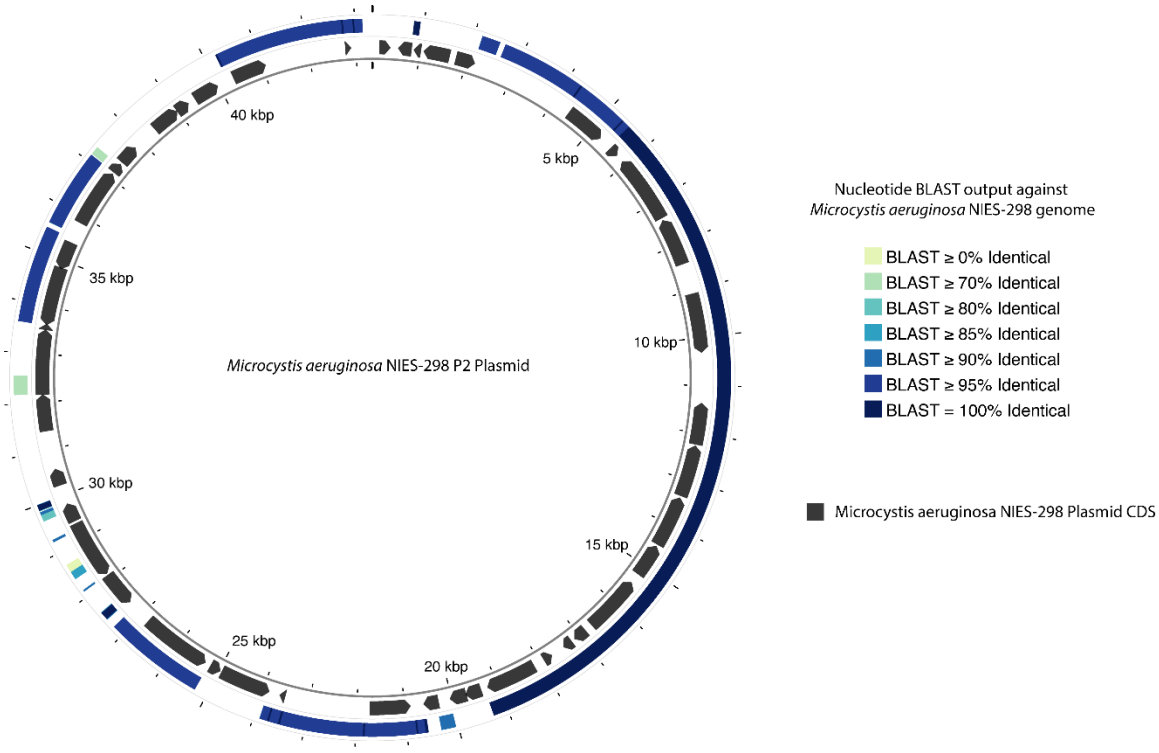

Figure 4.15 BLAST of NIES-298 genome against NIES-298 P2 plasmid

The *M. aeruginosa* NIES-298 P2 plasmid shares a lot of similarity to the NIES-298 genome, some regions being 100% identical to chromosomal regions. The nucleotide BLAST output are NIES-298 chromosome regions homologous to the NIES-298 P2 plasmid.

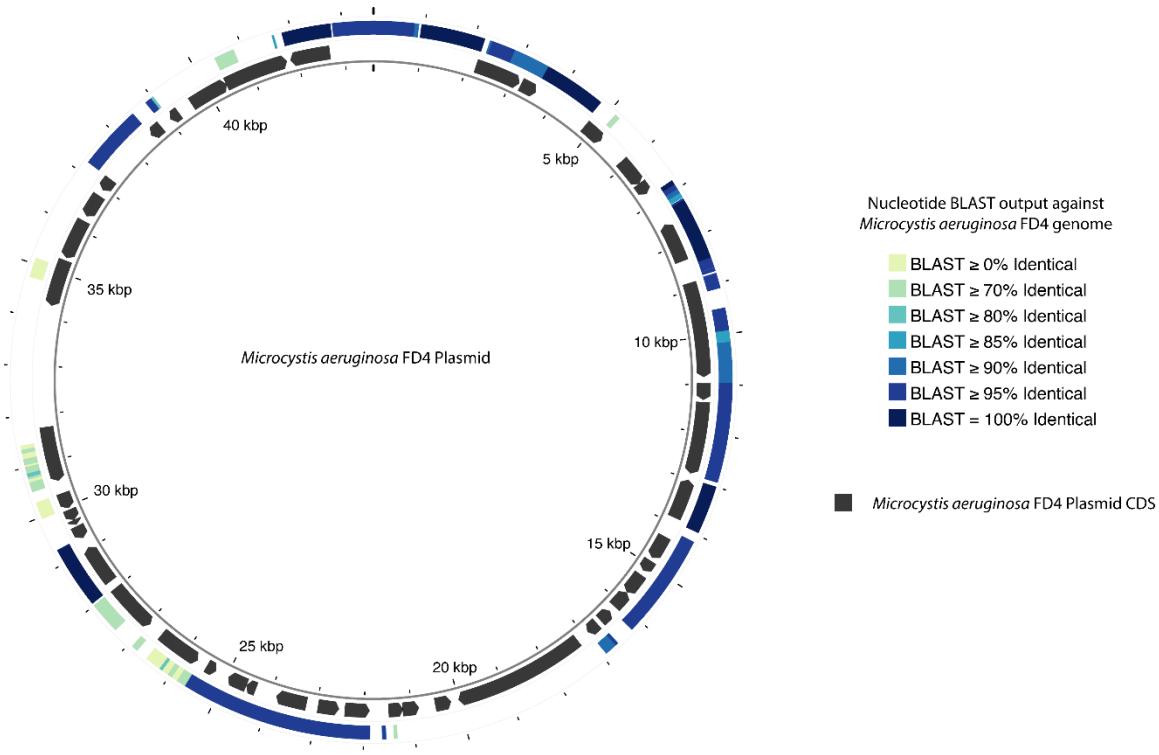

Figure 4.16 BLAST of *M.aeruginosa* FD4 genome against FD4 plasmid

The M.aeruginosa FD4 P1 plasmid shares a lot of similarity to it's host genome (*M. aeruginosa* FD4). The nucleotide BLAST output are FD4 chromosome regions homologous to the FD4 P1 plasmid.

