A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/89873

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Viral infection of marine picoplankton under nutrient depletion conditions: pseudolysogeny and magic spot nucleotides
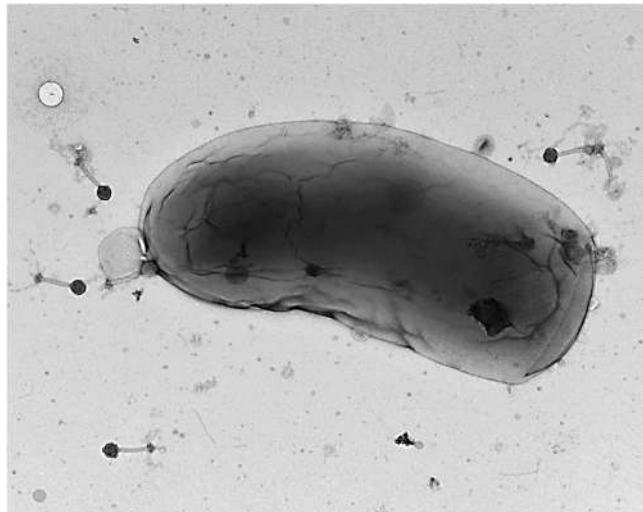
Branko Rihtman

Supervisors

Prof. David J. Scanlan

Prof. Martha R.J. Clokie

A thesis submitted for the degree of Doctor of Philosophy

School of Life Sciences, University of Warwick

September 2016

Contents

Acknowledgements

I would like to thank my supervisor Prof. Dave Scanlan for the incredible support he has given me over the course of my PhD, support that has started from my first email to him and hasn't stopped since. It is an incredible privilege to start a PhD degree with an absolute knowledge that your supervisor will do everything to navigate you to the end in the best possible way. Thank you for the opportunity to learn from you and work with you over the past few years. You have taught me a lot.

Additional thanks goes to Prof. Martha Clokie, who may have not been physically close to the lab but her support and guidance was most definitely felt from Leicestershire. I appreciate your advice, all your help and your friendship.

I am forever grateful to the University of Warwick and the Chancellor's International Scholarship for my PhD funding, hoping that my future career will justify the trust you have put in me and my research.

Even though my name is on the front page of this work, there are many people who made sure I completed it. Firstly, my "little sister" Blanca – thank you for being you, for all our talks, for staying with me through overnight experiments, for being my professional, moral and mental support. You can never be too much Latina ☺. I am grateful for Fran's pick-me-up talks, her vast knowledge and calm confidence and her ability to pave her own path, where none existed before. Rich, who has both been there for my beginnings and arrived in time to save me in the end, I do not know how to thank you enough, mate. Joseph, Mar, Tamsin, Sabine, Victoria, Carolina, Monica, Chris, Bex, Sally, Emily, Joe, Holly, Alexia, Tom, Gemma and all the other present and past members of C126 and MIU – you are all inseparable part of my PhD, my support group and my friends and I am forever thankful for that.

To my doppelganger Andy, I am forever grateful for all your tremendous help and support. I am very glad you were not true to your word from my first week of the PhD. In spite of your advanced age, you have managed to be an incredible friend and mentor and I am looking forward to our future endeavours.

Dragi mama, tata i Ančo, bez snažnih korijena, ni drvo se ne bi moglo uzdići u visine. Vi ste moje korijenje i bez vas ovoga ne bi ni bilo. Hvala za svu ljubav i pordršku.

To my Motke, thank you for traveling with me on this bumpy road. I could not have wished for a better life companion and friend. It takes a lot of patience to support a scientist and you have proven to have enough of it for both of us. I love you always, 8 🌼 +

Finally, to my sweet Noah and Naomi, everything I do, I do with the two of you in front of my eyes. I am a better scientist because of the two of you and I am grateful for your patience and understanding of the need to share your tata with his thesis. I hope you will want to read it some day and see the beauty and wonder that has captivated and amazed me for the last 4 years and will hopefully continue to do so like it does now.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submittedin any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author, except in the cases outlined below:

- Library preparation and MiSeq sequencing of cyanophage genomes described in Chapter 4 – performed by Dr. Andrew Millard, Warwick Medical School.

- Purification of *Synechococcus* WH7803 MazG overexpressed protein from insoluble fraction described in Chapter 5 – performed by Dr. Sabine Bowman-Grahl, School of Life Sciences, University of Warwick.

- Temporal clustering and Silhouette analysis done in Matlab, described in Chapter 6 – performed by Dr. Richard Puxty, School of Life Sciences, University of Warwick.

Summary

Cyanobacteria are major players in marine biogeochemical processes, primarily CO_2 fixation via oxygenic photosynthesis, and nitrogen cycling. These phototrophs occupy a variety of oceanic niches, with their distribution and abundance being shaped by a range of abiotic (e.g. temperature, light, nutrients) and biotic factors (e.g. grazing and virus infection). Viruses infecting cyanobacteria are termed cyanophage. During evolution these cyanophages have acquired an arsenal of 'auxilliary metabolic genes' (AMGs), often horizontally acquired, which can influence the metabolism of the infected host thereby optimising viral production. Cyanophage infection of their host under sub-optimal growth conditions can lead to deleterious effects on infection success. For example, cyanophage S-PM2 infection of *Synechococcus* under P-deplete conditions causes a delayed latent phase and decreased burst size – a process termed pseudolysogeny.

In this thesis I set out to provide a molecular understanding of pseudolysogeny, at the same time hypothesising that specific cyanophage AMGs help to avoid the negative effects of sub-optimal host growth conditions on the infection process. In support of this, infection by cyanophages that possess putative P-stress related AMGs do not show signs of delay (Chapter 3), and the presence of these genes in cyanophage genomes correlates well with the prevailing P conditions in the temporal and spatial niches from which these cyanophages were isolated (Chapter 4). Meanwhile, in a cyanophage lacking such nutrient stress genes, and thus entering pseudolysogeny during infection under P-deplete conditions, transcriptional profiling showed retardation in the timing of known cyanophage temporal gene expression clusters (Chapter 6). Moreover, a significant increase in expression of several cyanophage early genes involved in DNA replication was also observed under these P stress conditions compared to infection of a P-replete host. Quantitation of intracellular cyanophage DNA showed that while levels were generally lower under P-deplete conditions, the rate of DNA replication between P-replete/deplete conditions was similar. The observed increased expression of cyanophage genes involved in DNA replication during the early stages of infection may thus be an evolved response to compensate for decreased levels of intracellular phosphate experienced under these conditions (Chapter 6).

Overlaid on top of specific bacterial nutrient stress responses is the 'stringent' response, mediated by the alarmone molecule (p)ppGpp, a process which occurs under prolonged nutrient stress and in late stationary phase. A bacterial gene *mazG*, encodes a pyrophosphatase which participates in (p)ppGpp homeostasis. Interestingly, a *mazG* orthologue is found as part of the cyanomyovirus core genome, suggesting that cyanophages attempt to alter intracellular signalling during the course of infection. The stringent response has been shown to have a particularly negative effect on phage replication, with (p)ppGpp levels in a cyanobacterial host being previously shown to be dramatically reduced under phage infection. In this thesis I show that the *Synechococcus* host *mazG* is dispensable for growth under normal laboratory conditions. However, this *Synechococcus mazG* mutant shows a modified cyanophage infection profile, slower and less productive, compared to the WT, under P-deplete conditions (Chapter 5). Furthermore, comparison of enzymatic activity of host and cyanophage MazG showed that the viral orthologue exhibits an increased affinity towards GTP, compared to the host protein and a general preference towards G and C nucleotides (Chapter 5), possibly reflecting the low GC content of cyanophage genomes. Thus, the cyanophage and host MazG may have additional functions in phosphate metabolism and controlling DNA integrity, a hypothesis strengthened by experimental evidence for the cyanophage *mazG* being over-expressed under P-deplete conditions (Chapter 6).

Taken together, data presented in this thesis demonstrates a general strategy by cyanophages to acquire host genes involved in modification of central metabolism or that regulate host signalling. Furthermore, once acquired, cyanophage genes appear to have evolved divergent functions to suit specific differences in genome content, compared with their host, as well as mechanisms to regulate transcription of these genes in response to external nutrient stimuli. Thus, this study expands our view of lytic phages, and suggests sophisticated mechanisms occur for overpowering their hosts under a range of infection conditions. This new information provides a mechanistic understanding of viral infection in a ubiquitious primary producer under environmetally relevant conditions, and will undoubtedly improve our ability to understand and model biogeochemical cycling performed by these key marine phototrophs in a more accurate manner.

List of abbreviations:

ABC	ATP-binding complex	Mbp	Megabase pairs
AMG	Auxilliary metabolic gene	MOI	Multiplicity of infection
ASW	Artificial Sea Water	MPN	Most probable number
ATP	Adenosine triphosphate	mRNA	Messenger RNA
BAM	Binary alignment map	NCBI	National Centre for Biotechnology Information
BLAST	Basic Local Alignment Search Tool	OD	Optical density
bp	Base pair	ORF	Open reading frame
CDS	Coding DNA sequence	PCR	Polymerase chain reaction
Ct	Crossing threshold	PEG	Polyethylene glycol
CTP	Cytidine triphosphate	PFU	Plaque forming units
DCMU	3-(3,4-dichlorophenyl)-1,1-dimethylurea	Pi	Inorganic phosphate
DMSO	Dimethyl Sulfoxide	p-NP	*para*-nitrophenyl
DNA	Deoxyribonucleic acid	p-NPP	*para*-nitrophenyl phosphate
FDR	False discovery rate	PSII	Photosystem II
Fm	Maximal fluorescence	qPCR	Quantitative PCR
Fv	Variable fluorescence	RNA	Ribonucleic acid
g	Grams	RNASeq	RNA sequencing
g	Relative centrifugal force	SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
GTP	Guanosine tri-phosphate	SNP	Single nucleotide polymorphism
IAA	Iso-amyl alcohol	TSS	Translation starting point
JTT	Johnson-Taylor-Tamura model of amino acid substitutions	UTP	Uridine tri-phosphate
LB	Lysogeny broth	VLP	Virus-like particle
		WT	Wild-type

List of Figures and Tables

List of Appendices

Chapter 1:

Introduction

1.1 Importance of Cyanobacteria

In recent years it has become apparent that the microbial world is the dominant player in Earth's biosphere, both in terms of abundance and in the prominence of microorganisms in shaping biotic and abiotic parameters. When considering the influence that microbes exert on abiotic conditions on Earth, not many microbial phyla have played, and are still playing, such a significant role as Cyanobacteria.

Cyanobacteria are thought to be responsible for the current composition of the atmosphere, likely being responsible for the advent of oxygenic photosynthesis which radically changed the redox conditions on Earth, enabling the evolution of a cornucopia of metabolic processes that we find today in all living organisms. They are considered to be the oldest known organisms on Earth, dating back more than 3 billion years (Schopf, 1993). Although they remain important players in photosynthetic processes in the ocean today, they have also played a key role in the evolution of photosynthesis in other organisms. Thus, the endosymbiotic hypothesis attributes the origin of plastids in higher plants to the incorporation of a cyanobacterial ancestor into early eukaryotic cells, thus paving the way for the emergence of modern plants (Ochoa de Alda, *et al.*, 2014).

Cyanobacteria can be found in almost all ecological niches, ranging from extreme environments (Miller, *et al.*, 2007; Jungblut, *et al.*, 2010), to soils (Hagemann, *et al.*, 2015) and the global ocean (Scanlan and West, 2002), where they shape major biogeochemical processes, providing the chemical 'engines' for nutrient fluxes and influencing mineralization of macro- and micro nutrients.

Aquatic cyanobacteria are predominant in both freshwater and saline environments, where they are important primary producers and major players in the global carbon cycle, as well as performing an important role in nitrogen metabolism. The latter involves participation both in nitrogen assimilation (Flores and Herrero, 1994; Capone, *et al.*, 2005; Luo, *et al.*, 2012; Martínez-Pérez, *et al.*, 2016), but more importantly in anoxic nitrogenase-dependent fixation of atmospheric dinitrogen, occurring either in specialized cellular compartments – heterocysts (Bohme, 1998; Flores, *et al.*, 2015) or via temporal separation of oxygenic and anoxic metabolic processes (Capone, *et al.*, 1997; Mohr, *et al.*, 2013). In freshwater environments, cyanobacteria potentially have an even more prominent role by directly influencing biological diversity, since many freshwater cyanobacterial taxa (such as *Microcystis* and *Aphanizomenon*) are capable of producing powerful toxins that affect the composition and

dynamics of algal blooms (reviewed in (Paerl and Otten, 2013), and even having a significant impact on human health (Carmichael, 2001).

The primary influence of marine cyanobacteria on global biogeochemical processes is via their participation in photosynthetic carbon dioxide fixation (Pierce and Omata, 1988; Jardillier, *et al.*, 2010; Flombaum, *et al.*, 2013). Cyanobacteria are considered to be the most numerous photosynthetic organisms on Earth, with global numbers estimated at $3x10^{27}$ and $7x10^{26}$ cells for the two main marine cyanobacterial genera – *Prochlorococcus* and *Synechococcus*, respectively (Flombaum, *et al.*, 2013). Indeed, it has been estimated that half of the global primary production occurs in the oceanic photic zone with the cyanobacterial contribution estimated at 25% in certain regions (Burkill, *et al.*, 1993; Veldhuis, *et al.*, 1997; Jardillier, *et al.*, 2010).

1.2 Marine picocyanobacteria: Prochlorococcus and Synechococcus

Prochlorococcus and *Synechococcus* represent the major marine picocyanobacteria (comprising organisms <2µm in size). The most numerous genus, *Prochlorococcus*, dominates the tropical/sub-tropical, and mostly oligotrophic, waters found between 40°N and 40°S (Bouman, *et al.*, 2006) whilst *Synechococcus* is virtually ubiquitous, present in waters ranging from 2-3°C to >30°C, but dominating temperate, mesotrophic niches (Li, 1998; Tai and Palenik, 2009; Scanlan, 2012; Farrant, *et al.*, 2016) (Figure 1.1).

In some ecosystems, succession patterns between *Prochlorococcus* and *Synechococcus* mirror the cycling between the oligotrophic conditions of water stratification and nutrient-rich conditions of deep mixing (Lindell and Post, 1995), further stressing the speciation to different nutrient conditions these two genera have undergone throughout their evolution. In terms of local speciation within each genus, *Prochlorococcus* have adapted to take advantage of the different environmental conditions found down the vertical profile of a water column, with low light-, and high light-adapted ecotypes discernible (West and Scanlan, 1999; Johnson, *et al.*, 2006). In contrast, *Synechococcus* clade abundance corresponds to horizontal differences in nutrient conditions, with three major sub-clusters divided into more than 10 separate clades (Figure 1.3) (Zwirglmaier, *et al.*, 2008; Scanlan, *et al.*, 2009; Mazard, *et al.*, 2012; Sohm, *et al.*, 2016). These clades roughly correspond to the different ecological niches that they occupy – clade III being most abundant in ultra-oligotrophic waters, while clades I and IV dominate nutrient-rich coastal, and mesotrophic open-ocean, waters (Palenik, *et al.*, 2003; Palenik, *et al.*, 2006), for example. Such 'speciation' has left its mark on the genomic content of each genus, with differences in the genetic arsenal available to each, particularly for dealing with
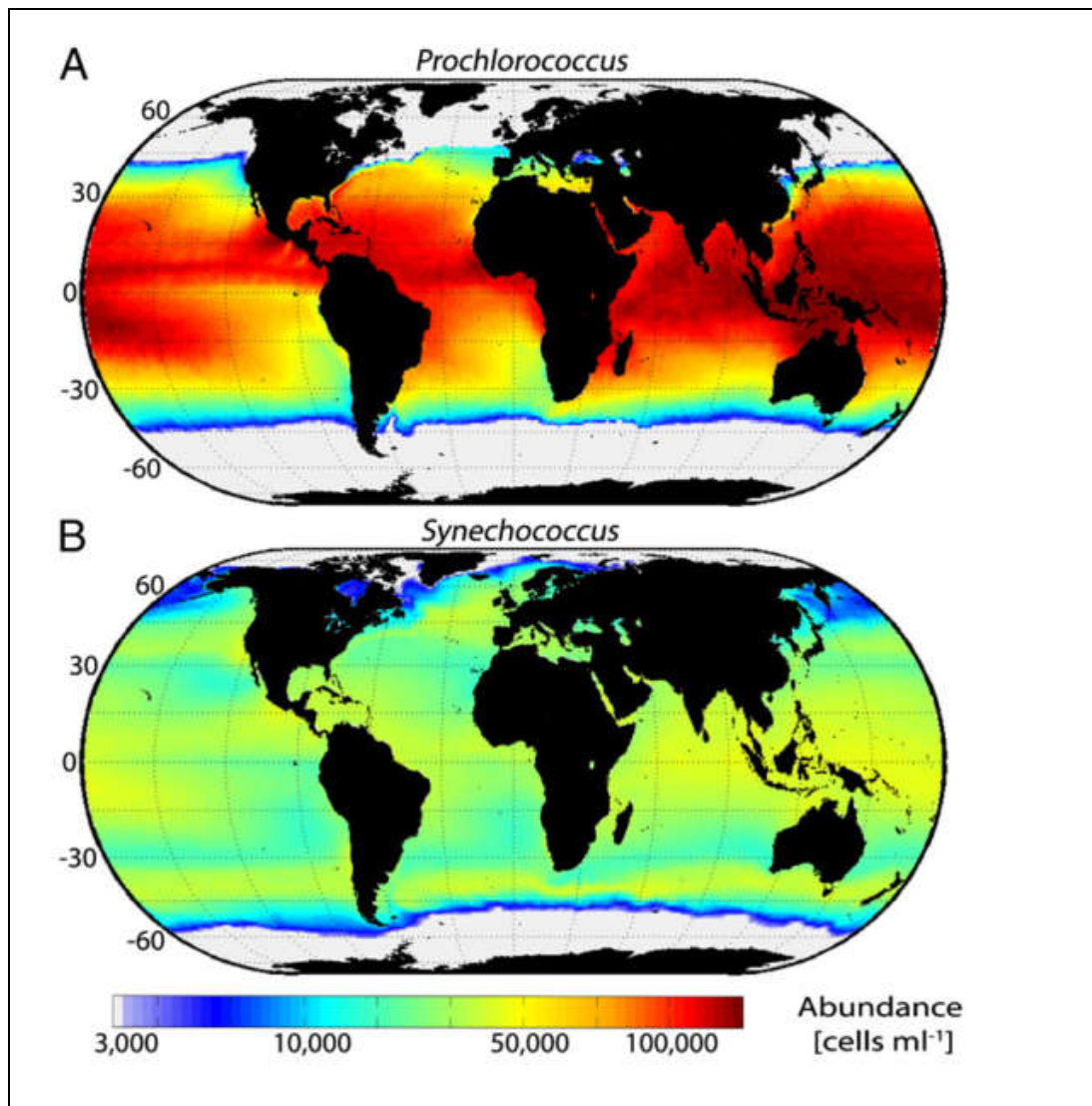
Figure 1.1: *Prochlorococcus* (A) and *Synechococcus* (B) global abundance. From (Flombaum*, et al.*, 2013)

the specific nutrient stresses they are exposed to in the various niches they occupy (Palenik*, et al.*, 2003; Palenik*, et al.*, 2006; Kettler*, et al.*, 2007; Dufresne*, et al.*, 2008;). These differences are particularly pronounced in *Prochlorococcus*, where 'speciation' to low-nutrient environments has resulted in extreme genome streamlining via the shedding of genes involved in e.g. light acclimation and nutrient acquisition, among others (Rocap*, et al.*, 2003; Dufresne*, et al.*, 2005; Partensky and Garczarek, 2010).

This has led to *Prochlorococcus* genomes being the smallest of any free-living photosynthetic organism on Earth (Dufresne*, et al.*, 2003; Kettler*, et al.*, 2007), akin to its heterotrophic bacterial counterpart *Pelagibacter* (in particular the SAR11 clade) (Giovannoni*, et al.*, 2005), meaning that *Prochlorococcus* research is critical for understanding the minimal genomic requirements for a photoautotrophic lifestyle.

As well as ecological differences, *Prochlorococcus* and *Synechococcus* also differ in their light-harvesting strategies. Thus, while *Synechococcus* possesses a phycobilisome light harvesting antennae typical of most other cyanobacteria (Scanlan*, et al.*, 2009), *Prochlorococcus* lacks this structure and with the light harvesting apparatus containing unusual divinyl derivatives of chlorophyll *a* and chlorophyll *b* (Chla_2 and Chlb_2 (Chisholm*, et al.*, 1992; Steglich*, et al.*, 2003). This enables *Prochlorococcus* to take advantage of the blue light wavelengths that dominate oligotrophic waters, particularly at the bottom of the photic zone. However, the existence of both high light- and low light-adapted *Prochlorococcus* ecotypes demonstrates its adaptability to take advantage of the whole photon flux spectrum throughout the photic layer of the water column (Moore*, et al.*, 1998).

Phycobilisome complexes in *Synechococcus* also vary within the genus, being dividing into three main types based on the pigment content: Type 1 which lacks phycoerythrin, Type 2 possessing one kind of phycoerythrin and Type 3 containing two types of phycoerythrin and with all Type 3 strains possessing both phycoerythrobilin and phycourobilin chromophores (Six*, et al.*, 2007). These differences in the chromophore content of the antennae are responsible for the striking range of colours seen in different *Synechococcus* strains when grown in pure culture (Figure 1.2). In addition to differences in appearance, particular pigments in their antennae provide specific *Synechococcus* strains with the ability to take advantage of distinct wavelengths of light, allowing them to thrive at different depths and under different light regimes (Palenik, 2001).

Figure 1.2: Representative cultured strains of the major pigment types (1-3) and subtypes (3a-d) of marine *Synechococcus*. Numbers represent different pigment types (adapted from (Six*, et al.*, 2007).

1.3 The marine environment and picocyanobacterial niche partitioning

Even though it may seem counterintuitive, the ocean is a highly diversified ecosystem, with distinct ecological niches, differing in conditions both vertically and horizontally. Factors like wind speed, differential solar radiation, proximity to land, ocean currents, etc., all influence the abiotic parameters that, in turn, define the composition of organisms that are present in different oceanic niches (Mann and Lazier, 1996). Marine bacteria are influenced by all these factors, but not all of them shape biological communities equally. Similarly, different taxonomic groups are influenced by discrete abiotic parameters, creating tremendous marine microbial diversity (Sunagawa, *et al.*, 2015; Farrant, *et al.*, 2016).

The abiotic factors that are considered to have the strongest influence in shaping marine microbial communities in general, and marine cyanobacteria in particular, are macro- and micronutrient levels, the intensity and quality of light, temperature and predation by bacteriovores or parasites (Scanlan, 2003). As mentioned above, *Prochlorococcus* has undergone 'speciation' based on the differing light intensities down the vertical gradient of the photic zone, as well as the temperature gradient (Moore, *et al.*, 1998; West and Scanlan, 1999; Johnson, *et al.*, 2006; Ribalet, *et al.*, 2015; Chandler, *et al.*, 2016; Larkin, *et al.*, 2016). *Synechococcus*, on the other hand, has evolved into specific genetic lineages based on the horizontal gradients of temperature, nutrients and light (see Zwirglmaier et al., 2008, 2009; Dufresne et al., 2008; Mazard et al., 2012; Sohm et al., 2016). These abiotic factors not only vary geographically, but also temporally, with almost all ocean areas undergoing interchanging periods of deep mixing, during which nutrients from the benthic ocean are brought upwards towards the photic zone, and stratification, a time when the photic layer is rapidly depleted of nutrients (Cushing, 1989) primarily due to primary production (Lindell and Post, 1995). The importance of understanding biogeochemical processes occurring in nutrient poor areas of the ocean also stems from their sheer size – almost 75% of the total ocean area is occupied by oligotrophic waters (Lewis, *et al.*, 1986) in which *Prochlorococcus* and *Synechococcus* populations dominate the primary producer community (Flombaum, *et al.*, 2013).
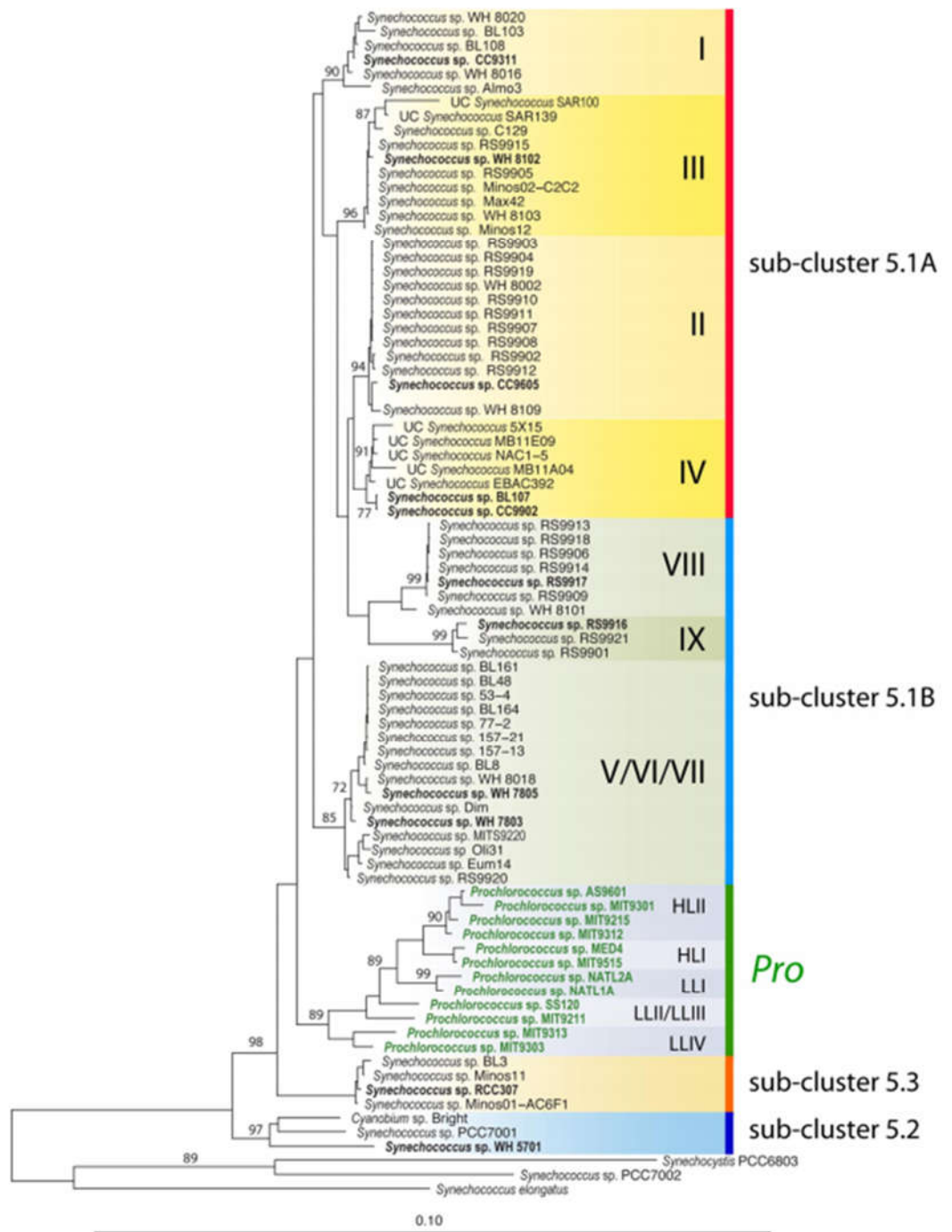
Figure 1.3: 16S rRNA-based neighborhood-joining tree, showing the major *Synechococcus* and *Prochlorococcus* clades and sub-clusters. Adopted from Scanlan*, et al.*, (2009)

1.4 Phosphorus in the marine environment

Of all nutrients required for growth in the marine environment, the availability of nitrogen (N), phosphorus (P) and iron (Fe) has been found to have the most profound impact on the ecological success of marine cyanobacteria in oceanic niches (Falkowski, 1997; Mills, *et al.*, 2004; Martiny, *et al.*, 2006; Scanlan, *et al.*, 2009; Rusch, *et al.*, 2010; Sohm, *et al.*, 2016). While N and Fe pools are replenished from external sources - N via diazotrophic bacteria capable of fixing atmospheric nitrogen (Capone, *et al.*, 1997; Bohme, 1998), and Fe mostly originating from continental dust via aeolian or fluvial deposition (Poulton and Raiswell, 2002) – sources of phosphorus are scarce. In coastal areas, P is supplied to the ocean via fluvial deposition of anthropogenic chemicals (mostly fertilisers, (Bennett, *et al.*, 2001)), while in the open ocean, and especially in the ocean gyres, these sources are not available. In situ measurements have found extremely low levels of inorganic P in some areas of the ocean, e.g. as low as 0.2 nM in the North Atlantic gyre (Wu, *et al.*, 2000) (Figure 1.4). Therefore, P plays an important role in shaping the structure of microbial communities in these vast ocean areas (Mills, *et al.*, 2004). Deposition of Fe and temperature increases have encouraged stable stratification of the photic zone in the Atlantic Ocean, conditions which favour the occurrence of nitrogen fixation (Karl, *et al.*, 1997; Sachs and Repeta, 1999). This observation hints towards tipping the N:P balance heavily towards high nitrogen availability, which would in turn make P the main limiting macronutrient in such environments (Sañudo-Wilhelmy, *et al.*, 2001; Ammerman, *et al.*, 2003). These ideas were confirmed both by in-situ enrichment experiments (Vaulot, *et al.*, 1996; Labry, *et al.*, 2002; Van Wambeke, *et al.*, 2002), as well as by measurements of N and P fluxes in the same ecosystem (Krom, *et al.*, 2004). This has demonstrated that primary production across a wide expanse of oligotrophic, stratified waters (e.g. in the Sargasso Sea, Mediterranean Sea, and North Atlantic gyre) is mainly limited by P availability.

1.4.1 Microbial responses to P stress

Cellular P requirements in marine cyanobacteria are mostly within genetic material (DNA, RNA) and membrane phospholipids (e.g. see van Mooy et al., 2006; Mazard et al., 2012). Thus, P content measurements in *Synechococcus* and *Prochlorococcus*, based on genome size, suggest that as much as half of the total cellular P can be found in DNA, with the remainder divided between RNA, phospholipids and other phosphorylated molecules (phosphorylated proteins, nucleotide signalling and energy metabolism molecules, etc.) (Bertilsson, *et al.*, 2003; Heldal, *et al.*, 2003).

Figure 1.4: Annual average sea-surface levels of inorganic phosphate. Image from http://www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html
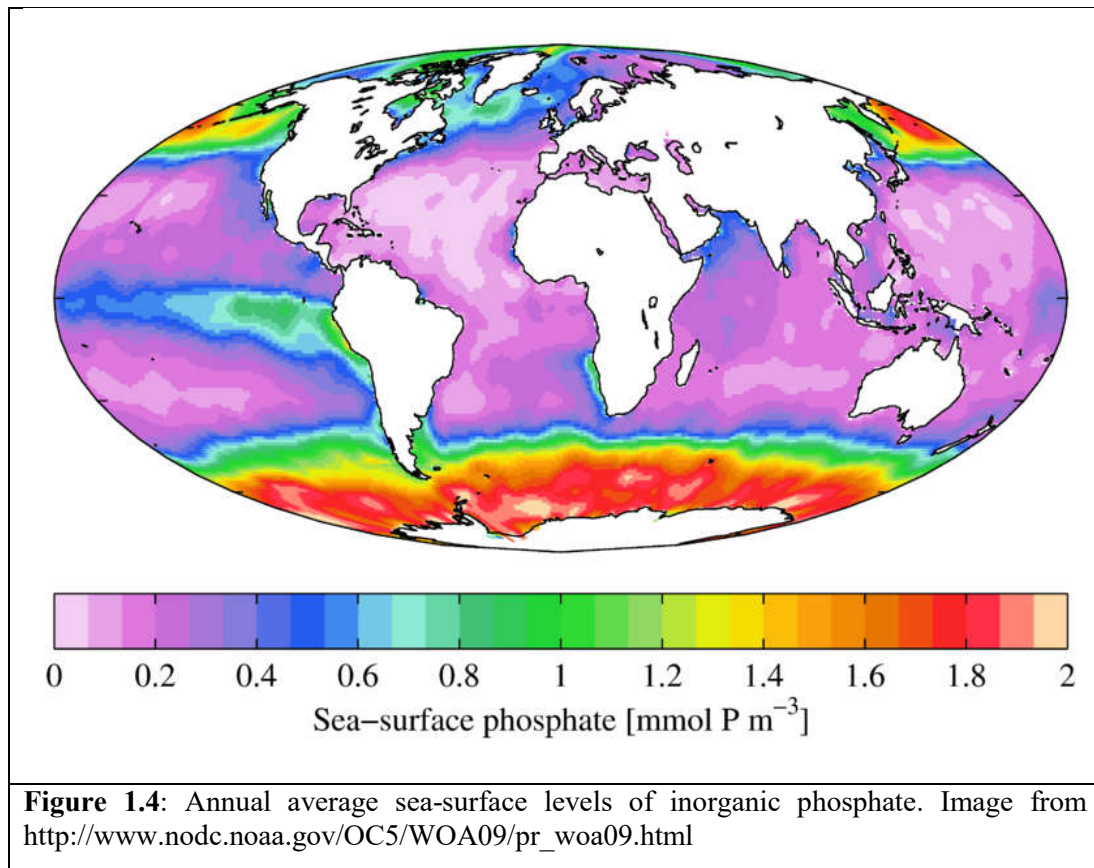
Primary producers obtain most of their P requirements from the inorganic pool of P, mostly found as orthophosphate (Paytan*, et al.*, 2003). However, they are also capable of incorporating organic P moieties (e.g. see Mazard *et al*., 2012). Adaptation to low P levels in oligotrophic and hyper-oligotrophic waters has resulted in alterations to cellular processes and even in the structure of the cell itself. For example, one strategy to reduce cellular P requirements both in primary producers and marine heterotrophic bacteria is via modifying the content of membrane lipids, switching from P-demanding phospholipids to non-P containing lipids, such as betaine, glyco- or ornithine-containing lipids in heterotrophic bacteria (Carini*, et al.*, 2015; Sebastian*, et al.*, 2016) or sulfolipids in picocyanobacteria (Van Mooy*, et al.*, 2006).

In general, bacteria have four different strategies at their disposal to deal with P stress:

1.   **Import** – increasing the import and scavenging of inorganic phosphate (Pi) in the environment. Most bacteria perform this by inducing the expression of high-affinity Pi transport systems. In cyanobacteria, this system is encoded by the pst transporter, comprising PstS – the periplasmic Pi binding protein and *pstABC* – encoding the membrane bound and ATPase components of the high affinity Pi-specific ABC transporter (Scanlan*, et al.*, 1993).

2.   **Use of stored phosphate** – when P is abundant in the environment bacteria transport more P into the cell than is required for cellular metabolism and growth. This excess P is used to synthesise polyphosphate (a linear polymer of orthophosphate units linked by phophoanhydride bonds) bodies via the activity of polyphosphate kinase (Albi-Rodríguez and Serrano, 2008). When P subsequently becomes depleted from the environment, cells use the *ppx* gene encoding an exopolyphosphatase to recycle Pi from the polyphosphate bodies and make it available for cellular metabolism (Gomez-Garcia*, et al.*, 2003).

3.   **Use of alternative P sources** –organic molecules containing Pi, typically phosphoesters, can be hydrolyzed outside the cell by secreted alkaline phosphatases or imported into the cell and hydrolyzed by intracellular phosphatases (Carini*, et al.*, 2014). Another type of organic P molecule are phosphonates which can be imported into the cell and hydrolyzed using a C-P lyase enzyme in an energetically taxing process (Dyhrman*, et al.*, 2006). However, utilization of these compounds by marine microbes has been relatively poorly studied, certainly in comparison to the potential importance of these organic P moieties in the ocean. Thus, some studies estimate that the vast majority of the marine P pool found in the photic zone (~75%-80%) is present as dissolved organic phosphate (DOP) (Benitez-Nelson, 2000), with phosphoesters and phosphonates making up 80%-85% and 5%-10% of the DOP pool, respectively (Young and Ingall, 2010).

4.   **Reducing cellular P requirements** – as mentioned above, *Prochlorococcus* and *Pelagibacter* have greatly reduced their genome size (Giovannoni*, et al.*, 2005; Kettler*, et al.*, 2007), thus lowering the amount of P necessary for replication of high P-content DNA and RNA molecules. Additional mechanisms include reducing the number of active ribosomes, which require P via the incorporated rRNA (Karpinets*, et al.*, 2006) and, as mentioned above, modifying the content of membrane phospholipids to non-P containing forms (Van Mooy*, et al.*, 2006).

Genes involved in the response to P-depletion are controlled by a two-component system, PhoB-PhoR. Here, PhoR, a histidine kinase, is involved in sensing low levels of ambient Pi, which in turn phosphorylates the DNA-binding protein PhoB, causing induction of the Pho regulon – comprising a battery of genes encoding proteins involved in the above outlined mechanisms of response to low Pi levels in the environment (Bachhawat*, et al.*, 2005). It is noteworthy that while phylogenies of marine cyanobacterial genes belonging to core metabolic pathways, such as those involved in photosynthesis, largely correspond to the phylogenetic classification based on 16S rRNA genes, both the abundance and phylogeny of

Pho regulon genes correlates better to the Pi availability of their environment, suggestive of some of these genes being acquired by horizontal gene transfer (Martiny, *et al.*, 2006). This adds to the evidence that P availability is an important factor shaping marine picocyanobacterial community structure and genome composition.

1.5 Bacteriophages

Bacteriophages (viruses infecting bacteria) are considered to be the most abundant biological entity on Earth, with numbers in marine waters estimated to be about 10^7 VLP/ml (on average over the global ocean, (Suttle, 2005). Ever since their discovery in the early 20th century (Twort, 1915; d'Herelle, 1917) bacteriophages have been an important model organism in microbiology, studies of which have provided us with insights into a large number of microbial processes. Bacteriophages are natural predators of bacteria, serving as both population control factors and diversity amplifiers (Williams, 2013). They infect bacteria via adsorption to a variety of membrane components (Rakhuba, *et al.*, 2010) and injection of viral genetic material into the host cell. From that point, the virus life cycle can take two major paths – lysogenic or lytic (Figure 1.5)
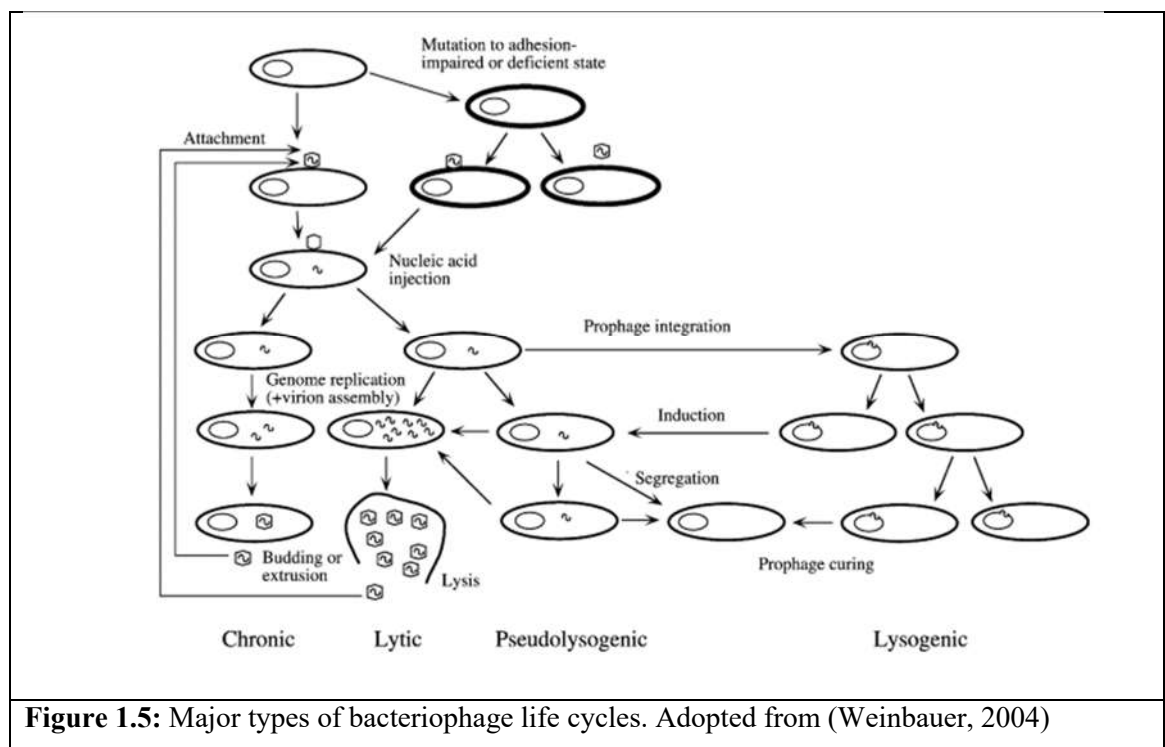
Figure 1.5: Major types of bacteriophage life cycles. Adopted from (Weinbauer, 2004)

During the lysogenic life cycle, viral genetic material gets incorporated (Freifeld and Meselson, 1970; Weinbauer, 2004) into the host genome. There it can lay dormant until the conditions are ripe for the production of a new generation of phages, a process that is tightly

controlled (Weinbauer, 2004), and start transcribing phage genes. These genes encode bacteriophage structural elements, into which the viral genetic material is packaged and new virions burst out from the infected host cell, causing cell lysis.