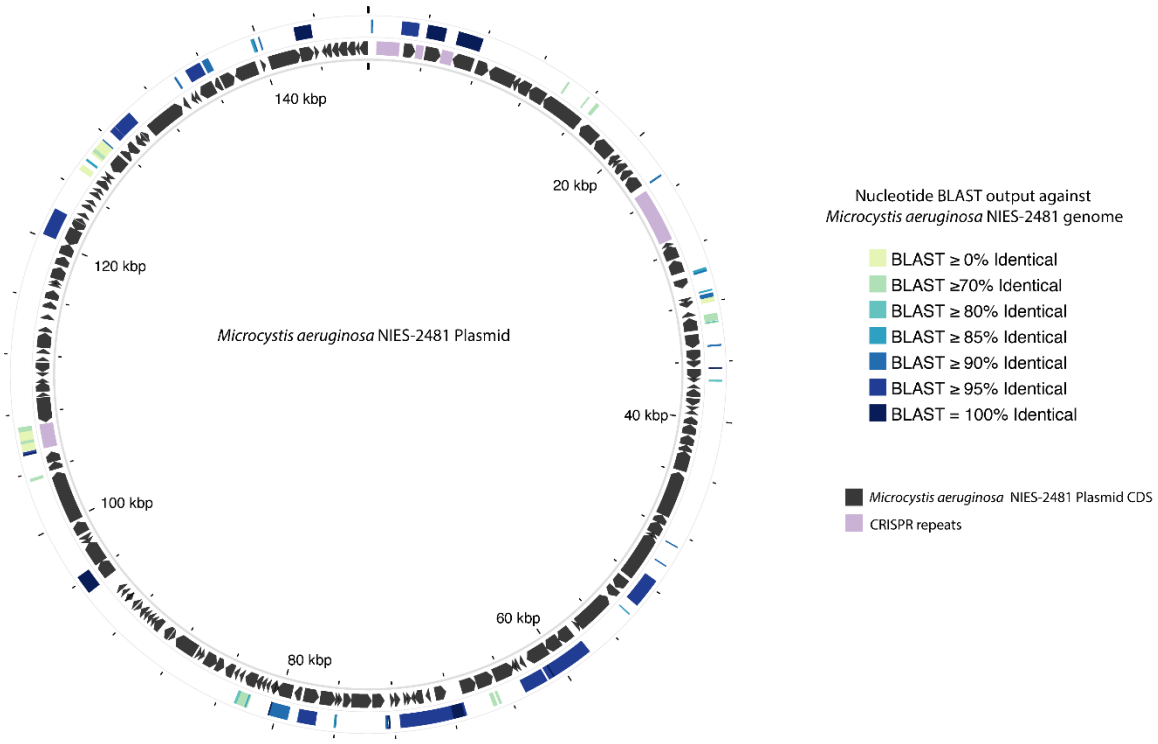

Figure 4.17 BLAST of *M.aeruginosa* NIES-2481 genome against NIES-2481 P1 plasmid

The *M.aeruginosa* NIES-2481 P1 plasmid shares a lot of similarity to it's host genome (*M. aeruginosa* nies-2481). The nucleotide BLAST output are NIES-2481 chromosome regions homologous to the nies-2481 P1 plasmid.

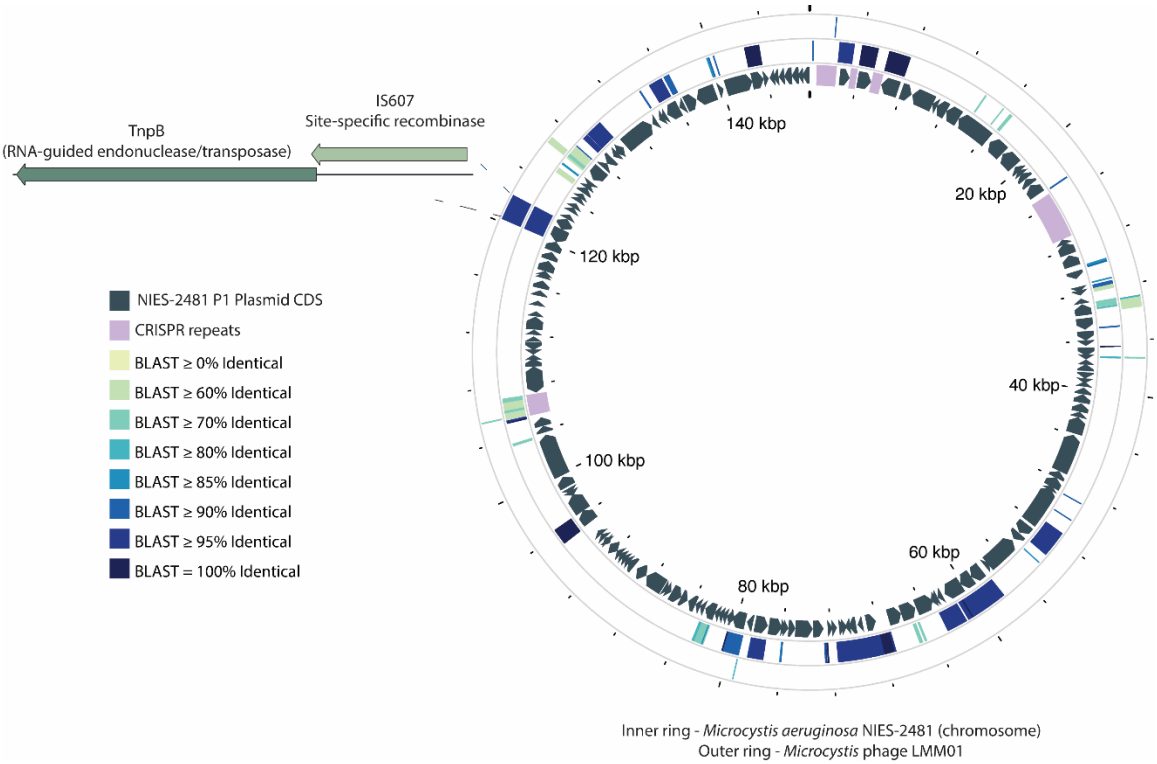

Figure 4.18 TnpB and IS607 recombinase that are 99% identical between NIES-2481 P1 plasmid, the NIES-2481 genome and *Microcystis phage* Ma-LMM01

The *M.aeruginosa* NIES-2481 P1 Plasmid has an area with two transposable elements (*tnpB* and IS607) that are between 98-99.9% identical between the NIES-2481 P1 plasmid (backbone), the NIES-2481 genome (inner ring) and the Microcystis phage Ma-LMM01genome.

CHAPTER FIVE:
CONCLUSIONS

When it comes to understanding *Microcystis* bloom dynamics and ecology, there is still much for researchers to learn. To develop models which will accurately predict bloom toxicity, we need controlled studies which help to elucidate abiotic and biotic factors that impact microcystin production. One of the best ways of doing this is developing non-toxic mutants and comparing growth dynamics between mutant and wildtype strains. Chapter 2 provides a better understanding of the potential ecological role microcystin may play in temperature acclimation. Chapter 2 initially set out to explore the hypothesis that microcystin aids in oxidative stress tolerance, a hypothesis that has dominated *Microcystis* literature since 2011[1]. Through growth in continuous cultures, we were able to trace the increase of the toxin, and subsequently observed how the mutant strain acclimated to these same conditions. While the mutant was able to acclimate, we did see various alternative electron pathways and a glutathione-dependent peroxiredoxin become upregulated and found more ROS damage in the mutant compared to the wildtype. We surmised from our results that microcystin likely does play some role in the oxidative stress response/redox regulation, which many others have also seen under different stressors (iron limitation and high light) [2, 3]. However, our results generated the new hypotheses/questions:

1. Do certain types of ROS agents influence microcystin production?

There are various forms of ROS generated by photosynthetic organisms ($OH\cdot$, H_2O_2,O_2^- ,$^1O_2$). It is also known that certain ROS forms are more destructive than others [4]. Since we only saw the differential expression of a single glutathione-dependent peroxiredoxin (and slight upregulation of thioredoxin-dependent peroxiredoxin), more focus should be put on understanding how different types of ROS-induced damage influence microcystin production, instead of the blanket statement that all ROS may be a factor in microcystin production.

2. Are the mutants strategies less effective/more costly on fitness than the wildtypes? If so, under what scenarios is this relevant?

Whether the mutants strategies in dealing with cold acclimation are more costly to fitness compared to increasing toxin production still needs exploration, which is what the experiments in Appendix B are set up to explore. We figure the best way of exploring which metabolic strategy is more effective during cold growth would be to compete the two strains at suboptimal growth conditions, and see which strategy wins out over the other. However, ecologically, when considering fitness in the context of cyanobacterial blooms, it may make more sense to consider total biomass (cell volume and concentration) and not just growth rate/cell division, which would be hard to do with mixed co-cultures. In the future, the effects of compounded stressors on these two growth strategies during cold growth would also be useful to explore. It may be that in ideal chemostat conditions the mutant/wildtype strains are capable of acclimating to the cold, but with the addition of some outside stress one strategy is more cost effective than the other (microcystin v. metabolic "rewiring"). These serve as future directions for Chapter 2.

Chapters 3 and 4 should be considered as cautionary tales for all *Microcystis* researchers that use the PCC 7806 wildtype and mutant system for physiology work. It was only after the experiment for Chapter 2 was done that we realized it was performed under the assumption the mutant and wildtype were genetically identical, aside from the chloramphenicol cassette insertion in *mcyB* [5]. While Δ*mcyB* and the PCC 7806 wildtype have served as type strains in numerous research papers to understand toxin production over the last 20 years, how genetically identical Δ*mcyB* was to its toxic wildtype strain was never established. This is particularly troubling, because the plasticity of the *Microcystis* genome has been known since 2007 and is a pervading topic in most *Microcystis* genomics literature. Appendix A includes our genome sequencing efforts for Δ*mcyB* . In Chapter 3 we detailed movement of five different families of transposases which caused genomic re-arrangements in the mutant and wildtype, leading to differential expression of genes between the strains. This chapter serves as a stark reminder that genome plasticity and evolution is happening in real-time with cultures, and before large omics-based studies, *Microcystis* strains should be sequenced. This is especially true when doing comparisons of the mutant and wildtype. Additionally, while we only found

five families of transposases that lead to changes in gene expression, these are not the only transposases in the genome. From this chapter we generated the questions:

1. Are there more families of transposases that are active and that can influence gene expression in *Microcystis*?

The amount of transposases which influenced gene expression between the mutant and wildtype seems surprisingly small, considering >10% of the genome is composed of repetitive regions/transposons. While we saw five families of transposases that contributed to changes in gene expression, the transposons within each family were identical to one another and did not encompass a large range/diversity of transposases within the transposase family.