The alternative pathway – the lytic cycle – does not include incorporation of the phage genetic material into the host chromosome, but rather viral genes are expressed immediately after infection, new virions form in the cytoplasm and the progeny burst out of the bacterial host, killing it in the process (Mandeville, *et al.*, 2003). The effect of lysogeny on the host life cycle (even without proceeding to lysis) is self-evident since every infection would provide the host cell with a new arsenal of genes encoded from the phage genome potentially affecting various metabolic processes (Mann, *et al.*, 2003), including increasing virulence (Canchaya, *et al.*, 2003) or providing antimicrobial resistance capabilities to the infected host (Chan, *et al.*, 1972). Furthermore, every integration into the host genome, followed by excision, is an opportunity to transfer host genes horizontally between closely related (in the case of narrow host-range infecting phage) or evolutionary distant bacterial hosts (Canchaya, *et al.*, 2003).

Bacteriophages can shape bacterial communities in a variety of ways:

1.   Influencing the assemblage of the resident microbial community via:

a.   **Reducing the number of the most successful species** – infection of bacteria by its phage is a function, among other things, of a probability that the phage encounters its host in the environment. Therefore, the more successful (and hence more numerous) a bacterial host is in a certain niche, the higher the chances of it encountering a phage capable of infecting and lysing it. This observation serves as one of the possible explanations to the huge variety of microbial species in every environment – in a "kill the winner" model, phages serve as a factor which keeps the successful bacterial species from overtaking the niche (Thingstad, 2000; Rodriguez-Valera, *et al.*, 2009; Koskella and Brockhurst, 2014).

b.   **By selecting for hypermutators** – one of the main characteristics of phage-host co-evolution is development of resistance. The rate at which resistance evolves in bacterial populations will be counterbalanced by the cost which this resistance brings (Lennon, *et al.*, 2007; Koskella and Brockhurst, 2014). This arms race between the phage and the host not only results in selection of phage-resistant populations, but also increases the proportion of cells capable of acquiring resistance-inducing mutations relatively quickly, thus changing the makeup of bacterial communities (Pal, *et al.*,

2007).

 c. **Serving as vectors of horizontal gene transfer** – as mentioned above, temperate phages which undergo integration into the host genome, often carry genes from the infected host, thus transferring them to cells infected by the next generation of phages (Canchaya*, et al.*, 2003).

 d. **Altering the composition of the dissolved organic nutrient pool** – lysis of bacterial hosts by the infecting phages, can provide an influx of organic material into the aquatic environment, which may in turn fuel the productivity of subsequent microbial populations (Wilhelm and Suttle, 1999; Fuhrman, 1999)

2. **Via influencing bacterial metabolism** – obligately lytic bacteriophages have a short time window to recruit bacterial genes via recombination. However, they can still possess genes which have seemingly originated from their host. These genes, termed Auxiliary Metabolic Genes (AMGs), provide the phage with the capability to influence host metabolism, via enhancing some, and blocking other, metabolic pathways in a way that best benefits the bacteriophage goal of maximizing the number of phage progeny (Breitbart*, et al.*, 2007). These genes can play a direct role in biogeochemical cycles (see section 1.5) and given the relatively high infection rates of environmentally important bacteria (Suttle, 2007) such phage-encoded genes may influence metabolic processes of global importance (Lindell*, et al.*, 2005; Clokie*, et al.*, 2006).

1.6 AMGs in bacteriophages

The variety of AMGs, metabolic pathway-affecting genes found in phage genomes – largely corresponds with the variety of metabolic pathways used by the bacterial host to produce energy (Hurwitz and U'Ren, 2016). The first AMGs in marine bacteriophages were discovered in *Roseobacter*-infecting viruses, in which a potential P metabolism gene, *phoH*, was identified, containing a putative Pho box upstream i.e. a PhoB binding site (Rohwer*, et al.*, 2000). As more marine bacteriophage genomes have been sequenced a clear pattern of the potential function of AMGs has begun to emerge. Thus, it seems that bacteriophages will acquire and maintain within their genomes metabolic genes, most probably originating from their host, which will influence those metabolic steps that represent the bottleneck in the process of producing progeny phage during the course of infection (Angly*, et al.*, 2006;

Sullivan*, et al.*, 2006). Specifically, for cyanophage, there are several key cyanobacterial metabolic processes that are thought, or known, to be altered by infecting cyanophages:

1.   **Photosynthesis and carbon metabolism AMGs** – one of the first marine cyanophages to be described and sequenced, S-PM2, was characterized by possession of the *psbA* and *psbD* genes (Mann, *et al.*, 2003), encoding the D1 and D2 proteins of photosystem II (PSII), respectively (Golden, 1995). These proteins are susceptible to photodamage, and an active repair cycle takes place in which these proteins, and especially D1, are rapidly degraded and replaced with newly synthesised protein. Net loss of PSII activity occurs when the rate of damage exceeds that of repair resulting in photoinhibition (Keren*, et al.*, 1997; Silva*, et al.*, 2003). It is thought that the phage encoded *psbA* and *psbD* genes maintain the photosynthetic competence of the infected host at a time when host transcription and protein synthesis is largely arrested, thus maintaining energy generation in the infected cell and in turn allowing the production of progeny phage. Since their discovery in cyanophage S-PM2, the *psbA*/*psbD* genes have been found in many marine cyanophages (Mann*, et al.*, 2003; Sullivan*, et al.*, 2003; Millard*, et al.*, 2009), are known to be expressed (Clokie*, et al.*, 2006) and translated (Lindell*, et al.*, 2005) during the course of infection and are thought to be influencing marine carbon flux on a global scale (Puxty*, et al.*, 2015a; Puxty*, et al.*, 2016). In addition to PSII genes, additional electron transport genes (Philosof*, et al.*, 2011), Calvin cycle genes (Thompson*, et al.*, 2011), and a complete set of photosystem I genes (Roitman*, et al.*, 2015) have been found either in cultured cyanophage or environmental phage metagenomes. In almost all of the above cases, viral AMGs are slightly different to their host orthologues, a difference which hints towards an altered activity of the resulting cyanophage protein or towards differential control of cyanophage gene expression and/or cyanophage mRNA stability (Puxty*, et al.*, 2015a) (Figure 1.6).
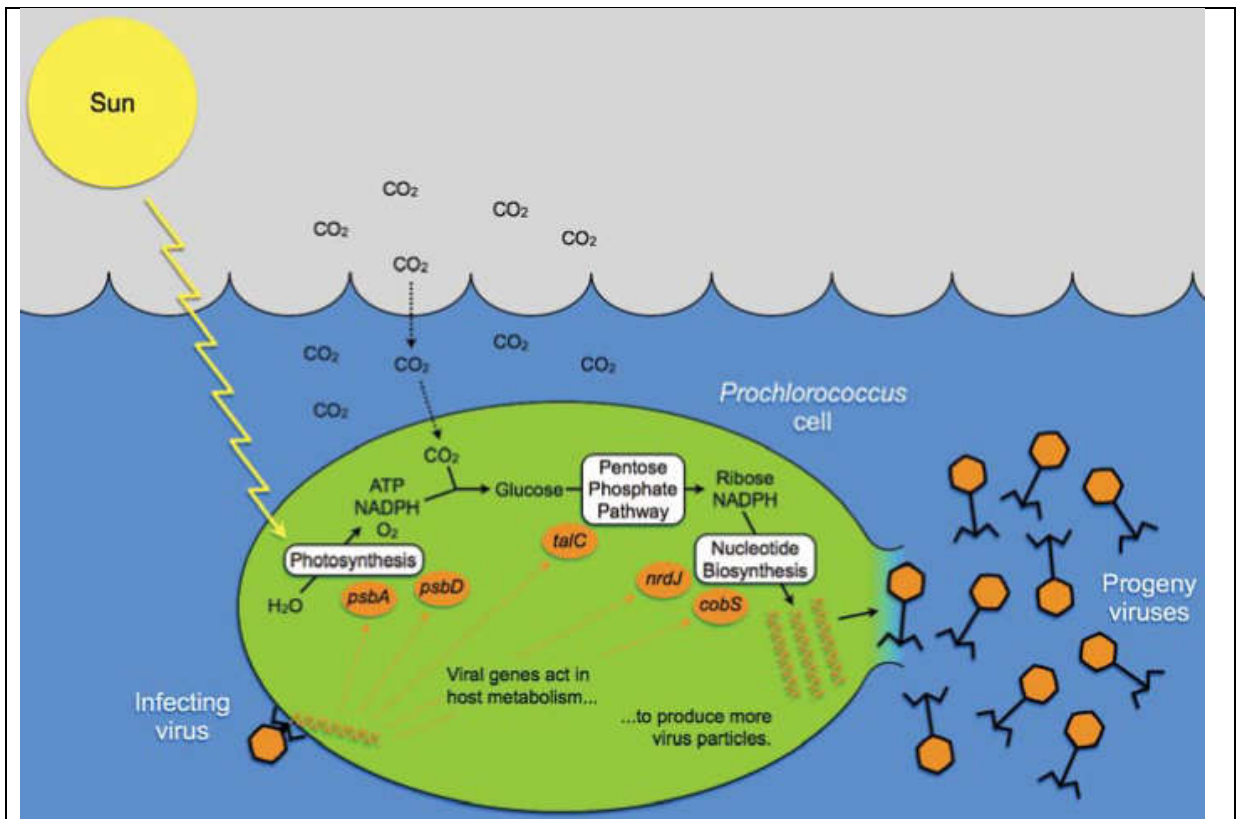
Figure 1.6: Scheme describing phage AMGs affecting different stages of photosynthesis and carbon fixation. Phage AMGs affecting the bacterial carbon metabolism are shown in orange. From (Breitbart, *et al.*, 2007)

2. **Phosphate metabolism** – as outlined above, the availability of Pi is a major environmental factor shaping cyanobacterial community structure (Mills, *et al.*, 2004). Therefore, it is of little surprise that infecting cyanophages will possess AMGs relevant to P metabolism. Until recently, genomic and metagenomic data has been the only line of evidence available to evaluate the importance of P-related AMGs to phage infectivity (Kelly, *et al.*, 2013; Sullivan, *et al.*, 2010). Indeed, the sequencing of cyanophage genomes infecting both *Prochlorococcus* and *Synechococcus* has revealed the presence of several types of P-metabolism related genes:

a. *pstS* **orthologue** – *pstS* is considered to be a P-stress marker in marine *Synechococcus* (Scanlan, *et al.*, 1993). It encodes a periplasmic Pi binding protein responsible for shuttling Pi across the periplasm from the outer membrane porins to the cytoplasmic membrane located ABC-type Pi transporter. PstS expression levels in the environment correlate with low Pi concentrations eliciting P-stress dynamics in picocyanobacterial populations (Fuller, *et al.*, 2005).

b.   ***phoA* orthologue** –encoding a putative alkaline phosphatase thought to be responsible for scavenging Pi from organic P sources and, together with *pstS*, is one of the most up-regulated genes under P-stress conditions in both *Synechococcus* (Tetu*, et al.*, 2009; Ostrowski*, et al.*, 2010;) and *Prochlorococcus* (Martiny*, et al.*, 2006).

c.   ***phoH* orthologue** – a homologue of this gene encodes a putative ATPase, encompassing a component of the P-stress response in *E. coli* (Wanner, 1990). *phoH* is ubiquitous in phage genomes and hence has been suggested to be used as a signature gene for bacteriophages in marine systems (Goldsmith*, et al.*, 2011). The function of the cyanobacterial PhoH homologue is not clear, as its expression is either repressed in *Synechococcus* (Tetu*, et al.*, 2009) or unchanged in *Prochlorococcus* under P-depletion conditions (Martiny*, et al.*, 2006).

Metagenomic studies have found a strong correlation between the presence of P metabolism-related AMGs in viral metagenomes and the P-status of the environment from which they were isolated, with an increased presence in P-depleted waters (Kelly*, et al.*, 2013). *In silico* analysis of promoter sequences upstream from the P-related AMGs in sequenced cyanophage genomes have suggested the presence of *pho* boxes – PhoB binding sequences – hinting towards host P-status as a controlling factor for the transcription of P-related viral AMGs (Sullivan*, et al.*, 2010). Furthermore, when *Prochlorococcus* sp. NATL2A was infected with the T4-like cyanophage P-SSM2 under P-deplete conditions, phage *pstS* and *phoA* (but not *phoH*) were shown to be over-expressed and the host PhoB bound to a *pho* box found in the promoter region of these two genes (Zeng and Chisholm, 2012). Recent RNA-sequencing of the same host-virus pair has shown that *pstS* and a gene of unknown function – *g247* – located upstream and in the same operon as *pstS*, were the only two cyanophage genes significantly over-expressed during infection under P-deplete conditions during a single infection cycle (Lin*, et al.*, 2016).

In contrast to the above *Prochlorococcus* experiments, in *Synechococcus* sp. WH7803 infected with cyanophage S-PM2, a phage that does not possess any of the putative P-regulon related AMGs (Wilson*, et al.*, 1993), the infection was seriously impaired, with the burst size reduced by ~80%, and the burst delayed by 18 hours compared to the regular latent period length of 9 hours under P-replete conditions (Wilson*, et al.*, 1996) (Figure 1.7).

Figure 1.7: Parameters of cyanophage S-PM2 infection of *Synechococcus* sp. WH7803 grown under P-stress conditions. A – the number of plaques appearing over the course of a 8 day period. Bars represent the number of plaques produced following infecting of the *Synechococcus* host grown under P-replete (white) or P-deplete (black) conditions. B – One step growth curve of *Synechococcus* sp. WH7803 infected by cyanophage S-PM2 under P-replete (■) or P-deplete conditions ( ). Adopted from Wilson, *et al.*, 1996.

This delay and reduction in the burst size during P deplete infection places further significance on P as a key environmental factor shaping cyanophage population structure and dynamics. This is especially so since S-PM2 was isolated from relatively nutrient rich coastal waters (Wilson, *et al.*, 1993), and has therefore probably not evolved to be able to circumvent the limitations imposed upon it by P-starvation. The same authors showed that S-BM1, a cyanophage isolated from P-deplete waters off Bermuda (Wilson, *et al.*, 1993), could overcome the P-starvation induced limitation on burst size and length of the latent phase. These results provide further support to the hypothesis that P-related viral AMGs help bacteriophages overcome, through metabolic hijacking, the problems imposed on both the infected host and infecting phage by P-limitation. Since S-PM2 (similar to all other marine cyanophages isolated to-date) is an obligately lytic phage it is interesting to note, that under P-starvation *Synechococcus* sp. WH7803 and the infecting phage seem to be found in an intermediate state between the lytic and lysogenic cycles. Thus, the cyanophage does not integrate into the host chromosome, while the production of the progeny phage is significantly delayed. This intermediate state has been termed pseudolysogeny.

1.7 Pseudolysogeny

In addition to obligately lytic and lysogenic lifestyles, bacteriophages can also exist in an intermediate state between these two options – post-infection, the bacteriophage genome can theoretically be located in the cytoplasm of the infected host, not incorporated into the host genome, subsequently delaying the assembly of the next generation of virions and/or burst of the host cell. This state is often referred to as pseudolysogeny (Ripp and Miller, 1997). While the majority of phage research is performed under laboratory conditions – with optimal growth and nutrient conditions (Ripp and Miller, 1997), the conditions in nature are far from ideal for bacterial growth (Roszak and Colwell, 1987). Studies in T4 phages, infecting a stationary phase *E. coli* host, have shown that viral progeny production in nutrient depleted cells is halted at the middle-stage of phage production, where host DNA has been broken down and nucleotides incorporated into viral DNA, while the synthesis of phage structural components is delayed until nutrients become available (Bryan*, et al.*, 2016). As described above, this is particularly true for marine environments where vast areas of the ocean are oligotrophic, only sustaining the growth of well-adapted microorganisms, with occasional blooms occurring after periodic injections of nutrients (Karl*, et al.*, 1992; Coale*, et al.*, 1996). It was previously shown that under nutrient deplete growth *Pseudomonas* bacteriophage infection is significantly reduced, as measured by reduced burst size, latent phase prolongation and reduction of virulence of lytic phages (Kokjohn*, et al.*, 1991). Additionally, the effective half-life of the phages produced under nutritional starvation conditions was shown to be significantly reduced (Ogunseitan*, et al.*, 1990). Finally, the majority of marine cyanophages isolated so far, as well as those identified from metagenomics samples (Sullivan*, et al.*, 2010) are obligately lytic viruses, both witnessed by observations of their exclusively lytic life cycle in culture and the absence of genes required for lysogeny in their genomes. Therefore, because of the oligotrophic nature of the majority of marine environments, the impact that nutrient starvation has on infection dynamics, and the persistence of bacteriophages in the environment due to the seemingly obligately lytic nature of marine cyanophages, it is of great interest to understand the underlying principles and mechanisms of pseudolysogeny within infected cells.

1.8 The bacterial stringent response

As outlined above, the marine cyanobacteria *Synechococcus* and *Prochlorococcus* thrive in oligotrophic waters, areas characterized by relatively low primary production rates but whose areal extent is large, hence these regions are key to global carbon cycling. In order to deal with

nutrient depletion, cyanobacteria have evolved various responses and adaptations, usually characterized by intricate transcriptional networks and/or post-translational regulation processes (Schwarz and Forchhammer, 2005). Since activation of specialized genes requires energy and cellular resources, the transcriptional control of these responses is usually very tightly regulated, with overlapping transcriptional regulatory networks and cross-talk between different factors and signals (Rodriguez-Garcia, *et al.*, 2009). However, in the majority of bacteria, there are overarching stress response mechanisms, which are activated when the prolonged starvation response e.g. of amino acids, affects the most basic of cellular processes. One such mechanism is called the bacterial stringent response.
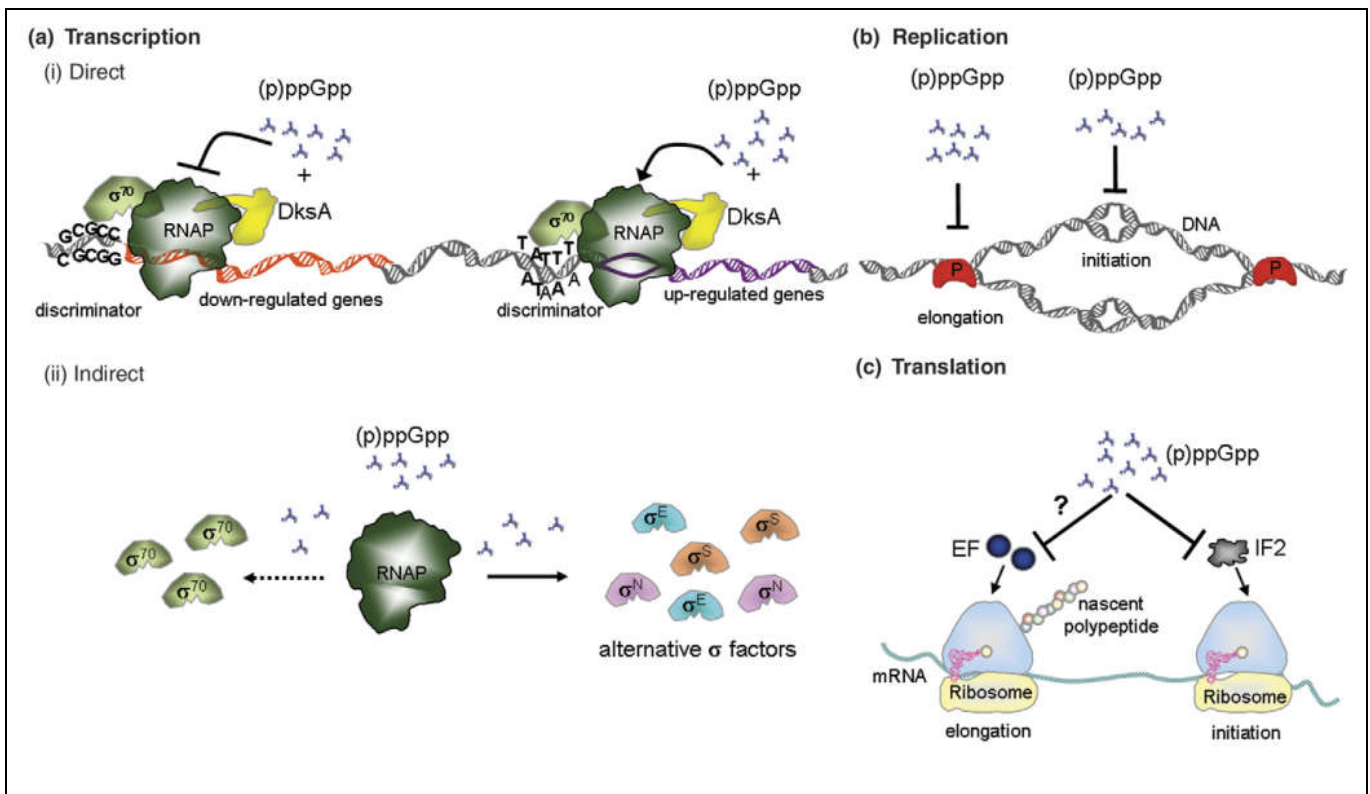
Figure 1.8: The various ways in which (p)ppGpp affects fundamental cellular processes: (a) binding of (p)ppGpp to the RNAP-σ^{70} complex causes destabilisation of the transcription apparatus across a large number of genes, redirecting the transcriptional complexes to stress-related genes; (b) (p)ppGpp affects genome replication via destabilization of open DNA complexes and binding of elongation factors. (c) (p)ppGpp influences peptide translation via binding to elongation and initiation factors, thus reducing their translation. Adapted from (Srivatsan and Wang, 2008).

In Gram negative bacteria, the stringent response is initiated upon nutrient depletion, with the trigger for activation being the stalling of uncharged tRNAs inside the ribosomal complex (Haseltin and Block, 1973; Wendrich, *et al.*, 2002; English, *et al.*, 2011). Due to amino acid depletion (caused by nutrient starvation), uncharged tRNA molecules in the ribosomal A site induce the binding of the RelA synthase protein to the ribosome, where it initiates synthesis

of an 'alarmone' molecule guanosine tetraphosphate (ppGpp) or guanosine pentaphosphate (pppGpp) by transfer of pyrophosphate from ATP to GDP or GTP, respectively (Avarbock, *et al.*, 2000; English, *et al.*, 2011). Synthesis of (p)ppGpp can occur under conditions other than amino acid starvation via a different pathway. Thus, under phosphate, iron, carbon or fatty acid depletion, SpoT, which can serve both as a (p)ppGpp synthase and hydrolase, shifts its activity towards alarmone synthesis and causes accumulation of this stringent response signal in the starved cell (Xiao, *et al.*, 1991). In the case of fatty acid starvation, the shift from the hydrolytic to synthetic activity of SpoT occurs due to binding of the acyl carrier protein (ACP) which modulates the activity of SpoT. SpoT has been observed to interact with G-coupled proteins (Jiang, *et al.*, 2007) which hints towards a versatile role of SpoT in monitoring the metabolic status of bacterial cells and relaying the signal, via (p)ppGpp synthesis, to various stress response mechanisms.

1.8.1 The effects of (p)ppGpp on cellular processes

The alarmone molecule (p)ppGpp affects cellular processes in a variety of ways (Figure 1.8). The most immediate effect of this molecule is on transcription – via binding of the bacterial RNA polymerase it can redirect transcription towards genes involved in the stringent response. Operating in synergy with an additional transcriptional regulator, DksA, binding of (p)ppGpp causes structural changes in the RNAP-σ^{70} complex, which increases its affinity towards promoters of genes responsible for the stringent response (such as tRNA and amino acid biosynthesis promoters) (Paul, *et al.*, 2005). Since under normal growth conditions the open complexes at rRNA promoters are less stable than other genes, (p)ppGpp has an even more destabilising effect on transcription of rRNA (Haugen, *et al.*, 2006), which creates an indirect and negative effect on protein synthesis under stringent response conditions. Additionally, binding of (p)ppGpp to RNAP induces transcription of alternative σ factors, such as stationary phase-specific σ^S (Gentry, *et al.*, 1993), σ^E which is responsible for the outer membrane porin mis-folding response (Costanzo and Ades, 2006) or σ^N, responsible for the nitrogen stress response (Laurie, *et al.*, 2003). While the phenomenon is ubiquitous, the exact mechanism for transcriptional regulation by (p)ppGpp may vary among bacterial phyla. For example, in *Bacillus subtilis*, accumulation of (p)ppGpp inhibits production of GTP, which in turn down-regulates transcription initiation from rRNA promoters (Krasny and Gourse, 2004).

Metabolic control by (p)ppGpp extends beyond transcription and onto translation. In addition to the already mentioned de-regulation of transcription from rRNA promoters, (p)ppGpp has also been shown capable of binding different elongation factors, thus preventing successful formation of initiation complexes (Milon, *et al.*, 2006). The stringent response also affects

genome replication, though with apparently conflicting effects. Thus, in *E. coli* and *B. subtilis*, (p)ppGpp inhibits DNA replication initiation and elongation, respectively (Levine*, et al.*, 1991; Schreiber*, et al.*, 1995), processes which initiates the SOS response. In contrast, in *E. coli* , (p)ppGpp can also preserve genome integrity by resolving transcription/replication conflicts that occur when RNAP complexes stall at DNA lesion locations. (p)ppGpp will destabilize the stalled RNAP, release the DNA and make it available for repair (Trautinger*, et al.*, 2005). It is clear that the role of (p)ppGpp in stress response processes in bacteria is multi-layered and multifaceted representing fertile ground for further studies.

1.8.2 The effects of (p)ppGpp on bacteriophage infection dynamics

Since the bacterial stringent response has such a pervasive effect on various cellular processes, it is likely that bacteriophage infection dynamics will also be affected. Increased genome replication and transcription (due to phage genome replication), together with the increased amino acid requirement (stemming from synthesis of capsid proteins), will make infection by bacteriophages a convoluting factor in the stringent response. Conversely, the bacteriophage infection rate is expected to be reduced in cells undergoing a stringent response, due to the transcriptional and translational changes outlined above. Indeed, it has been demonstrated that intracellular levels of ppGpp and pppGpp are significantly reduced in the freshwater cyanobacterium *Anacystis nidulans* infected with AS-1 cyanophage upon transfer from light to dark conditions, compared to an uninfected control (Borbély*, et al.*, 1980) (Figure 1.9). Recent metabolomics experiments have also shown that (p)ppGpp levels are affected in *Pseudomonas aeruginosa* following phage infection, although the different phages examined had varying effects on intracellular (p)ppGpp levels (De Smet*, et al.*, 2016). A direct effect of (p)ppGpp levels on bacteriophage replication has been demonstrated in *E. coli* ppGpp^0 mutants, in which phage DNA replication of shiga-toxin containing bacteriophages was significantly increased under amino acid-starvation conditions, compared to infected WT. In the same experiment, infecting the WT under conditions inducing the stringent response produced smaller plaques relative to the ppGpp^0 mutant (Nowicki*, et al.*, 2013a).

Figure 1.9: Relative levels of ppGpp (●) and pppGpp (▲) in uninfected *Anacystis nidulans* (A) or infected with cyanophage AS-1, 1 hour (B) or 2 hours (C) post-infection (from (Borbély, *et al.*, 1980).
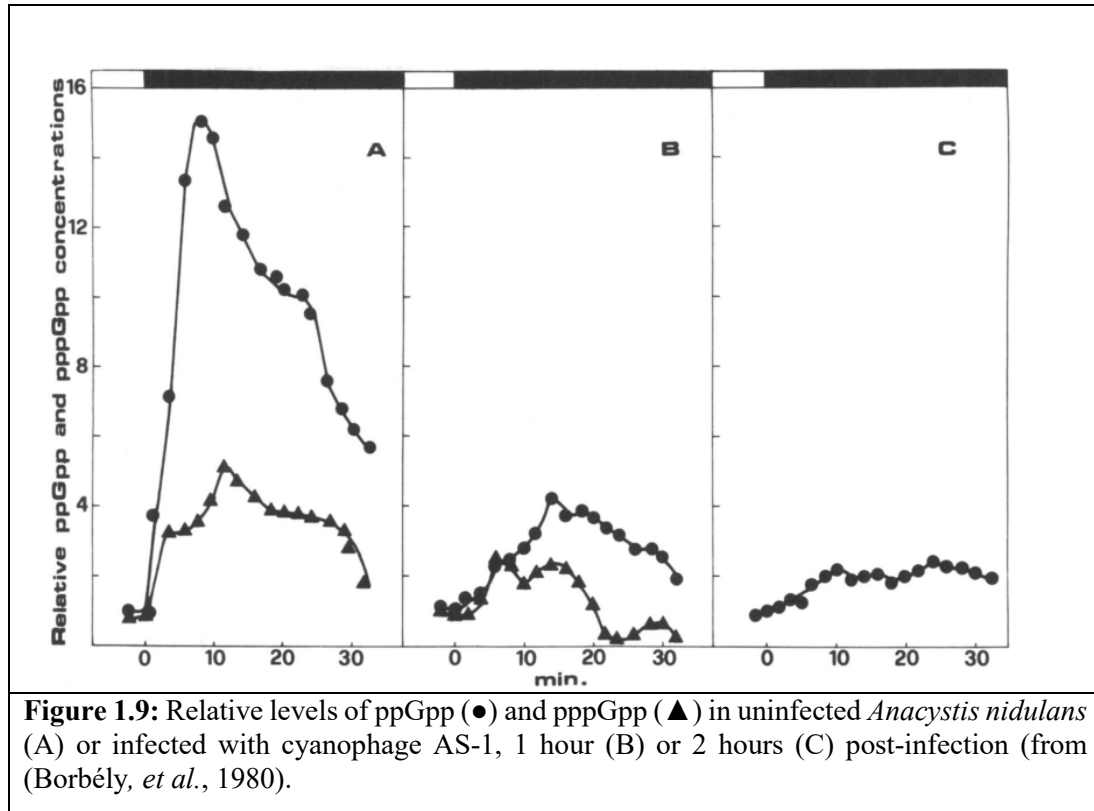
The above described examples demonstrate the effect the stringent response may have on the bacteriophage infection cycle. It is perhaps unsurprising then to discover AMGs in bacteriophage genomes which potentially affect the homeostasis of the host cell undergoing the stringent response. The most common such gene found in sequenced bacteriophage genomes is *mazG*.

mazG encodes a small pyrophosphohydrolase with a wide variety of functions. Most of its characterization has been done in *E. coli* (Gross, *et al.,* 2006) although it has been shown to perform a similar function in *Mycobacterium tuberculosis* (Lu, *et al.*, 2010) and in *Thermotoga maritima (*Zhang, et al., 2003*)*. It is thought to function as an NTPase, with the ability to hydrolase all eight conventional ribo- and deoxyribonucleotides (Zhang and Inouye, 2002). It is also induced under oxidative stress conditions, inferring that it has an active role in eliminating abnormal (d)NTPs which can arise as a result of oxidative stress-induced damage (Moroz, *et al.*, 2005). Indeed, activity assays in both *E. coli* and *Mycobacterium tuberculosis* have shown that MazG is capable of hydrolysing abnormal nucleotides, such as dUTP, 2-hydroxy-dATP and 8-oxo-dGTP (Lu, *et al.*, 2010). Its role in the stringent response was demonstrated via lack of induction of *relA* under oxidative stress conditions in *mazG⁻ E. coli* mutants (Lu, *et al.*, 2010). Furthermore, MazG has been shown to play an active role in

reducing ppGpp levels in amino acid starved *E. coli* , further confirming its role in the stringent response (Gross, et al., 2006).

mazG has been found in the majority of sequenced cyanophage genomes and is defined as part of the core genome (Sullivan, et al., 2010). In addition to cyanophages, it has been found in other marine bacteriophage genomes, such as *Roseobacter* and *Pseudoalteromonas* phages (Duhaime, et al., 2011) and SAR116 phages (Kang, et al., 2013). Metagenomic studies also suggest that the viral *mazG* can be found in the majority of marine environments and has potentially been transferred between different hosts via horizontal gene transfer (Bryan, et al., 2008). In contrast to numerous studies investigating MazG function and its role in nutrient stress metabolism in other bacterial phyla, its role in cyanobacteria and the potential role the cyanophage homologue plays during infection have yet to be clarified.

1.9 Aims and Objectives of my PhD Thesis:

The major aims of the present study were as follows:

1.   To investigate the molecular processes taking place during phage infection of the marine cyanobacterium *Synechococcus* sp. WH7803 during growth under different environmental stress conditions (primarily P depletion but also Fe depletion, N depletion, etc.)

2.   To understand the biochemical function of the cyanophage MazG protein as well as the environmental and physiological conditions under which the cyanophage copy of *mazG* is expressed.

Specific questions:

1.   What is the mechanism underlying the phage-infected delay of lysis of marine *Synechococcus* under P-deplete conditions?
2.   What is the underlying physiology of cyanophage during "pseudolysogeny"? Can we differentiate between states in which partially ready virions are found within the host cell, waiting for additional nutrients to become available, as opposed to the release of incomplete virions, unable to infect additional host cells?
3.   What are the regulatory processes guiding the cellular response to viral infection under P-limited conditions?

4.   Which phage genes are up-regulated or down-regulated in the state of "pseudolysogeny"?

5.   Is there cross-talk between the transcriptional regulation networks of the phage and that of the host during infection under P-limiting conditions?

6.   Does the burst size and burst delay seen during cyanophage infection of P-deplete *Synechococcus* occur during host depletion of other nutrients/environmental conditions?

7.   How does the viral *mazG* expression profile change under nutrient limitation conditions?

8.   Is there a difference in kinetic activity between viral and host MazG proteins?

Chapter 2

General Materials and Methods

2.1 Strains and plasmids

Strain	Genotype	Purpose	Reference
Synechococcus sp. WH7803	WT	Main model organism of this study	(Waterbury, *et al.*, 1986)
E. coli DH5α (K12 derivative)	F⁻ *endA1 gln*V44 *thi⁻*1 *recA1 relA1 gyr*A96 *deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsd*R17(rK⁻ mK⁺), λ⁻	Plasmid propagation and cloning	(Taylor, *et al.*, 1993)
E. coli MC1061	F⁻ Δ(*ara-leu*)7697 [*araD*139]_{B/r} Δ(*codB-lacI*)3 *galK*16 *galE*15 λ- e14- *mcrA*0 *relA*1 *rpsL*150(*str^R*) *spoT*1 *mcrB*1 *hsd*R2(*r⁻m⁺*)	Bi-parental conjugation with *Synechococcus* sp. WH7803	(Casadaban and Cohen, 1980)
E. coli BL21 Star™ DE3	F⁻ *ompT hsd*S_B (r_B⁻, m_B⁻) *galdcmrne*131 (DE3)	Protein over-expression	(Invitrogen Inc.)

Table 2.1: Bacterial strains used in this study

Name	Type	Isolation location	Comment	Reference
S-PM2	*Myoviridae*	Plymouth Sound, English Channel, United Kingdom (50°18'N, 4°12'W)	-	
S-BM1	*Myoviridae*	coastal water from the Sargasso Sea just outside the Bermuda Biological Station for Research, Bermuda	-	(Wilson, *et al.*, 1993)
S-WHM1	*Myoviridae*	Woods Hole Oceanographic Institute harbour, Woods Hole, Massachusetts, USA.	-	
Syn9	*Myoviridae*	Woods Hole Oceanographic Institute harbour, Woods Hole, Massachusetts, USA.	-	(Weigele, *et al.*, 2007)
S-SM1	*Myoviridae*	Atlantic Ocean, (38°17`N 73°15`W), 0m	Kindly provided by Dr. Matthew Sullivan,	(Sullivan, *et al.*, 2010)
S-SM2	*Myoviridae*	Atlantic Ocean, (38°17`N 73°15`W), 15m		

			University of Arizona, USA	
S-RSM3-88	*Myoviridae*	Gulf of Aqaba, Red Sea, Station A (29°28'N 34°55'E)	85 environmental phage isolates tested for presence of *pstS*	(Millard and Mann, 2006)

Table 2.2: Cyanophage strains used in this study

Name of plasmid	Genotype	Comment	Reference
pET-151d	Amp^R, pBR322 *ori,* T7*lac* promoter, N-terminal His•Tag and TEV protease cleavage site	Used for protein overexpression	Thermofisher
pET-28a	Kan^R, T7*lac* promoter, N-terminal His•Tag/thrombin/T7•Tag configuration in addition to an optional C-terminal His•Tag sequence	Used for protein over-expression. Kindly provided by Dr. Yin Chen, University of Warwick, UK	Novagen
pET-22b	Amp^R, T7 promoter, N-terminal *pelB* signal sequence for potential periplasmic localization, plus optional C-terminal His•Tag	Contains the *mazG* gene from *E. coli* . Kindly provided by Prof. Guo-Ping Zhao, Shanghai Institutes for Biological Sciences of the Chinese Academy of Sciences, China	(Lu*, et al.*, 2010), Novagen
pRL153	RSF1010 derivative, Km^R, *oriV mobC mobB mobA repB repA repC*	Used for biparental conjugation in *Synechococcus* sp. WH7803 transformation	(Elhai and Wolk, 1988)
pGEM	Ap^R Km^R	Used for different PCR product propagation	Promega

Table 2.3: List of plasmids used in this study

Name	Sequence 5' – 3'	Comment
WH7803 MazG KO1 F	GAATTCACGACCATCAGCGACAAACT	Amplification of a gene fragment used to knock out the *mazG* gene in *Synechococcus* sp. WH7803
WH7803 MazG KO1 R	GCCTCCTTGAGCTCATCGAGGAATTC	
7803_MazG_ExpF	CACCATGGCACAGCATGCAACCGT	Amplification and cloning of the *mazG* gene from *Synechococcus* sp. WH7803
7803_MazG_ExpR	CTAGCTGCTCTGGGTGTTTTC	
SPM2_MazG_ExpF	CACCATGAGCAAAGTGAACT	Amplification and cloning of the *mazG* gene from cyanophage S-PM2
SPM2_MazG_ExpR	CCAGGTACCGCGATCATCTG	
M13F	GTAAAACGACGGCCAGT	Used in colony PCR for verification of the presence of the right insert in the plasmid.
M13R	AGCGGATAACAATTTCACACAGGA	
T7F	TAATACGACTCACTATAGGG	
T7R	GCTAGTTATTGCTCAGCGG	
pRIL153 NheI F	GGTGCTGGCCGAGGAGATTA	Used to confirm the size of insert into the pRL153 plasmid
pRL153 NheI R	CAGATTATCCGGCTCCTCCAT	
16S-27F	AGAGTTTGATCMTGGCTCAG	Used to amplify and sequence 16S rRNA gene fragment for heterotrophic bacteria contaminant identification
16S-1492R	TACGGYTACCTTGTTACTACTT	
PstS_Deg_F	TGGTTCAGATCCCGATGACT	Degenerate primers designed to identify the presence of *pstS* in cyanophage isolates
PstS_Deg_R	CACATAGTTRCCTGCRTTGT	
CPS1	GTAGWATTTTCTACATTGAYGTTGG	Used to amplify a fragment from gp20-encoding gene in cyanophage genomes (Fuller, *et al.*, 1998)
CPS2	GGTARCCAGAAATCYTCMAGCAT	
psbA_SPM2 F	CTGGTCTGGGTATGGAGGTG	Primers used for qPCR enumeration of S-PM2 cyanophage in one step infection experiments
psbA_SPM2 R	TGTCGGACGCTTATTCCTGT	

Table 2.4: List of primers used in this study

2.2 Growth media and growth conditions

2.2.1 LB growth media (pH 7.0)

Compound	g L^{-1}
Bacto Tryptone	10g
Bacto-yeast extract	5g
NaCl	10g
pH adjusted to 7.0 using 5M NaOH	

Table 2.5: LB media composition (Maniatis, *et al.*, 1989)

2.2.2 2xYT growth media

Compound	g L^{-1}
Bacto Tryptone	16g
Bacto-yeast extract	10g
NaCl	5g
pH adjusted to 7.0 using 5M NaOH	

Table 2.6: 2xYT media composition (Maniatis, *et al.*, 1989)

2.2.3 M9 minimal growth media

Compound	Amount in 1L
M9 salt solution (x5)	200 ml
MgSO$_4$ (1M)	2 ml
CaCl$_2$ (1M)	0.1ml
M9 Salt solution (x5)	
Compound	g L^{-1}
Na$_2$HPO$_4$·7H$_2$O	64g
KH$_2$PO$_4$	15g
NaCl	2.5g
NH$_4$Cl	5.0g

Table 2.7: M9 minimal growth media composition (Maniatis, *et al.*, 1989)

Antibiotic	Working concentration
Ampicillin	50 μg/ml
Chloramphenicol	35 μg/ml
Tetracycline	15 μg/ml
Kanamycin	50 μg/ml
Spectinomycin	100 μg/ml

Table 2.8: Working concentrations of antibiotics used in this study

2.3 Growth of *Synechococcus* sp. WH7803

2.3.1 *Synechococcus* sp. WH7803 liquid culture growth

Synechocococcus sp. WH7803 was grown in artificial seawater (ASW) which is a defined medium (Table 2.3) commonly used for growth of marine cyanobacteria in laboratory conditions (Wilson*, et al.*, 1996). Nutrient deplete medium was prepared in the same way, while omitting $FeCl_3$ (ASW-Fe), $NaNO_3$ (ASW-N) or K_2HPO_4 (ASW-P). The final medium was autoclaved for 20 min at 121ºC, 15 psi. Cultures were grown at 23º C under permanent illumination at a light intensity of ~10 μmol photons m^{-2} s^{-1} with constant shaking at ~220 rpm.