2. How are these different families of transposases regulated in *Microcystis*?

Again, while the plasticity of the genome is well known, there have not been many studies that have investigated the regulation of transposases in *Microcystis,* and subsequently confirmed transposition through genome sequencing. Most often, it is reported in passing that transposases are up/down regulated. In the future, more controlled studies should be done to understand how different stressors/growth conditions impact different transposase family's regulation/movement in the *Microcystis* genome. This is relevant for better understanding *Microcystis* evolution, diversity, and stress responses.

From our genome re-sequencing we were surprised to find a closed, extrachromosomal contig in both the mutant and wildtype PCC 7806 genomes. In general, *Microcystis* plasmids are not discussed, even in "mobilome" research, which usually focuses on transposable elements and phage. Most literature concerning *Microcystis* plasmids is from the 1980-90's. In Chapter 4, we found an identical plasmid in both the mutant and wildtype strain, that has not been reported to date, even though PCC 7806 was isolated from the Netherlands in 1972. This plasmid was transcriptionally

active and was found on a contig from 2007 PCC7806 sequencing data and was in another *Microcystis* strain isolated in Canada. Therefore, this plasmid may play some important cellular role, and appears to be present in other geographically distant strains. This leads up to our first question:

1. Are conserved plasmids a common feature in *Microcystis*?

We began investigating this question with *M.aeruginosa* NIES-298 in chapter 4, but our sampling size is still too small to say this is a "common" feature. However, NIES-298 does appear to have an unchanging plasmid based on sequencing from 2017-2024. We also made a second discovery while investigating this plasmid, which was finding how similar it was to a bacteriophage contig. In phage research, marker genes are often used to taxonomically identify a contig as "phage". This plasmid has a gene putatively encoding a conserved phage-tail tip protein present, which brings up the troubling questions:

1. Have we mistakenly labeled plasmids as phage through the marker gene approach?
2. How big of a role do plasmids play in bloom ecology, especially if they are being mistaken for phage?

Phage have previously been implicated/correlated in *Microcystis* bloom collapses through bioinformatics-based investigations [6, 7]. However, plasmids often are not considered at all in many of these investigations, and it is clear from Chapter 4 that it is possible for there to be overlap between *Microcystis* phage, and *Microcystis* plasmids. While this is a troubling thought, this is information which needs future consideration in all bioinformatics-based cyanobacterial bloom investigations. Additionally, while *M.aeruginosa* NIES-298 is used in *Microcystis* phage infection work, the two plasmids present in the strain, and their physiological relevance have been ignored. Therefore, for future *M.aeruginosa* NIES-298 phage work, we ask the question:

1. Do the plasmids present in NIES-298 participate in plasmid-phage "warfare"?

We began investigating this question in Chapter 4, by using publicly available transcriptomes which were created during a phage infection [8]. While we saw some evidence of plasmid genes having changes in transcript abundance during infection, more detailed work would be needed to understand any phage-plasmid interactions. Future studies should consider plasmid gene expression, and any gene transfer events from plasmid-phage-chromosome that may occur in response to phage infection. We were particularly interested in whether the IS5-family/TnpA transposons present on one of the NIES-298 plasmids may participate in some sort of phage-defense, as many of them were more transcriptionally active during phage infection. Thus, *M.aeruginosa* NIES-298 would serve as a good system to investigate gene flow between phage, mobile elements and the genome. This may be informative for investigating phage-host evolutionary trajectories and phage-resistance, which we have seen develop numerous times (unpublished observations) with *M.aeruginosa* NIES-298 and it's phage.

While this dissertation began with investigating the intracellular role of microcystin during cold temperature acclimation, it developed into two chapters which serve as "cautionary tales" to other *Microcystis* researchers. These two tales need to be told for everyone who studies *Microcystis*. The knowledge from these chapters will help to create better controlled, reproducible experiments. In *Microcystis* physiology studies, strain variations and extrachromosomal elements do not garner much attention. It may be because of these reasons that reproducibility in the *Microcystis* field has been an issue, with contradictory results among researchers as to what environmental parameters may influence toxin production. Understanding how different lab strains may be from one another, despite being the same "strain", and establishing the roles or impacts plasmids may have on host-strain physiology will bring researchers two steps closer to creating better models forecasting bloom toxicity and improving our understanding of bloom ecology.

References

1. Zilliges, Y., et al., *The Cyanobacterial Hepatotoxin Microcystin Binds to Proteins and Increases the Fitness of Microcystis under Oxidative Stress Conditions.* PLoS ONE, 2011. **6**(3): p. e17615.

2. Alexova, R., et al., *Specific global responses to N and Fe nutrition in toxic and non-toxic Microcystis aeruginosa.* Environmental Microbiology, 2016. **18**(2): p. 401-413.

3. Kaebernick, M., et al., *Light and the Transcriptional Response of the Microcystin Biosynthesis Gene Cluster.* Applied and Environmental Microbiology, 2000. **66**(8): p. 3387-3392.

4. Latifi, A., M. Ruiz, and C.-C. Zhang, *Oxidative stress in cyanobacteria.* FEMS Microbiology Reviews, 2009. **33**(2): p. 258-278.

5. Dittmann, E., et al., *Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium Microcystis aeruginosa PCC 7806.* Molecular Microbiology, 1997. **26**(4): p. 779-787.

6. Steffen, M.M., et al., *Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH.* Environmental Science & Technology, 2017. **51**(12): p. 6745-6755.

7. McKindles, K.M., et al., *Dissolved microcystin release coincident with lysis of a bloom dominated by Microcystis spp. in Western Lake Erie attributed to novel cyanophage.* Applied and Environmental Microbiology, 2020. **86**(22).

8. Morimoto, D., et al., *Transcriptome Analysis of a Bloom-Forming Cyanobacterium Microcystis aeruginosa during Ma-LMM01 Phage Infection.* Frontiers in Microbiology, 2018. **9**.

APPENDIX:

APPENDIX A:
SEQUENCING THE MUTANT GENOME

Publication note

This appendix was published in 2023 as a microbial resource announcement and has been adapted for this dissertation.

Citation:
Stark, G. F., Truchon, A. R., Dittmann, E., & Wilhelm, S. W. (2023). Closed circular genome sequence of a *Microcystis aeruginosa* PCC7806 ΔmcyB (UTK) non-toxic mutant. Microbiology resource announcements, 12(11), e0070023.
https://doi.org/10.1128/MRA.00700-23

Abstract

Here we report the complete, closed genome of the non-toxic *Microcystis aeruginosa* PCC7806 Δ*mcyB* mutant strain. This genome is 5,103,923 bp long, with a GC content of 42.07%. Compared to the published wildtype genome (*Microcystis aeruginosa* PCC7806SL), there is evidence of accumulated mutations beyond the inserted chloramphenicol resistance marker.

Announcement

Microcystis aeruginosa PCC7806 Δ*mcyB* is a non-toxic strain available at the Pasteur Culture Collection (1). This mutant is widely used in comparative studies with the toxic wildtype strain PCC 7806 (2-15). Except for a chloramphenicol cassette inserted in the *mcyB* gene, the mutant and wildtype genomes were thought to be identical, although this has not been confirmed. Whole genome sequencing revealed that our strain, which has undergone live-passage and cryopreservation (-150° C) for the last seven years after an aliquot was transferred from Potsdam (Germany) to Knoxville (USA), has accumulated mutations relative to the published 7806SL genome – we have added the "UTK" annotation to signify our specific isolate. Of note, our mutant strain has an additional putative transposase gene (*tpnA*), not seen in the 7806SL wildtype, inserted in an SLC13-family permease gene. There is also a region of the genome that assembled in a manner suggesting a large inversion (~2.5 Mbp) occurred when aligned against the 7806SL genome. This MRA serves as a cautionary tale for labs to routinely re-sequence strains.

Axenic cultures of *Microcystis aeruginosa* PCC7806 *ΔmcyB*(UTK) were cultured in 50mL of modified CT media (16) (B-glycerophosphate was replaced with 1mL of Na_2HPO_4 (23.19 mg/mL), and grown under a 14/10hr day/night cycle (~35 μmol photons $m^{-2}s^{-1}$). Late-log phase cells were pelleted (15,000 xg, 10 min) and resuspended in 400 μL TE buffer. RNase (0.6 μL of 10 mg/mL) and 120 μL of lysozyme (20 mg/mL) were added, and cells were incubated at 37° C for 30 min. Then, 6 μL of proteinase K (20 mg/mL) in 3 mM $CaCl_2$ and 200 mM Tris buffer, and 39.5 μL of 20% SDS were added

before incubation at 55° C for two hours. DNA was extracted *via* a phenol-chloroform method (17). Half the extracted DNA was used for long-read sequencing (DNA was not sheared), and the other half for Illumina short-read sequencing. Long reads were sequenced in-house, using a MinION R9.4.1 flow cell (Oxford Nanopore Technologies). High-molecular-weight DNA was prepared for sequencing using the SQK-LSK110 kit (ONT) per manufacturer protocols, which enriched for reads >3kb, but no formal size selection was performed. Long-read sequencing resulted in 320,921 raw reads (N_{50}= 17.55kb). For Illumina sequencing, DNA was sent to SeqCenter (Pittsburgh, PA). Sample libraries were prepared using an Illumina DNA prep kit and IDT 10 bp UDI indices, and processed on an Illumina NextSeq 2000, generating 8,278,500 151-bp paired-end reads. Demultiplexing, quality control and adapter trimming was performed with bcl-convert, using default parameters (v3.9.3) (Illumina).

For genome assembly, default parameters were used for all software. Bases were called using Guppy (v. 6.01) (18), with the dna_r9.4.1_450bps_fast.cfg config file. Adapter sequences were trimmed using Porechop (v.0.2.4) (19), then filtered using NanoFilt at a quality of 9 and a length of 500 (v.2.8.0) (20). Illumina reads were trimmed using CLC Genomics Workbench (v. 21.0.4). Genome assembly using long and short reads was performed *de novo* using Unicycler (v.0.4.9b) (21). A circular contig was assembled, circularized, trimmed, and rotated with Unicycler. The final genome assembly was annotated using NCBI Prokaryotic Genome Annotation Pipeline (annotation software revision: 2022-12-13.build6469), and genomic GC content determined with QUAST (v4.4)(22).