Compound	g L^{-1}	Final concentration
$NaNO_3$	0.75	8.8 mM
$MgCl_2 \cdot 6H_2O$	2	10 mM
$MgSO_4 \cdot 7H_2O$	3.5	14 mM
KCl	0.5	6.7 mM
$CaCl_2 \cdot 2H_2O$	0.5	3.4 mM
Tris	1.1	9.08 mM
$K_2HPO_4 \cdot 3H_2O$	0.03	0.172 mM
NaCl	25	428 mM
Trace metal stock	1 ml	-
Trace metal stock solution composition		
$FeCl_3 \cdot 6H_2O$	3	11.10 μM
$Co(NO_3)_2 \cdot 6H_2O$	0.005	0.02 μM

CuSO$_4$·5H$_2$O	0.008	0.03 µM
Na$_2$MoO$_4$·2H$_2$O	0.39	1.61 µM
ZnSO$_4$·7H$_2$O	0.222	0.77 µM
MnCl$_2$·4H$_2$O	1.81	9.15 µM
H$_3$BO$_3$	2.86	46.26 µM
EDTA	0.5	1.71 µM

Table 2.9: Composition of ASW medium

Cultures were grown in volumes of 100ml, 500ml and 1L ASW medium in 250ml, 1L and 2L Erlenmeyer flasks, respectively. When grown at larger volumes (7L and 10L), autoclaved ASW medium was supplemented with filter-sterilised 10 mM NaHCO$_3$ and aerated with filtered air. When reaching stationary phase, cultures were re-subbed into fresh medium at 10% v/v concentration.

2.3.2 Growth of *Synechococcus* on ASW solid media

For the purpose of single colony isolation or plaque assay, *Synechococcus* was regularly grown on solid media plates prepared in the following manner: 0.5 L of 2X concentrated ASW was prepared based on the ASW composition outlined in Table 2.7. Separately, clean agar, prepared as outlined in (Millard, 2009) was added to 0.5L of MilliQ purified water to the concentration of 0.8% (w/v). Dissolved agar and 2X ASW medium were autoclaved separately for for 20 min at 121ºC, 15 psi and mixed together aseptically. The resulting ASW-agar was either poured on Petri plates or left to cool down for the plaque assay purposes.

2.3.3 Monitoring of *Synechococcus* growth

Growth was routinely monitored either via OD$_{750}$ measurements, using an Ultrospec 3000 pro (Biochrom, Cambridge, UK) spectrophotometer, or via cell counting using a FACScan flow cytometer (Becton Dickinson). Each sample count was normalised to counts of multifluorescent beads (Polysciences), the concentration of which was prior estimated via fluorescent microscopy. Red and orange fluorescence, accounting for chlorophyll/allophycocyanin and phycoerythrin, respectively, were measured through

FL3 (650nm) and FL2 (585/42nm) filters respectively. Cell and bead counts were collected using CellQuest software (Becton Dickinson, UK).

2.3.4 Assessment of *Synechococcus* culture contamination

Prior to re-subbing, the axenic status of *Synechococcus* cultures was verified by streaking *Synechococcus* onto ASW contamination medium i.e. ASW medium containing 1.5% (w/v) agar, supplemented with 2% (w/v) glucose and 0.15% (w/v) yeast extract. Agar plates were incubated at 23° C in the dark for at least 72 hrs. Cultures displaying no visible heterotroph growth were deemed axenic and used for further work.

2.4 *Synechococcus* sp. WH7803 metabolic assays

2.4.1 Measurement of phosphate (Pi) concentration

Extracellular Pi levels were measured according to (Chen*, et al.*, 1956). Briefly, 1ml sample was mixed with 1ml phosphate assay reagent containing 0.5% ammonium molybdate (w/v), 2% ascorbic acid (w/v) and 1.2M sulphuric acid, and incubated for 1.5hrs at 37° C. Absorbance was then measured at 820nm using a 5cm pathway glass cuvette against a deionised water blank. Pi concentration was determined based on a standard curve of KH_2PO_4, ranging in concentration between 0.1-200µM.

2.4.2 Alkaline phosphatase assay

Alkaline phosphatase activity was adapted from (Bessey*, et al.*, 1946), based on the rate of conversion of *para*-nitrophenyl phosphate (p-NPP) to *para*-nitrophenyl (*p*-NP) in the following manner: 40 µl 18 mM *p*-NPP in 1mM TRIS pH 8.0 was added in triplicate to 160 µl sample culture, making the final *p*-NPP concentration 3.6 mM. Formation of *p*-NP was measured over 4 hours by reading the absorbance at 415 nm using a iMark Microplate reader (Biorad). Reads were normalised to a control well containing 18mM *p*-NPP in 1 mM Tris pH8.0 without sample culture. The amount of *p*-NP produced was estimated from a standard curve measuring the absorbance of known concentrations of *p*-NP ranging between 0 – 100 µM. Alkaline phosphatase activity was measured from the linear section of the plot showing formation of *p*-NP over time, normalised to cell count (see section 2.2) and expressed as mol *p*-NP cell^{-1} min^{-1}.

2.4.3 Assessment of photosystem II (PSII) photophysiology

The effective quantum yield of PSII was estimated using the methods outlined in (Garczarek*, et al.*, 2008). Briefly, 500 µl *Synechococcus* sample culture was incubated 5 min in the dark in order to completely oxidise the pool of primary electron acceptor Q_A. Cells where then transferred to a glass cuvette in a pulse amplitude-modulated fluorometer (PhytoPAM, Walz, Effeltrich, Germany) and diluted to 2ml with fresh ASW medium. The sample was then exposed to low-level modulated light (520nm, ~1 µmol photons m^{-1} s^{-1}) and F_0 measurements taken. Subsequently, a volume of 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) to a final concentration of 100 µM was added, and actinic light (1300 µmol photons m^{-1} s^{-1}) was applied, causing the fluorescence levels to rise, followed by the final pulse of saturating light (2600 µmol photons m^{-1} s^{-1}) for 200ms, and the measurement of maximum fluorescence (F_m) was taken. The effective quantum yield of photosystem II was calculated as:

$$F_v/F_m = \frac{F_m - F_0}{F_m}$$

2.5 Purification and concentration of cyanophages

2.5.1 Plaque assay purification of phages

Each of the cyanophages used in this study was produced via triple plaque purification from a lysate stored at 4º C in dark. 10-fold serial dilutions of lysate were obtained by diluting the lysate in ASW. These dilutions were used to perform plaque assays, as described below in 2.6.1. After 3-4 days, plaques became visible and a single plaque was picked using a Pasteur pipette. An agar plug containing the plaque was resuspended in 500 µl ASW and incubated in the dark at 4º C for 1 hr. The resuspended phages were serially diluted in ASW and used for a new round of plaque assay. This procedure was repeated 3 times. After plaque purification, phages were used to propagate on exponentially growing *Synechococcus* sp. WH7803 cultures.

2.5.2 Phage purification and PEG precipitation

Plaque assay purified cyanophages were propagated on *Synechococcus* sp. WH7803 either from liquid or plate lysates. In the case of liquid medium extraction, previously obtained lysate, kept at 4ºC in the dark, was added to an exponentially growing *Synechococcus* culture and shaken as described above until cell lysis was visible. Lysate

was subjected to centrifugation at 4º C at 3000xg for 10 min in order to remove cell debris.

In the case of plate lysis, an exponentially growing culture of *Synechococcus* sp. WH7803 was concentrated 50x and incubated with 1ml of previously obtained lysate for 1 hr at 23º C, to allow phages to adsorb. The infected culture was then mixed with 0.27% (w/v) clean agar in ASW medium at a ratio of 1:1 (v/v) and poured into a Petri dish. Clean agar was prepared as in (Millard, 2009). The infected culture was incubated at 23º C at a light intensity of 10 µmol photons m^{-1} s^{-1} until confluent lysis was visible. The agar was collected and re-suspended in 5ml ASW medium containing 10% (v/v) chloroform, shaken vigorously and incubated at room temperature for 1hr. The suspension was then centrifuged at 8000g for 10 min at 4º C and the supernatant collected. From this point onwards liquid and plate lysates were treated identically.

The collected supernatant was incubated for 1 hr at room temperature with 10 ng ml^{-1} (final concentration) RnaseA (Invitrogen) and 0.25 SU ml^{-1} DNAseI (New England Biolabs) prior to centrifugation at 8400 rpm using a Beckman Coulter Avanti JLA 10.500 rotor for 30 min at 4ºC. The supernatant was collected and incubated overnight with 100g L^{-1} PEG 8000 (Fisher Bioreagents) at 4ºC. Cyanophage were pelleted by centrifugation for 30 min at 4ºC at 13,000xg. Pellets were re-suspended in appropriate ASW medium (ASW, ASW-P, ASW-N or ASW-Fe, as described in section 2.1). Re-suspended cyanophage were washed in an equal volume of chloroform twice, in order to remove the remaining PEG from the solution. The top layer was collected and concentrated either via centrifugation using an Amicon centrifugal ultrafiltration device, with a 100kDa nominal molecular weight cut-off filter (Amicon, Millipore), or via CsCl concentration and dialysis (see below).

2.5.3 CsCl purification and dialysis of cyanophage

Cyanophage were re-suspended in ASW medium and layered onto a CsCl gradient. The gradient was prepared in an ultracentrifugation tube using 2ml, 3ml and 3ml CsCl in ASW medium at densities of ρ=1.45, 1.5 and 1.7 g ml^{-1}, respectively. The gradient was then subjected to ultracentrifugation at 217,290xg at 4ºC for 3 hr. The resulting phage band (Figure 2.1) was collected and dialysed overnight in 12-14 kDa cut-off dialysis membranes (Visking Medicell, UK) against 2L of appropriate ASW medium. Concentrated cyanophages were stored in 50ml Falcon tubes at 4ºC in the dark.
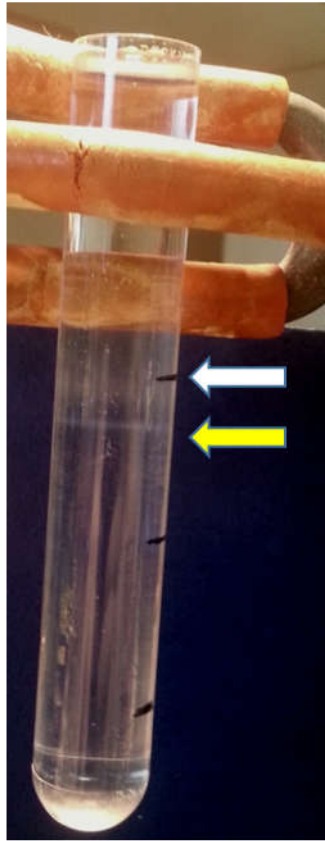
Figure 2.1: CsCl gradient after ultracentrifugation. The white arrow marks the proximity of the 1.45-1.5 CsCl density phase. The yellow arrow points to the concentrated phage band

2.6 Enumeration of cyanophages

After purification and concentration, cyanophages were enumerated either via plaque assays or via the Most Probable Number (MPN) method.

2.6.1 Plaque assays

Serial 10-fold cyanophage dilutions were prepared by diluting the purified phage stock in ASW medium. Exponentially growing *Synechococcus* WH7803 cultures (~5×10^7 cells ml^{-1}, as estimated by flow cytometry, described in 2.1.2) were concentrated 50X by centrifugation at 4000xg at 23° C for 15 min. 200 µl of concentrated cells were incubated with 200 µl of each of the 10-fold phage dilutions at 23° C and a light intensity of 10 µmol photons m^{-1} s^{-1} in triplicate for 1 hr to allow phage adsorption. After the incubation, each of the cultures infected by the different dilutions were mixed with 1.6

ml 0.4% (w/v) ASW clean agar and poured into a 6-well microtitre plate. Plates were sealed with parafilm, kept in sealed bags to prevent desiccation, and incubated at 23ºC at a light intensity of 10 µmol photons m^{-1} s^{-1} until plaques were visible and no new plaques appeared (usually 5-6 days). The cyanophage concentration in the initial stock was calculated according to the following formula:

$$number\ of\ p\ ages\ ml^{-1}\ =\ number\ of\ plaques\quad 5\quad dilution\ factor$$

2.6.2 The Most Probable Number (MPN) method for cyanophage enumeration

The MPN method was also used for cyanophage enumeration as follows: *Synechococcus* cells were grown in liquid ASW medium to the exponential growth phase, as described above, concentrated 50x by centrifugation at 4000xg at 23º C for 15 min and 200 µl concentrated culture added to 96-well plates. Purified cyanophage extract of unknown concentration was serially diluted 10-fold in ASW medium and 50 µl of each dilution added in triplicate to concentrated *Synechococcus* cells. 50 µl ASW
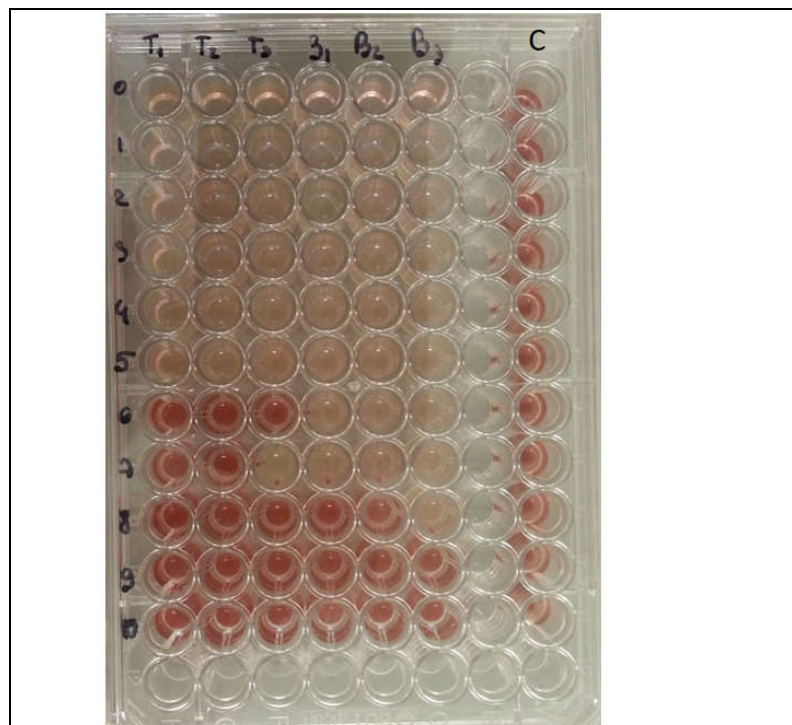
Figure 2.2: Photograph of an MPN plate. Rows 0-10 represent the serial 10-fold dilutions of unknown cyanophage sample. Columns T1-3 and B1-3 are triplicates of two different phage samples. Column labelled C contains the no-phage negative control sample.

medium without phage were used as a negative control. Plates were incubated at 23ºC

under constant illumination at 10 μmol photons m^{-1} s^{-1} and monitored for lysis. When lysis occurred (usually after 5-8 days), phage number was estimated, judging by the highest dilution replicate at which lysis of *Synechococcus* cells occurred (Figure 2.2).

2.7 Molecular biology methods

2.7.1 Genomic DNA isolation

2.7.1.1 Chromosomal DNA isolation from *Synechococcus* sp. WH7803

Synechococcus sp. WH7803 chromosomal DNA was isolated using phenol-chloroform extraction in the following manner: *Synechococcus* sp. WH7803 was pelleted via centrifugation at 10,000 rpm for 30 min at 4ºC using a Beckman Coulter Avanti JLA 10.500 rotor. Pelleted cells were re-suspended in 2ml Tris-HCl pH 8.0, added to the Ribolyser tube (Lysing Matrix E, MP Bioproducts) and lysed in Tissue Lyser (Qiagen, Retsch) by shaking at 3000 rpm for 30 sec, followed by incubation on ice for 1 min. This was repeated 3 times and the resulting lysate subjected to centrifugation at 16,000xg for 10 min at 4º C for cell debris removal. The supernatant was washed with 1 volume phenol/chloroform/iso-amyl alcohol (IAA) (25:24:1) and subjected to centrifugation at 16,000xg for 5 min at 4º C. The resulting supernatant was washed with 1 volume chloroform/IAA (24:1), centrifuged at 16,000xg for 5 min at 4º C and the supernatant collected. The DNA from the resulting aqueous phase was precipitated using 1 volume isopropanol/7.5M ammonium acetate (3:2 v/v) and incubated overnight at -20ºC. The precipitated DNA was pelleted by centrifugation at 10,000g for 45 min at 4º C and the resulting pellet washed with 1 volume of 70% (v/v) ethanol. After evaporation of ethanol, the pellet was re-suspended in MilliQ water and stored at -20ºC.

2.7.1.2 Cyanophage DNA isolation

Cyanophage DNA was isolated exactly as described for host chromosomal DNA isolation above, with the culture pelleting step being replaced with a phage concentration protocol as described in section 2.6.2.

2.7.2 Cloning and transformation

2.7.2.1 Preparing chemically competent *E. coli*

Chemically competent *E. coli* was prepared using the calcium chloride method (Maniatis*, et al.*, 1989) as follows: the relevant *E. coli* strain was streaked onto solid LB medium i.e. containing 1.5% (w/v) Bactoagar, and incubated overnight at 37ºC. A single colony was re-suspended in 5 ml LB medium and grown overnight shaking at 200 rpm at 37ºC in a C24 incubator orbital shaker (New Brunswick Scientific). 1 ml of the overnight starter culture was then used to inoculate 10ml LB medium and incubated shaking at 200 rpm at 37ºC for 2-3 hr. Culture density (OD_{600}) was measured every 30 min using an Ultrospec 3000 Pro (Biochrom, Cambridge, UK) spectrophotometer until the OD_{600} reached 0.8-1.0. Cells were then cooled for 10 min on ice, and pelleted by centrifugation using a Beckman Coulter Avanti JA 25.50 rotor at 6,000 rpm for 10 min at 4ºC. The supernatant was removed and the cell pellet re-suspended in 10ml ice-cold 0.1M $CaCl_2$. Re-suspended cells were incubated on ice for 20 min, pelleted by centrifugation as above, and the supernatant replaced with 5ml ice cold 0.1M $CaCl_2$/15% (v/v) glycerol. Competent cells were divided into 50 µl aliquots, snap-frozen in liquid N_2 and stored at -80ºC until use.

2.7.2.3 Transformation of *E. coli*

Competent cells prepared as described in section 2.7.2.1 were thawed on ice and mixed with 5 µl plasmid. Cells were incubated on ice for 20 min, then transferred to a 42ºC heated bath for 45 sec and returned to ice for an additional 5 min. 1ml LB medium was added to the transformed cells and incubated at 37ºC shaking at 200 rpm for 2 hr in an orbital shaker (Gallenkamp). *E. coli* cells were then streaked onto LB solid medium containing the appropriate antibiotic and incubated at 37ºC overnight, after which transformant colonies were further analysed.

Chapter 3

The presence of *pstS* in cyanophage genomes: a connection to pseudolysogeny?

3.1 Introduction

Oligotrophic environments present challenges to both bacterial hosts (Scanlan, *et al.*, 2009; Flombaum, *et al.*, 2013) and their phages (Kokjohn, *et al.*, 1991; Wilson, *et al.*, 1996; Scanlan, *et al.*, 2009; Flombaum, *et al.*, 2013). Lack of nutrients should theoretically have a higher impact on viruses, due to their high demand for building blocks during the intense period of replication and virion assembly. Therefore, it is not surprising to discover that cyanophages have appropriated bacterial host genes, incorporating them into their own genetic arsenal and using them as a way to overcome nutrient limitation and/or redirect essential metabolic processes towards products which they find most beneficial to the purpose of maximising viral progeny (Breitbart, *et al.*, 2007; Puxty, *et al.*, 2015a; Hurwitz and U'Ren, 2016).

As described in section 1.3, phosphorus (P) is one of the main limiting nutrients for primary production in vast areas of the global ocean (Mills, *et al.*, 2004). Since marine cyanobacteria *Synechococcus* and *Prochlorococcus* are the dominant primary producers in marine oligotrophic waters (Flombaum, *et al.*, 2013), strategies they have developed for dealing with P-limiting conditions have been a focus of many previous studies (Moore, *et al.*, 2002; Scanlan, 2003; Bertilsson, *et al.*, 2003; Fuller, *et al.*, 2005 ; Martiny, *et al.*, 2006; Kettler, *et al.*, 2007; Mazard, *et al.*, 2012). As cyanophages infecting these two genera are found in high concentrations in these waters (Brum, *et al.*, 2015), they will probably be subject to the same limitations like their cyanobacterial hosts. Indeed, sequencing of both cultured cyanophage isolates (Sullivan, *et al.*, 2010) as well as viral fractions from oligotrophic areas of the ocean (Kelly, *et al.*, 2013), have shown a prevalence of P-related genes and a high correlation with low P-levels in their site of isolation.

Previous studies have shown that P limitation has a large impact on infection success (measured via burst size and the length of the latent period) of some cyanophages (Wilson, *et al.*, 1996). In addition to finding marked infection differences between P-replete/P-deplete conditions, this study also found that not all phages respond to P stress in the same manner. While cyanomyovirus S-PM2 was found to be most affected, P-depletion had a much smaller effect on the infection capacity of cyanophage S-BM1. Interestingly, even though both S-BM1 and S-PM2 were isolated on the same host – *Synechococcus* sp. WH7803, they were isolated from different geographical and environmental niches – S-BM1 was isolated from the Northern Atlantic gyre, an extremely oligotrophic environment (Wu, *et al.*, 2000), while S-PM2 was isolated from a coastal area of the English Channel, characterised by nutrient-replete conditions (Wilson, *et al.*, 1993). This study, together with metagenomic sequencing performed by (Kelly, *et al.*, 2013), showcases possible correlations between environmental conditions, the presence of P-stress related genes in cyanophage genomes and the ability to

overcome P-stress imposed limitations on the success of infection. In order to examine the possible presence of P-stress related genes in cyanophage, this study examined the infective success of cyanophages isolated from the Gulf of Aqaba, Red Sea under P-deplete conditions.

The Gulf of Aqaba represents a marine environment which, due to its geological structure, climate, and its proximity to desert, undergoes interchanging cycles of deep water column mixing during the winter and strong stratification during the summer (Wolf-Vecht, *et al.*, 1992). Winter deep mixing brings nutrients to the photic layer of the water column, enabling primary production. Primary producers, in turn, deplete the photic layer of nutrients during the stratification period, creating oligotrophic conditions in the upper layer of the water column. These interchangeable nutrient levels in the Gulf of Aqaba, make it an interesting study site, since the nutritional status of phytoplankton also changes with changing nutrient conditions, cycling between summer oligotrophs, adapted to low nutrient conditions and winter dominated mesotrophs (Lindell and Post, 1995). Resident bacterial populations respond to stratification and subsequent nutrient depletion, by overexpressing genes that will enable them to either optimize and increase acquisition of the depleted nutrient or by reducing their cellular requirements for this specific nutrient. Previous studies in the Gulf of Aqaba have shown that there is a high correlation between the P-status of the water column and expression of the cyanobacterial phosphate stress-related gene *pstS* (Fuller, *et al.*, 2005) (Figure 3.1).

Figure 3.1: Annual measurement of soluble reactive phosphate (SRP) and levels of cyanobacterial PstS in the Gulf of Aqaba (Fuller, *et al.*, 2005).
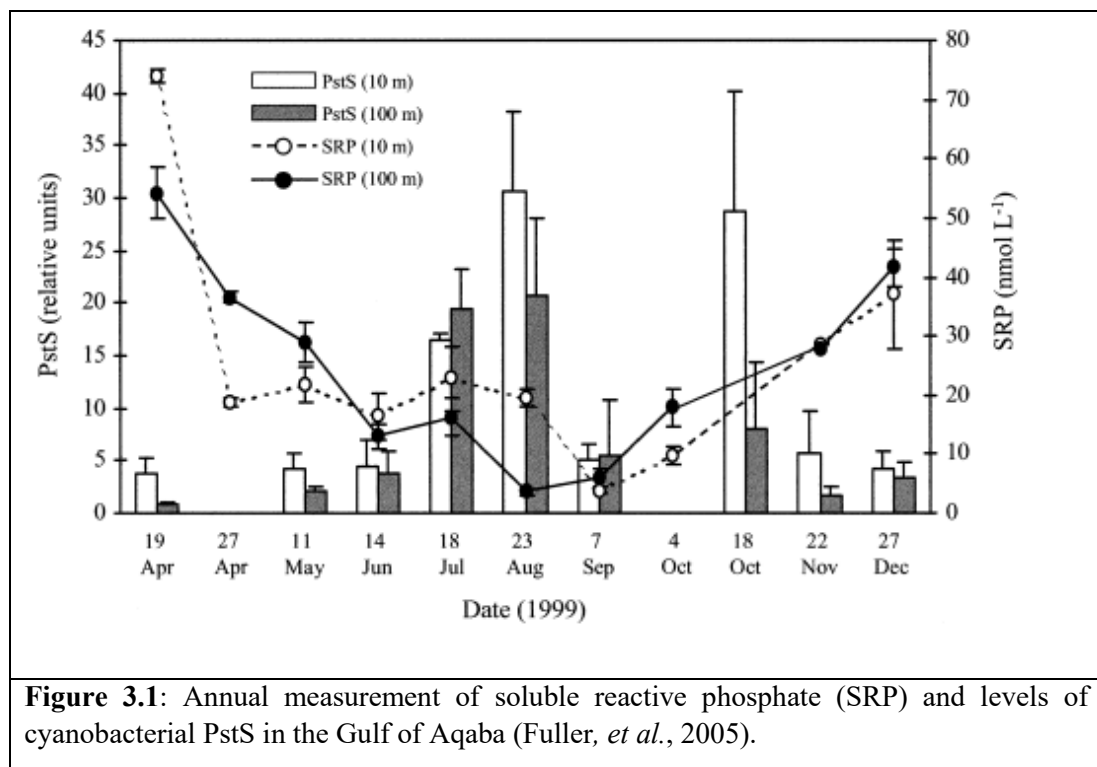
The constantly increasing number of genomes in public databases is enabling us to assess the genetic capabilities of cyanophages to deal with environmental stress. Relying on nucleotide similarity, previous studies have identified three such possible genes which could potentially help cyanophages deal with persistent P-stress – *phoA* – a gene encoding a putative alkaline phosphatase; *phoH* – a gene of currently unknown function but which has previously been associated with the P-stress response in *E. coli* (Wanner, 1990) and *pstS* – a gene encoding a periplasmic inorganic phosphate binding protein, previously used as an indicator of the P-stress status of cyanobacterial host cells (Scanlan*, et al.*, 1993). A study examining the transcriptional response of cyanophage and a *Prochlorococcus* host during infection under P-limited conditions, has shown that the most over-expressed phage gene is a *pstS* orthologue (Lin*, et al.*, 2016), transcription of which is putatively under the control of the host P-stress transcriptional regulator PhoB-PhoR (Zeng and Chisholm, 2012).

All of the above findings point towards the cyanophage copy of *pstS* playing an important role in dealing with P-stress during the infection process. In light of this, the goals of this chapter were as follows:

1. To investigate whether cyanophage possession of *pstS* confers a 'selective advantage' during the infection process under P-limited conditions.

2. To test the hypothesis that novel cyanophages isolated from P-deplete waters, will possess *pstS* in their genomes.

3. To investigate the phylogenetic relationship between the different copies of host marine picocyanobacterial PstS, and PstS orthologues from sequenced cyanophages.

3.2 Methods

3.2.1 Enumeration of phages

Infection parameters were measured either via plaque assays following (Wilson, *et al.*, 1996), or via one-step growth experiments performed in 96-well plates, as described in the General Methods (see section 2.5).

3.2.2 Polymerase chain reaction (PCR).

The presence of *pstS* in phage isolates was checked via PCR amplification, using degenerate primers listed in the General Methods section, Table 2.4. Degenerate primers were designed using the HYDEN program (Linhart and Shamir, 2005). As an input for HYDEN script, sequences of putative *pstS* genes from 6 sequenced cyanophages were used (Table 3.1).

Cyanophage	Sequence ID
S-SM1	GU071094.1
Syn30	HQ634189.1
S-SM5	GU071095.1
Syn19	GU071106.1
S-SKS1	HQ633071.1
Syn2	HQ634190.1
Table 3.1 – List of *pstS* sequences from sequenced marine cyanophages that were used for degenerate primer design	

Amplification of phage *gp20* was done using CPS1 and CPS2 primers, outlined in Table 2.4 in the General Methods chapter and described in (Fuller, *et al.*, 1998). PCR reactions were performed using MyTaq™ Mix (Bioline) according to manufacturer's instructions, over 35 amplification cycles at the annealing temperature of 55ºC.

3.2.3 Phylogenetic analysis

Phylogenetic analysis of *pstS* was performed using MEGA v6.0 (Tamura*, et al.*, 2013). Cyanophage and cyanobacterial *pstS* sequences were extracted from the NCBI nr database, Cyanorak (http://abims.sb-roscoff.fr/cyanorak/) and Cyanobase (http://genome.microbedb.jp/cyanobase/), using BLAST. Sequences were aligned using MUSCLE (Edgar, 2004) and the alignment trimmed using TrimAl (Capella-Gutierrez*, et al.*, 2009). The evolution model was selected using jmodeltest (Manuel Santorum*, et al.*, 2014) and a maximum likelihood tree built in MEGA, using 100 bootstrap replications to determine branch lengths and support values.

3.3 Results

3.3.1 Screening of cyanophages for the presence of *pstS*

In order to establish whether *pstS* is present in yet unsequenced cyanophages isolated from the Gulf of Aqaba, degenerate primers were designed and tested against previously sequenced cyanophages known to possess *pstS*. Previous work has shown that there is a significant decrease in levels of phosphate in the photic zone during the summer months, when stratification is at its most pronounced. This decrease occurs congruently with an increase in levels of expression of cyanobacterial PstS *in situ* (Fuller*, et al.*, 2005) . The cyanophages in

Figure 3.2: PCR products using degenerate *pstS* primers with different phage isolates. L- 100bp DNA ladder (Invitrogen); 1 – S-RSM60; 2 – S-RSM61; 3 – S-RSM62; 4 – S-RSM63; 5 – S-RSM68; 6 – S-RSM69; 7 – S-RSM70; 8 – S-RSM71; 9 – S-BM1; 10 - S-SSM1; 11 – S-SSM2; 12 – S-PM2; 13 – S-BM3; 13A – S-BM3 with CPS1 and CPS2 primers; 14 – S-BM4; 14A – S-BM4 with CPS1 and CPS2 primers.
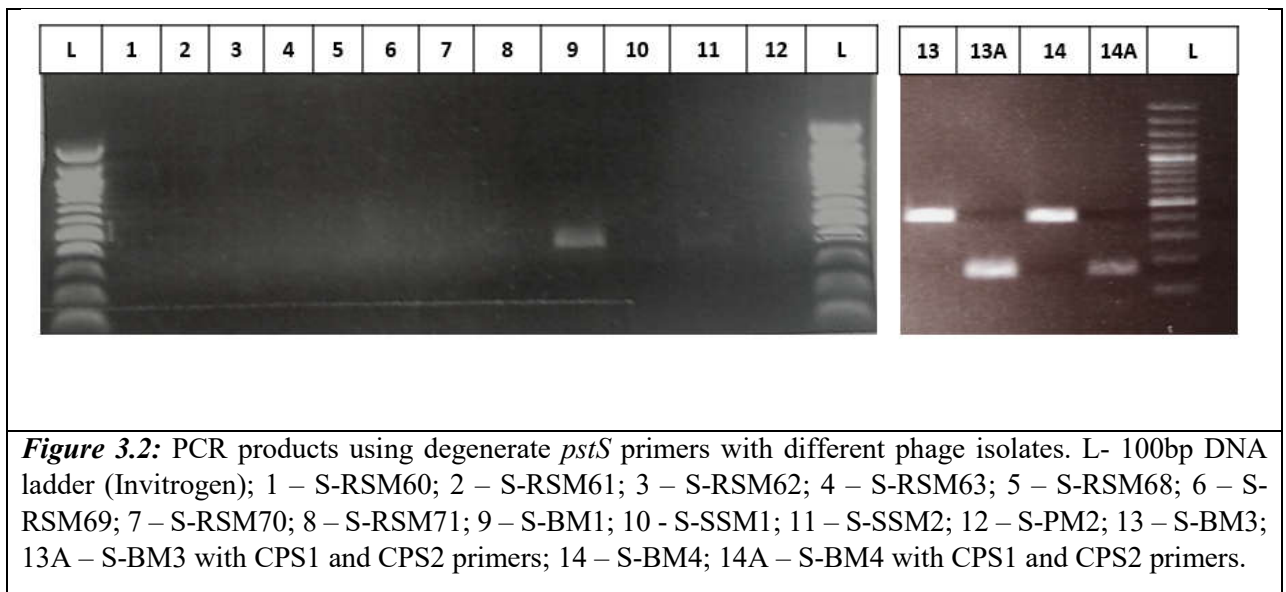
our collection were isolated from different depths and over the course of a whole year from station A in the Gulf of Aqaba (Millard and Mann, 2006). Taking the ambient P levels and

expression of *pstS* into account, we chose to screen those viral isolates that were isolated from the photic zone (top 100m) and during the months in which the cyanobacterial population was P-stressed. PCR amplification using *pstS* degenerate primers produced a fragment of ~450 bp using various cyanophages isolated from the Gulf of Aqaba as template (see Table 3.2). In the case of a positive result, the PCR product was sent for Sanger sequencing and the sequence analysed using BLAST. Sanger sequencing and BLAST analysis confirmed that the amplified fragments are indeed part of novel phage *pstS*. Examples of PCR reactions with amplified *pstS* fragments can be seen in Figure 3.2

The list of Red Sea cyanophages assessed for the presence of *pstS*, and the result of the PCR analysis can be found in Table 3.2.

Cyanophage	Type	Host of isolation	Date of isolation (1999)	Depth of isolation (m)	*pstS* presence
S-RSM60			18/07	30	-
S-RSM61					-
S-RSM62				150	-
S-RSM63			23/08	80	-
S-RSM64			07/09	30	-
S-RSM65				80	-
S-RSM66	*Myoviridae*	*Synechococcus* sp. WH7803	18/10	125	-
S-RSM67					-
S-RSM68			18/07	150	-
S-RSM69				125	-
S-RSM70			23/08	10	-
S-RSM71			18/10	30	-
S-RSM72			18/07	10	-

Table 3.2*:* List of cyanomyoviruses isolated from the Gulf of Aqaba that were tested for presence of viral *pstS* using degenerate *pstS* primers

In addition to the Red Sea cyanophage isolates, degenerate *pstS* primers were used to test additional cyanophages from our collection. The results of these PCR detection tests are shown in Table 3.3

Cyanophage	Type	Location of isolation	*pstS* presence
S-PM2	*Myoviridae*	Plymouth Sound, English Channel, UK	-
Syn9	*Myoviridae*	Woods Hole Oceanographic Institute harbour, Woods Hole, Massachusetts, USA	-
S-WHM1	*Myoviridae*		-
S-SSM1	*Myoviridae*	Atlantic Ocean, (38°17`N 73°15`W)	+
S-SSM2	*Myoviridae*		+
S-BM1	*Myoviridae*	Coastal water from the Sargasso Sea	+
S-BM3	*Myoviridae*		+
S-BM4	*Myoviridae*		+
S-BM6	*Myoviridae*		+

Table 3.3: List of additional cyanophages tested for the presence of viral *pstS* using degenerate *pstS* primers

Surprisingly, no *pstS* was detected in any of the cyanophage isolates from the Gulf of Aqaba, contrary to the initial hypothesis predicting that phages containing *pstS* will have an advantage over cyanophages that don't possess *pstS* and will dominate in P-depleted environments. Possible reasons for this discrepancy are discussed (see section 3.4).

3.3.2 The ability of cyanophage to overcome P stress-induced pseudolysogeny

Previous work has shown that cyanophage S-PM2, isolated from a nutrient-rich marine environment, exhibits a reduced infection success when infecting *Synechococcus* sp. WH7803 grown under P-limiting conditions (Wilson*, et al.*, 1996). The same work also showed that S-BM1, a cyanophage isolated from the P-deplete Sargasso Sea, does not suffer from the same deleterious effect under the same conditions. Examination of the S-PM2 genome, together with the above described PCR amplification using degenerate *pstS* primers, shows that S-PM2 does not possess *pstS* in its genome, while S-BM1 does (see Fig. 3.2, Table 3.3). In order to examine whether the presence of *pstS* in cyanophage genomes will define whether it is able to successfully infect a P-depleted host, a series of infection experiments was conducted, examining a panel of cyanophages. Two phage enumeration methods were used – plaque assay

and one-step infection in liquid culture. Examples of results of the plaque assay can be seen in Figure 3.3. P-stress induced pseudolysogeny was identified by a reduction in the final titre of cyanophages infecting a host grown under P-stress conditions, as can be seen in Table 3.4.

Phage	PFU/ml in +P conditions	PFU/ml in -P conditions	% reduction in yield
S-PM2	2.95×10^6	1.30×10^6	55.93%
S-BM1	9.00×10^3	1.15×10^4	-27.78%
S-WHM1	9.00×10^3	5.00×10^3	44.44%
S-RS60	7.52×10^7	4.96×10^7	34.07%
Syn9	1.93×10^7	1.34×10^5	99.31%

Table 3.4*: Differences in the burst size of different cyanophage strains infecting *Synechococcus* sp. WH7803 under P-replete and P-deplete conditions. A negative percentage of yield reduction observed in the case of S-BM1 is indicative of an increase in average yield under -P conditions, compared to infection in P-replete medium.