Data Availability

All sequencing data is available on NCBI under the BioProject ID PRJNA995633. Raw reads can be found at SRA under accession numbers SRX21194812 (Nanopore) and SRX21194811(Illumina). The annotated genome assembly can be found under GCA_030553035.1.

References

1. Dittmann E, Neilan BA, Erhard M, Von Döhren H, Börner T. 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. Molecular Microbiology 26:779-787.

2. Barchewitz T, Guljamow A, Meissner S, Timm S, Henneberg M, Baumann O, Hagemann M, Dittmann E. 2019. Non-canonical localization of RubisCO under high-light conditions in the toxic cyanobacterium *Microcystis aeruginosa* PCC7806. Environmental Microbiology 21:4836-4851.

3. Zuo J, Hu L, Shen W, Zeng J, Li L, Song L, Gan N. 2021. The involvement of α-proteobacteria *Phenylobacterium* in maintaining the dominance of toxic *Microcystis* blooms in Lake Taihu, China. Environmental Microbiology 23:1066-1078.

4. Mallia V, Verhaegen S, Styrishave B, Eriksen GS, Johannsen ML, Ropstad E, Uhlig S. 2020. Microcystins and *Microcystis aeruginosa* PCC7806 extracts modulate steroidogenesis differentially in the human H295R adrenal model. PLoS One 15:e0244000.

5. Tillett D, Dittmann E, Erhard M, Von Döhren H, Börner T, Neilan BA. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. Chemistry & Biology 7:753-764.

6. Hesse K, Dittmann E, Börner T. 2001. Consequences of impaired microcystin production for light-dependent growth and pigmentation of *Microcystis aeruginosa* PCC 7806. FEMS Microbiology Ecology 37:39-43.

7. Jähnichen S, Ihle T, Petzoldt T, Benndorf J. 2007. Impact of Inorganic Carbon Availability on Microcystin Production by *Microcystis aeruginosa* PCC 7806. Applied and Environmental Microbiology 73:6994-7002.

8. Briand E, Bormans M, Quiblier C, Salençon M-J, Humbert J-F. 2012. Evidence of the Cost of the Production of Microcystins by *Microcystis aeruginosa* under Differing Light and Nitrate Environmental Conditions. PLoS One 7:e29981.

9. Zilliges Y, Kehr J-C, Mikkat S, Bouchier C, De Marsac NT, BöRner T, Dittmann E. 2008. An Extracellular Glycoprotein Is Implicated in Cell-Cell Contacts in the

Toxic Cyanobacterium *Microcystis aeruginosa* PCC 7806. Journal of Bacteriology 190:2871-2879.

10. Zilliges Y, Kehr J-C, Meissner S, Ishida K, Mikkat S, Hagemann M, Kaplan A, Börner T, Dittmann E. 2011. The Cyanobacterial Hepatotoxin Microcystin Binds to Proteins and Increases the Fitness of *Microcystis* under Oxidative Stress Conditions. PLoS One 6:e17615.

11. Van De Waal DB, Verspagen JM, Finke JF, Vournazou V, Immers AK, Kardinaal WEA, Tonk L, Becker S, Van Donk E, Visser PM, Huisman J. 2011. Reversal in competitive dominance of a toxic versus non-toxic cyanobacterium in response to rising CO2. The ISME Journal 5:1438-1450.

12. Meissner S, Steinhauser D, Dittmann E. 2015. Metabolomic analysis indicates a pivotal role of the hepatotoxin microcystin in high light adaptation of *Microcystis*. Environmental Microbiology 17:1497-1509.

13. Urrutia-Cordero P, Agha R, Cires S, Lezcano MA, Sanchez-Contreras M, Waara KO, Utkilen H, Quesada A. 2013. Effects of harmful cyanobacteria on the freshwater pathogenic free-living amoeba *Acanthamoeba castellanii*. Aquat Toxicol 130-131:9-17.

14. Dittmann E, Börner T, Von Döhren H, Kaebernick M, Erhard M, Neilan BA, Scheler C. 2001. Altered expression of two light-dependent genes in a microcystin-lacking mutant of *Microcystis aeruginosa* PCC 7806. Microbiology 147:3113-3119.

15. Alexova R, Dang TC, Fujii M, Raftery MJ, Waite TD, Ferrari BC, Neilan BA. 2016. Specific global responses to N and Fe nutrition in toxic and non-toxic *Microcystis aeruginosa*. Environmental Microbiology 18:401-413.

16. Wilhelm SW. 2017. *CT Media V.2*. doi:10.17504/protocols.io.iizccf6.

17. Martin RM, Wilhelm SW. 2018. Phenol/chloroform extraction of DNA from cyanobacteria v.2. doi:10.17504/protocols.io.ptndnme.

18. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biology 20.

19. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial Genomics 3.

20. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666-2669.

21. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

22. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

APPENDIX B:
EXPERIMENTAL DEVELOPMENT OF CO-CULTURE CHEMOSTAT COMPETITION ASSAYS

Publication Note

The methods development from this appendix may be used in a future publication. The current contributors to this appendix are; Gwendolyn F. Stark, Laura E. Smith, Bofan Wei, Jordan A. Cannon, Gregory Boyer, and Steven W. Wilhelm.

GFS set up, ran and maintained the chemostats. GFS and LES sampled the chemostats. GFS did qPCR. JAC created the plasmids with the qPCR amplicons of interest. BW and GB did LC-MS. GFS did all data analysis. GFS and SWW designed the experiment.

Background

Previously, our lab found that growing *Microcystis aeruginosa* NIES 843 at suboptimal growth temperature (18°C) nearly doubled toxin quota per cell [1]. This was also first observed in continuous cultures of *Microcystis aeruginosa* PCC 7806 at 19°C by Martin et al. [2]. Since then, this observation has been reproduced PCC 7806 in both continuous and batch culture growth, and different forms of growth media [3, 4]. To see if microcystin provided the toxic wildtype strain a fitness advantage at cold growth temperature, we replicated the Martin et al. [2] experiment, with the addition of two, non-toxic mutant *Microcystis aeruginosa* PCC 7806 *ΔmcyB* chemostats, and did additional assays, comparing growth, photophysiology and reactive oxygen species (ROS) production between the mutant and wildtype at 26°C (warm) and 19°C (cold). We found that both strains were able to acclimate to the 19°C treatment, and cell counts for both strains stabilized around 192 hours in chemostat growth [3]. However, evidence from our transcriptomic time series and photophysiology readings show a metabolic change between the two strains, beginning at 192 hrs. 192 hrs is also when microcystin quota double (and stayed elevated) in the wildtype. In general, our ROS assays also showed that the mutant had more ROS damage then the wildtype during the 19°C treatment [3].

The results from our monoculture experiments brought about a new hypothesis, that microcystin may serve as a better long-term mechanism to acclimate to cold conditions, compared to the mutant's need to upregulate alternative electron pathways and ROS defenses. From an ecological perspective, it's been shown that *Microcystis* blooms have more toxic genotypes present earlier in the season, when water temperatures are cooler (~18°C) [5]. We wondered if cold temperatures inhibited non-toxic strains growth during re-seeding in the spring, where they become outcompeted by toxic genotypes. We decided the best way to test this hypothesis would be to compete the strains in chemostats at optimal growth 26°C and suboptimal growth temperatures 19°C.

Methods

Continuous Cultures

Initially, wildtype *Microcystis aeruginosa* PCC 7806 and *Microcystis aeruginosa* PCC 7806 Δ*mcyB* were grown in monoculture 1-L chemostat vessels, each with an n of 2. Chemostat set up and design was as follows in [3]. Chemostat cultures were grown at 26°C, with a 24hr light cycle, in modified CT media. The modified CT media protocol can be found in [3]. Dilution rate for the chemostats was set to $0.25d^{-1}$.

Flow Cytometry

Cell counts were taken daily on a Beckman Coulter Cytoflex. 250 μL of undiluted culture was collected from each chemostat and loaded into a 96-well plate. Cells were gated on red fluorescence and side scatter.

Collecting samples for qPCR

For this experiment, daily 1-mL samples were collected from chemostats into 1.5mL microcentrifuge tubes and spun down into a pellet. This was done in triplicate for each chemostat. Pellets were washed twice with molecular grade water and left at -80°C for resuspension in molecular grade water for qPCR analysis. Troubleshooting was done with 100μL re-suspensions, however, later work showed better results with 1mL after qPCR was done to find plasmid copy # (Chapter 4). Additionally, methods of spinning down cells varied, as phenotypic changes occurred throughout the length of the experiment, in some instances samples stuck to the sides of the tube and would need multiple rounds of scraping/centrifuging to pellet. Disrupting the gas vesicles through pressure before centrifuging worked in some cases but was inconsistent.

Initial mixing the Chemostats & Sampling Schedule

Cells were allowed to acclimate for ~3 weeks, from there, chemostats were sampled daily for ~week. When cell counts remained relatively steady, cultures were mixed in a 9:1 ratio. Meaning, 10% of the population in the 1-L growth vessel was removed and filled with the equivalence of 10% of the other strain. All pumps were

stopped during the removal and transfer of all cultures. After all transfers were complete, pumps were restarted. After the mixing of the chemostats, they were monitored daily for 13 days, then we switched to sampling every two days for ~ a week. From there, sampling was done every 7-10 days. In all, samples remained mixed in the continuous cultures for 46 days before we started the nitrogen depletion treatments.

Nitrogen depletion treatments

Before the nitrogen depletion part of the experiment, the chemostat growth media contained 370 μM of nitrogen, in the form of KNO_3 and $Ca(NO_3)$. For the first N depletion, we halved this amount, and grew the mutant and wildtype mono and co-cultures for two weeks in 181μM of nitrogen. After two weeks, we decreased nitrogen to 81μM and monitored the chemostats daily for a week. Recipe for N replete for 1-L is as follows: 2mL of 0.2g/mL TAPS, 1mL of 13.76 mg/mL KNO_3 , 1mL $Ca(NO_3)$ 22.02mg/mL, 1mL $MgSO_4$ stock at 40mg/mL, 1mL K_2HPO_4 stock of 2.845 mg/mL and 3mL trace metals solution.

qPCR Primers

To target both Δ*mcyB* and the wildtype, primers were made for the phycocyanin intergenic spacer region[6]. To target only Δ*mcyB*, primers were made to amplify an area of the chloramphenicol cassette [6]. Primer sequences are the same as those used in [6], except the phycocyanin primers were switched (Table 5.1) [6]. Primer sequences were verified in our lab sequenced genomes [7] to still be present (unmutated) and single copy.