The infection dynamics of cyanophages used in this study was also assessed via one-step infection in liquid culture in 96-well plates (see section 2.6.2). In this experimental setup, approximately 5×10^7 cells of exponentially growing *Synechococcus* sp. WH7803 culture growing in +P/-P conditions were infected with $\sim 10^8$ PFU of purified cyanophage. This high titre of cyanophages ensured that all of the cells were infected at once, meaning that the rate at which the OD_{750} in each well decreased was indicative of the real infection kinetics and not a product of late or second round infections. Since the initial OD_{750} of +P and -P grown *Synechococcus* cultures was different in some cases, and for the sake of comparative clarity of results, growth was calculated as the percentage of initial OD_{750} values over time for each replicate separately and an average was calculated. A representative one-step infection dataset can be seen in Figure 3.4.
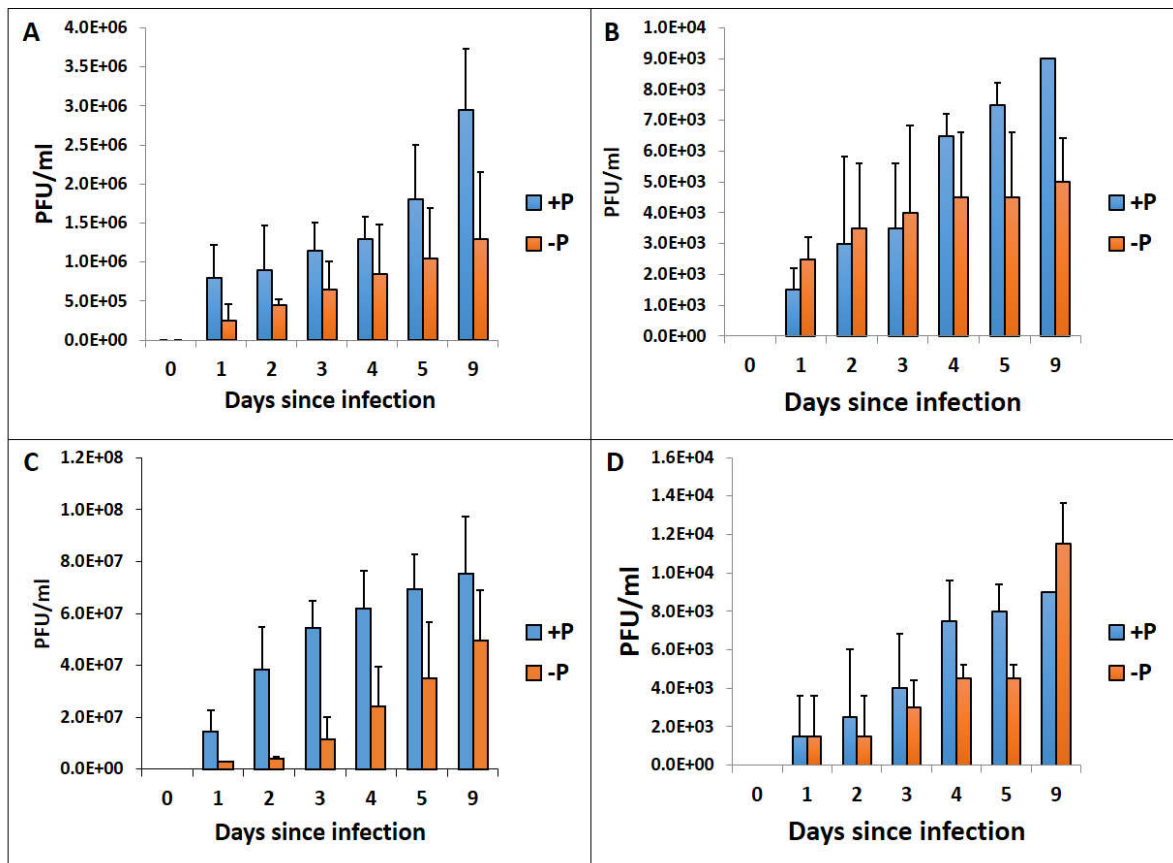
Figure 3.3: Phage infection of *Synechococcus* sp. WH7803 grown under +P/-P conditions assessed using plaque assays. Cyanophages used in this experiment were as follows: A – S-PM2, B – S-WHM1, C- S-RSM60, D – S-BM1. Error bars represent the standard error of an average of 3 biological replicates.

In order to examine whether possession of *pstS* in cyanophage genomes enables successful infection under -P conditions, *Synechococcus* sp. WH7803 was infected by all of the cyanophages used in this study under -P/+P conditions, using either the plaque assay method or one-step infection method. Some cyanophages were tested using both methods. The results of these experiments, with the notation of the method used to test for successful infection, are shown in Table 3.5
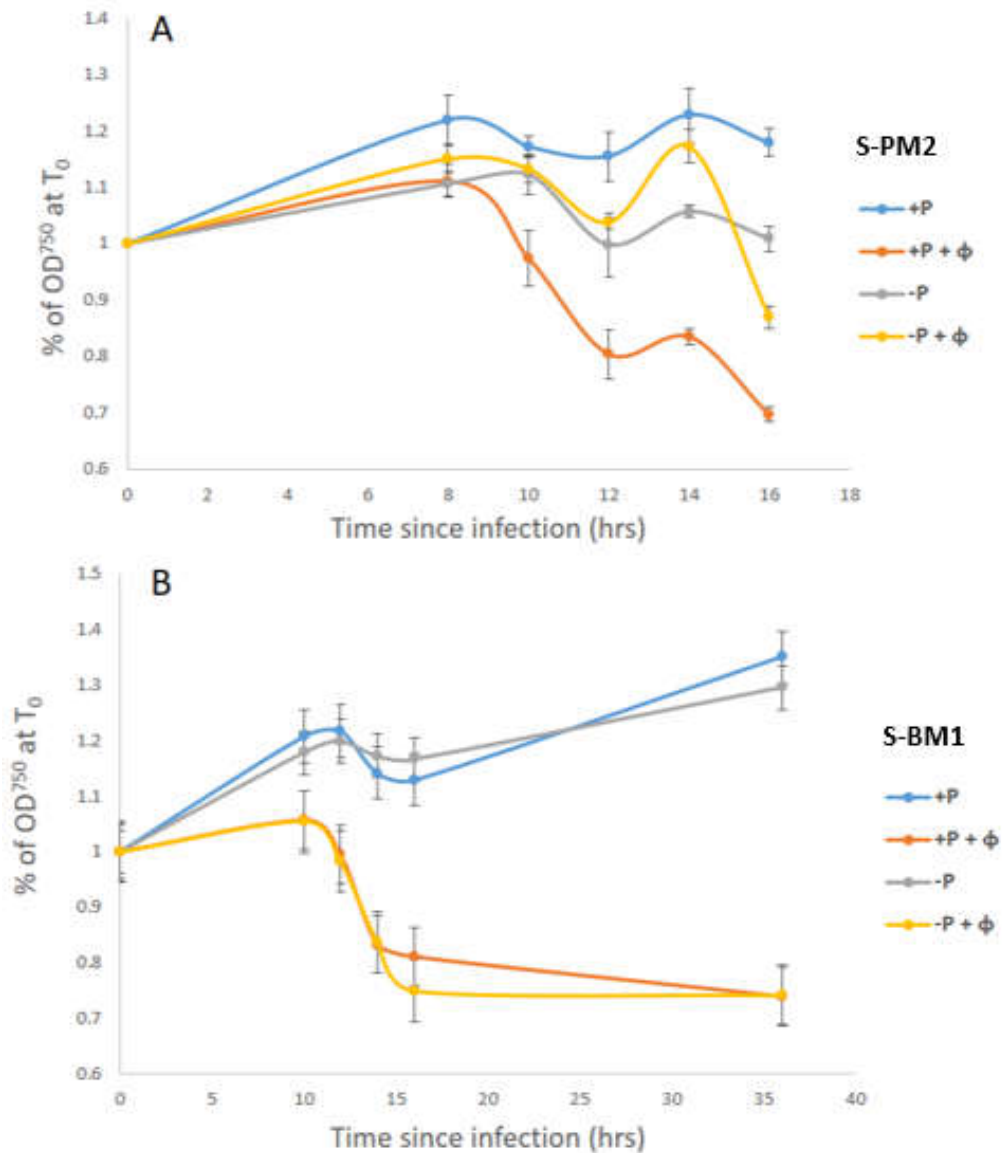
Figure 3.4: One step infection results, represented as the percentage of OD_{750} change over time. In both panels "+P" and "-P" indicates a non-infected culture grown under replete/deplete conditions respectively. "+P + φ" and "-P + φ" indicates the infected P-replete/deplete cultures, respectively. Panel A shows the infection kinetics of S-PM2 and is indicative of phosphate depletion induced pseudolysogeny, as characterised by an approximate 6-hour delay in lysis, while panel B shows infection by S-BM1 and is indicative of infection not affected by P-depletion. Error bars represent the standard error of an average of 3 biological replicates.

Cyanophage	Environment from which the cyanophage was isolated	Method of infection used		Shows pseudolysogeny-like infection phenotype during infection under P-limiting conditions	Presence of *pstS* (examined using PCR)
		Plaque assay	One-step infection		
S-PM2	Eutrophic	+	+	+	-
S-WHM1	Eutrophic	+	+	+	-
Syn9	Eutrophic	+	+	+	-
S-RSM60	Oligotrophic	+	+	+	-
S-RSM61	Oligotrophic	-	+	+	-
S-RSM62	Oligotrophic	-	+	+	-
S-RSM63	Oligotrophic	-	+	+	-
S-RSM64	Oligotrophic	-	+	+	-
S-RSM65	Oligotrophic	-	+	+	-
S-RSM66	Oligotrophic	-	+	+	-
S-RSM67	Oligotrophic	-	+	+	-
S-RSM68	Oligotrophic	-	+	+	-
S-RSM69	Oligotrophic	-	+	+	-
S-RSM70	Oligotrophic	-	+	+	-
S-RSM71	Oligotrophic	-	+	+	-
S-RSM72	Oligotrophic	-	+	+	-
S-BM1	Oligotrophic	+	+	-	+
S-BM3	Oligotrophic	-	+	-	+
S-BM4	Oligotrophic	-	+	-	+
S-BM6	Oligotrophic	-	+	-	+

Table 3.5*: The presence of *pstS* in marine cyanophage and their ability to infect *Synechococcus* sp. WH7803 under +P and -P conditions

The above results show that the majority of cyanophages isolated on *Synechococcus* sp. WH7803 show a significant decrease in infection efficiency under -P conditions. This handicap is, however, seemingly countered by those cyanophages that possess *pstS* in their genome, even though both *pstS* possessing cyanophages and cyanophages lacking *pstS* were isolated from P-deplete environments. These results were consistent when using two fundamentally different experimental methods, confirming the validity of results.

3.3.3 Phylogenetic relationships between cyanophage and host PstS proteins

In order to assess whether the cyanophage and host *pstS* share the same evolutionary ancestor, a neighbour-joining phylogenetic tree was created, based on PstS amino acid sequences from all known marine and freshwater cyanobacterial strains whose sequences were available in public databases and possessed one or two copies of *pstS* in their genomes (Figure 3.5). In addition, PstS sequences from known *Synechococcus*- and *Prochlorococcus*-infecting cyanophages were also compared. While the phylogeny of cyanobacterial *pstS* has been previously studied, at the time there were almost no cyanophage *pstS* sequences available. In this study, 20 marine cyanophage *pstS* sequences were included in the analysis, including sequences from S-BM1 and S-BM6 which were sequenced during the course of this work. Additionally, and in line with previously published work, phylogenetic analysis included cyanobacterial sequences of another PstS homologue – SphX, the function of which has not yet been elucidated. Finally, the tree was rooted using the sequence of the *E. coli* PstS orthologue.

3.4 Discussion

Almost all isolated and sequenced cyanophages possess auxiliary metabolic genes (AMG) – genes which are orthologous to host metabolic genes and which are thought to interfere with host metabolism during the infection, in order to maximise phage yield. The most investigated AMGs in cyanophages are those which are suspected to participate in carbon metabolism (Mann*, et al.*, 2003; Millard*, et al.*, 2004; Lindell*, et al.*, 2005; Thompson*, et al.*, 2011). By augmenting photosynthetic energy metabolism and amplifying those reactions in carbon metabolism which will increase the production of pentose sugars, cyanophages are hijacking host metabolism and harnessing it for its own needs (Puxty*, et al.*, 2015a). Generally, the nature of AMGs found in different cyanophages will be expected to reflect specific environmental constraints and metabolic bottlenecks which could severely limit production of the next generation of cyanophages (Hurwitz and U'Ren, 2016). Therefore, it is of no surprise that phosphate metabolism-related genes are found in genomes of cyanophages isolated from oligotrophic marine environments, since primary production is P-limited in those waters (Ammerman*, et al.*, 2003). The only data we have regarding how P-limitation may affect cyanophage infection kinetics, has shown that under P-depleted conditions cyanophage S-PM2 infects *Synechococcus* sp. WH7803 with reduced burst size and prolonged latent period
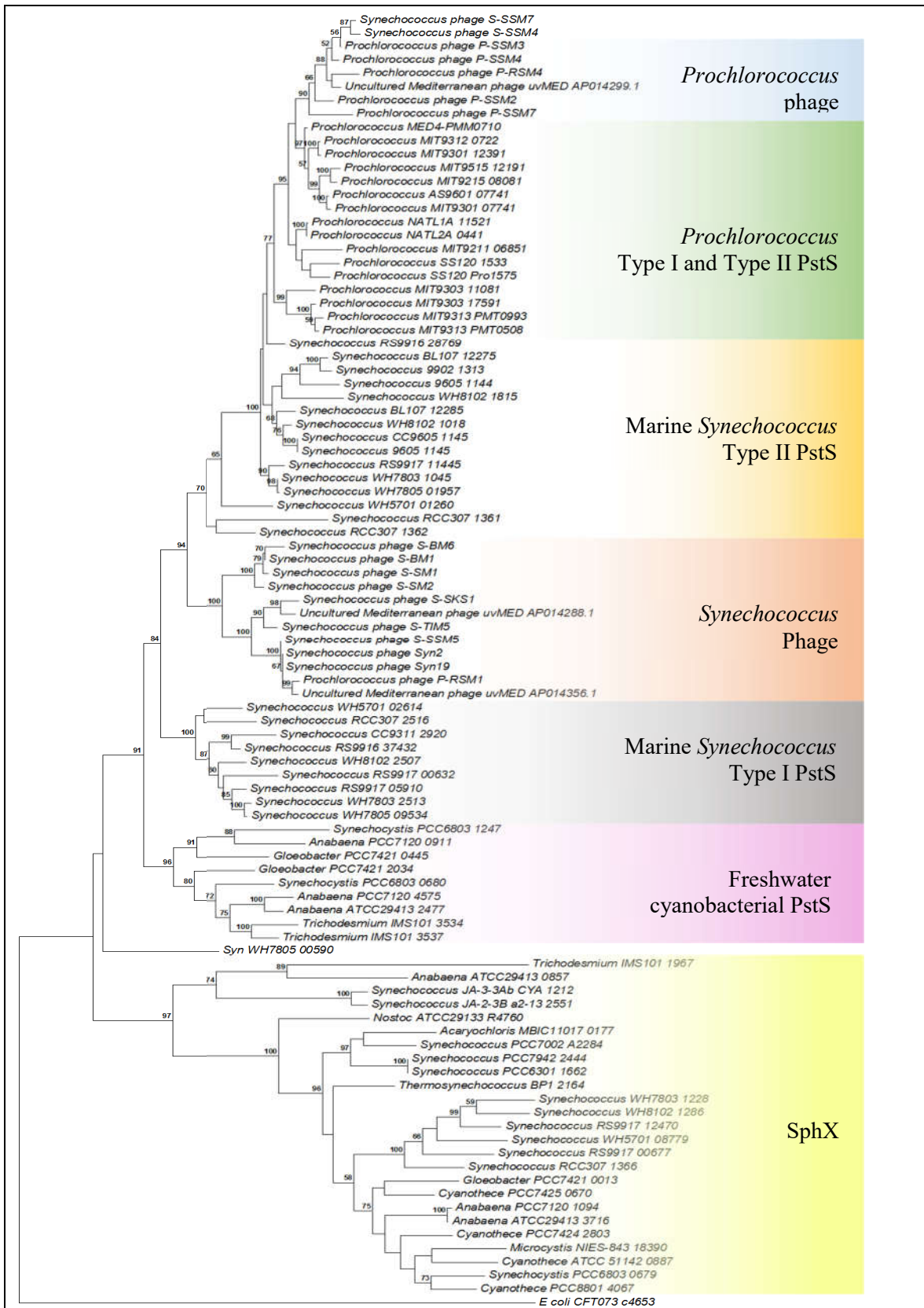
Figure 3.5: legend on the next page

Figure 3.5: Neighbour-joining tree based on the amino acid sequences of cyanobacterial and cyanophage PstS and SphX proteins. The tree was created from amino acid sequences aligned with MUSCLE (Edgar, 2004) The *pstS* sequences were extracted from the NCBI nr database, Cyanorak (http://abims.sb-roscoff.fr/cyanorak/) and Cyanobase (http://genome.microbedb.jp/cyanobase/), using BLAST. S-BM1 and S-BM6 PstS sequences were produced in this study. Sequences were aligned using MUSCLE (Edgar, 2004) and a neighbour-joining tree built in MEGA (Tamura*, et al.*, 2013), using 100 bootstrap replications to determine branch lengths and support values. The scale bar represents 10 substitutions per 100 amino acids.

(Wilson*, et al.*, 1996). S-PM2 was isolated from nutrient-rich waters and its genome sequence has shown that it doesn't possess any apparent P-stress related AMGs (Mann*, et al.*, 2005). Therefore, it is reasonable to expect that it's infection will be sub-optimal under conditions for which it hasn't evolved, judging by the lack of genes which could enable it to overcome this limitation. Correlation between the presence of P-related AMGs in viral metagenomes and the low level of ambient P in the environments from which these metagenomes were sequenced, further supports this assumption (Kelly*, et al.*, 2013).

Figure 3.6: Signal-P v4.1 output for S-BM1 PstS sequence. C-score describes the probability of a cleavage site, S-score estimates the location of signalling peptide and Y score represents a combination of the C-score and the slope of the S-score, producing a more accurate prediction of the exact cleavage site.
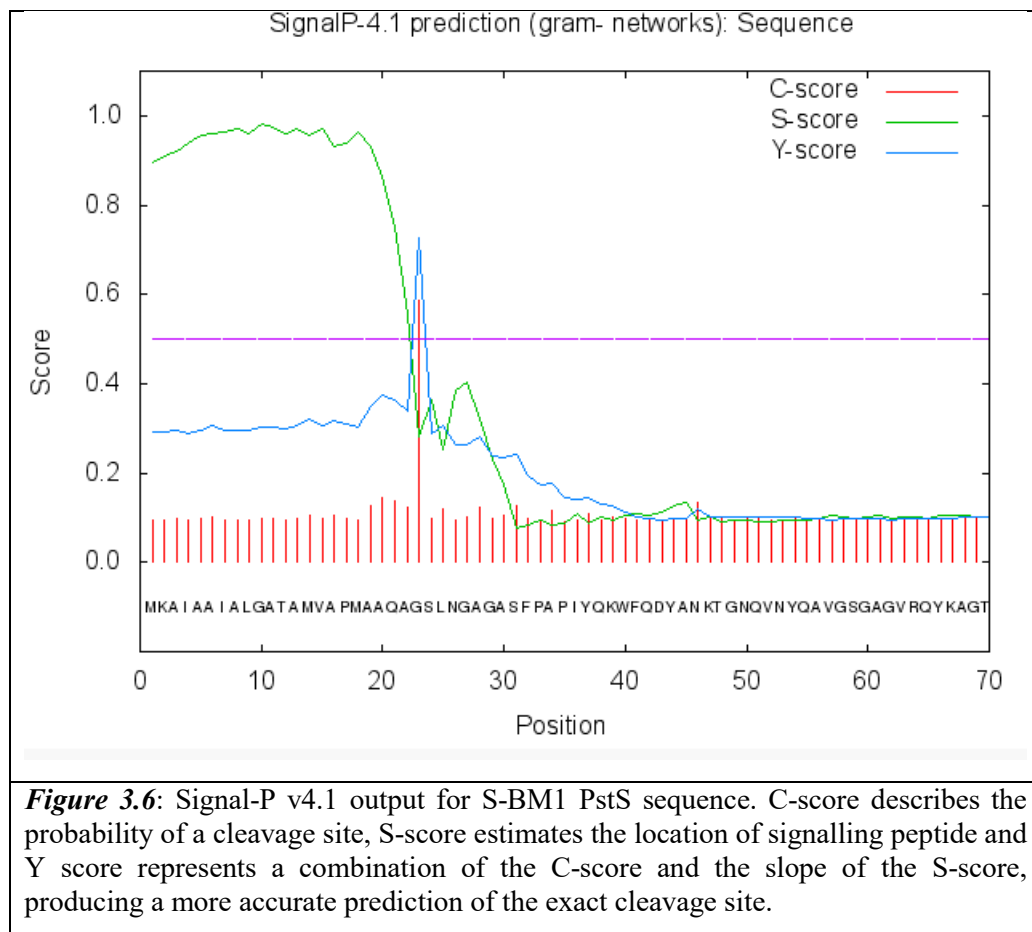
Due to these previous results, this work set out to test the hypothesis that cyanophages isolated from P-deplete waters will possess at least one of the AMGs suspected to be involved in host P metabolism. Furthermore, those cyanophages which were found to possess these genes in

their genomes, were hypothesised to be able to overcome the infection kinetics limitation that are imposed on cyanophages not possessing P-limitation related AMGs, like S-PM2. As mentioned above, sequenced cyanophages possess 3 AMGs which were thought to be involved in P-metabolism – *pstS* (encoding a periplasmic Pi binding protein), *phoA* (encoding a putative alkaline phosphatase) and *phoH* (a gene of unknown function but thought to be involved in P-metabolism in *E. coli*). Out of these 3 genes, *pstS* was chosen to be examined as a marker for the ability to successfully infect a cyanobacterial host under P-limiting conditions. The reasons for this choice were two-fold: firstly, high levels of cyanobacterial PstS are considered to be a marker for P-limitation in the environment (Fuller*, et al.*, 2005). Intracellular levels of PstS increase immediately after the onset of P-limitation in *Synechococcus* sp. WH7803 (Scanlan*, et al.*, 1993). Secondly, RNA sequencing of cyanophage infecting *Prochlorococcus* under P-limiting conditions shows that the cyanophage *pstS* is one of the two over-expressed genes, compared to the +P control (Lin*, et al.*, 2016). Results of this work seems to support the initial hypothesis, since those phages that were found to possess *pstS* in their genomes, were able to overcome the infection handicap imposed by P-limited conditions. While the function of bacterial PstS has been experimentally determined, this function has not yet been elucidated for the viral PstS orthologue. Using the online signal peptide prediction tool SignalP v4.1 (Petersen*, et al.*, 2011) with amino-acid sequences of viral PstS, shows a high probability of a signal peptide at the N-terminus, with predicted cleavage site at position number 22 (Figure 3.6).

The presence of a signal peptide at the N-terminus of cyanophage PstS proteins suggests that, like its host orthologue, it is probably exported to the periplasmic space. This however does not provide a hint towards its function. We already know that the cyanobacterial PstS is over-expressed immediately following the onset of P-limiting conditions. Therefore, it is not clear what potential contribution the product of the viral orthologue would provide to the P-limited infected host. It is possible that the viral protein has a higher affinity towards inorganic phosphate and thus improves the ability of infected hosts to scavenge ambient phosphate. Another possibility is that viral PstS sequesters inorganic P in the periplasm but shuttles it towards a P-transport system other than PstABC, so that the transported P is used preferentially for viral needs. Finally, it is possible that the viral copy is expressed consistently over the course of infection, while the host mRNA expression declines with the onset of the latent phase of infection. Recent work has showed that host mRNA is degraded through the action of virally-induced RNase E, while non-coding viral RNA protects viral mRNA from degradation (Stazic*, et al.*, 2016). Since PstS is important for providing the infected host (and thus the virus) with essential P from the environment, phage PstS would substitute the mRNA-depleted host's PstS activity and keep the supply of inorganic P constant. Further work is

required to understand the difference between host and phage PstS and understand the mode of action of the viral orthologue.

Clustering of cyanobacterial PstS orthologues fits the previously published joined-neighbour trees (Scanlan*, et al.*, 2009), with marine *Synechococcus* and *Prochlorococcus* occupying a separate branch from the freshwater cyanobacteria. Within the marine sub-group, the majority of *Synechococcus* isolates possess two copies of *pstS,* with one of them (*pstS*II*)* being clearly up-regulated under P-limiting conditions. The phylogenetic tree shows a clear distinction between these two copies, with PstSI clustering separately from PstSII. Additionally, *Prochlorococcus* isoforms cluster together with *Synechococcus* PstSII, suggesting similar functionality. This isoform has been extensively studied and shown to be heavily over-expressed under P-limiting conditions (Scanlan*, et al.*, 1993). Combined with the fact that *Prochloroccocus* has undergone specific adaptations to oligotrophic environments, this indicates possession of PstSII as a way of adaptation to nutrient-depleted conditions. Therefore, it is no surprise that cyanophage PstS orthologues cluster together with the PstSII isoform. It is interesting to note that the sequences originating in *Prochlorococcus*-infecting phages, although closely related, cluster separately to *Synechococcus*-infecting phages (except for *Synechococcus*-infecting S-SSM4 and S-SSM7 which group with *Prochlorocccus* phages and *Prochlorococcus*-infecting P-RSM1 which clustered with *Synechococcus*-infecting phages). Although expected, it is notable that PstS found in *Prochloroccus* phages seems to be more closely related to both its host orthologue and PstSII found in *Synechococcus* genomes, while the *Synechococcus*-infecting phage copy clusters separately from all the rest. This could potentially hint towards novel functionality or expression regulation of PstS found in genomes of cyanophages infecting *Synechococcus.*

In the course of previous work performed in the lab, 88 cyanophages were isolated from the Gulf of Aqaba over the course of a year (Millard and Mann, 2006). The Gulf of Aqaba undergoes interchanging periods of water column stratification in summer (during which the photic zone is depleted of macronutrients) and deep winter mixing (during which nutrients from deep waters are brought to the surface via convective mixing (Wolf-Vecht*, et al.*, 1992). Therefore, our hypothesis predicted that cyanophages isolated from the photic zone during the summer months will be better adapted to P-limiting conditions, a quality which was predicted to be demonstrated via possession of *pstS* in the viral genome and the ability to successfully infect the cyanobacterial host under P-limiting conditions. Surprisingly, this was not found to be true. All the Red Sea phage isolates that were isolated during the summer months of water column stratification were apparently not found to possess *pstS* in their genomes nor were they able to successfully infect their host under P-limiting conditions. On the other hand, phages that were isolated from the Bermuda coast in the Sargasso Sea, an area known to be extremely oligotrophic and severely P-limited, were found to possess *pstS* orthologue in their genomes

and had similarly successful infection parameters under P-limiting and P-replete conditions. The lack of *pstS* in the summer RS phage isolates, comes as a surprise, both due to the fact that a *Prochlorococcus* phage, P-RSM4, that was isolated from the same sampling location and during the same time period, does possess a *pstS* orthologue in its genome (Henn*, et al.*, 2010). Furthermore, it was shown that bacterial PstS levels during those summer months are significantly high *in situ (*Fuller*, et al.*, 2005).

A possible reason for this gap between expected and actual results, may lay in the identity of the host on which the Red Sea cyanophages were isolated. All the RS isolates were isolated on *Synechococcus* sp. WH7803 – a strain belonging to clade V which is mostly found in eutrophic waters and is therefore not adapted to prolonged P-deplete conditions. In addition, during the summer months and after water column stratification sets in, previously dominant *Synechococcus* populations are replaced by *Prochlorococcus* which becomes dominant in the photic zone of the water column (Lindell and Post, 1995). The *Prochlorococcus* genus contains members which have been strongly selected for survival under oligotrophic conditions (Bouman*, et al.*, 2006). Therefore, cyanophages isolated on eutrophic *Synechococcus* during the stratified summer months, will be representative of the population that infects cyanobacteria which thrives under the nutrient-rich conditions of water-column mixing. In order to increase the chance of isolating *Synechococcus*-infecting phage, that does possess a *pstS* orthologue from the Gulf of Aqaba, cyanophages would need to be isolated on *Synechococcus* strains which become dominant during the stratified months. Previous work has identified that *Synechococcus* strains that dominate the photic zone during the stratified conditions belong predominantly to clade III, while deep mixing months showed the prevalence of clade XII. *Synechococcus* sp. WH7803, which was used as a model organism in this study, belongs to clade V, members of which were found in the photic zone during the transition months between mixing and stratification (Fuller*, et al.*, 2005; Post*, et al.*, 2011). Therefore, in order to isolate cyanophages that would possibly be adapted to the low-P conditions, the host of isolation should be from clade III, such as *Synechococcus* sp. WH8102. These results should provide guidance for future attempts of isolation of environmental cyanophages, underlining the importance of choosing the host of isolation that would be representative of the dominant population both from temporal and biogeochemical aspects.

Chapter 4:

Genomic analysis of cyanophages isolated from the Gulf of Aqaba, Red Sea and the Sargasso Sea, Bermuda Archipelago

4.1 Introduction

Viruses are the most abundant biological entities on the planet (Suttle, 2005), influencing the entirety of life via important roles in shaping the composition of microbial communities (Dalmasso*, et al.*, 2014) and modulating major biogeochemical cycles (Puxty*, et al.*, 2016), amongst others. Such ecological roles are in stark contrast to the dearth of information regarding the genomic composition of bacteriophages and the function of specific phage genes. Indeed, the majority of genes found in viral metagenomes are unknown and are classed as hypothetical, with many not being found even in sequenced bacteriophage genomes (Willner*, et al.*, 2009; Hurwitz and Sullivan, 2013). Therefore, it is of great importance to continue increasing the number of sequenced bacteriophage genomes, which will expand our ability to classify these unknown genes to specific phage families and get closer to understanding their role in the infection cycle.

In spite of their importance in biogeochemical cycling and marine microbial community shaping, there is a surprisingly low number of complete sequenced genomes of isolated marine bacteriophages. At the time of writing, out of 2010 sequenced bacteriophage genomes submitted to the EBI phage database (http://www.ebi.ac.uk/genomes/phage.html), only 482 are marine phage genomes (Figure 4.1). This disproportion may be a product of the relative difficulty of isolating and culturing marine microbes (Moore*, et al.*, 2007; Carini*, et al.*, 2013), a step required for isolating phages that infect them.

Metagenomic sequencing of global oceanic viral populations shows to an even larger extent the gap in knowledge about viral genomes – out of 5476 predicted viral populations in the upper ocean layers, only 39 could be assigned to viral families belonging to cultured isolates (Brum*, et al.*, 2015).

In addition to increasing our understanding of the prevalence of unknown genes in viral metabolism, sequencing phage isolates from known cultured bacteria has even greater importance due to our ability to understand the dynamics of phage populations in the context of the environmental parameters prevalent at the location of isolation of both phages and their bacterial hosts. Previous sequencing of marine bacteriophages has contributed to our understanding of the biology and metabolic requirements of both phages and their bacterial hosts (Sabehi*, et al.*, 2012; Kang*, et al.*, 2013; Zhao*, et al.*, 2013; Chan*, et al.*, 2014). Combined with knowledge of the dynamics of bacterial populations over temporal and geographic scales (Lindell and Post, 1995; Farrant*, et al.*, 2016) and changes in the viral metagenomic genetic arsenal in relation to the prevailing nutrient conditions at the isolation location (Kelly*, et al.*, 2013), it is clear why sequencing of additional marine bacteriophage genomes is important for understanding of the marine microbiome and its effect on biogeochemical cycling.

Figure 4.1: The number of bacterial genomes, bacteriophage genomes and marine bacteriophage genomes submitted to the International Nucleotide Sequence Database Collaboration (INSDC), per year. Adapted from (Perez-Sepulveda, *et al.*, 2016).
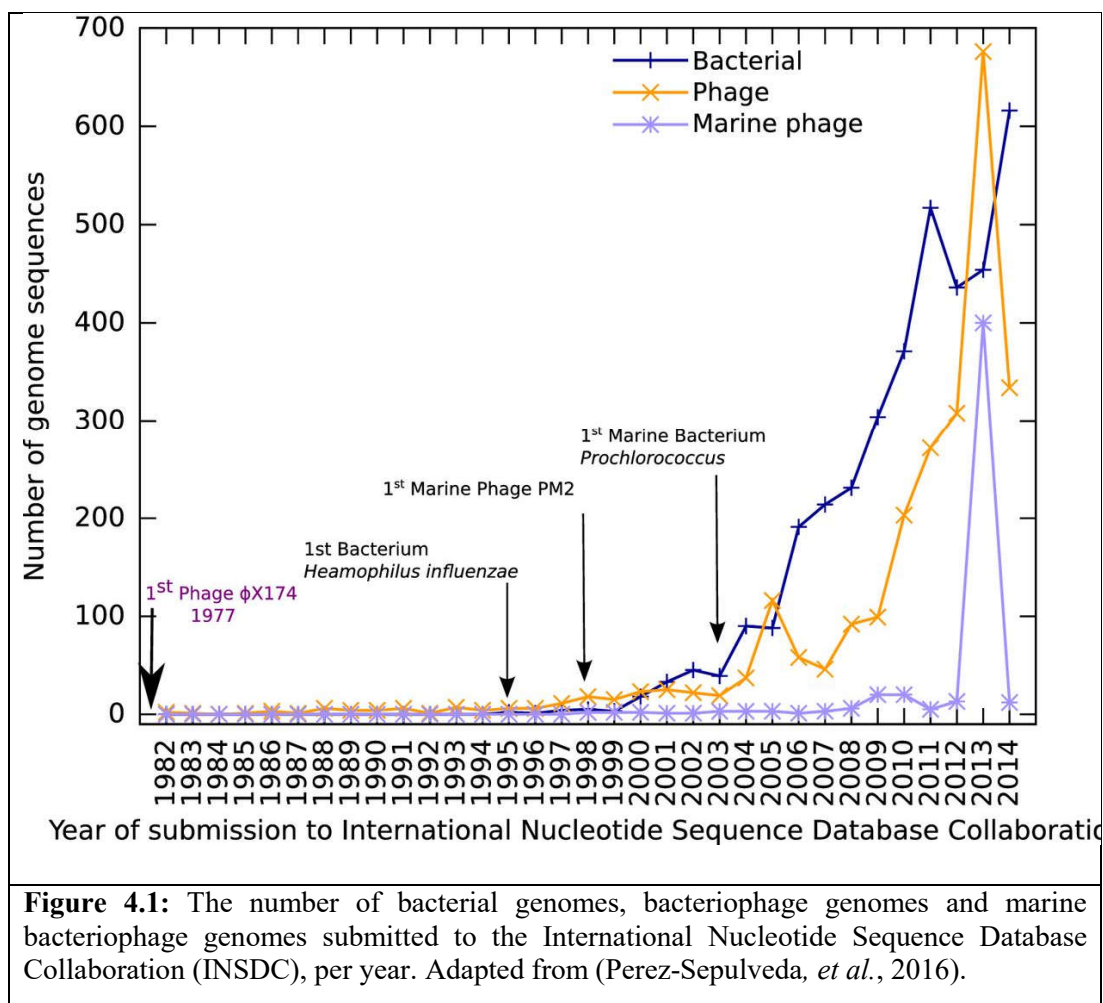
In this chapter, I describe the results of sequencing 6 cyanophages, isolated from two geographically separate locations. Three of the phages sequenced here (S-RSM61, S-RSM70, S-RSM78) were previously isolated from the Gulf of Aqaba, Red Sea, as part of a larger cyanophage isolation effort, which produced close to 100 phage isolates that have not yet been sequenced (Millard and Mann, 2006). As described in Chapter 3, the availability of environmental data, paired with the cyanobacterial abundance data at the time and location of isolation of these phages (Fuller, *et al.*, 2005), has the potential to test hypotheses relating to the role of cyanophages in marine nutrient cycling. The second group of three phages (S-BM1, S-BM3, S-BM6) were isolated from oligotrophic waters of the Sargasso Sea, in the vicinity of the Bermuda Biological Station for Research (Wilson, *et al.*, 1993) (Figure 4.2). As described in chapter 3, some of the cyanophages isolated from these waters, possess an orthologue of the periplasmic binding protein component of the phosphate transport operon, PstS (Scanlan, *et al.*, 1993), while the Red Sea phages have not shown the presence of this gene in their genome, nor the ability to avoid P-stress induced negative effect on infection physiology (Chapter 3).

59

Figure 4.2: Map showing global chlorophyll *a* concentrations as captured by SeaWIFS satellite technology (http://seawifs.gsfc.nasa.gov/) with the location of isolation sites of cyanophages sequenced in this study indicated by asterisks.
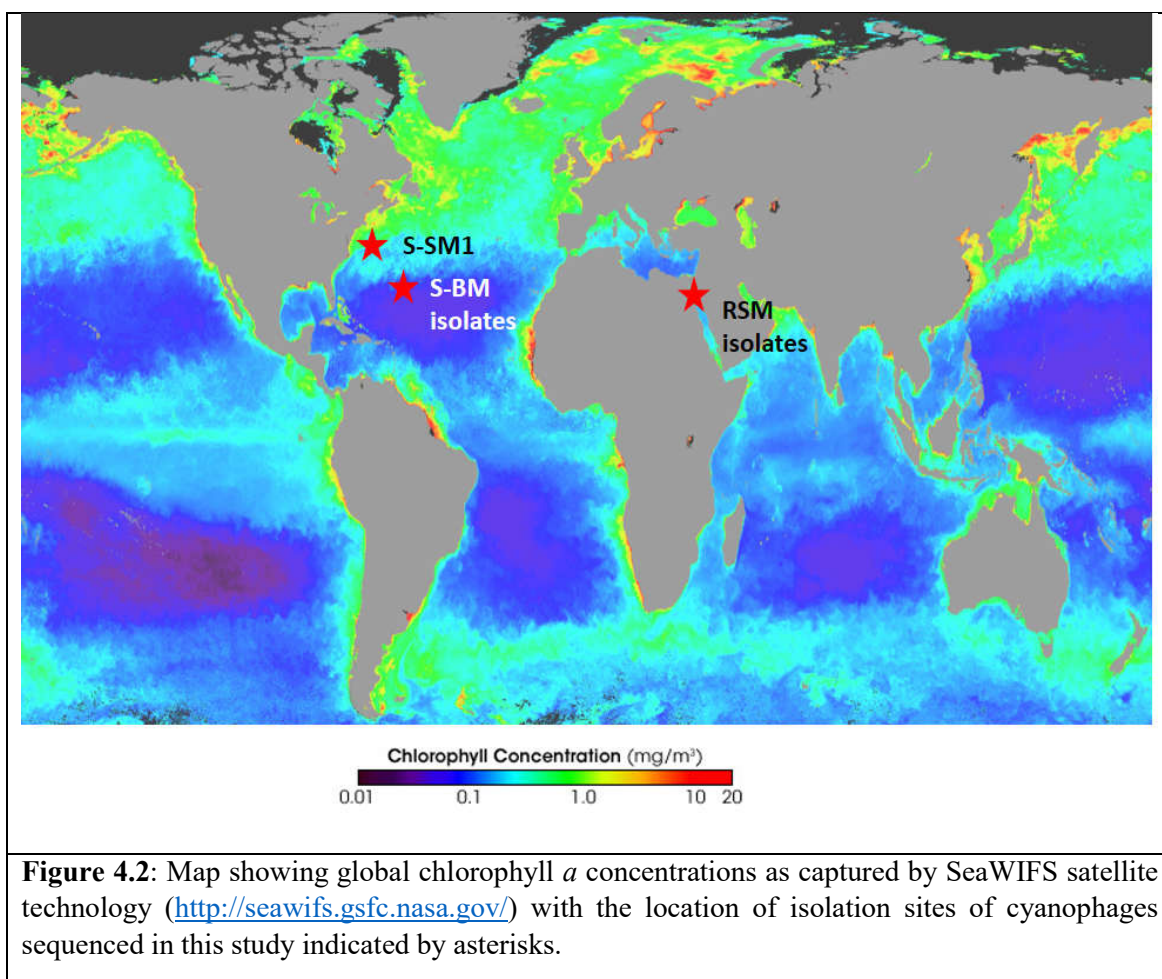
4.2 Aims and Objectives

In order to further understand cyanophage metabolic capability, and potentially correlate this capacity with environmental conditions prevalent at the time and location of isolation and examine the variability amongst isolates, I sequenced and analysed several cyanophages isolated from the Gulf of Aqaba, Red Sea, and Sargasso Sea off Bermuda.

The objectives of this chapter were:

1. To perform genomic analysis of six cyanophage isolates.

2. To correlate the specific gene content of these sequenced cyanophages with the environmental conditions prevalent at the time and location of their isolation.

4.3 Materials and methods

4.3.1 Cyanobacterial host growth conditions

All the cyanophages used in this study were originally isolated on *Synechococcus* sp. WH7803 which was used for propagation of the cyanophages for the purpose of DNA isolation and sequencing. *Synechococcus* sp. WH7803 was grown on ASW medium, under continuous illumination conditions, as described in section 2.1.

4.3.2 Cyanophage infection and purification

Cyanophages were propagated on *Synechococcus* sp. WH7803 as described in section 2.6.1. Briefly, 0.5L of exponentially growing *Synechococcus* sp. WH7803 culture was infected with each cyanophage isolate and kept incubated under regular growth conditions until visible lysis. Lysate was obtained by centrifugation at 4° C at 3,000 x g for 10 minutes in order to remove cell debris. Collected supernatant was incubated for 1 hr at room temperature with 10 ng ml^{-1} of RNase A (Invitrogen) and 0.25 SU ml^{-1} of DNaseI (New England Biolabs) and centrifuged at 13,000 x g for 30 mins at 4ºC. Supernatant was collected and incubated overnight with 100g L^{-1} of PEG 8000 (Fisher Bioreagents) at 4° C. Cyanophages were pelleted by centrifugation using the Beckman Coulter Avanti JLA 10.500 for 30 minutes at 4° C at 13,000 x g. Pellets were resuspended in ASW medium and washed in an equal volume of chloroform twice in order to remove any remaining PEG from the solution.

4.3.3 Cyanophage DNA purification and sequencing

Resuspended PEG-concentrated cyanophages were washed with 1 volume of phenol/chloroform/iso-amyl alcohol (IAA) (25:24:1) and centrifuged. The resulting supernatant was washed with 1 volume of chloroform/IAA (24:1) and supernatant collected. The DNA from the resulting aqueous phase was precipitated using 1 volume of isopropanol/7.5M ammonium acetate (3:2) and incubated overnight at -20° C. The precipitated DNA was pelleted by centrifuging at 10,000xg for 45 minutes and the resulting pellet washed with 1 volume of 70% ethanol (v/v). After evaporation of ethanol the pellet was resuspended in MilliQ water, quantified using Qubit (Life Technologies) and stored at -20° C. Prior to sequencing, cyanophage DNA was thawed and diluted to 0.2 ng µl^{-1}. Libraries were prepared using the NexteraXT (Illumina) protocol, according to the manufacturer's instructions and sequenced using the MiSeq (Illumina) platform at the Warwick Medical School.