Cloning phycocyanin and chloramphenicol amplicons

For cloning, fragments of genomic DNA were PCR amplified *from Microcystis aeruginosa* PCC 7806 using the forward primers PIGS-F and PIGS-R for the reverse primer of the Phycocyanin IGS region (Table 5.1). Primers included restriction sites for XbaI on the forward and PstI on the reverse. Primers containing the same restriction sites were designed for the chloramphenicol cassette region and are denoted CmC-F and CmC-R for the forward and reverse primers, respectively. Amplified fragments were PCR

Table A2.1 Primers used for qPCR of co-cultures

Primer	Sequence
Phycocyanin IGS Forward	TCCTACGGTTTAATTGAGACTAGCC
Phycocyanin IGS Reverse	GCTACTTCGACCGCGCC
Chloramphenicol Forward	GTTTATTGACTACCGGAAGCAGTG
Chloramphenicol reverse	CACGGGGAGAGCCTGAGC
PIGS-F	aaaaaa**TCTAGA**CTAGCTGAGAGCGTTGATAGC
PIGS-R	aaaaaa**CTGCAG**TGCCAGCTACTTCGACC
CmC-F	aaaaaa**TCTAGA**CGTGGCCAATATGGACA
CmC-R	aaaaaa**CTGCAG**TACATCTGTATTAACGAAGCG

purified and digested with the XbaI and PstI, then ligated into the plastic pECE732-P_{veg} [8] digested with the same restriction enzymes. Ligated plasmids transformed, propagated, and screened for successful insertion using *E. coli* DH5α (NEB). The size of the phycocyanin plasmid is 6971bp, and 6882bp for the chloramphenicol plasmid.

qPCR

Whole cell templates were used for qPCR troubleshooting, and we planned on using this method for this co-culture experiment. Primer concentrations of 0.15μM for the phycocyanin primers, and 0.1μM for chloramphenicol primers, with DNA concentrations <1.0 ng/μL were determined to give the best standard curves with batch culture tests. Total working volumes for each reaction were 20μL. 2μL of whole cell template was used. For the standard curves, plasmid stocks were diluted fivefold, from 10^{-2} to 10^{-6}. A Quantstudio 3 was used to run samples. The program on the Quantstudio was set to the following: a 50°C 2m hold stage and 95°C 15m initial denaturation. From there, 40 PCR cycles were run with 95 °C 15s denature, a 60°C 30s annealing, and 72°C 10s elongation steps. Melt curve analysis was done to ensure a single Tm for each primer.

Microcystin Toxin Measurements

Samples for microcystin were collected in the same manner as [3]. 50mL was collected from the chemostats and filtered onto 47-mm glass fiber filters (GF-75, Advantec) via vacuum filtration and immediently flash frozen in liquid nitrogen. All samples were stored at -80°C until processing. Quantification of microcystin and the congener breakdown was done by LC-MS according to the methods of Boyer et al. dx.doi.org/10.17504/protocols.io.bck2iuye.

Photophysiology Readings

A Phyto-PAM II was used to take Fv/Fm and run rapid light curves. All phyto-PAM II settings were kept the same as published in [3]. If needed, further explanation for adjusting Phyto-PAM settings and troubleshooting samples can be found in the Stark et. al. protocol at (dx.doi.org/10.17504/protocols.io.eq2ly7jqelx9/v1).

References

1. Peng, G., et al., *Seasonally relevant cool temperatures interact with N chemistry to increase microcystins produced in lab cultures of Microcystis aeruginosa NIES-843.* Environmental Science & Technology, 2018. **52**(7): p. 4127-4136.

2. Martin, R.M., et al., *Episodic decrease in temperature increases mcy gene transcription and cellular microcystin in continuous cultures of Microcystis aeruginosa PCC 7806.* Frontiers in Microbiology, 2020. **11**.

3. Stark, G.F., et al., *Microcystin aids in cold temperature acclimation: Differences between a toxic Microcystis wildtype and non-toxic mutant.* Harmful Algae, 2023. **129**: p. 102531.

4. Roy, S., A. Guljamow, and E. Dittmann, *Impact of temperature on the temporal dynamics of microcystin in Microcystis aeruginosa PCC7806.* Frontiers in Microbiology, 2023. **14**.

5. Dick, G.J., et al., *The genetic and ecophysiological diversity of Microcystis.* Environmental Microbiology, 2021. **23**(12): p. 7278-7313.

6. Briand, E., et al., *Evidence of the Cost of the Production of Microcystins by Microcystis aeruginosa under Differing Light and Nitrate Environmental Conditions.* PLoS ONE, 2012. **7**(1): p. e29981.

7. Stark, G.F., Truchon, A. R. , Dittmann, E. , Wilhelm, S. W. , *Closed circular genome sequence of a Microcystis aeruginosa PCC7806 -mcyB (UTK) nontoxic mutant.* Microbiology Resource Announcements, 2023.

8. Cannon, J.A. and T.B. Reynolds, *Synergistic Mutations Create Bacillus Subtilisin Variants with Enhanced Poly-l-Lactic Acid Depolymerization Activity.* Biomacromolecules, 2023. **24**(3): p. 1141-1154.

9. Salomon, E., et al., *Balancing photosynthetic electron flow is critical for cyanobacterial acclimation to nitrogen limitation.* 2013. **1827**(3): p. 340-347.

VITA

Gwendolyn F. Stark spent a majority of her youth in Bethlehem, PA. After high school, Gwen explored various occupations, having briefly worked at Disney World, before joining the Conservation Corps of Minnesota. From her time spent with the Conservation Corps, Gwen decided she wanted to go to college and study environmental biology. This was how she ended up at SUNY-ESF, where she eventually got her Bachelor of Science in Conservation Biology in 2017. After college, she had a few more odd jobs before joining Dr. Steven Wilhelm's lab at the University of Tennessee in 2019. Outside of lab, Gwen enjoys cooking, bodies of water, and outdoor beverages.