4.3.4 Bioinformatic analysis

Obtained sequences were trimmed using Sickle trimming tool (https://github.com/najoshi/sickle) with the following parameters: -t sanger -l 150 -q 25 -n 1. Trimmed sequences were then assembled using SPAdes v3.1 (Bankevich, *et al.*, 2012) with default parameters. Assembled contigs were visually examined and the contig with the appropriate size (100-200kb) and high coverage was chosen as the assembled cyanophage genome. The genomes were then annotated using Prokka v1.11 (Seemann, 2014) against a T4-like protein database (Sullivan, *et al.*, 2010) using the following parameters: --addgene.

Original reads were remapped to the newly assembled genome in order to assess quality and coverage with bwa-mem (Li, 2013; Li and Durbin, 2009) and SAMtools (Li, *et al.*, 2009). The resulting report was produced using qualimap (Garcia-Alcalde, *et al.*, 2012). Genomes that were suspected of being identical were tested for SNPs and indels against each other using SAMtools mpileup using default parameters and Varscan v.2.3.9 (Koboldt, *et al.*, 2009).

In order to examine the presence of cyanophage core genes, as previously defined (Sullivan, *et al.*, 2010), pangenomes were created and analysed using Roary (Page, *et al.*, 2015). Genome alignment were visualised using Easyfig (Sullivan, *et al.*, 2011) or BRIG (Alikhan, *et al.*, 2011).

4.4 Results

The assembly of each of the sequencing reads produced a number of contigs for each cyanophage genome. However, in each set of contigs, a single contig stood out both in terms of the length and the read coverage. The results of the Qualimap report on each of the sequenced genomes can be seen in Table 4.1 and Figure 4.3.

Name	Number of sequencing reads	Genome size	Mean Fold Coverage	GC%
S-RSM61	2,062,611	187,109	2,073.81	38.33%
S-RSM70	377,470	194,581	270.29	41.3%
S-RSM78	505,986	186,871	575.75	38.96%
S-BM1	177,010	175,222	244.77	42.42%
S-BM3	497,052	169,344	725.72	41.38%
S-BM6	297,397	175,287	402.12	43.71%

Table 4.1 – Sequencing statistics of the phage genomes sequenced in this study

Name	Host of isolation	Genome size (kb)	# ORFs	%GC	Site of isolation	# of tRNAs	Reference
S-RSM4	*Synechococcus* sp. WH7803	194.5	238	41.1%	Station A, Gulf of Aqaba, Red Sea (29º28'N, 34 º55'E)	12	(Millard and Mann, 2006)
S-RSM61		187.1	226	38.33%		26	
S-RSM70		194.6	232	41.3%		12	
S-RSM78		186.9	225	38.96%		26	
S-SM1	*Synechococcus* sp. WH6501	178.5	234	41.0%	Atlantic Slope Waters, 0m (38°10'N, 73°09'W)	6	(Sullivan, *et al.*, 2010)
S-BM1	*Synechococcus* sp. WH7803	175.2	217	42.42%	Sargasso Sea, Bermuda Biological Station for Research (32°22'N, 64°41'W)	6	(Wilson, *et al.*, 1993)
S-BM3		169.3	212	41.38%		11	
S-BM6		175.3	216	43.71%		6	

Table 4.2: Properties of the sequenced cyanophage isolates. S-RSM4 and S-SM1 were sequenced in previous studies and are included here for the sake of comparison.

SNP and indel analysis of the sequenced cyanophage genomes revealed that the S-RSM70 genome was identical to a previously sequenced genome, S-RSM4 (see (Millard, *et al.*, 2009). Similarly, the genomes of cyanophage S-RSM61 and S-RSM78 were found to be identical, showing that even though these pairs of cyanophages were isolated on different dates and from different depths in the Gulf of Aqaba, Red Sea, they could still be found and isolated from different environmental samples. The dates and depths of isolation of these 4 cyanophages can be seen in Table 4.3.

Name	Date of isolation	Depth of isolation	Sequenced
S-RSM4	27/04/1999	50m	(Millard, *et al.*, 2009)
S-RSM61	18/07/1999	30m	This study
S-RSM70	23/08/1999	10m	This study
S-RSM78	18/07/1999	150m	This study

Table 4.3: Dates and depths of isolation of cyanophages sequenced in this study. Data adopted from (Millard and Mann, 2006)

Similar genomic analysis of cyanophages isolated from the Sargasso Sea in the vicinity of the Bermuda Biological Station for Research showed that the S-BM1 and S-BM6 genomes were also identical, while S-BM3 was quite different. Due to these genomes being identical, the remainder of the analysis was done only on the S-BM1, S-BM3 and S-RSM61 genomes. An alignment of the genome sequences of the S-BM and S-RSM isolates can be seen in Figure 4.4.

Figure 4.3: Qualimap coverage maps of the genome sequences of the four isolated cyanophages. The graph under each coverage map shows the %GC content across the genome.
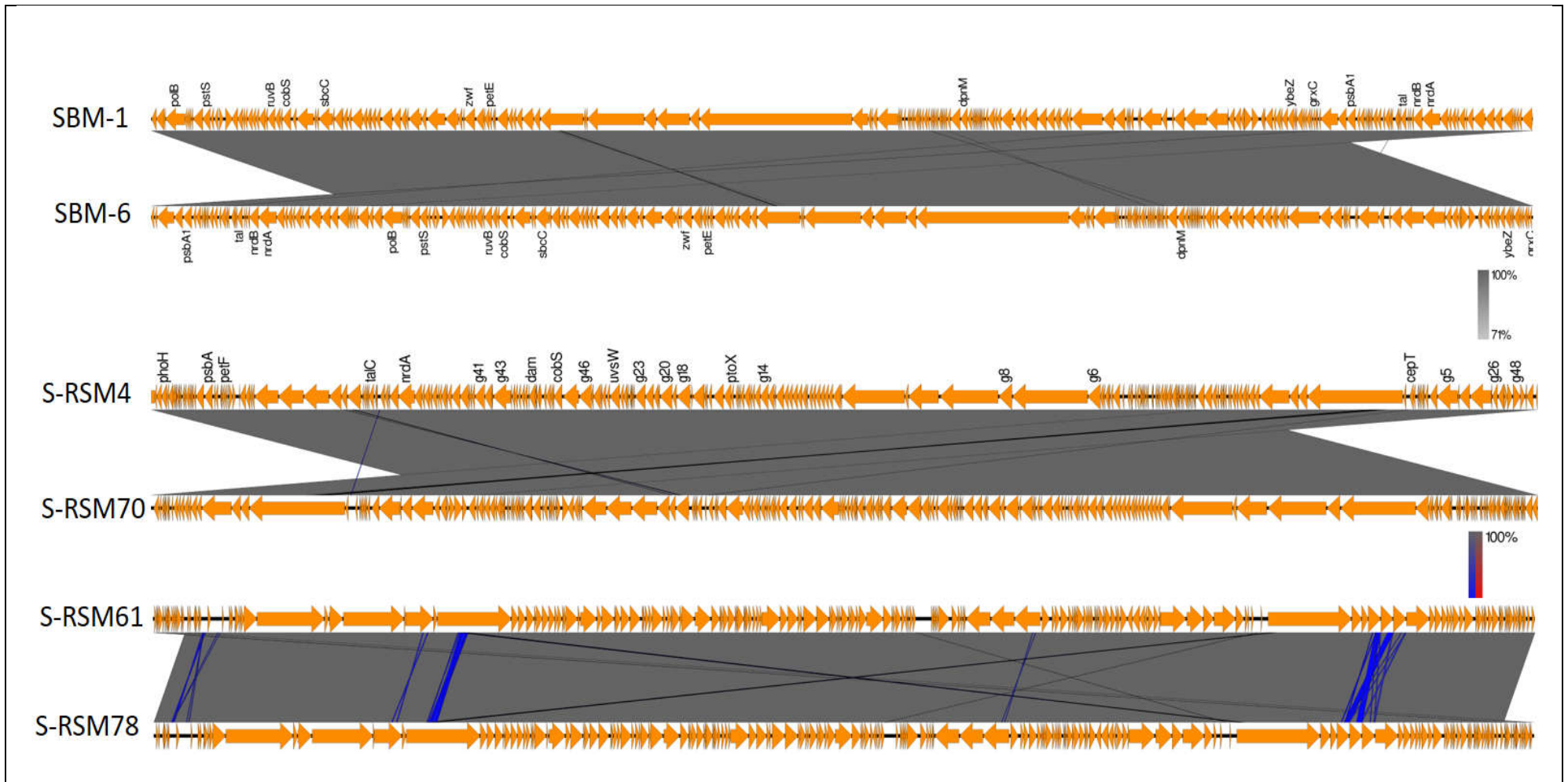
Figure 4.4: Genome comparison of cyanophage isolates S-BM1 and S-BM6, S-RSM4 and S-RSM70, and S-RSM61 and S-RSM78

An interesting observation that arose from initial analysis of the coverage maps is that in the case of each of the sequenced cyanophages, there is an area of the genome in which coverage is pronouncedly higher than the rest of the genome (Figure 4.3). This observation has been made in the past with genome sequences of cyanophages S-PM2 and R-SM4 (Puxty, *et al.*, 2015b; Rihtman, *et al.*, 2016), although no unique properties were found to be characteristic of these regions. The %GC content of these regions is similar to the rest of the genome and there are no apparent repeats that could account for the possible over-representation of these fragments in the sequencing reads.

Another feature observed following genome analysis was the difference in tRNA content between S-RSM61 and the rest of the newly sequenced cyanophages. While the majority of sequenced cyanophages have been shown to possess between 6-12 tRNAs in their genomes, S-RSM61 has 26. Interestingly, the only other cyanophage that was found to contain a larger number of tRNAs in its genome, S-PM2, contains 25 tRNAs (Mann, *et al.*, 2005). Currently it is not known what selective advantage an increased tRNA inventory provides to the infecting cyanophage, but it is interesting to observe a range of cyanophages isolated from the same environment with such differences in tRNA content.
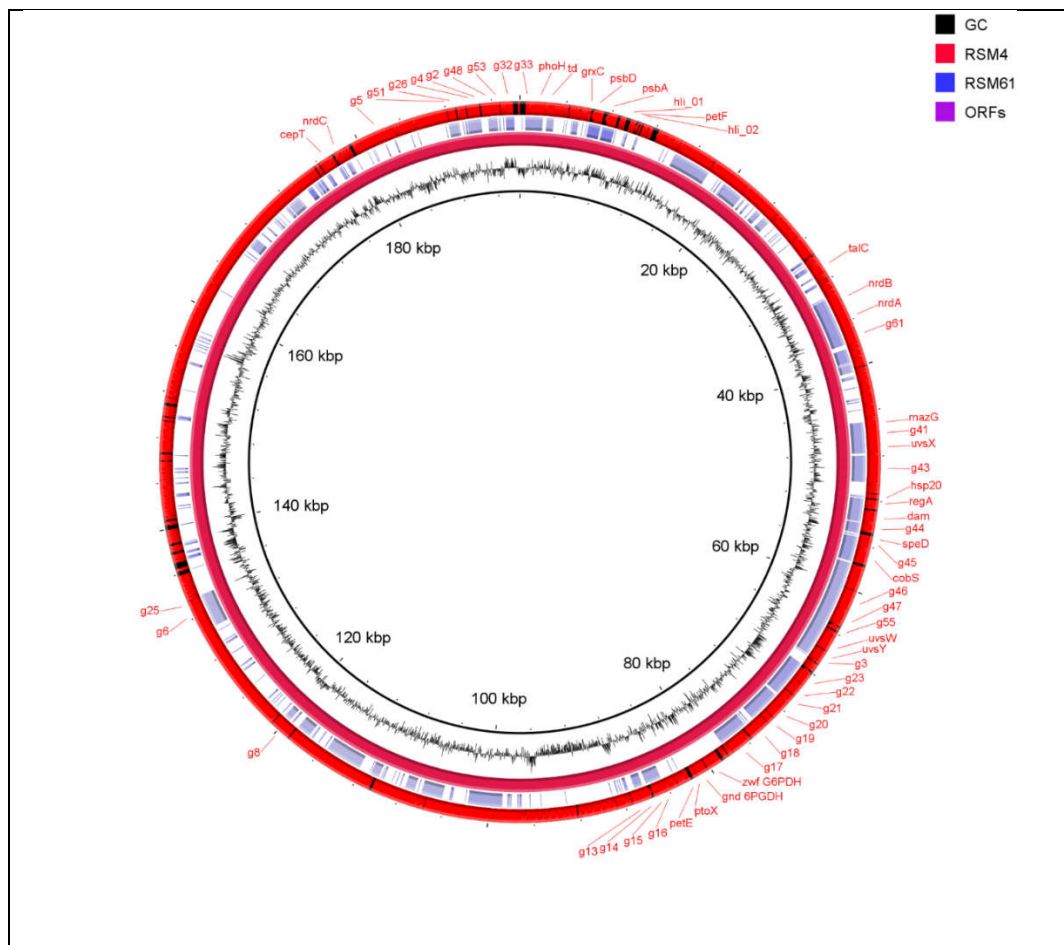
Figure 4.5: Genome alignment of S-RSM61 against the genome of S-RSM4

An interesting aspect of cyanophage genome analysis is evaluation of their auxiliary metabolic gene (AMG) content. Inspection of the differences in AMG arsenal between the different cyanophages sequenced in this study, shows that the environmental conditions are not the only parameter that can be used to predict the metabolic capabilities of cyanophages. Firstly, as expected, all of the sequenced cyanophages possess the photosynthetic genes which are the hallmark of marine cyanophages isolated thus far – *psbA* and *psbD*. Interestingly, all of the S-BM cyanophages contain the carbon metabolism-related AMGs 6-phosphogluconate dehydrogenase (*gnd*), glucose 6-phosphate dehydrogenase (*zwf*), transaldolase (*talC*) and CP12, the latter the Calvin cycle inhibitor also present in plants and cyanobacteria (Tamoi*, et al.*, 2005) (Table 4.4), while S-RSM61 does not contain any of these four genes.

Figure 4.6: Genome alignment of S-BM1 and S-BM3 against the genome of S-SM1
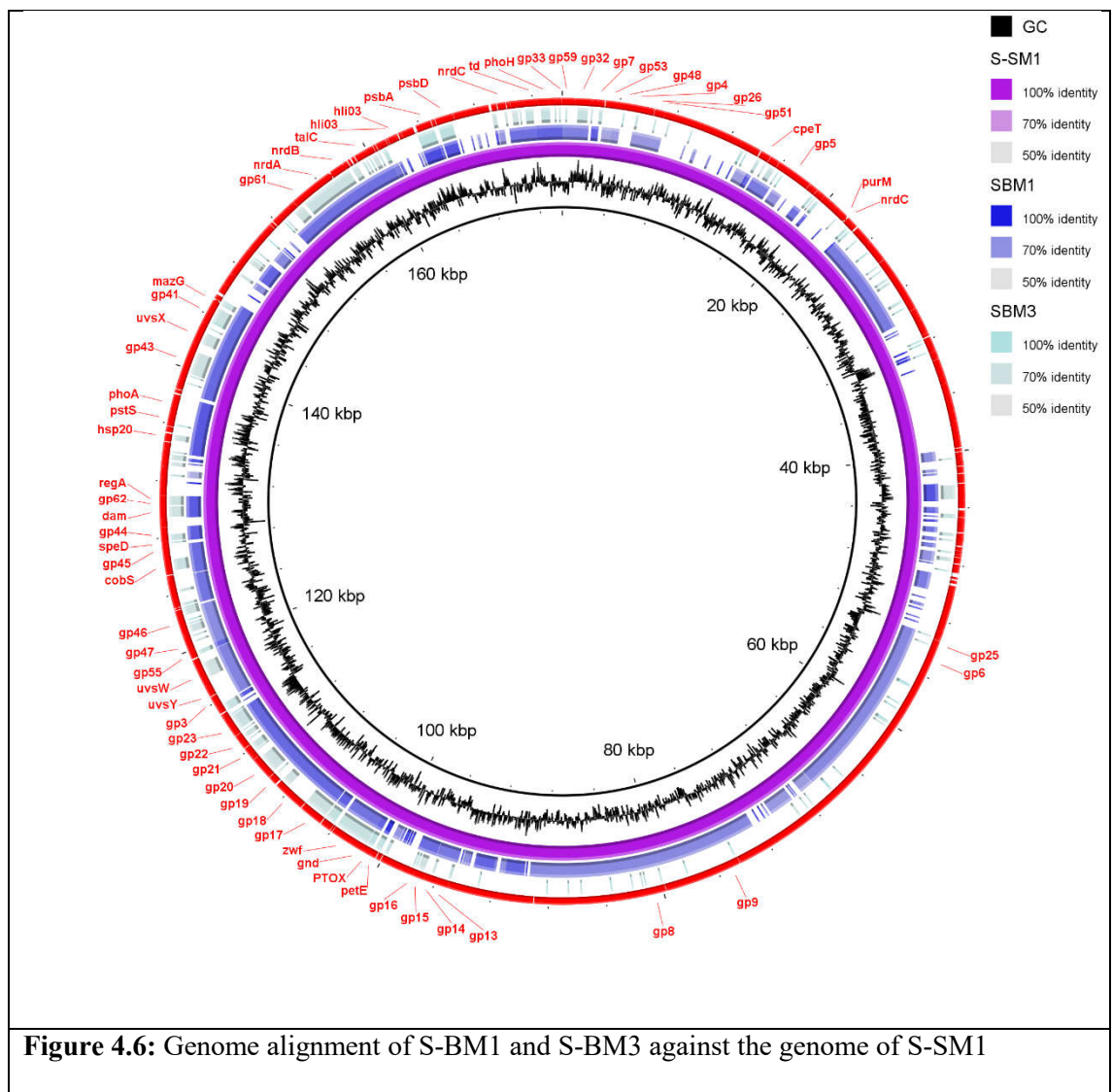
An additional group of genes that are notably missing from some of the newly sequenced cyanophage genomes are those genes related to phosphate metabolism (see Table 4.4). Thus, both *phoA* and *pstS* are absent from the S-RSM and S-BM3 cyanophages but are present in cyanophages S-BM1 and S-BM6. As has been discussed in Chapter 3, the reasons for the absence of *phoA* and *pstS* from the S-RSM phages are probably related to the seasonal community shifts that the niche from which the phages were isolated experiences over the course of a year. More surprising is the absence of P-metabolism related genes from the genome of S-BM3. Even though this cyanophage was isolated from the same geographic location and at the same time as S-BM1, the absence of *pstS* and *phoA* from this phage hints towards a variety of conditions under which cyanophage infection occurs in the ocean, where the absence of otherwise beneficial genes, may "clear a place" in the phage genome for acquisition of different sets of genes, thus providing a potential advantage under different environmental conditions, and driving the evolution of cyanophage sub-populations towards larger speciation.

Gene	Gene product	S-SM1	S-BM1	S-BM3	S-RSM4	S-RSM61
Photosynthesis-related						
psbA	Photosystem II D1 protein	+	+	+	+	+
psbD	Photosystem II D2 protein	+	+	+	+	+
hli	High light-inducible protein	-	-	-	-	-
PTOX	Plastoquinol terminal oxidase	+	+	+	+	-
petE	Plastocyanin	+	+	+	+	-
petF	Ferredoxin	-	-	+	+	-
speD	S-adenosyl-methionine decarboxylase	+	-	-	+	+
cpeT-like	Phycocyanobilin biosynthesis	+	-	+	-	-
pebS	Phycoerythrobilin biosynthesis	-	-	-	-	-
pcyA	Phycobilin biosynthesis	-	-	-	-	-
ho1	Haem oxygenase	-	-	-	-	-
Hypothetical	Ferrochelatase domain-containing protein	+	-	-	-	+
Hypothetical	2OG-Fe(II) oxygenase superfamily protein	+	+	+	+	+
Carbon metabolism						
gnd	6-phosphogluconate dehydrogenase	+	+	+	+	-
zwf	glucose-6-phosphate dehydrogenase	+	+	+	+	-
cp12	Thioredoxin-mediated regulation of the Calvin–Benson cycle.	+	+	+	+	-
talC	Transaldolase	+	+	+	+	-
Phosphate metabolism						
phoA	Alkaline phosphatase	+	+	-	-	-

pstS	Substrate-binding component of ABC-type phosphate transporter	+	+	-	-	-
phoH	P-starvation inducible protein	+	+	+	+	+
Other						
prnA	Tryptophan halogenase	-	-	-	-	+
Hypothetical	S-layer domain protein	+	-	-	-	-
Hypothetical	Carboxylesterase	-	-	-	-	-
Hypothetical	HN-haemaglutanin neuraminidase	-	-	-	-	-
Hypothetical	Carbamoyl-transferase	-	-	-	-	-
Hypothetical	tRNA ligase	-	-	-	-	-
mazG	Pyrophosphatase	+	+	+	+	+
cobS	Porphyrin biosynthetic protein	+	+	+	+	+
Hypothetical	Phytanoyl-CoA-dioxygenase domain	+	+	+	-	+

Table 4.4: The presence of commonly found auxiliary metabolic genes, potentially involved in host metabolic reactions in the newly sequenced cyanophage genomes. S-SM1 and S-RSM4 are listed for comparison sake only and were not sequenced in the course of this study.

4.5 Discussion

Sequencing the genomes of phages from different environments provides an opportunity to investigate the correlation between the genome content of bacteriophages, the metabolic capabilities of their hosts and the environmental parameters (such as nutrients, light intensities, etc.) which shape the microbial communities in the niches from which such phages were isolated. In this case, cyanophage genomes from two geographically distant locations (Sargasso Sea off Bermuda and the Gulf of Aqaba, Red Sea) were sequenced. Even though these two marine environments are geographically separate, their oligotrophic nature formed the basis of a hypothesis in which nutrient depletion in general, and low levels of inorganic phosphate in particular, were thought to be responsible for shaping the evolution, and gene content of, the cyanophage genomes isolated from those niches. Previously isolated cyanophages from those environments have provided hints that this indeed was the case: *Prochlorococcus*-infecting myovirus P-RSM4, isolated from the same location as the the S-RSM cyanophages analysed here (Sullivan*, et al.*, 2010), was shown to possess P-stress related genes. In a different *Prochlorococcus*-infecting cyanophage, P-SSM2, isolated from the Sargasso Sea, near to where the S-BM cyanophages were isolated, phosphate-related auxiliary metabolic genes (AMGs) were significantly over-expressed when infecting a P-stressed cyanobacterial host ((Lin*, et al.*, 2016). Cyanophage S-BM1 phage, also isolated from the oligotrophic Sargasso Sea in the North Atlantic Gyre, has been shown in my own work to be able to evade P-stress induced burst size limitation and delay (Chapter 3). Due to these findings, it is a testimony to the local variety of isolated cyanophages that there are pronounced differences even between phages isolated from the same environment. The fact that S-BM3 does not contain any of the putative phosphate stress-related AMGs, points towards the possibility of another level of speciation of phages to hosts that are experiencing short periods of nutrient availability, possibly following bursts of co-existing bacteria infected by different phages. Another notable difference in genome content between S-BM3 and S-BM6 is the lack of photosynthesis-related orthologues *cpeT* and *petF,* encoding a putative phycoerythrin metabolism component and a putative ferredoxin, respectively, in S-BM6. With the caveat that these phage AMGs were never functionally characterised, it would be interesting to examine whether there is an effect on the infection kinetics between these two phages under different light intensities and whether the presence of these photosynthesis-related AMGs in S-BM3 provides it with any advantage over S-BM6 under these experimental conditions. Additionally, it would be interesting to examine the kinetics of infection of these phages under different nutrient conditions as well as perform competition experiments and examine phage population dynamics prior to and following phosphate increase bursts under laboratory conditions.

The absence of P-stress related genes in S-RSM phages was discussed previously (see chapter 3). The interesting finding is that a certain stability seems to exist amongst cyanophage genotypes isolated from the same area over a range of isolation dates. This contradicts previous findings showing that a large variation in cyanophage lineages exists between samples isolated from the same location on different dates (Dekel-Bird, *et al.*, 2015). However, this lack of genetic variation in *Synechococcus*-infecting cyanophage genotypes is most probably a result of the small sample size sequenced in this study as well as the fact that all of the cyanophages investigated here were isolated on a single host – *Synechococcus* sp. WH7803 (Millard and Mann, 2006).

In addition to the absence of P-stress related AMGs in the S-RSM cyanophages, S-RSM61 seems to be unique in its lack of carbon metabolism-related AMGs. Except for the photosynthesis genes, no known carbon related genes were found in the S-RSM61 genome, in stark contrast to both S-RSM4 and both the S-BM phages. Carbon metabolism-related genes in cyanophages are thought to participate in the Calvin cycle and Pentose Phosphate pathway, where they influence the relative abundance of intermediate products of both cycles, tilting the metabolic balance towards production of NADPH, reduction of ATP hydrolysis and metabolism of nucleotide precursors (Thompson, *et al.*, 2011) (Figure 4.7).
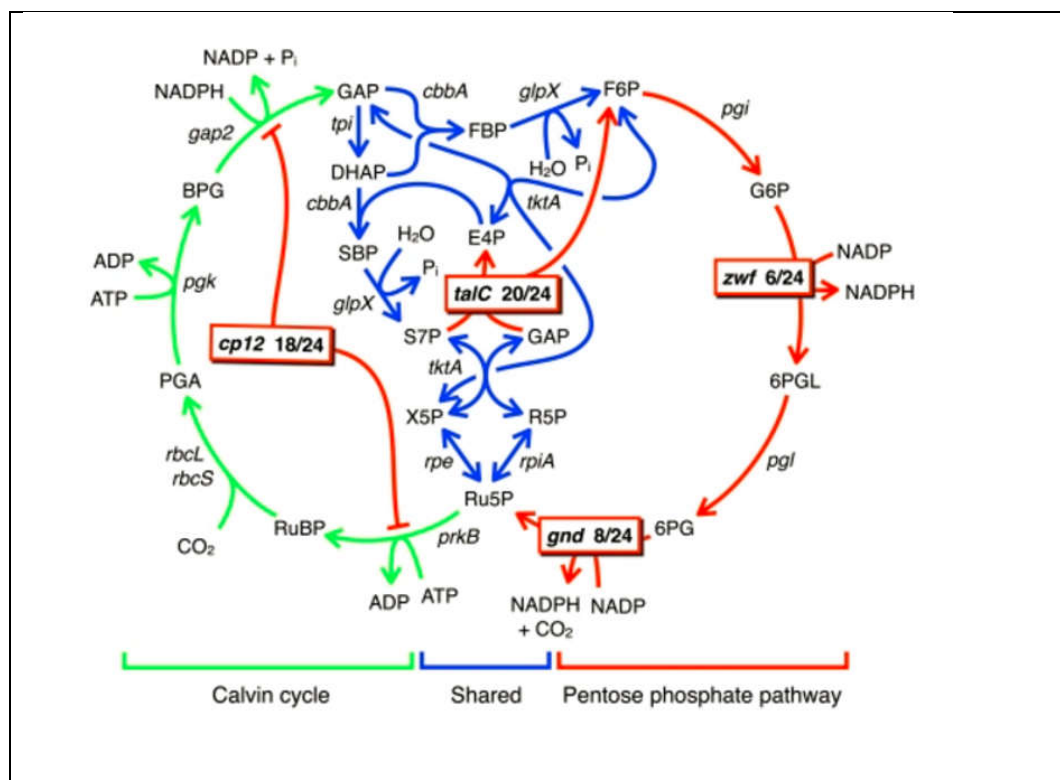
Figure 4.7: Schematic representation of the cyanobacterial Calvin cycle and Pentose Phosphate Pathway. Boxed genes represent gene orthologues found in cyanophages with the number of cyanophage genomes in which these genes were found indicated (at the time of publication of (Thompson, *et al.*, 2011) from which this image was adapted.

Since carbon related host orthologues play an important role in energy and carbon metabolism during cyanophage infection (Puxty, *et al.*, 2016), it would be interesting to further investigate the metabolic changes occurring in the *Synechococcus* host during infection by S-RSM61 compared to the closely related S-RSM4 which possesses all of the carbon metabolism-related AMGs in its genome.

Another interesting finding arising from this study is the division between two groups of phages based on the number of tRNA-encoding genes. One group, comprising the majority of currently sequenced cyanomyoviruses, contains 6-12 different tRNA-encoding genes in their genomes. An outlier from this group was cyanophage S-PM2 which encodes 26 tRNA genes (Mann, *et al.*, 2005). Similarly, the majority of cyanophages sequenced in this study contain 6-12 tRNA-encoding genes. In contrast, S-RSM61 possesses 26 putative tRNAs in its genome. It has not yet been experimentally shown that possession of more tRNAs provides a selective advantage for cyanophages in their environment. According to one theory, carrying their own tRNA genes enables cyanophages to maintain a lower GC% content of their genomes, relative to their cyanobacterial hosts (Limor-Waisberg, *et al.*, 2013). A lower %GC content of cyanophage genomes may increase their fitness, since high %AT containing DNA molecules require less energy to separate the DNA strands, thus enabling faster replication and increasing general phage fitness (Miller, *et al.*, 2003b). Furthermore, it is possible that by carrying a large number of tRNA genes, cyanophages increase their host range without altering their %GC content, being capable of infecting both GC-rich *Synechococcus* and AT-rich *Prochlorococcus* (Enav, *et al.*, 2012). It would be interesting to investigate whether a significantly higher number of putative tRNA genes confers any advantage to S-RSM61, including, but not limited to, an increased host range.

Finally, the results of this work point towards the importance of whole genome sequencing of cyanophages when characterising phage populations from environmental samples. Firstly, there are a large number of studies which rely on PCR-based techniques to characterise single cyanophages from environmental samples (Fuller, *et al.*, 1998; Millard and Mann, 2006; Dekel-Bird, *et al.*, 2015). While performing the work described in Chapter 3, the polymerase chain reaction, using degenerate primers targeting the cyanophage *pstS* gene, identified S-BM3 as a cyanophage possessing *pstS* in its genome. Sequencing of the PCR-amplified product showed this *pstS*-containing phage as different from other cyanophages isolated from the same environments – S-BM1/S-BM6. However, whole genome sequencing has now shown that cyanophage S-BM3 – does not contain *pstS*. This strongly suggests that the original phage sample contained more than one cyanophage i.e. was not clonal, with a minor cyanophage component containing *pstS* in its genome (minor since no *pstS* reads were obtained from the 725.72-fold coverage of the S-BM3 genome). Since the propagation of cyanophages on *Synechococcus* sp. WH7803 was performed

under nutrient replete conditions, it is possible that the *pstS*-containing phage was less adapted to nutrient-replete conditions and was outcompeted by S-BM3. It would be interesting to investigate whether phage propagation under different nutrient deplete conditions would produce different phage populations propagated on the same host. This outcome is similar to a recent discovery in which deep sequencing of a seemingly homogenous population of plaque-isolated S-PM2 phage, propagated over a large number of generations on *Synechococcus* sp. WH7803, revealed the low level presence of a novel, previously undetected *Caulobacter* phage. (Rihtman*, et al.*, 2016). Therefore, when sequencing phage isolates from environmental samples, it is not enough to separate between isolates using PCR-based techniques, but efforts should also be made to test phage populations produced via infection of the same host under different nutrient conditions (or even infection of different hosts). Moreover, sequencing to a large depth should be performed in order to discover potential contaminating populations of different phages that could be present in the environmental sample.

4.6 Future directions

A number of possible future experiments arise from the results of the work described in this Chapter:

- Propagation and genome sequencing of the remaining Red Sea cyanophage isolates described in (Millard and Mann, 2006), comparison of their genome content, and identification of possible correlations between the environmental conditions prevalent at the time and depth of isolation and the presence/absence of specific cyanophage AMGs.

- Infection of the same host with the same cyanophage sample under different nutrient deplete conditions and deep sequencing of phage progeny produced under each of the conditions, with the purpose of characterising different co-occurring cyanophage isolates, adapted for different nutrient regimes.

- Host range investigation of cyanophages containing different numbers of tRNA genes, with the purpose of investigating the hypothesis that an increasing number of tRNA genes in cyanophage genomes increases their host range.

Chapter 5:

Characterisation of the activity of the cyanophage MazG orthologue and a *Synechococcus* *mazG⁻* mutant

5.1 Introduction

The bacterial stringent response is an overarching mechanism for dealing with stress in which the affected cell initiates a wide series of cascade pathways in order to, on the one hand, minimise cellular production and thus ease the burden of depleted building blocks and, on the other hand, initiate a genome-wide transcriptional response, via altering the activity of RNA polymerase (Chatterji and Ojha, 2001). This response, if the nutrient stress is not relieved, ends in programmed cell death (Gross, *et al.*, 2006). Due to its severity, the process itself, as well as the signals facilitating it are tightly regulated (Srivatsan and Wang, 2008). One such signal molecule is (p)ppGpp – an alarmone molecule which, together with additional transcriptional activators, directly binds sigma factor-RNA-polymerase complex, altering its affinity for house-keeping gene promoters, and directing transcription to general stress response genes (Srivatsan and Wang, 2008).

Transcriptional regulation and transcript integrity are one of the ways in which bacteriophages ensure successful infection of their bacterial hosts (Zeng and Chisholm, 2012; Stazic, *et al.*, 2016) Furthermore, since the stringent response affects DNA replication (Schreiber, *et al.*, 1995), successful phage replication will be compromised in the case of initiation of the stringent response. Previous work has shown that in case of its onset, phage DNA replication is indeed most significantly affected (Nowicki, *et al.*, 2013b). Therefore, it is expected that over time, phages will evolve a capability to interfere with the bacterial stringent response, mitigating its deleterious effect on phage progeny. Due to the pervasive effect the stringent response has on bacterial cell metabolism, similar to phage interference of the homeostasis of metabolic processes via AMGs, it would be reasonable to expect that bacteriophages would attempt to mitigate the severity and outcome of the stringent stress response. Indeed, previous work on freshwater cyanophages infecting the cyanobacterium *Anacystis nidulans* has shown that intracellular levels of (p)ppGpp are significantly reduced in infected cells that are transferred from light to dark conditions, compared to uninfected control cells (Borbély, *et al.*, 1980). This suggests that bacteriophages are indeed affected by the stringent response of host cells, that this effect is exerted via the alarmone signalling molecule (p)ppGpp and that some phages have the capability to alleviate the restrictions bestowed upon the infected cell through the stringent response potentially via reduction of intracellular (p)ppGpp levels.

Examination of sequenced cyanophage genomes points towards a *mazG* orthologue as the possible AMG candidate that interferes with the host stringent response in the manner described above. *mazG* knockout mutants have confirmed the role of this gene in hydrolysing (p)ppGpp in *E. coli* (Gross, *et al.*, 2006), whilst characterisation of MazG in *Mycobacterium smegmatis* demonstrated its capacity to hydrolyse other nucleotides and deoxyribonucleotides (Lu, *et al.*,

2010). This makes MazG a candidate participant in additional cellular processes other than regulating the stringent response via (p)ppGpp levels, such as removal of aberrant nucleotides which could cause random mutations in the bacterial genome via their incorporation into the nascent DNA strand during replication (Galperin*, et al.*, 2006). Interestingly, a *mazG* orthologue has been found in all sequenced cyanophage genomes and is part of the core cyanophage genome (Sullivan*, et al.*, 2010). The pervasiveness of this gene in cyanophage genomes, as well as its potential effect on the cyanobacterial nutrient stress response, led me to functionally characterise both the marine *Synechococcus* sp. WH7803 host, and cyanophage encoded, *mazG* proteins during the framework of this thesis.

5.2 Goals and Objectives

In order to gain a better understanding of the function of the cyanobacterial MazG protein, here, I created a $mazG^0$ knockout mutant in the marine cyanobacterium *Synechococcus* sp. WH7803 and examined its growth rate under a number of environmentally relevant nutrient stress conditions – P, N and Fe-deplete medium – compared to the WT strain. Additionally, I examined the photosynthetic efficiency of the *mazG* knockout mutant and WT growing under each of the above conditions. Finally, I infected both the *mazG* knockout mutant and WT strain growing under different nutrient stress conditions with cyanophage S-PM2 and assessed the infection kinetics in a one-step growth experiment.

In order to investigate the potential difference in enzymatic activity of the cyanobacterial host MazG and its cyanophage orthologue, I also cloned, over-expressed and purified both *Synechococcus* sp. WH7803 host and cyanophage S-PM2 MazG proteins and tested their enzymatic activity, using different nucleoside triphosphates and deoxynucleotide triphosphates under standard conditions.

The aims of this chapter were thus:

1.   To assess any physiological differences between a *Synechococcus* sp. WH7803 $mazG^0$ mutant and the WT, growing under different, environmentally relevant, nutrient stress conditions.

2.   To compare the infection kinetics of S-PM2 cyanophage infecting both the *Synechococcus* sp. WH7803 $mazG^0$ mutant and WT strain growing under environmentally relevant nutrient stress conditions.

3.   To biochemically compare the enzymatic activities of both cyanobacterial and cyanophage-encoded MazG orthologues, using a variety of deoxynucleotide and nucleotide substrates.

5.3 Materials and Methods

5.3.1 Construction of a cyanobacterial *mazG* knockout mutant

A *Synechococcus* sp. WH7803 *mazG*^{0} mutant was created using the protocol described in (Brahamsha, 1996), amended in the following manner:

A 305bp fragment of the *Synechococcus* sp. WH7803 *mazG* gene, located at coordinates 2280705 - 2281009 was amplified via PCR using primers *WH7803 mazG KO1 F* and *WH7803 mazG KO1 R* outlined in Table 2.4 (see section 2.3). Both primers contain a GAATTC sequence at their 5' end which is a recognition sequence for restriction enzyme *Eco*RI. The fragment was cloned into the pGEM plasmid (Table 2.3), which was used to transform *E. coli* DH5α (Table 2.1). After overnight growth at 37º C, the plasmid was purified using the Qiagen Miniprep kit and sent for Sanger sequencing (GATC) to confirm the sequence of the amplified fragment. The plasmid was then digested using *Eco*RI (Promega) according to the manufacturer's specifications. The resulting *mazG* fragment was separated from the digested plasmid via agarose gel electrophoresis, cut out from the gel and purified using the Promega Gel purification kit. *Eco*RI was also used to digest the suicide plasmid pMUT100 (Table 2.3) and the purified *mazG* fragment was subsequently cloned into pMUT100. The ligated construct was used to transform *E. coli* MC1061 (Table 2.1) which serves as a bi-parental cloning strain for transformation of *Synechococcus* sp. WH7803. 1ml of overnight grown culture (grown in LB with Amp [100 µg/ml], Cm [10 µg/ml) and Kan [50 µg/ml]) of the transformed *E. coli* MC1061 was pelleted via centrifugation and washed twice with 1ml of antibiotic free LB medium. The cells were resuspended in 500 µl ASW -P medium (as described in section 2.1) using 2 mM NH$_4$Cl instead of NaNO$_3$ as a N source (ASW NH$_4^+$ -P). In parallel, 25 ml *Synechococcus* sp. WH7803 growing in ASW NH$_4^+$ -P medium at late exponential phase (OD$_{750}$= 0.35 – 0.37) was pelleted via centrifugation and resuspended in 150 µl ASW NH$_4^+$ -P medium. Varying volumes of the cyanobacterial cells and 50 µl transformed, concentrated *E. coli* MC1061 cells were mixed and spotted onto 0.6% (w/v) ASW NH$_4^+$ -P solid agar, prepared as described in General Materials and Methods chapter 2.2.3. The volumes of cyanobacterial culture were varied to create 1:1, 1:2, and 1:5 ratios of *Synechococcus*: *E. coli*. The spots were left to dry and incubated at 23º C under 10 µmol photons m^{-1} s^{-1} of constant light for 48 hours. The spots were then excised from the

agar, re-suspended in ASW NH_4^+ -P medium and incubated in an orbital shaker at 23º C under 10 µmol photons $m^{-1} s^{-1}$ overnight to re-suspend the cells. The cells were then mixed with 0.3% (w/v) ASW agarose containing 50 µg/ml Kan as the selective agent and incubated at 23º C under 10 µmol photons $m^{-1} s^{-1}$ until colonies appeared (3-8 weeks). *Synechococcus* sp. WH7803 colonies were picked from the agar plates, resuspended in ASW medium with Kan (50 µg/ml) as the selective agent, left to grow in a rotary shaker at 23º C under 10 µmol photons $m^{-1} s^{-1}$ until reaching exponential phase and analysed by PCR for the presence of *mazG* insert using pMut100US F, mazGF and mazGR primers (Table 2.4). Possible conformations of the plasmid insertion and the resulting PCR products can be seen in Figure 5.4. Colonies which confirmed insertion of the plasmid into the cyanobacterial genome in the correct location were kept for further experiments. Since *Synechococcus* sp. WH7803 is an oligoploid organism, estimated to possess 2-4 chromosomes (Griese*, et al.*, 2011), every colony with a confirmed insertion of the *mazG* fragment potentially also possessed an uninterrupted copy of the gene, assumed to be located on a different chromosome. This is the reason why PCR of transformed colonies produced an expected size of the product both when using the plasmid primer combinations (Figure 5.5B, lane 1) and the WT mazGF/mazGR primers (Figure 5.5B, lane 3). Unsegregated mutant strains were grown in the presence of higher concentrations of kanamycin (up to 1mg/ml) in ASW medium, in order to force segregation. Full segregation was periodically examined, using the 7803_MazG_ExpF/7803_MazG_ExpR primers (Table 2.4 in General Methods and Materials Chapter) and experiments on the mutant strain were conducted only when full segregation was confirmed.

5.3.2 Growth and physiological analysis of the *Synechococcus* sp. WH7803 *mazG⁰* mutant

Synechococcus sp. WH7803 *mazG⁰* mutants were inoculated in triplicate into ASW, ASW-P, ASW -N and ASW -Fe media, prepared as described in the General Methods and Materials chapter, section 2.2.2. Cultures were grown using an orbital shaker at 23º C under 10 µmol photons $m^{-1} s^{-1}$ and growth was monitored daily, via OD_{750} measurements, using an Ultrospec 3000 pro (Biochrom, Cambridge, UK). The PSII yield was assessed 8 days after culture inoculation, using a pulse amplitude-modulated flourometer (PhytoPAM, Walz, Effeltrich, Germany), as described in General Methods and Materials chapter, section 2.4.3.

5.3.3 Cyanophage infection of the *Synechococcus* WH7803 mazG⁰ mutant

15 ml of *Synechococcus* sp. WH7803 *mazG* mutant as well as WT *Synechococcus* sp. WH7803, growing in ASW, ASW-P, ASW-N, and ASW-Fe medium (see General Materials and Methods chapter, section 2.2.2) were collected on day 8 and concentrated 20X via centrifugation. 200 µl concentrated *Synechococcus* WT and *mazG* mutant cells were infected with cyanophage S-PM2, in triplicate, following resuspension of cells in appropriate medium (ASW, ASW-P, ASW-N and ASW-Fe) at an MOI of 3 (as previously determined by plaque assays and cell counts, performed using flow cytometry) in triplicate in 96-well plates, as described in General Methods and Materials chapter, section 2.6.2. Infection kinetics were measured via monitoring of optical density (OD_{750}) of the infected cultures every 2 hours for 17 hours, using the iMark Microplate reader (Biorad).

5.3.4 MazG protein activity assays

Synechococcus sp. WH7803 and cyanophage S-PM2 MazG proteins were over-expressed and purified in the following manner:

5.3.4.1 Over-expression of *Synechococcus* sp. WH7803 and cyanophage S-PM2 MazG proteins

Synechococcus sp. WH7803 and cyanophage S-PM2 *mazG* gene were codon-optimized for *E. coli* over-expression using the GeneArt gene synthesis service (Thermofisher Scientific™). The sequence with optimised codons can be seen in Figure 1 in the Appendix. The gene was PCR amplified from the supplied plasmid, using the 7803_MazG_ExpF-7803_MazG_ExpR or SPM2_MazG_ExpF - SPM2_MazG_ExpR primer combination for the cyanobacterial and phage genes, respectively. The PCR was performed using MyTaq™ Mix (Bioline) according to the manufacturer's instructions and the following PCR program:

Initial denaturation step – 95º C for 1 min

30 cycles of the following steps:

      95º C for 15 sec – denaturation

      55º C for 15 sec – annealing

      72º C for 15 sec – extension

Final extension step – 72º C 5 min

Storage at 4º C

Amplified PCR products were cloned into different over-expression plasmids and transformed into a variety of *E. coli* strains which were then tested for over-expression under a variety of conditions, growing on different media. The list of plasmids, *E. coli* strains, growth media and growth conditions tested for over-expression can be seen in Table 5.1. Described below is the experimental approach that resulted in purification of active MazG protein from both host *Synechococcus* sp. WH7803 and cyanophage S-PM2:

Plasmid pET151d-TOPO® containing the amplified PCR product (codon-optimized *mazG* from either *Synechococcus* sp. WH7803 or S-PM2) was created according to the pET151d-TOPO manufacturer instructions (ThermoFisher Scientific) and used to transform *E. coli* BL21 Star™(DE3), as described in General Materials and Methods chapter, section 2.6.2. A map of the pET151d-TOPO plasmid containing the *mazG* gene from *Synechococcus* sp. WH7803 is shown in Figure 5.1. A similar construct containing the *mazG* gene from cyanophage S-PM2 was also created. Transformed *E. coli* was then plated onto an LB agar plate containing ampicillin (50 µg ml^{-1}) and single colonies analysed using colony PCR with primers designed to amplify a fragment contained within the multiple cloning site of the plasmid (Table 2.1). Colonies that contained an insert of the expected size were further analysed by Sanger sequencing (GATC Biotech). Clones containing the correct product were used further for protein over-expression.

Cloned gene	Plasmid	Expression strain	Growth medium	Temperature of induction	Time of induction	Soluble protein
S-PM2 *mazG*		*E. coli* BL21 Star™(DE3)	MagicMedia *E. coli* Expression Medium (ThermoFisher Scientific™)	30º C	o/n	Yes
Synechococcus WH7803 mazG	pET151d-TOPO	*E. coli* BL21 Star™(DE3)	MagicMedia *E. coli* Expression Medium (ThermoFisher Scientific™)	30º C	o/n	No
				37º C	o/n	No
			Lysogenic Broth Medium (LB)	19º C	o/n	No
				21º C	o/n	No
				30º C	o/n	No
				37º C	4 hrs	No
			M9 minimal medium	19º C	o/n	No
				21º C	o/n	No
			2xYT medium	19º C	o/n	No
				21º C	o/n	No
		E. coli Takara4	Lysogenic Broth Medium (LB)	19º C	o/n	No
				21º C	o/n	No
				30º C	o/n	No
				37º C	4 hrs	No
			M9 minimal medium	19º C	o/n	No
				21º C	o/n	No
			2xYT medium	19º C	o/n	No
				21º C	o/n	No
	pET28-b	*E. coli* BL21 Star™(DE3)	MagicMedia *E. coli* Expression Medium (ThermoFisher Scientific™)	30º C	o/n	No
				37º C	o/n	No
			Lysogenic Broth Medium (LB)	19º C	o/n	No
				21º C	o/n	No
				30º C	o/n	No
				37º C	4 hrs	No
			M9 minimal medium	19º C	o/n	No
				21º C	o/n	No
			2xYT medium	19º C	o/n	No
				21º C	o/n	No

			Lysogenic Broth Medium (LB)	19º C	o/n	No
				21º C	o/n	No
				30º C	o/n	No
				37º C	4 hrs	No
	E. coli Takara4		M9 minimal medium	19º C	o/n	No
				21º C	o/n	No
			2xYT medium	19º C	o/n	No
				21º C	o/n	No

Table 5.1–Expression strains, cloning plasmids and expression conditions used in the protein over-expression experiments. Description of *E. coli* strains, plasmids and media can be found in the Methods and Materials Chapter 2. o/n = overnight.

Protein over-expression was performed in the following manner: *E. coli* strain BL21 containing the over-expression plasmid was incubated overnight in an orbital shaker at 37ºC at 300 RPM, in LB with 50 µg/ml ampicillin and 1% (w/v) glucose (used to repress leaky expression of the lac promoter). This starter culture was used to inoculate 1L of *E. coli* MagicMedia Expression Medium (ThermoFisher Scientific™) which was incubated overnight at 30º C in an orbital shaker at 300 RPM. The resulting culture was then harvested via centrifugation, the pellet resuspended in binding buffer (20 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 10% (v/v) glycerol, 10 mM imidazole) and lysed using a French Press (Aminco). Lysate was separated into soluble and insoluble fractions via centrifugation at 13,000 x g and assessed for the presence of the over-expressed protein via SDS-PAGE based on polypeptide size. In the case of cyanophage S-PM2 MazG, the over-expressed protein was soluble, and was purified using the soluble protein purification protocol outlined below, while the *Synechococcus* sp. WH7803 MazG protein was found in the insoluble phase and was purified using the inclusion body purification protocol outlined in section 5.3.4.3.

Figure 5.1 –Map of pET151/D-TOPO® plasmid containing the *mazG* gene from *Synechococcus* sp. WH7803
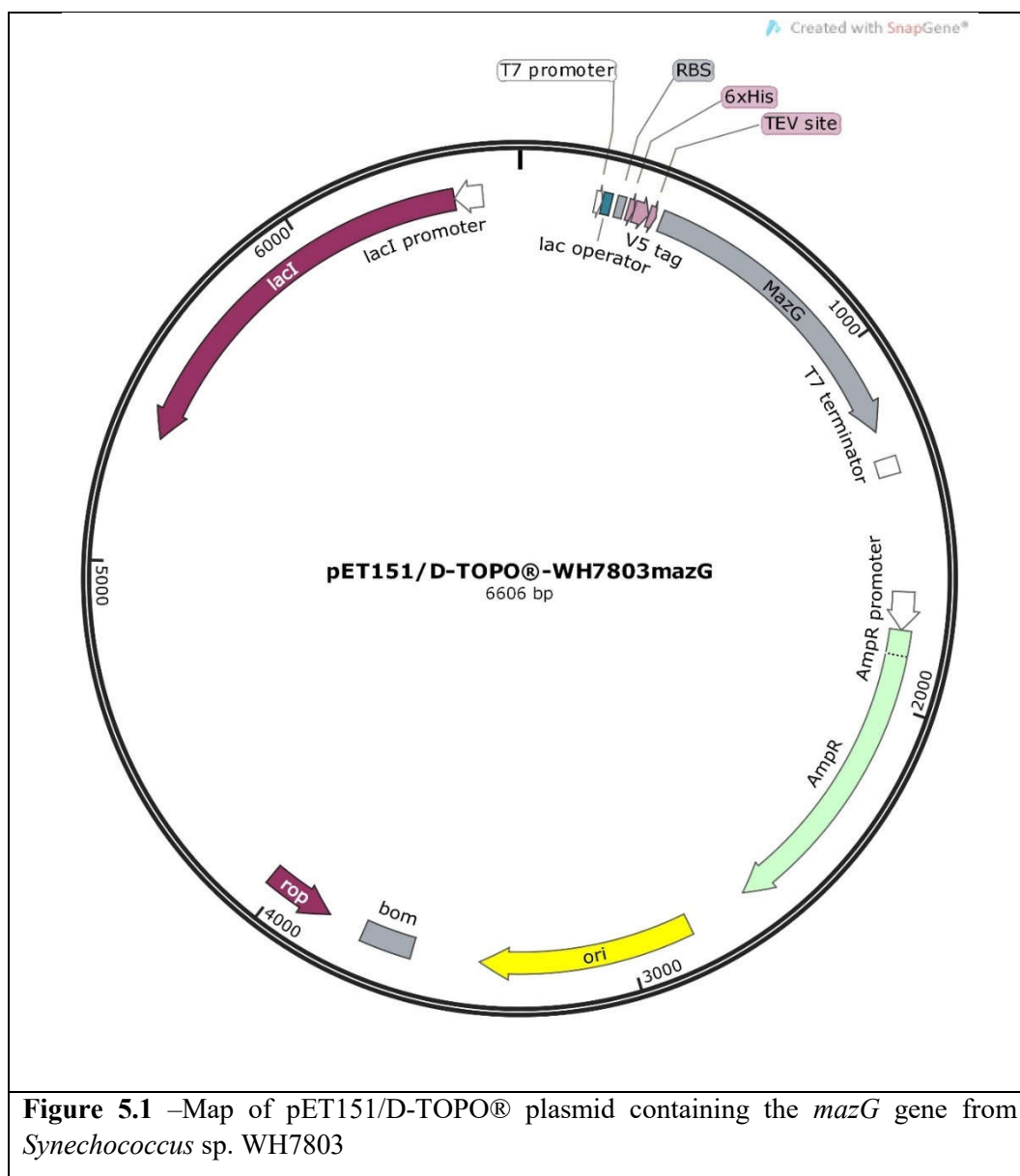
5.3.4.2 Purification of the soluble cyanophage S-PM2 MazG protein

Over-expressed cyanophage S-PM2 MazG protein extract was resuspended in 5 ml of binding buffer (20 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 10% (v/v) glycerol, 10 mM imidazole) and applied to a HisTrap HP Ni^{++} affinity column (GE Healthcare) and washed with 5 volumes of binding buffer (25 ml). The protein was then eluted with stepwise-increasing concentrations of imidazole in binding buffer. Concentrations of imidazole ranged from 30-300 mM, as shown in Figure 5.10. Finally, the column was stripped with stripping buffer (500 mM NaCl, 50 mM EDTA, pH 8.0). An aliquot of each fraction was used for verification of the presence of the expected protein, using SDS-PAGE protocol as described in 5.3.4.4.

5.3.4.3 Purification of the insoluble *Synechococcus* sp. WH7803 MazG protein from inclusion bodies

After separating from the soluble phase via centrifugation at 13,000xg, the insoluble phase was resuspended in 10 ml/g 20 mM Tris-HCl, pH 8; 150 mM NaCl, 5 mM MgCl$_2$ and treated with 100 µg ml^{-1} DNAseI (New England Biolabs) for 15 min at 37° C. Cellular debris was pelleted at 20,000xg, for 15 min at 4°C. The pellet was denatured in 50 ml/g Buffer A (6 M guanidine hydrochloride, 20 mM Tris-HCl, pH 8; 500 mM NaCl, 10 mM imidazole). Protein was loaded onto a HisTrap HP (GE Healthcare) Ni^{++} affinity column and washed with 3 column volumes (CV) of Buffer A. *Synechococcus* sp. WH7803His$_6$-MazG was refolded in Buffer B (as Buffer A, without guanidine hydrochloride) over 10 CV and finally gradually eluted in Buffer C (as Buffer B, with 500 mM imidazole) over 5 CV. Fractions containing His$_6$-MazG were pooled and dialysed against 20 mM Tris-HCl, pH 8; 100 mM NaCl, 20% (v/v) glycerol o/n at 4°C. The protein was concentrated to ~1 µg/µl using an Amicon ultra centrifugal filter (Amicon, MWCO 3,000) and stored at -20°C.

5.3.4.4 SDS-PAGE

Protein sample (either whole cell extracts or soluble/insoluble fraction described in 9.1) was added to 3X Sample Buffer (30% v/v Glycerol, 6% w/v Sodium Dodecyl sulphate, 0.15M Tris-HCl pH 6.8, 1.5M Dithiothreitol, 2% w/v Bromophenol Blue) at 2:1 (protein sample:sample buffer) ratio and boiled at 100° C for 5 minutes. It was then loaded on to 10% acrylamide gel and ran at 200 mA in 10% Bis-Tris acrylamide protein gels (NuPAGE Novex) for 120 min in 1X running buffer (NuPAGE® MOPS SDS Running Buffer at room temperature. Protein gel was then removed and stained using either SimplyBlue™ SafeStain (Invitrogen) or by Silver stain as described in (Maniatis*, et al.*, 1989).

5.3.4.5 Mass spectrometry analysis

Since purification of the cyanophage S-PM2 MazG protein produced two visible bands (Figure 5.10) as assessed by SDS-PAGE, a larger, fainter band was excised from the gel and sent for analysis at the Proteomics Unit of the School of Life Sciences, University of Warwick. Mass spectrometry of selected peptides was performed on a Thermo Orbitrap Fusion mass spectrometer (Q-OT-qIT, Thermo Scientific). The larger size band was also confirmed to be the cyanophage S-PM2 MazG protein, probably a product of dimerization of a single polypeptide.

5.3.5 MazG activity assays

Activity assay of the purified *Synechococcus* host and cyanophage MazG proteins was performed as described in (Lu*, et al.*, 2010). Briefly, purified protein (1 µg) in Binding Buffer (20 mM Tris-HCl, pH 8; 100 mM NaCl, 20% (v/v) glycerol) was incubated in triplicate with increasing concentrations of different substrate nucleotides at 37ºC for 10 minutes, followed by an inactivation step at 70ºC for 10 minutes. The potential nucleotide hydrolysis reaction was then assessed for the presence of released inorganic phosphate, using the PiPER pyrophosphate assay kit (ThermoFisher Scientific), as per the manufacturer's specifications. In the assay, the presence of inorganic phosphate is quantified via measurement of fluorescent by-products of incorporation of phosphate into glucose, as described in Figure 5.2. Fluorometric measurements were performed using a ClarioStar plate reader (BMG-Labtech) using an excitation wavelength of 550nm and an emission wavelength of 590nm. The amount of available phosphate resulting from cleavage of pyrophosphates from substrate nucleotides was estimated based on the standard curve produced using 0 µM – 100 µM of inorganic phosphate supplied with the PiPER kit. Enzymatic kinetic parameters were calculated in the following manner:

$$V_{max} = P_{max} \quad C^{-1} \quad T^{-1}$$

where P_{max} represents the maximum amount of free phosphate over the experimental range of substrates, produced by the enzymatic activity, C – amount of protein used in each reaction (1 µg) and T – time of the enzymatic reaction. (10 min).
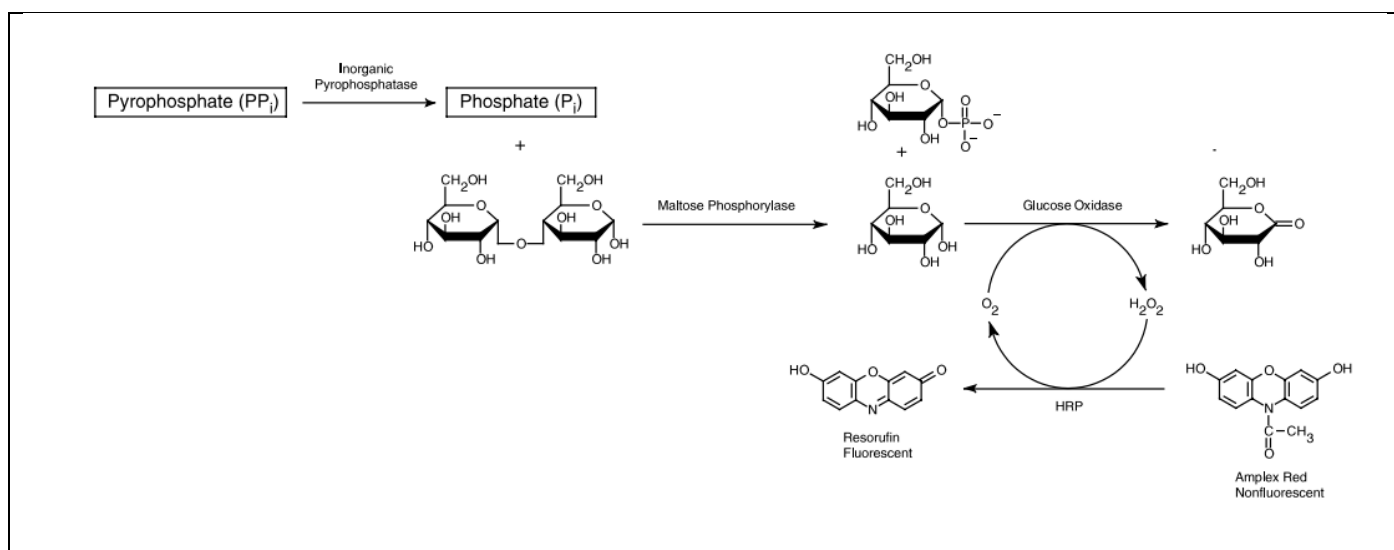
Figure 5.2: The outline of reactions performed in the framework of PiPer pyrophosphatase activity assay. Phosphate produced via activity of pyrophosphatase is used by maltose phosphorylase to convert maltose to glucose-1-phosphate and glucose. Glucose is converted by glucose oxidase to gluconolactone and H_2O_2. Hydrogen peroxide is used by Horseradish Peroxidase (HRP) to convert nonfluorescent Amplex Red to fluorescent Resorufin. The produced fluorescence is proportional to the initial amount of pyrophosphate present in the reaction

K_m (the Michaelis constant) was calculated as the substrate concentration for which ½ V_{max} was measured and K_{cat} was calculated as

$$K_{cat} = \frac{V_{max}}{E}$$

where E is the total concentration of the enzyme in mM.

5.3.6 Phylogenetic analysis

Phylogenetic analysis of marine *Synechococcus* and *Prochlorococcus* MazG, selected freshwater cyanobacteria MazG and cyanophage MazG proteins was performed using the maximum likelihood method based on the Whelan and Goldman model (Whelan and Goldman, 2001). The tree with the highest log likelihood (-26719.1598) score is shown. The percentage of trees in which the associated taxa clustered together is shown above the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.0667)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 162 amino acid sequences aligned with MUSCLE (Edgar, 2004). The cyanobacterial and cyanophage MazG sequences were extracted from the NCBI nr database, Cyanorak (http://abims.sb-roscoff.fr/cyanorak/) and Cyanobase (http://genome.microbedb.jp/cyanobase/), using BLAST. S-BM1, S-BM6, S-BM4, S-WHM1 and S-RSM70 MazG sequences were produced in this study. The maximum-likelihood tree was built in MEGA (Tamura*, et al.*, 2013), using 100 bootstrap replications to determine branch lengths and support values.

5.4 Results

5.4.1 Phylogenetic relationships amongst different cyanobacterial and cyanophage MazG orthologues

In order to examine the evolutionary relationships between the MazG orthologues found in cyanobacteria and cyanophages, a phylogenetic analysis was carried out (see section 5.3.4.6). A maximum-likelihood tree was created using amino-acid sequences of all the cyanobacterial sequenced genome submitted to Cyanobase (http://genome.microbedb.jp/cyanobase/) and Cyanorak (http://abims.sb-roscoff.fr/cyanorak/) Additionally, MazG sequences from cyanophage genomes submitted to NCBI, as well as orthologues from cyanophage genomes sequenced during the course of this work, were included to examine the phylogenetic relationships and potential common ancestry of the *mazG* gene. The tree was rooted using the MazG sequence from *Escherichia coli* K12 substr. MG1655. The resulting phylogenetic tree is shown in Figure 5.3. From the phylogenetic tree it is clear that there is a clear evolutionary separation between the viral and the host *mazG* orthologues. Unlike the *pstS* phylogeny (Figure 3.5) where the phage orthologues clustered with the host family that they are capable of infecting, here the different host families and genera are clustering completely separately from the phage orthologs. Additional notable feature of the phylogenetic tree is that there is no distinct clustering between the *Prochlorococcus*- and *Synechococcus*-infecting phages. This could hint towards exchange between these two groups of phages, possibly via a common host.

Cyanophage (*Synechococcus* & *Prochlorococcus*)

Prochlorococcus

HL

LL

Synechococcus

Prochlorococcus LL

Synechococcus

Cyanobium

Synechococcus

Freshwater cyanobacteria
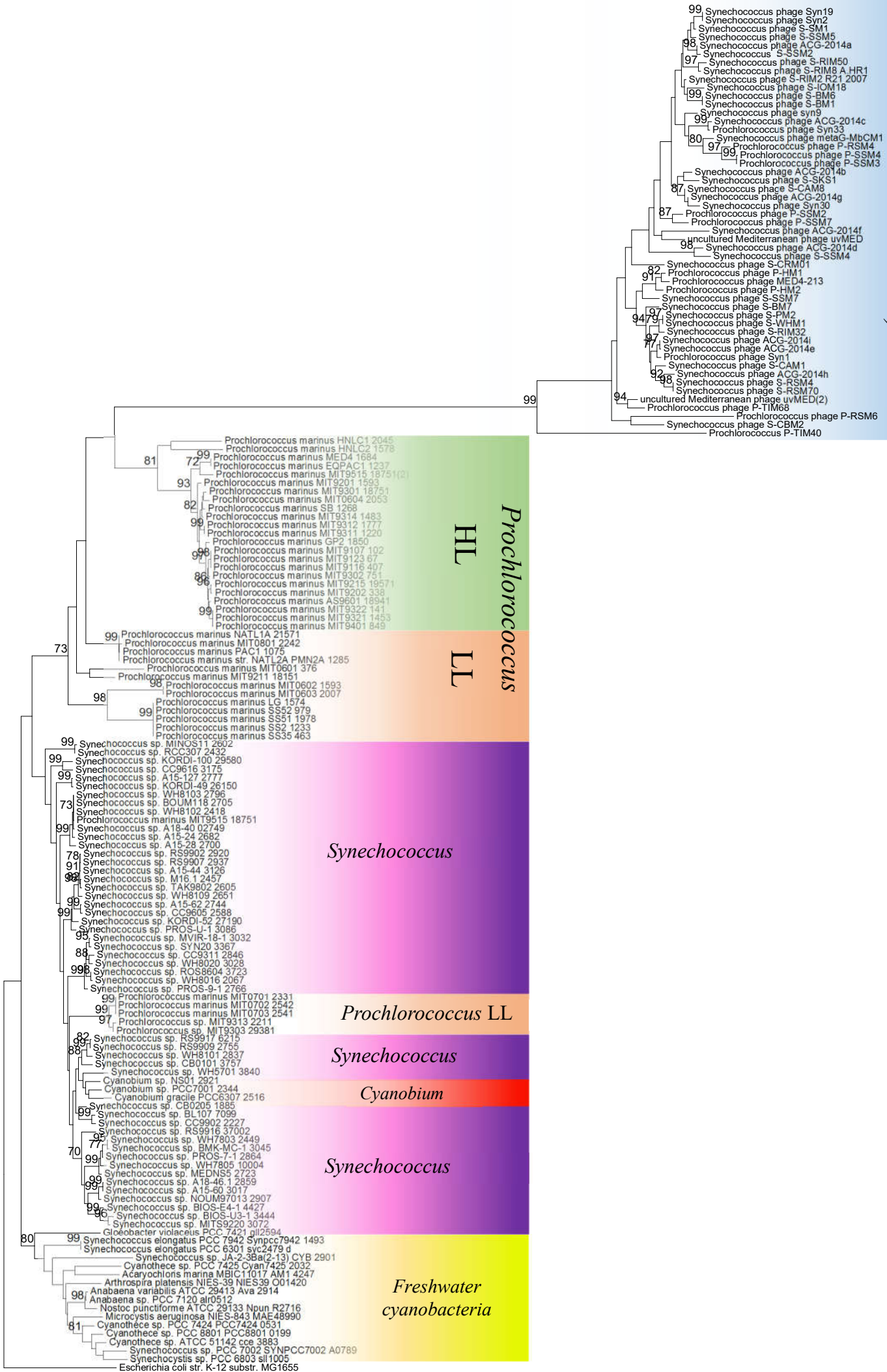
Figure 5.3 Phylogenetic relationships amongst cyanobacteria and cyanophage MazG proteins

5.4.2 Growth and physiological analysis of the *Synechococcus* WH7803 *mazG⁰* mutant

The *Synechococcus* WH7803 $mazG^0$ mutant strain was constructed as described in section 5.3.1 and complete segregation was monitored via PCR using primers mazGF and mazGR (see Table 2.4). Segregation was shown to be complete when no PCR product was obtained using the MazGF–MazGR primer pair (see Figure 5.4).

Figure 5.4 – A: Schematic representation of the *mazG* gene with the location of mazGF and mazGR primers and the amplified region (marked in red) serving as a target for plasmid insertion (1). Since the genome insertion is a product of a single cross-over event, plasmid can be inserted in either of the two conformation, resulting in a duplication of the inserted region (red) flanking a plasmid sequence (green). Two possible conformations with the orientations and positions of relevant primers can be seen in schemes 2 and 3.

B: Agarose gel electrophoresis of the PCR products obtained using different combinations of pMUT100USF primer and mazGF/mazGR primers. 1 – pMUT100USF + mazGF; 2 – pMUT100USF + mazGR; 3 – mazGF + mazGR; G – negative control using mazGF + mazGR and purified WT *Synechococcus* sp. WH7803 genomic DNA as template.

In order to examine the physiological response of the *Synechococcus* sp. WH7803 $mazG^0$ mutant to different nutrient stress conditions both the mutant and the WT were grown in ASW medium lacking different essential nutrients (see section 5.3.2) and growth assessed via OD_{750} measurements on a daily basis. The resulting growth curves are shown in Figure 5.5.
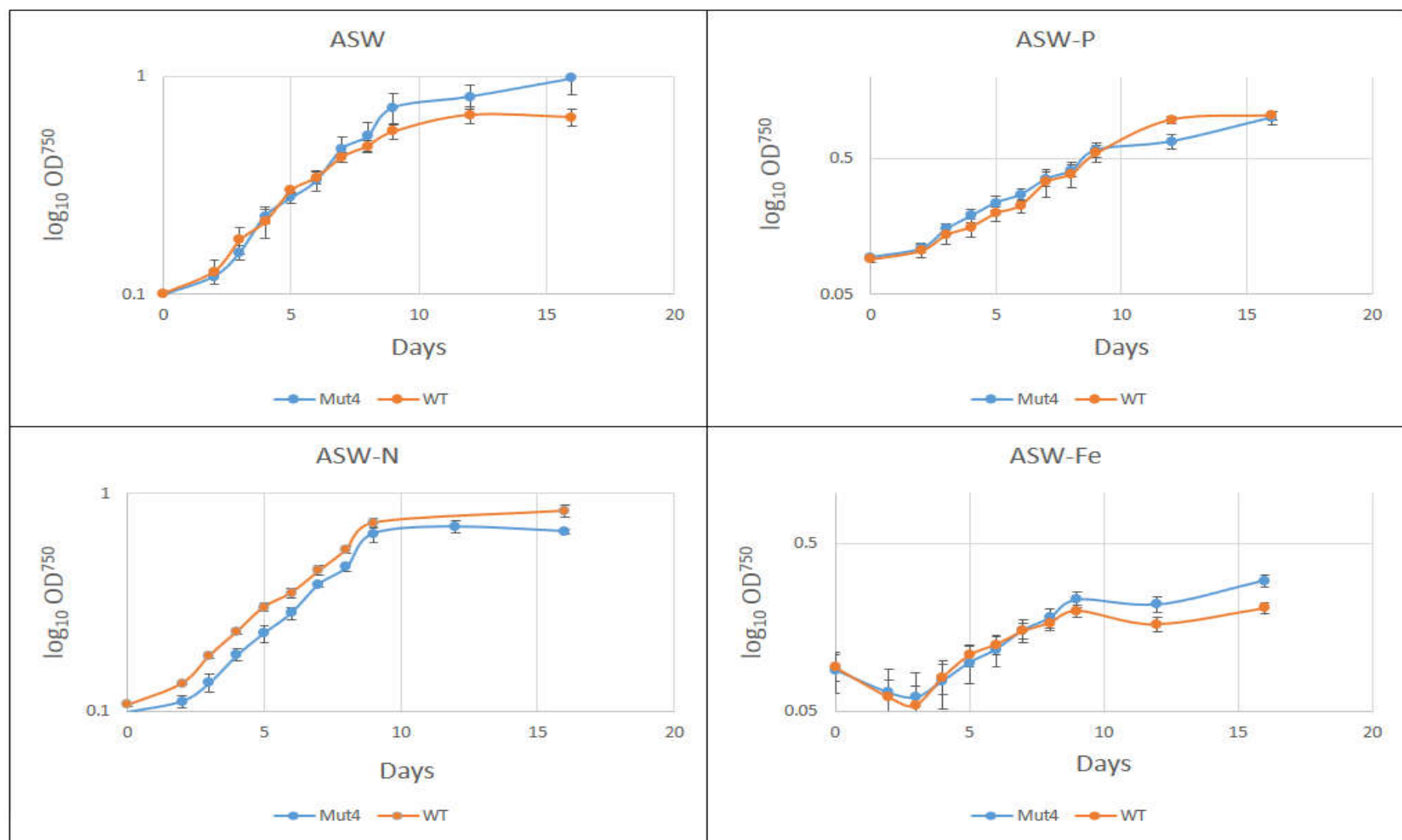
Figure 5.5 – Growth (semi-log plots) of the *Synechococcus* sp. WH7803 wild type and *Synechococcus* sp. WH7803 *mazG⁰* mutant *(Mut4)* under different nutrient stress conditions. Error bars represent the standard error of three biological replicates.

Additionally, specific growth rates for each of the strains was calculated for each of the 3 biological replicates for every conditions, based on the linear part of the growth curve (T$_3$ – T$_9$). The specific growth rate μ was calculated according to the following equation:

$$\mu = \frac{\ln \left(OD_{750\,[T9]} \big/ OD_{750\,[T3]} \right)}{\Delta T}$$

The resulting growth rates are shown in Figure 5.5 and Table 5.2.

Growth medium (ΔOD$_{750}$/day)	*Synechococcus* sp. WH7803	*Synechococcus* sp. WH7803 *mazG^0 (Mut4)*
ASW	0.1899 (+/- 0.0011)	0.2042 (+/-0.0125)
ASW-P	0.1715 (+/- 0.0100)	0.1646 (+/-0.0132)
ASW-N	0.1871 (+/- 0.0055)	0.2047 (+/-0.0086)
ASW-Fe	0.1882 (+/- 0.0095)	0.1626 (+/-0.0125)

Table 5.2 – Specific growth rates (day^{-1}) of the *Synechococcus* sp. WH7803 wild type and *Synechococcus* sp. WH7803 *mazG^0* mutant *(Mut4)* growing under different nutrient stress conditions.

Figure 5.6: Specific growth rates (day^{-1}) of the *Synechococcus* sp. WH7803 WT and *Synechococcus* sp. WH7803 *mazG^0* mutant *(Mut4)* growing under different nutrient stress conditions.
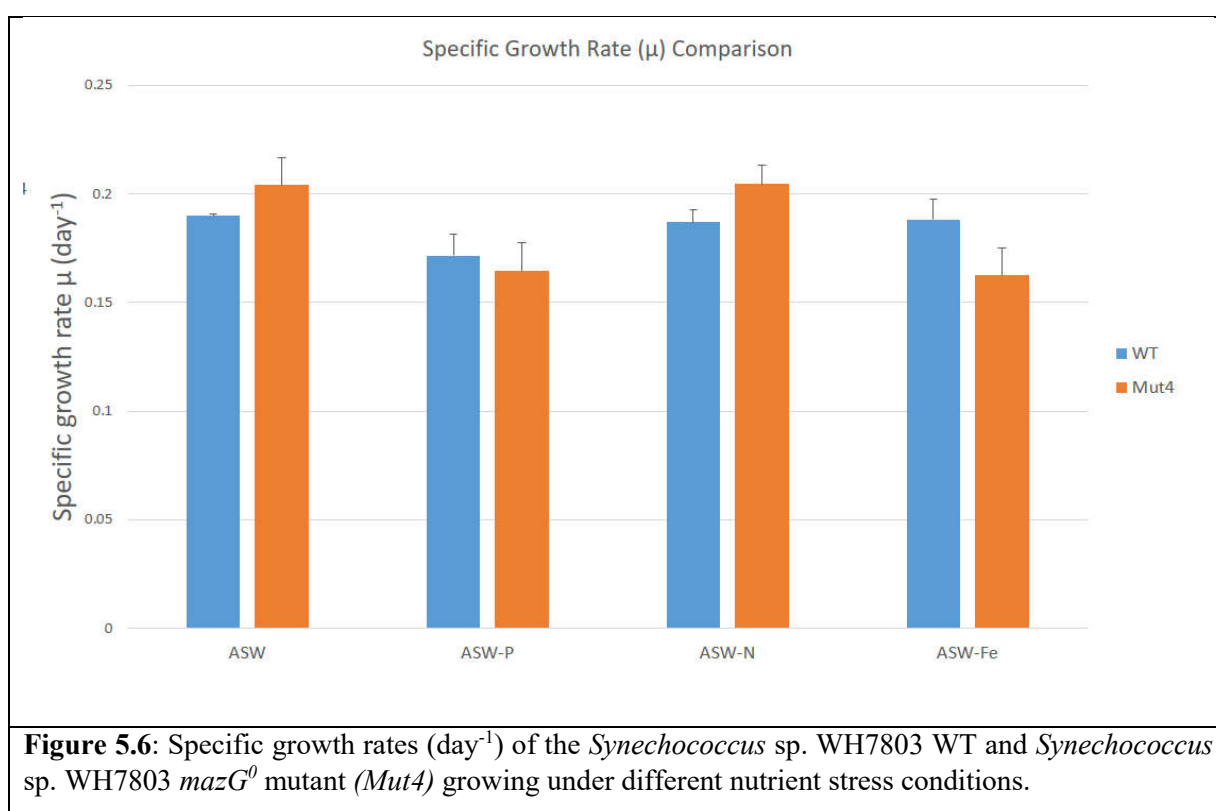
These results show that there is no significant growth rate difference between the WT and the mutant under any nutrient regime. Based on these results there is insufficient information to conclude that

mazG plays a role in different nutrient stress responses, since, hypothetically, there could be alternative mechanisms by which alarmone molecules are being removed from the system and MazG could serve only as a backup removal mechanism.

As well as growth rate assessment, photophysiology of the wild type *Synechococcus* and *mazG* mutant strain was also determined using F_v/F_m measurements (see General Materials and Methods chapter, section 2.4.3) as a proxy for photosystem II efficiency, under the different nutrient stress conditions described above. Cultures were sampled on day 8 of the experiment, representing the late exponential phase of growth. Results (Figure 5.7) showed that there is a significant difference in photosystem II efficiency in the mutant strain under P-deplete conditions (t = 5.0958, p-value = 8.04×10^{-3}) compared to the WT strain, the latter having a higher efficiency than the mutant. For the remainder of the conditions examined, there was no significant difference in photosystem II efficiency between the WT and the mutant strain.
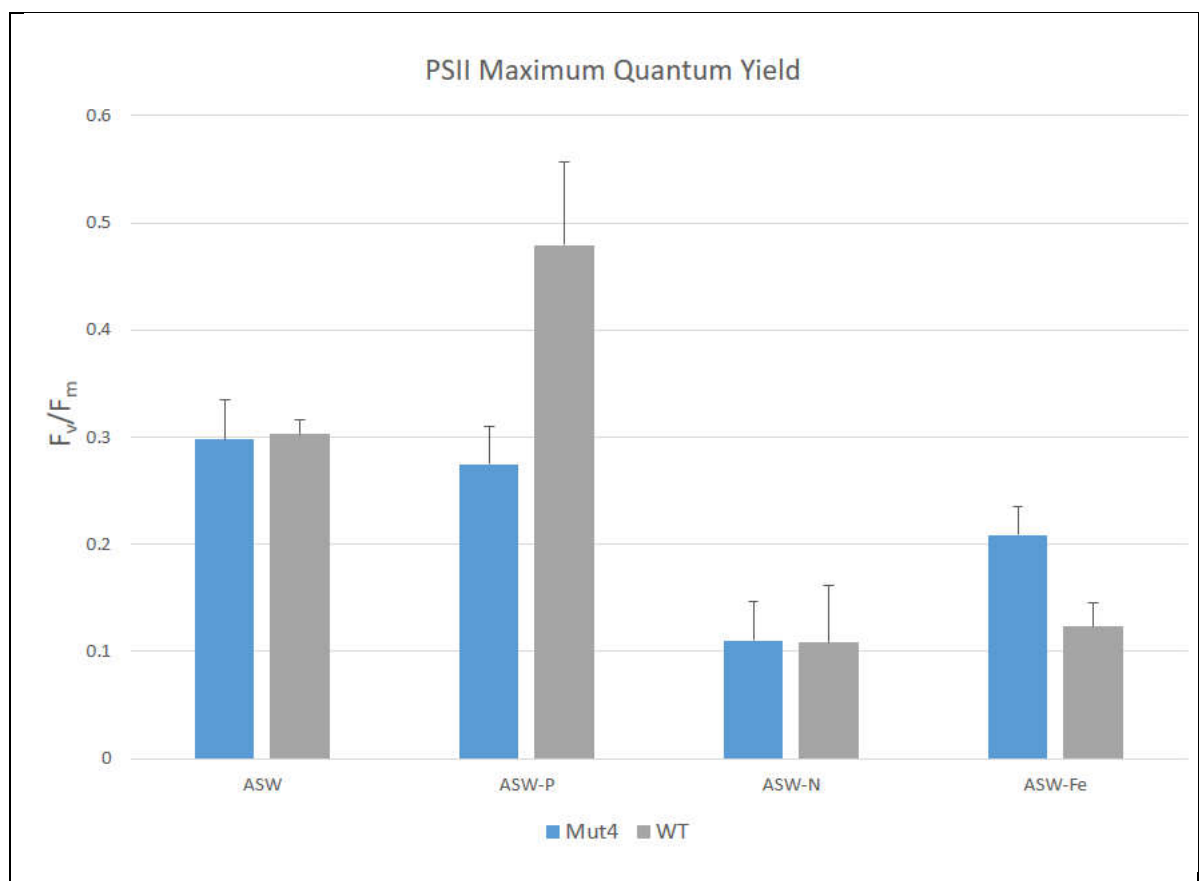
Figure 5.7: Comparison of photosystem II quantum yield of *Synechococcus* WH7803 and *Synechococcus* WH7803 *mazG[0]* *(Mut4)* growing under different nutrient stress conditions

5.4.3 Cyanophage S-PM2 infection dynamics of the *Synechococcus* sp. WH7803 wild type and *Synechococcus* sp. WH7803 *mazG[0]* mutant (Mut4)

In order to examine cyanophage infection kinetics in the *Synechococcus* sp. WH7803 *mazG* mutant, a one-step infection experiment was performed in which both the WT and *mazG* mutant strain were

infected by cyanophage S-PM2 in triplicate and the infection kinetics assessed via measurement of OD_{750}, as described in section 5.2.3 of this chapter. Since different cultures were at different cell densities at the time of infection (due to different yields stemming from the different nutrient stress conditions) and for the sake of comparison between different infection conditions, the infection kinetics are represented as a percentage of OD_{750} at T_0 thus representing a relative change in OD_{750} for each culture per each of the biological replicates. Infection kinetics (Figure 5.8) show that in the case of infection of cultures growing in ASW and ASW-P media, the infection of the mutant is slower and seemingly less productive, judging by the smaller degree of the culture lysis happening in mutant cultures infected by S-PM2. In the case of infection under -P conditions, the onset of lysis of the mutant culture also seems to be delayed, compared to the ASW-grown infected culture. No difference between the infection kinetics of the WT and the mutant growing in ASW-N medium can be observed. Interestingly no infection seems to be occurring in Fe-deplete conditions, neither for the WT nor for the Δ*mazG* mutant strain.

5.4.4 MazG over-expression and comparison of the enzyme activity of the *Synechococcus* sp. WH7803 MazG protein with the cyanophage S-PM2 orthologue

In order to compare differences in activity between the *Synechococcus* sp. WH7803 MazG and the cyanophage S-PM2 orthologue, genes encoding both proteins were cloned into *E. coli* , over-expressed and purified, as described in section 5.3.4. A number of over-expression conditions including the use of different plasmids, *E. coli* expression strains and induction conditions were used, as outlined in Table 5.1. SDS-PAGE analysis revealed the over-expressed *Synechococcus* host and cyanophage MazG proteins in either the soluble or insoluble fractions (see Figure 5.9).

Purification of over-expressed proteins resulted in elution of two bands (Figure 5.10, lane E4, marked by arrows). Both bands were excised from the SDS-PAGE gel and were analysed via mass spectrometry as described in section 5.3.4.4 of this chapter. Mass spectrometry analysis showed that both bands contain the over-expressed MazG protein and are probably a result of dimerization of the over-expressed protein during purification.

Following purification and estimation of protein concentration, enzyme activity assays were conducted (see section 5.3.5). Michaelis-Menten activity plots are shown in Figure 5.11.

Figure 5.8: Infection dynamics of the *Synechococcus* sp. WH7803 wild type and *Synechococcus* sp. WH7803 mazG^0 mutant (Mut4) by cyanophage S-PM2.
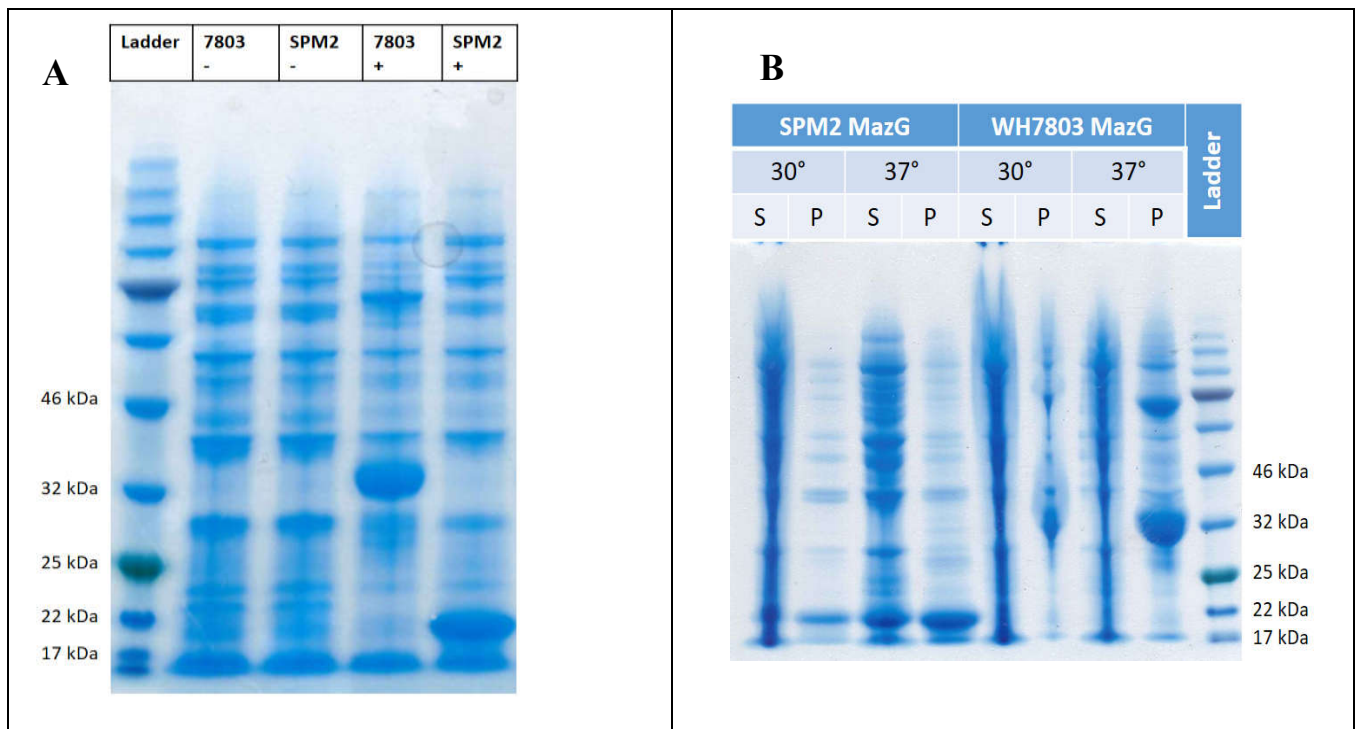
Figure 5.9 - SDS-PAGE analysis showing over-expression of the MazG protein from *Synechococcus* sp. WH7803 (expected size 35.1 kDa) and S-PM2 (expected size 19.5 kDa). A –IPTG-induced (+) and uninduced (-) whole cell protein extracts. B – separation of whole cell protein extracts into soluble (S) and insoluble (P) fractions, induced in MagicMedia *E. coli* Expression Medium at 30º C or 37º C.
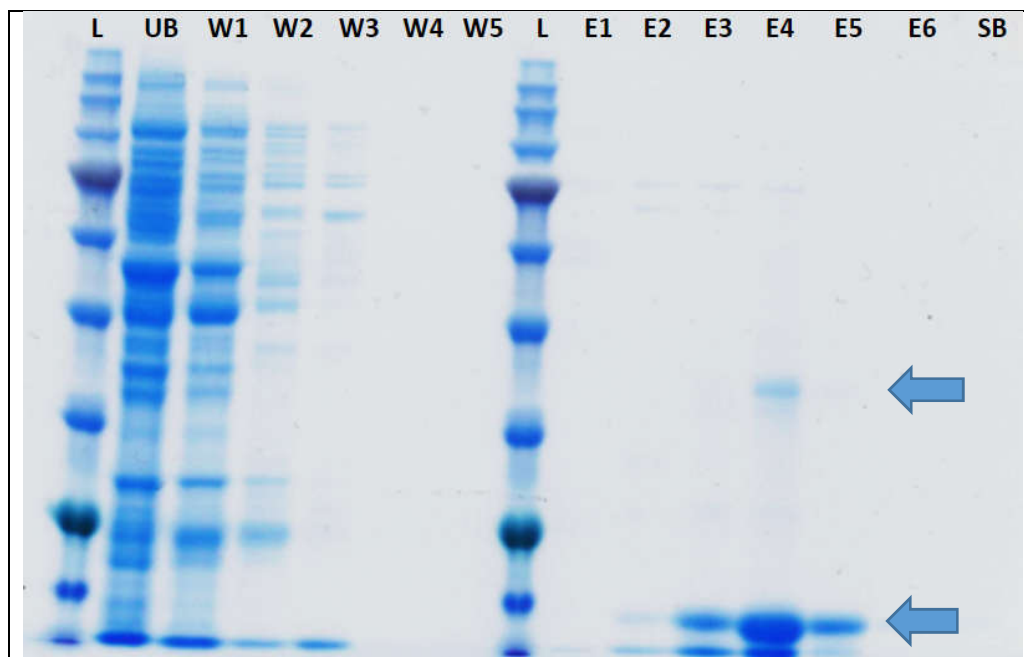
Figure 5.10: Coomassie-stained SDS-PAGE gel showing the products of protein purification of the cyanophage S-PM2 MazG protein. L – protein ladder; UB – unbound fraction (proteins that did not bind to the column). W1-W5 – washes of the column with binding buffer; E1-E6 – elution fractions containing increasing amounts of imidazole (30mM, 50mM, 100mM, 150mM, 200mM, 300mM accordingly); SB – stripping buffer. Arrows show the location of the two bands corresponding to the over-expressed MazG protein

Figure 5.11– Michaelis-Menten curves (top) and Lineweaver-Burk plots (bottom) of enzymatic activity of MazG orthologues from *Synechococcus* sp. WH7803 and cyanophage S-PM2 incubated with increasing concentrations of dATP and dGTP
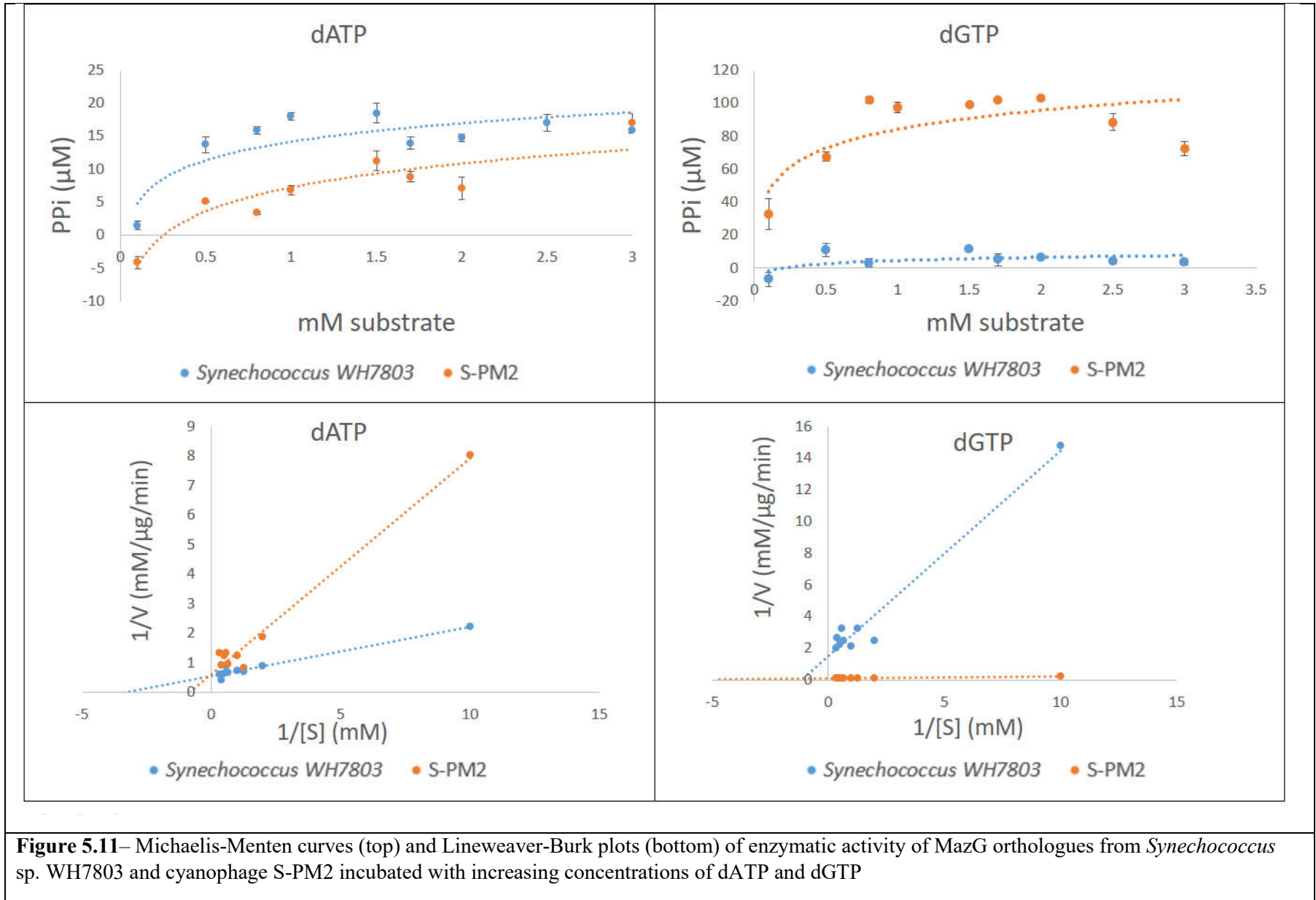
The results of enzyme assays were used to calculate kinetic parameters (V_{max}, K_m and K_{cat}) of the activity of each of the MazG orthologues (Table 5.3). This showed that the highest activity was observed with the S-PM2 protein, using dGTP as a substrate, while the lowest activity was observed in the case of the S-PM2 protein using GTP as a substrate. Since the enzyme activity of these *Synechococcus* or cyanophage MazG orthologues is the first obtained for cyanobacteria or their phages, there is no direct comparable data. However, the values below (Table 5.3) are in the same order of magnitude as those previously published values obtained for *E. coli* and *Mycobacterium tuberculosis* MazG activity (see Lu, *et al.*, 2010).

	V_{max} (nmol/µg/min)		K_m (mM)		K_{cat} (min^{-1})	
	Synechococcus sp. WH7803	S-PM2	*Synechococcus* sp. WH7803	S-PM2	*Synechococcus* sp. WH7803	S-PM2
dATP	1.8 (+/-0.28)	1.62 (+/-0.19)	0.3 (+/-0.09)	1.2 (+/-0.21)	126.12 (+/-19.35)	62.97 (+/-7.44)
dCTP	3.81 (+/-0.36)	8.86 (+/-0.2)	0.14 (+/-0.03)	1.16 (+/-0.04)	267.68 (+/-25.02)	344.68 (+/-7.72)
dTTP	6.57 (+/-0.19)	5.68 (+/-0.2)	N/D	1.23 (+/-0.06)	461.04 (+/-13.43)	221.00 (+/-7.78)
dGTP	0.64 (+/-0.25)	10.29 (+/-0.25)	0.85 (+/-0.07)	0.14 (+/-0.01)	45.16 (+/-17.6)	400.35 (+/-9.91)
ATP	2.55 (+/-0.35)	2.28 (+/-0.24)	0.63 (+/-0.23)	1.43 (+/-0.36)	179.27 (+/-24.4)	88.7 (+/-9.41)
CTP	1.96 (+/-0.14)	2.51 (+/-0.17)	1.2 (+/-0.21)	0.85 (+/-0.11)	137.81 (+/-9.81)	97.48 (+/-6.68)
GTP	0.7 (+/-0.13)	0.3 (+/-0.02)	0.26 (+/-0.02)	N/D	49.46 (+/-9.19)	11.67 (+/-0.6)
UTP	3.02 (+/-0.2)	3.31 (+/-0.18)	1.33 (+/-0.3)	0.6 (+/-0.37)	221.07 (+/-6.00)	128.75 (+/-7.12)

Table 5.3 – Kinetic parameters of enzymatic activity of the *Synechococcus* sp. WH7803 and cyanophage S-PM2 MazG proteins. The values in brackets represent standard error based on three replicates. N/D – not detected.

For the sake of comparison between the V_{max} of the host and cyanophage MazG proteins using different substrates, the highest V_{max} value in all experiments (in this case the V_{max} of the S-PM2 MazG with dGTP as a substrate) was represented as 100% activity and the rest of the V_{max} values were compared to this (see Figure 5.12).

Figure 5.12 – Relative maximal activity (V_{max}) of the *Synechococcus* sp. WH7803 and S-PM2 MazG proteins normalised to the activity of the S-PM2 MazG using dGTP as a substrate. Error bars represent standard error based on three replicate experiments.
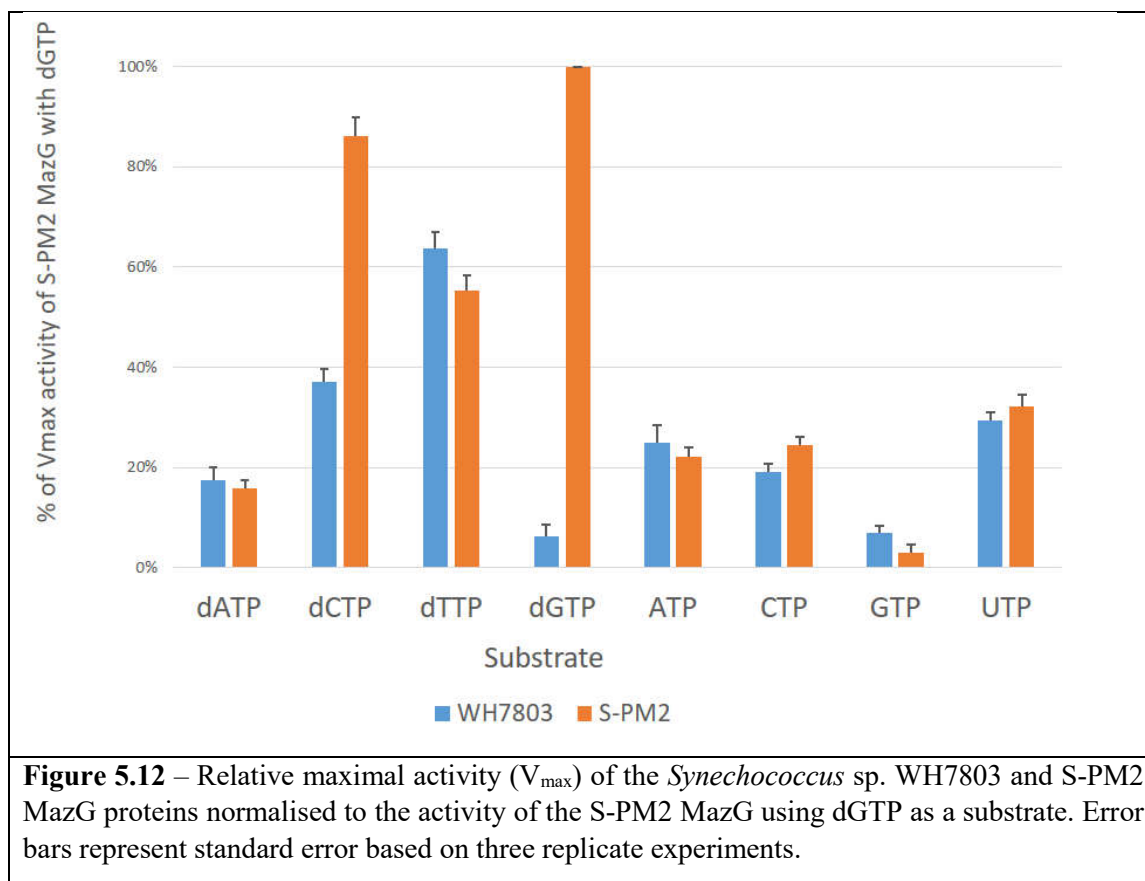
5.5 Discussion

5.5.1 Phylogenetic analysis of *Synechococcus* host and cyanophage MazG proteins

Phylogenetic analysis of MazG proteins (Fig. 5.4) showed the obvious separation of cyanophage sequences from all of the *Synechococcus* host orthologues. This is unlike the situation for host and cyanophage PstS orthologues (Fig 3.5), where the cyanophage proteins cluster together with their specific hosts. In contrast, for MazG, there is a large phylogenetic distance between the host and cyanophage proteins, pointing towards potential divergence, stemming either from divergence in function or lack of common ancestry (Figure 5.4).

Figure 5.13: Boxplot charts comparing the average molecular weight and size of bacterial and cyanophage MazG proteins, based on 110 cyanobacterial and 51 cyanophage MazG sequences used for generating the phylogenetic tree in Figure 5.2.
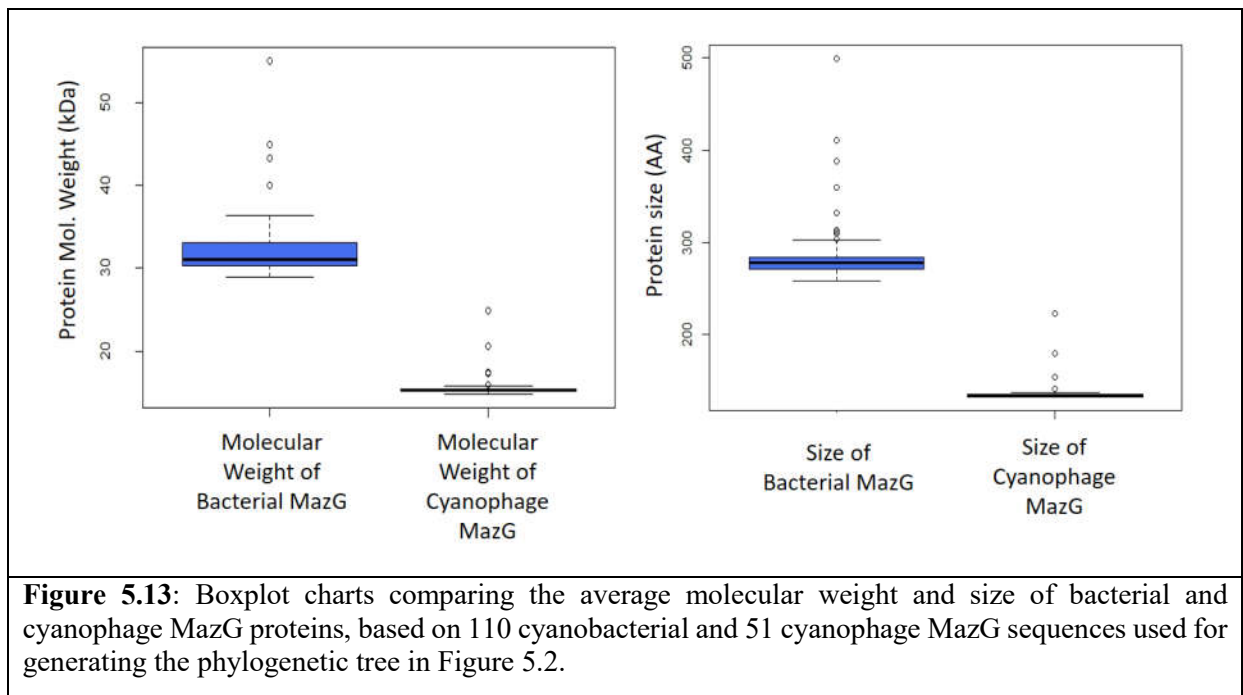
Characteristic differences between host genes and their bacteriophage orthologues is well documented (Breitbart, *et al.*, 2007; Hurwitz and U'Ren, 2016). This is particularly true when considering the size of phage auxiliary metabolic genes (AMGs) which are generally smaller than their host counterpart (Sullivan, *et al.*, 2010). Possible reasons for this size difference could stem from i) size constraints – smaller phage genomes are not able to contain the minimal number of required genes if the genes are of the same size as their host orthologues ii) a strategy to avoid regulatory constraints that are imposed on host orthologues of metabolic genes, via removal of protein domains on which regulatory modifications are exerted (Stazic, *et al.*, 2016).

Annotation of the cyanophage S-PM2 gene S-PM2d136 as *mazG*, similar to other cyanophage orthologues of *mazG*, has been exclusively made based on similarity via automatic annotation methods. While there is a high degree of similarity between the viral orthologues, the amino acid similarity between phage and cyanobacterial proteins is quite low. For example, a BLASTp alignment between the MazG protein from *Synechococcus* sp. WH7803 and cyanophage S-PM2 showed only 37% identity (E value 4e-11). This difference in predicted protein size and low identity levels likely explains the large phylogenetic distance between the cyanophage MazG orthologues and their cyanobacterial hosts.

As mentioned above, the phylogenetic analysis (Fig. 5.4) showed that whilst sequences from the same host genera clustered together, hinting at evolutionary pressure leading to diversification within the marine cyanobacteria, phage orthologues do not exhibit such clustering. This contrasts with cyanophage PstS orthologues, which show a clear

clustering between *Synechococcus* and *Prochlorococcus*-infecting phages (Figure 3.5). Additionally, phylogenetic trees based on another cyanophage core gene, *phoH*, show similar clustering based on the identity of the infected host from which the phage was isolated (Goldsmith*, et al.*, 2011; Rihtman*, et al.*, 2016). Distinct clustering correlating with the infected host genera was also observed in the case of the phage photosynthetic gene *psbA* (Chenard and Suttle, 2008*)*. The implications of a lack of genus-correlated clustering between cyanophages based on the MazG sequence can stem from several possible sources. One possibility is that there is a high rate of exchange of this gene between different families of phages, via viruses that can co-infect the same host or via lysogenic viruses (Canchaya*, et al.*, 2003). This possibility seems less likely given that lysogenic cyanophages have not yet been described whilst there is distinct clustering observed in the case of other genes. Another possibility is that different selection pressures apply on the cyanophage and cyanobacterial host *mazG* orthologues. The difference in selection pressure would imply differences in activity (a different role in host metabolism or different substrate preference) or being subject to different intracellular regulation. This scenario seems more likely in light of the low identity between the host and phage orthologues, as well as differences in kinetic parameters measured in the course of this study and discussed below (section 5.5.4).

5.5.2 Growth analysis of the *Synechococcus* sp. WH7803 *mazG⁰* mutant

While the *E. coli* MazG and its role in the cellular stress response, have been well characterised (Lee*, et al.*, 2008; Gross*, et al.*, 2006), there is little known of the activity of MazG orthologues in cyanobacteria. Therefore, it was important to investigate the physiological differences in cellular response to common nutrient stress conditions in a *mazG* mutant strain. Growth rates of both *mazG* mutant and *Synechococcus* sp. WH7803 WT were extremely similar in each growth condition (Figure 5.6 and Table 5.2). This observation is congruent with the fact that the main predicted role of *mazG* is to reduce the level of the alarmone molecule responsible for the stringent response and cell death (Srivatsan and Wang, 2008). Since the growth rate is measured during the exponential stage of bacterial growth, it is possible that *mazG* is still not required during this stage and therefore the mutant lacking *mazG* is not disadvantaged compared to the WT strain. The only physiological difference observed between the WT and the mutant strain was in the lower photosynthetic efficiency of the mutant compared to WT when grown in P-deplete medium (Figure 5.7). Under these growth conditions, the WT strain exhibits a much higher photosynthetic efficiency than the mutant strain as well as higher efficiency than the WT strain growing under normal conditions. The possible

connection between P-depletion and photosynthetic metabolism in marine cyanobacteria is unknown. However, it's possible that the intracellular pyrophosphatase activity of MazG provides the cell with additional inorganic phosphate required for metabolic processes. Previous work has shown that in the freshwater cyanobacterium *Microcystis aeruginosa*, P-depletion causes a decrease in light harvesting capacity (Kromkamp*, et al.*, 1989), but it is not clear how results of this study can explain this observation. Due to the strongly pronounced difference between the photosynthetic yield of the *mazG^0* mutant, it is possible that regulatory networks involved in the stringent response also play a role in the P-stress response. Previous studies have shown that there is a strong connection between the accumulation of (p)ppGpp and the ability to form intracellular polyphosphate storage bodies as a strategy to deal with P-depletion (Rao*, et al.*, 1998). Therefore, it would be interesting to investigate whether there is a pronounced and detectable difference between P-stressed cells compared to the P-stressed *mazG^0* mutant on a transcriptional and translational level.

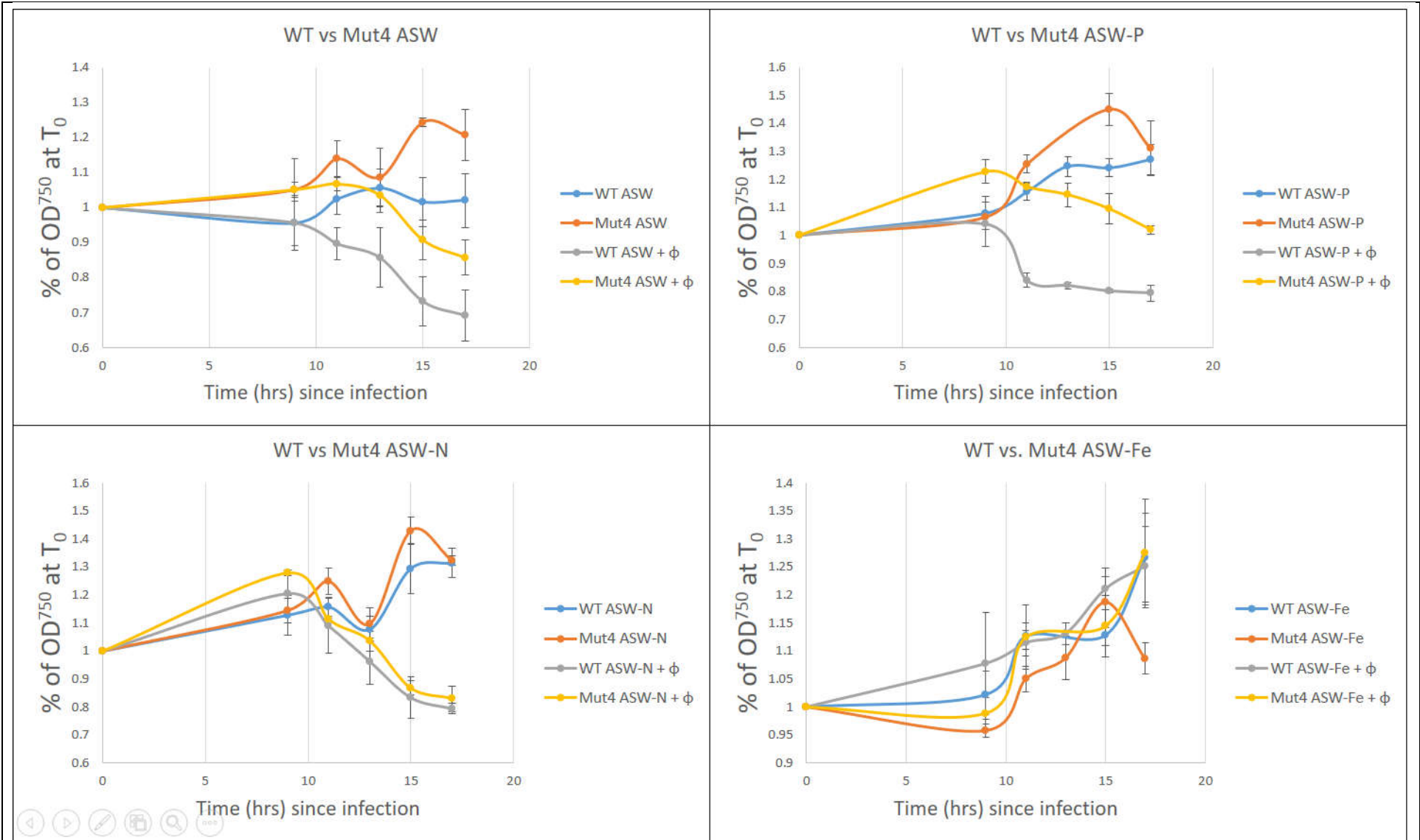### 5.5.3 Infection dynamics of the *Synechococcus* sp. WH7803 and *Synechococcus* sp. WH7803 *mazG^0* mutant (Mut4) by cyanophage S-PM2

An additional aspect of investigating the potential role that MazG may play in cellular metabolism is examination of the difference in infection kinetics between the WT and the *mazG^0* mutant. During the course of infection by a bacteriophage, there are significant changes in the levels of intracellular metabolites (De Smet*, et al.*, 2016). More specifically, during the infection of *Synechococcus* sp. WH7803 by cyanophage S-PM2, there is a diversion of carbohydrate metabolism away from CO_2 fixation towards production of metabolites required for phage progeny synthesis (Puxty*, et al.*, 2015a; Puxty*, et al.*, 2016). Therefore, it is possible that phage infection causes intracellular nutrient depletion, which in turn could lead to the stringent response. It has been shown that the stringent response has a specific negative effect on the bacteriophage life cycle – e.g. a preferential redirection of RNA polymerase to transcription of stress response genes (Paul*, et al.*, 2005) could cause reduced transcription of bacteriophage genes. Furthermore, it has been shown that phage DNA replication is specifically reduced under the stringent response in *E. coli* (Nowicki*, et al.*, 2013b). Therefore, even though during the course of infection, the host cell is depleted of nutrients, interfering with the host stringent response would provide a competitive advantage to phages capable of such interference. As described earlier, MazG is such a potential agent of interference, available to infecting phage. As previous work has shown, levels of the alarmone molecule (p)ppGpp are reduced in freshwater

cyanobacterial cells infected by cyanophage (Borbély, *et al.*, 1980). Hence, it was of interest to investigate whether the *Synechococcus* sp. WH7803 *mazG⁰* mutant would exhibit different infection kinetics compared to the WT. One-step infection experiments performed during the course of this work (Fig. 5.7) showed there are significant differences in the relative success of infection between these two strains. Moreover, the success of infection of both WT and mutant strains (as measured via the length of the latent period) was highly dependent on the type of nutrient stress under which the infection took place (Fig. 5.7)., Cyanophage S-PM2 infection of the mazG mutant strain under nutrient-replete conditions caused a prolongation of the latent period, as exhibited by the delayed decrease in the OD_{750} of the infected mutant culture, compared to WT. This effect is even more pronounced during P-deplete conditions, where there is only a slight decrease in the culture density of the infected mutant strain, even after 16 hours. Interestingly, there is almost no difference between the infection kinetics between the mutant and the WT strain growing under N- or Fe- deplete conditions. This result is particularly surprising since one of the most significant triggers for the stringent response is amino acid depletion (Srivatsan and Wang, 2008). One of the immediate effects of N-stress in cyanobacteria is reduced glutamate-glutamine synthesis (Garcia-Fernandez, *et al.*, 2004), which propagates further down metabolic pathways and negatively affects amino acid metabolism. Therefore, it was expected that N-depletion would result in an activated stringent response, a condition which would be significantly harder to alleviate in the strain lacking the *mazG* gene. However, similar infection kinetics of the mutant and WT strain could potentially hint towards the complementary role that the phage *mazG* may play in the course of infection, replacing the function of the missing cyanobacterial orthologue. In order to further investigate this possibility, complementary mutants should be created in which the cyanobacterial host gene is replaced by the cyanophage orthologue and physiological parameters compared between the WT, *mazG⁰* mutant and the complemented strain.

Examination of the infection kinetics of Fe-deplete cultures provided another interesting result. Even though there was no visible difference between the infection kinetics of WT and mutant strains, the complete lack of lysis of both strains under Fe-deplete conditions is surprising. While reduction in infection success under nutrient-deplete conditions has been previously observed (Wilson, *et al.*, 1996), in no case has this been a complete lack of lysis. This result points towards the importance of iron on cyanobacterial metabolism in general and its importance in securing a successful cyanophage infection cycle in particular. Iron has previously been identified as one of the most important factors limiting primary productivity in the oligotrophic marine

environment (Moore*, et al.*, 2013). This result shows that iron depletion also affects viral infection of primary producers in this environment and should be further investigated in order to better understand its effect on biogeochemical cycling in the ocean.

5.5.4 Comparison of enzymatic activity of *Synechococcus* WH7803 MazG protein with the cyanophage S-PM2 orthologue

In order to further understand the potential role that the viral MazG orthologue could play in the course of infection, metabolic assays were performed, examining the enzymatic activity of both host and viral proteins against a range of substrates. The results of these experiments point towards significant differences in the affinity towards different substrates between the bacterial and viral orthologues (Figure 5.12). While for all other nucleotides, the V_{max} of the cyanophage MazG is similar to that of the *Synechococcus* protein, for dCTP and dGTP, there is a significant difference in kinetic parameters between the two orthologues, where the cyanophage enzyme seems to exhibit much higher activity than the bacterial one. There are several potential reasons for this difference:

• It is possible that the high affinity (V_{max}, K_{cat}, K_m) of the cyanophage MazG towards dGTP stems from an increased cyanophage enzyme affinity towards the guanosine-based alarmone molecule (p)ppGpp. This possibility is congruent with the predicted function of the cyanophage enzyme which serves to reduce levels of (p)ppGpp during the stringent response, thus delaying cell death and increasing the time-window during which the phage progeny could be released from the cell. However, this result does not explain the low V_{max} observed in reactions using GTP as a substrate, since the (p)ppGpp anabolic pathway uses GTP and not dGTP as a precursor (Hauryliuk*, et al.*, 2015).

• It is possible that the role of the cyanophage MazG is unrelated to the stringent response, but rather has to do with removing mutagenic nucleotides from the cellular pool. Previous experiments using *E. coli* and *Mycobacterium* MazG enzymatic assays have shown that they are also capable of hydrolysing 8-oxo-dGTP, a mutagenic nucleotide that arises in cells during oxidative stress (Lu*, et al.*, 2010). MazG from *Sulfolobus* has been implicated in hydrolysis of 2-oxo-(d)ATP, another mutagenic non-canonical nucleotide (Galperin*, et al.*, 2006). Relative to other cyanophages, *Synechococcus*-infecting cyanophage have a lower %G+C content (Sullivan*, et al.*,

2010), which could be the source of the higher affinity of S-PM2 MazG for dGTP and dCTP observed here.

It is important to note that due to the relatively low-similarity between the host and cyanophage MazG, any conclusion about the potential role that the cyanophage MazG plays in cellular metabolism is rather speculative. (Cyano)bacterial MazG contains 2 putative catalytic domains, the function of which was confirmed by resolving its crystal structure (Lee, *et al.*, 2008). The *Synechococcus* sp. WH7803 MazG orthologue possesses the same two predicted catalytic domains as the *E. coli* protein. However, pairwise alignment of the cyanophage MazG protein shows that the area of alignment covers only the predicted C-terminal catalytic domain of the bacterial protein (Figure 5.13). Point mutations in *E. coli mazG* have demonstrated that the C-terminal domain is the only domain that has NTPase activity (Lee, *et al.*, 2008), which is entirely consistent with the cyanophage protein also possessing this catalytic capability.
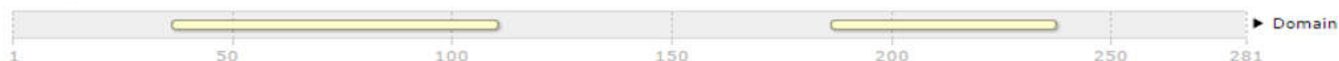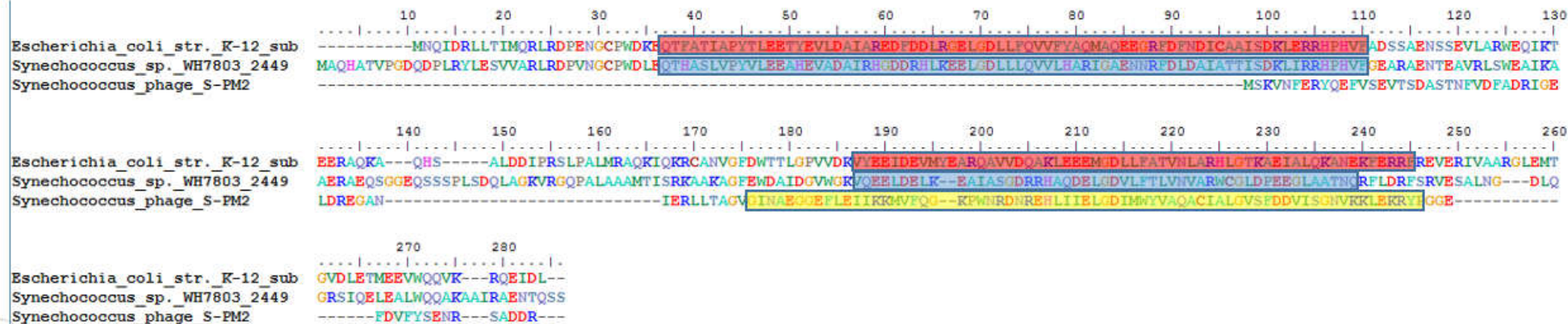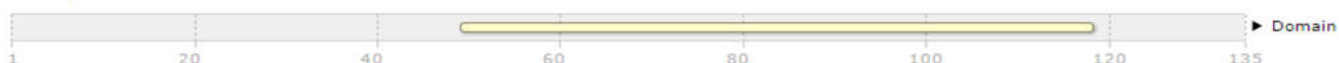
Figure 5.14: A - InterProScan5 (Jones, *et al.*, 2014)-predicted pyrophosphatase catalytic domain in *E. coli* and *Synechococcus* WH7803 MazG proteins. The numbers above each domain represent the position of amino acids in each of the domains. B – ClustalW pairwise alignment of *E. coli* , *Synechococcus* WH7803 and cyanophage S-PM2 MazG orthologues. Blue, red, and yellow shading represent the location of the predicted catalytic domains in *E. coli* , *Synechococcus,* and S-PM2 peptides, respectively.

5.5 Future work

While this work is the first attempt to understand the role of both host and cyanophage MazG proteins during the course of phage infection, several unanswered questions remain. Firstly, there are additional nutrient stress conditions under which $mazG^0$ mutant physiology could be investigated e.g. growth under different light conditions, different growth temperatures and lengths of stationary phase. Due to the putative role of MazG in alleviating oxidative stress damage to nucleotides, it would be particularly interesting to examine how the $mazG^0$ mutant copes with prolonged high light conditions.

Secondly, the fact that growth rates did not significantly differ between the *Synechococcus* sp. WH7803 mazG mutant and WT strain, probably stems from the role MazG plays in the late stages of cell growth i.e. during stationary and death phase. Therefore, it would be interesting to perform long term growth experiments comparing the behaviour of the mutant and WT strains over the stationary, death and even long-term stationary phases. Potential differences in growth rate should also be reflected in gene expression. Therefore, transcriptomics analysis using RNASeq and comparing gene expression in the WT and mutant under different nutrient stress conditions and at different growth stages should be able to shed more light on the role *mazG* may be playing in the stress response. Additionally, it would be interesting to replace the host *mazG* with the cyanophage orthologue and investigate whether the lack of the N-terminal catalytic domain in the phage MazG plays a role in regulation of enzymatic activity of MazG under stringent conditions.

Chapter 6

RNAseq analysis of P-stressed *Synechococcus* sp. WH7803 infected by cyanophage S-PM2

6.1 Introduction

While the transcriptional landscape of bacteriophage T4 has been previously thoroughly investigated (Luke, *et al.*, 2002; Miller, *et al.*, 2003b), RNA sequencing of the T4-like cyanophage infection process has only recently gained attention. The primary focus of past cyanophage gene expression studies were AMGs and their impact on infection dynamics under relevant growth conditions (Clokie, *et al.*, 2006; Lin, *et al.*, 2016; Thompson, *et al.*, 2011). However, a recent study focused on potential differential gene expression of a broad host-range cyanophage infecting different *Synechococcus* hosts (Doron, *et al.*, 2016).

While genetically and morphologically similar, the temporal transcriptional profile of cyanomyoviruses is slightly different to that found in T4-like bacteriophages. T4 genes were shown to be expressed in 3 main clusters – early, middle and late (Luke, *et al.*, 2002) – with the temporal differences stemming from different promoter classes and host RNA polymerase modifications performed by phage-encoded sigma factors and inhibitors (Mosig and Eiserling, 2006).

In gene expression studies of cyanomyoviruses, two temporal clusters were identified – early and late. This conclusion was based primarily on investigation of changes in expression levels of specific genes (Clokie, *et al.*, 2006; Thompson, *et al.*, 2011) as well as due to the fact that other myoviruses were found to be lacking both the phage T4-encoded σ^{70} orthologue *motA* and the host RNA polymerase inhibitor – Asi (Desplats, *et al.*, 2002; Miller, *et al.*, 2003a; Mann, *et al.*, 2005) – considered to be responsible for expression of the viral "middle" genes. More recent studies on some cyanopodoviruses (Lindell, *et al.*, 2007) as well as cyanomyoviruses (Doron, *et al.*, 2016; Lin, *et al.*, 2016) did show the presence of the "middle" gene cluster. However, in contrast to phage T4 no known transcription factors have been identified in the genomes of those cyanophages enabling them to facilitate "middle" gene expression.

Additional potential complexity in the regulation of cyanophage gene expression stems from the presence of AMGs. These genes, thought to originate from the infected host genome (Hurwitz and U'Ren, 2016), actively participate in host metabolic processes (Clokie, *et al.*, 2006; Thompson, *et al.*, 2011; Puxty, *et al.*, 2016). Due to their metabolic roles, it was perhaps to be expected that rather than responding only to temporal transcriptional signals, viral AMGs should also be responsive to the metabolic state of the infected host. Interestingly, cyanophage P-SSM2 infection of a P-limited *Prochlorococcus* host showed only two genes to be up-regulated relative to the P-replete control conditions. Here, a *pstS* gene orthologue encoding a well characterised periplasmic phosphate-binding protein component of a phosphate ABC transporter (Scanlan, *et al.*, 1993), as well as another gene (*g247*) of unknown function found in the same operon with *pstS* (Lin, *et al.*, 2016) were both up-regulated. Regulation of these

phage genes is via the host two component system PhoB/PhoR, with the *pstS* promoter possessing a Pho box to which the transcription factor PhoB binds (Zeng and Chisholm, 2012). As described in Chapter 3, infection of *Synechococcus* sp. WH7803 by cyanophage S-PM2 under P-limiting conditions, results in a prolonged latent period and reduced burst size (Wilson*, et al.*, 1996). Since cyanophage S-PM2 does not contain any putative AMGs relevant to P-metabolism, it is of interest to investigate whether this delay in the latent period is associated with differential gene expression, compared to infection under P-replete conditions. Also of interest is whether the temporal gene clusters are conserved, both in their timing as well as in the genetic make-up of each cluster. Additionally, it is not clear whether the delay in infection kinetics, as well as potential differences in temporal phage gene expression profile, are in any way reflected in differential host gene expression levels. The purpose of the work described in this chapter was to give answers to these questions.

6.2 Aims

The overall goal of the work described in this chapter was to understand the transcriptional response of cyanophage S-PM2 infecting the host *Synechococcus* sp. WH7803 under P-deplete conditions.

Specific aims were as follows:

• To identify the optimal sampling points, via qPCR, for measuring the levels of intracellular phage DNA replication over the course of a one-step infection cycle.

• To assess the transcriptional response, using RNAseq, of cyanophage S-PM2 over the course of infection of *Synechococcus* sp. WH7803 during host growth under P-replete and P-deplete conditions.

6.3 Materials and methods

6.3.1 One step infection experiment

In order to define the time points which will be used in the RNASeq experiment, a one-step infection experiment was performed in order to estimate the dynamics of intracellular phage DNA replication, in the following manner:

Synechococcus sp. WH7803 was grown in ASW medium (see section 2.3.1) at 23ºC under continuous illumination at an intensity of ~10 µmol photons m^{-2} s^{-1} with constant shaking at ~220 RPM. When the culture reached OD$_{750}$=0.35, it was transferred into 1L of fresh ASW (+P) and ASW -P (-P) media at 10% v/v concentration. The fresh cultures were grown under the same conditions as the initial starter culture until reaching OD$_{750}$ = 0.25. From this point, the P-deplete culture was tested daily for the occurrence of pseudolysogeny via a 96-well plate infection assay, as described below.

Upon detection of pseudolysogeny (see section 6.3.2), the cell concentration in both +P and -P cultures was measured via flow cytometry (see section 2.4.3) and the cultures were diluted to a concentration of 5x10^7 cells ml^{-1} in ASW/ASW-P accordingly and 5ml of diluted cultures were aliquoted into polycarbonate test tubes in triplicate and infected with cyanophage S-PM2 at a multiplicity of infection (MOI) of 10. In addition to infected +P/-P cultures, a triplicate aliquot of cultures were used as non-infected controls.

Sampling was performed as described in (Thompson*, et al.*, 2011). Briefly, at time points 0, 2, 4, 6, 8, 10, 12, 14, 16 and 18 hours from the start of infection, 200 µl from each replicate was diluted in 500 µl ASW medium and vacuum filtered on a 0.2 µm pore size polycarbonate filter (Isopore, Millipore, USA), mounted on a glass filter tower. Samples from all replicates were taken simultaneously. While still on the filter, the samples were washed with 1 ml preservation solution (100 mM EDTA, 500mM NaCl, 10mM Tris-HCl, pH 8.0) three times. The filters were then inserted into a ribolyser tube (Lysing Matrix E, MP Bioproducts, USA) and snap-frozen in liquid nitrogen. An additional 100 µl of infected culture was fixed using paraformaldehyde (1% (w/v) final concentration) and stored at 4º C for cell counting by flow cytometry (see section 2.4.3).

Intracellular phage DNA was extracted in the following manner: 650 µl Tris-HCl pH 8.0 was added to the filter-containing ribolyser tubes from the previous step and cells were lysed via 3 cycles of 30s shaking at 30Hz in a Reusch Tissue Lyser. The tubes were then centrifuged at 30s at 10,000g, the supernatant snap-frozen in liquid nitrogen and stored at -80ºC.

In order to estimate intracellular phage DNA replication, samples were diluted 1/10 (v/v). The qPCR reaction was performed in triplicate, with a negative control using 1 µl nuclease-free water instead of the template. 1 µl of diluted phage-DNA containing sample was used as the template in 20 µl final volume of the qPCR reaction, according to the manufacturer's specifications (SYBR Select Master Mix, Applied Biosystems). Primers targeting the S-PM2 *psbA* gene (Table 2.3) were used at a concentration of 250

nM and qPCR reactions were run on a 7500 Fast Real-Time PCR thermocycler (Applied Biosystems), using fast settings (95º C for 20s, followed by 40 cycles of 95º C for 3s and 60º C for 30s). The starting amount of phage DNA was estimated via comparison of Ct values from each sampling point (in triplicate) to the linear regression curve of Ct values of the standard curve obtained from wells containing purified, quantified S-PM2 genomic DNA.

6.3.2 Assessment of pseudolysogeny using 96-well plates

In order to assess whether *Synechococcus* sp. WH7803 was sufficiently P-depleted for the occurrence of pseudolysogeny, a one-step infection assay was performed in 96-well plates as follows: *Synechococcus* cell concentrations were measured via flow cytometry (section 2.4.3); subsequently, 200 µl of both P-replete and P-deplete *Synechococcus* sp. WH7803 was aliquoted into 96-well plates in triplicate. A stock of cyanophage S-PM2 of known concentration was diluted up to a 100 µl final volume in ASW and ASW-P and added to each *Synechococcus*-containing well so that the final multiplicity of infection was 10. 100 µl ASW or ASW-P were added to wells containing just the uninfected host for a no-virus control. The plate was incubated under continuous illumination at an intensity of ~10 µmol photons m^{-2} s^{-1}. OD$_{750}$ was measured every two hours after 8 hours of incubation, using the iMark Microplate reader (Biorad). Pseudolysogeny was established when there was a delay of more than 2 hrs of lysis of the P-deplete culture, compared to the P-replete control culture.

6.3.3 RNASeq analysis: experimental set-up

Synechococcus sp. WH7803 was grown under P-replete and P-deplete conditions and the presence and absence of infection with cyanophage S-PM2 in an identical manner to that described above for the one-step infection experiment (see section 6.3.1). When phosphate stress in the -P culture had reached levels inducing pseudolysogeny in the infected culture (as detected via the above-described 96-well plate assay, see section 6.3.2.), the +P and -P cultures were divided into six replicates. Three replicates of each of the -P conditions were infected with cyanophage S-PM2 suspended in ASW-P medium at a calculated multiplicity of infection of 8. The remaining three replicates of both +P and –P conditions were left uninfected as the no-phage control. The scheme for the experimental setup can be seen in visual format in Figure 6.1.

Figure 6.1: Outline of the RNAseq experiment: 1. The *Synechococcus* sp. WH7803 starter culture was grown until the exponential phase (OD$_{750}$≈0.3). 2. The starter culture was transferred into +P/-P ASW medium and the –P culture assessed for the occurrence of pseudolysogeny using 96 well plates 3. Division of +P/-P cultures into 6 replicates and infection of 3 of the replicates for each condition with cyanophage S-PM2.

The replicates were sampled according to the sampling chart shown in Table 6.1.

Time point (hours)	ASW + P - uninfected	ASW + P infected	ASW -P uninfected	ASW -P infected
0	+	-	+	-
3	+	+	+	+
6	+	+	+	+
9	+	+	+	+
12	+	-	+	+
15	+	-	+	+

Table 6.1: The strategy used to sample the infected and uninfected *Synechococcus* sp. WH7803 cultures for the RNAseq experiment.

At each time point, 15 ml sample was taken from each replicate and filtered onto 0.2 µm pore size polycarbonate filters, in a manner identical to the procedure performed in

the one-step infection experiment described above (section 6.3.1). The filters were washed three times using preservation buffer (100 mM EDTA, 500mM NaCl, 10mM Tris-HCl, pH 8.0), snap frozen in ribolyzer tubes in liquid nitrogen and stored at -80º C for RNA extraction.

6.3.4 RNA extraction

RNA was isolated using a modified phenol-chloroform method in the following manner: Filters in ribolyser tubes were resuspended in 400 µl Solution A (0.3M sucrose, 0.01M Sodium acetate) and 400 µl Solution B (0.01M Sodium acetate, 2% (w/v) SDS pH 4.5). An equal volume of phenol pH 4.5 was added, and cells were ribolysed as described above. Lysed cells were centrifuged at 10,000g at 4ºC for 5 minutes. The aqueous phase was extracted and washed with an equal volume of phenol pH 4.5 and centrifuged as before. The aqueous phase was then extracted and washed in an equal volume of phenol:chloroform (1:1) and centrifuged. The aqueous layer was removed and RNA was precipitated by adding 1/10th volume of 3.5 M Sodium Acetate pH 4.5 and an excess 100% (v/v) EtOH and incubating overnight at -80ºC. Precipitated RNA was pelleted for 30 min at 10,000g at 4ºC and washed in 70% (v/v) ice-cold EtOH. After removing the EtOH and air-drying the pellet, RNA was resuspended in 100 µL sterile MiliQ water. DNA was removed using the Ambion TurboDNA-free kit protocol (Life Technologies). After DNA removal, RNA was cleaned using Rybo-zymo columns according to the Rybo-Zymo RNA Clean and Concentrator™ protocol (Zymo Research Corp. CA, USA), concentrated in 50 µl sterile miliQ water and stored at -80ºC until further use.

6.3.5 RNA sequencing

Samples for RNA sequencing were sent to the Next Generation Sequencing Facility at the Leeds Institute of Biomedical and Clinical Sciences, St James University Hospital, University of Leeds. The libraries were prepared using the ScriptSeq Complete Kit (Bacteria) and ScriptSeq v2 RNA-seq Kit (Illumina). Ribosomal RNA was removed using the RiboZero Kit – Bacteria (Illumina). Paired end sequencing was performed on a HiSeq 3000 platform, producing 150bp long reads.

6.3.6 RNAseq data analysis

Paired end reads were mapped to both the cyanophage S-PM2 and *Synechococcus* sp. WH7803 genomes using BWA MEM script (Li, 2013) with default parameters. The

resulting SAM files were converted to BAM files and sorted using SAMtools (Li, *et al.*, 2009). In order to count the reads mapping to specific S-PM2 loci, sorted BAM files were first converted to BED files using Bedtools bamtobed script (Quinlan and Hall, 2010). BED files were then used to count the reads mapping to different S-PM2 genes, using Bedtools intersect script with '-c -bed -s' options. Reads were normalised using the RPKM model, giving an estimate of relative expression calculated according to the following equation:

$$RPKM = \frac{Read\ Count}{Gene\ lengt\quad 10^{-3}\quad Total\ Reads\quad 10^{-6}}$$

where Read Count represents the number of reads mapping to a specific locus, Gene length is the length of that locus and Total Reads represents the number of total reads from each sample mapping to the cyanophage S-PM2 and *Synechococcus* sp. WH7803 genomes.

6.3.7 Gene expression profile clustering

Different cyanophage S-PM2 genes were clustered according to the RPKM expression profiles in the MATLAB environment. Clusters were based on relative expression estimates where the maximum expression of each gene was defined as the time point in which the RPKM value for that gene was the highest. All other time point RPKM values were then normalised to the maximum value. For the purpose of clustering, dissimilarity matrices were generated using the Euclidian distance measure and trees were generated using the 'average' linkage method. To test the robustness of the generated clusters, *k-means* clustering was performed, with k values ranging from 0 to 10. The stability of the resulting clustering models was tested by silhouette generation (Rousseeuw, 1987).

6.3.8 Differential Gene Expression

In order to establish any changes in gene expression of cyanophage S-PM2 genes between the corresponding time points in +P and -P infected *Synechococcus* cultures, differential gene expression values were calculated using the DESeq2 R package (Love, *et al.*, 2014) in the R Studio environment. The gene expression fold change values, as well as statistical significance were calculated using the DESeqDataSetFromHTSeqCount function from the DESeq2 package. Only genes with a False Discovery Rate (FDR) p-value <0.05 were considered to be differentially expressed between the conditions.

6.3.9 Pho-box predicition

Promoters of S-PM2 genes identified as differentially expressed under -P conditions, were examined for presence of putative PhoB-binding motifs. Promoter sequences were analysed using the Pattern Locator script (Mrazek and Xie, 2006). The consensus sequence previously identified as a putative Pho box binding site in cyanobacteria - 5'-PyTTAAPyPyT/A-3' - (Su, *et al.*, 2007) was used to scan the upstream region of these S-PM2 genes and several potential binding sites were found.

6.4 Results

6.4.1 qPCR and cell counts analysis of intracellular phage dynamics during one-step infection of *Synechococcus* sp. WH7803

Due to the relatively high phage titer required for the transcriptomics experiment, as well as the time-consuming nature of producing sufficient amounts of phage for all the replicates, conditions and time points in this experiment, there was a need to accurately define the sampling points in a manner that would provide the best coverage across a relatively low number of sampling points. Therefore, qPCR analysis was performed (see section 6.3.1), with the purpose of estimating the dynamics of cyanophage S-PM2 infection via measurement of the amount of intracellular phage DNA. The results of this analysis are presented in Figure 6.2.
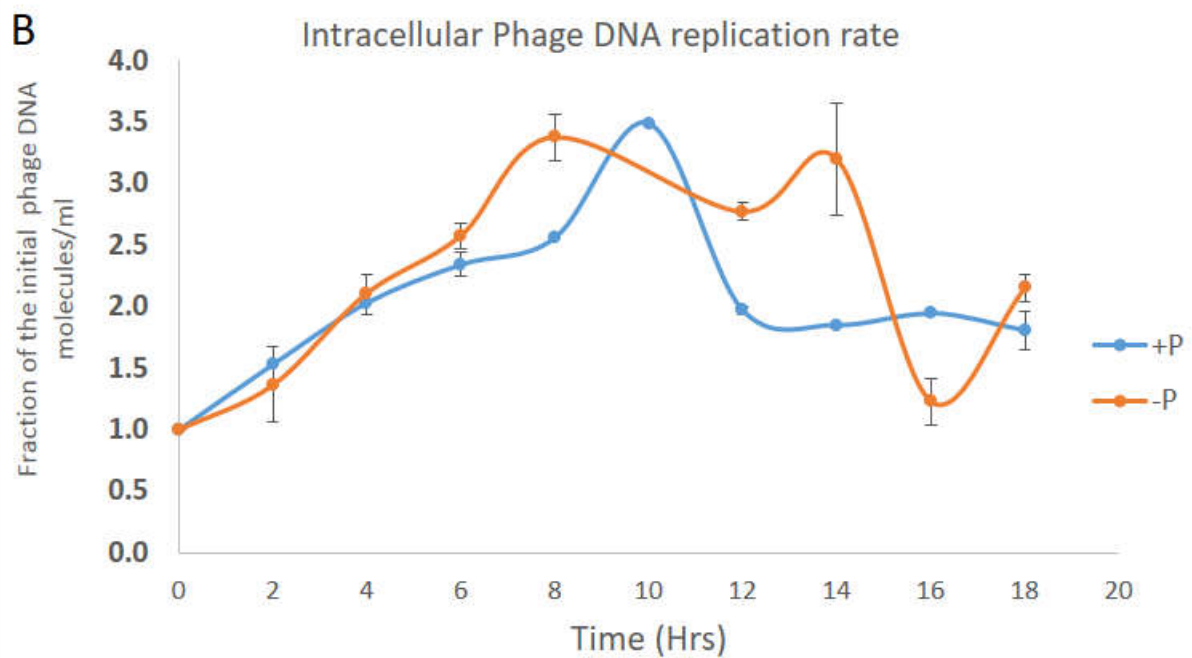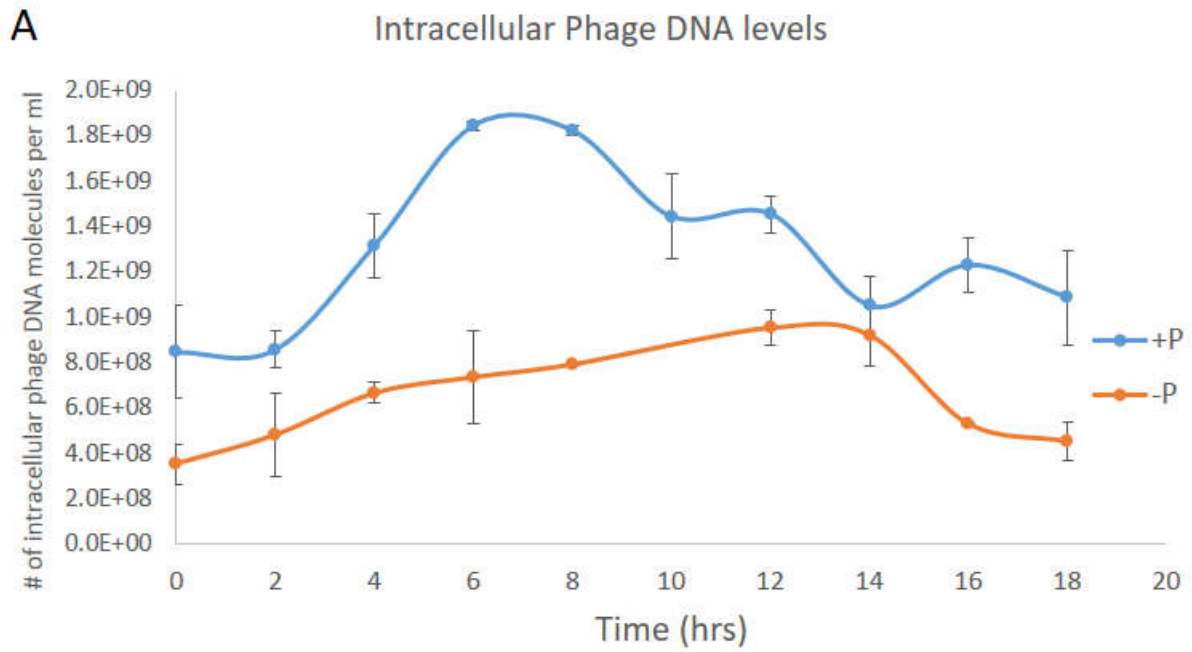
Figure 6.2: Amount of phage DNA inside infected *Synechococcus* sp. WH7803 over the course of infection. A – number of phage DNA molecules inside infected cells; B – fraction of the initial amount of intracellular phage DNA over the course of infection

qPCR showed that viral DNA replication occurred immediately following infection, continuing throughout the latent phase. During infection of the P-replete culture, the amount of intracellular viral DNA starts to decrease between 8–10 hours post infection, a value in good agreement with the previously reported length of the cyanophage S-PM2 latent period (9 hours) (Wilson*, et al.*, 1996). Interestingly, when examining the absolute number of cyanophage DNA molecules inside infected *Synechococcus* cells (Figure 6.2A), viral DNA

replication seems to be slower under P-deplete conditions. However, when normalising the amount of DNA to the value measured at T_0 (Figure 6.2B), the replication rates in +P and -P conditions seem to be similar, even though the total number of phage DNA molecules produced in P-deplete conditions is lower. These results suggest that cyanophages infecting a P-deplete host are still able to maintain a DNA replication rate similar to that observed under P-replete conditions. However, cyanophage S-PM2 DNA replication is clearly more prolonged extending to 14 hours post infection (Figure 6.2A). This is similar to that reported previously (Wilson, *et al.*, 1996), and agrees well with my own flow cytometry data from this experiment (Figure 6.3), where a clear delay in lysis of the infected P-deplete culture is visible compared to the infected P-replete control culture.

Figure 6.3: *Synechococcus* sp. WH7803 cell abundance, as measured using flow cytometry, in cyanophage S-PM2 infected and uninfected cultures, growing under +P or -P conditions over the course of a one-step growth experiment.
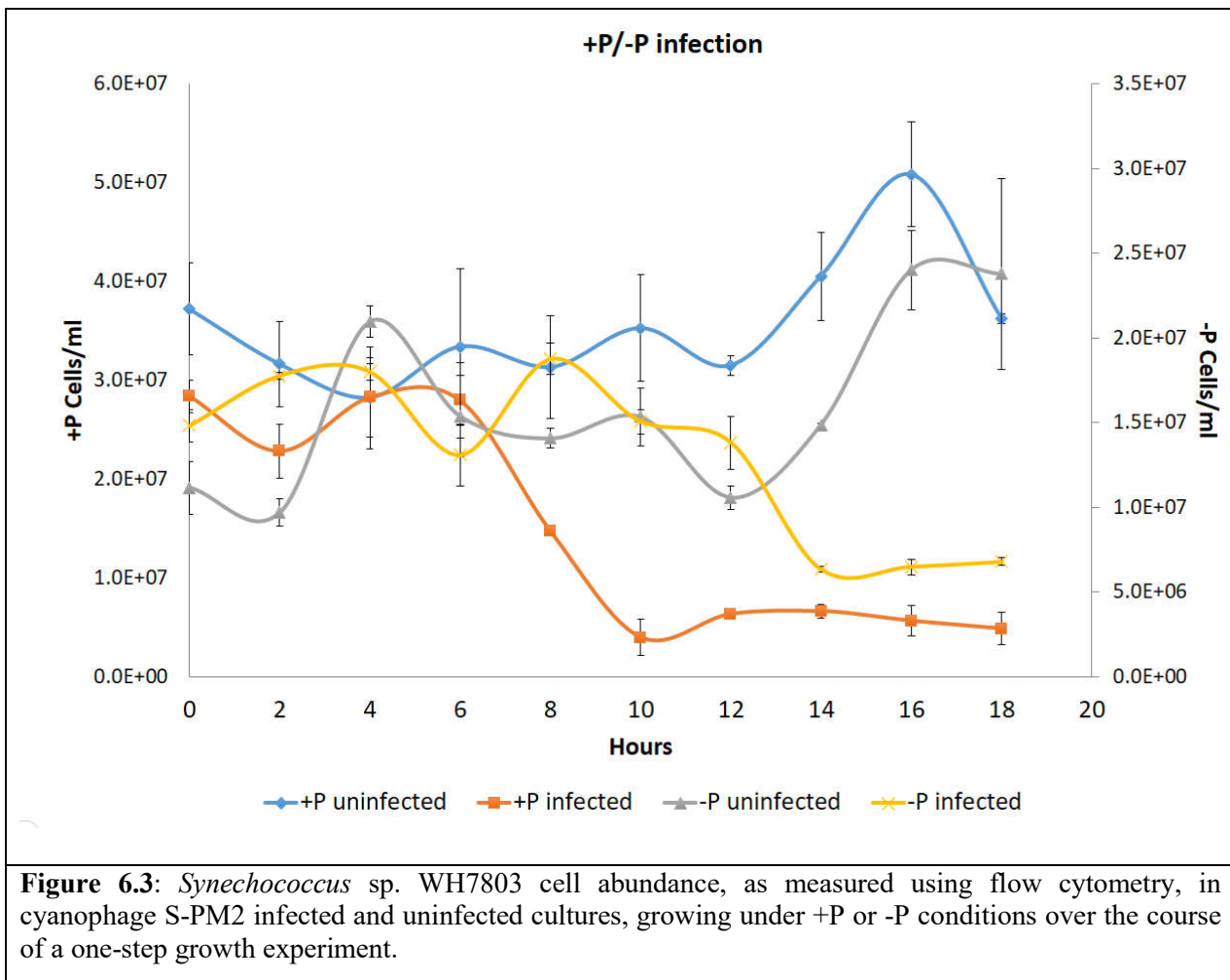
Based on these qPCR results, time points of 3, 6 and 9 hours were chosen for sampling of the +P infected culture, with the -P infected culture being sampled at an additional two time points, 12 and 15 hours post infection.

6.4.2 Transcriptomics analysis of cyanophage S-PM2 infection of *Synechococcus* sp. WH7803 during P-replete and P-deplete conditions: an RNAseq approach

6.4.2.1 RNASeq mapping and coverage statistics

The sequence coverage and mapping statistics of the RNAseq data is shown in Table 6.2. This demonstrates significant coverage of both cyanophage S-PM2 and *Synechococcus* sp. WH7803 host genomes across all time points and conditions. Interestingly, there is a significant difference between the average number of cyanophage reads extracted from the P-deplete and P-replete samples (Figure 6.4) – with the fraction of S-PM2 reads in the P-replete sample higher than the fraction of reads in the P-deplete sample (0.970 (+P) versus 0.924 (-P), t=4.2109, p-value = 0.005783).

Such data supports evidence obtained from other types of experiments, both previously reported in the literature (Wilson*, et al.*, 1996) and performed during the course of this study, showing that cyanophage S-PM2 infection of *Synechococcus* sp. WH7803 grown under P-deplete conditions suffers physiological 'setbacks', resulting in decreased viral progeny, a prolonged latent phase and reduced viral DNA replication.

A pronounced difference in the number of reads mapping to the host *Synechococcus* sp. WH7803 genome in the infected and uninfected samples was also observed (Figure 6.5). While the average percentages of mapped reads are similar between +P and -P uninfected (0.888 and 0.912 accordingly, t=-1.5771, p-value=0.1469) and between +P and -P infected samples (0.017 and 0.026 accordingly, t=-1.7925, p-value=0.1245), there is a stark difference between infected and uninfected samples. The low number of host reads in infected cultures is consistent with the phage-induced degradation of host transcripts during the infection process. Thus, recent work has shown that during the course of cyanophage P-SSP7 infection of *Prochlorococcus* sp. MED4, phage mRNA is selectively protected from degradation by the host-encoded RNase E enzyme, which is induced during the course of infection (Stazic*, et al.*, 2016).

The percentage of reads mapping to ribosomal RNA is particularly low in each sample (<1%), confirming that the ribosomal depletion step performed during the course of the library preparation, which used the Illumina ScriptSeq Complete Kit (Bacteria), was very efficient.
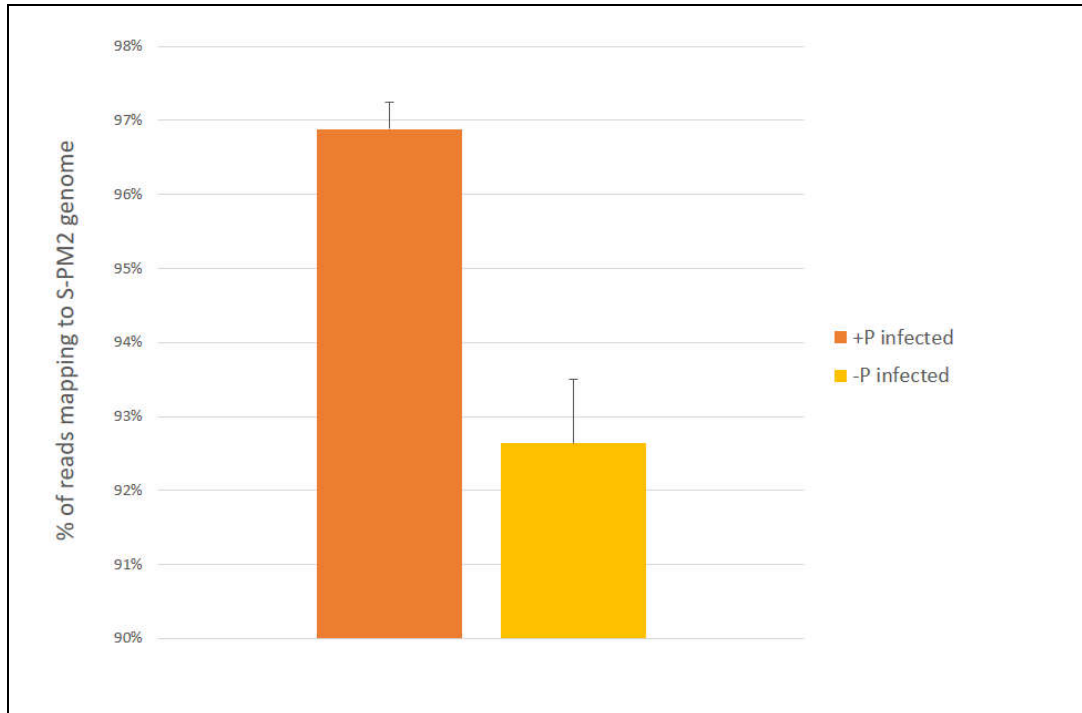
Figure 6.4 – Average percentage of reads mapping to the cyanophage S-PM2 genome in P-deplete vs. P-replete infected *Synechococcus* cultures.
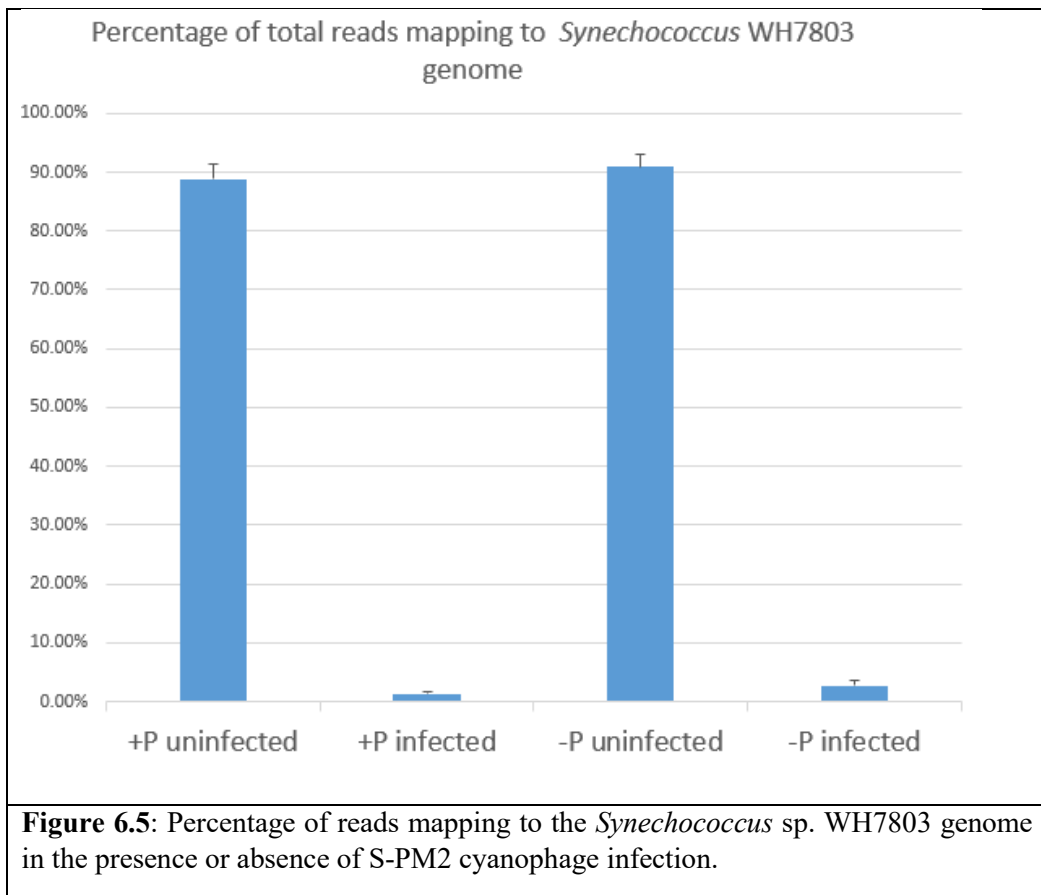
Figure 6.5: Percentage of reads mapping to the *Synechococcus* sp. WH7803 genome in the presence or absence of S-PM2 cyanophage infection.

121

Sample		Total reads	Aligned reads S-PM2	Aligned reads WH7803 (total)	Aligned reads rRNA	Computed fold coverage S-PM2	Computed fold coverage WH7803
+P uninfected	T0	45,835,981	90,667 (0.20%)	39,044,970 (85.18%)	294,458 (0.64%)	72.83	2,474.35
	T3	42,955,097	92,998 (0.22%)	37,044,630 (86.24%)	176,731 (0.41%)	74.70	2,347.59
	T6	45,437,628	376,272 (0.83%)	41,267,449 (90.82%)	138,930 (0.31%)	302.25	2,615.20
	T9	46,563,971	864,069 (1.86%)	41,856,958 (89.89%)	208,41 (0.45%)	694.08	2,652.55
	T12	46,578,331	315,017 (0.68%)	41,378,285 (88.84%)	269,553 (0.58%)	253.04	2,622.22
	T15	57,689,177	191,652 (0.33%)	53,064,767 (91.98%)	282,354 (0.49%)	153.95	3,362.81
+P infected	T0	NS	NS	NS	NS	NS	NS
	T3	64,708,906	62,589,646 (96.72%)	1,094,275 (1.69%)	14,548 (0.02%)	50,276.58	69.35
	T6	45,611,926	44,773,750 (98.16%)	417,687 (0.92%)	5,854 (0.01%)	35,965.55	26.47
	T9	61,605,022	59,001,150 (95.77%)	1,043,947 (1.69%)	25,903 (0.04%)	47,394.03	66.16
-P uninfected	T0	51,571,921	90,077 (0.17%)	47,883,066 (92.85%)	290,222 (0.56%)	72.36	3,034.44
	T3	52,868,975	204,113(0.39%)	48,935,389 (92.56%)	202,218 (0.38%)	163.96	3,101.13
	T6	56,975,818	703,449 (1.23%)	52,139,206 (91.51%)	219,097 (0.38%)	565.06	3,304.16
	T9	49,340,614	1,785,182 (3.62%)	42,690,551 (86.52%)	217,053 (0.44%)	1,433.99	2,705.38
	T12	51,479,973	850,823 (1.65%)	46,528,201 (90.38%)	215,280 (0.42%)	683.44	2,948.58
	T15	57,485,017	366,150 (0.64%)	52,698,190 (91.67%)	334,278 (0.58%)	294.12	3,339.58
-P infected	T0	NS	NS	NS	NS	NS	NS
	T3	51,162,704	45,515,772 (88.96%)	2,257,859 (4.41%)	64,086 (0.13%)	36,561.59	143.08
	T6	59,600,741	56,204,379 (94.30%)	1,647,048 (2.76%)	27,496 (0.05%)	45,147.46	104.38
	T9	56,776,022	53,626,948 (94.45%)	1,231,052 (2.17%)	25,571 (0.05%)	43,077.08	78.01
	T12	64,304,977	59,837,063 (93.05%)	1,344,363 (2.09%)	81,317 (0.13%)	48,065.50	85.19
	T15	53,647,162	49,583,494 (92.43%)	1,255,848 (2.34%)	44,638 (0.08%)	39,829.09	79.59

Table 6.2: RNASeq data: sequencing and mapping statistics. Fold coverage is calculated as the number of reads, multiplied by the read length (150 bp) divided by the genome size. NS – no sample was taken for this time point and condition.

Surprisingly though, there are a number of reads mapping to the S-PM2 genome in uninfected samples. The source of these reads is unknown but could be due to i) cross-contamination during sample processing (since all samples were filtered on the same filter tower, and despite the fact that this was cleaned between each sample being processed, cross contamination between samples is a possibility) or ii) cross-contamination during HiSeq sequencing, (i.e. between different lanes during the course of the Hiseq sequencing). Since PCR amplification of cDNA is one of the steps involved in the library preparation, it is also possible that a small amount of phage RNA from another sample(s) has contaminated the uninfected samples. Since different normalisation methods were used in both the temporal clustering analysis and the differential expression analysis, the presence of this small proportion of reads should have no bearing on the conclusions drawn from this RNAseq experiment.

6.4.2.2 Temporal clustering of cyanophage S-PM2 gene expression

The number of reads mapping to each gene of the cyanophage S-PM2 genome was transformed into RPKM values for each gene according to the equation outlined in section 6.3.6. RPKM values were then normalised to the maximal value for each gene across all time points and clustering analysis performed for +P and -P infected samples. The resulting clustergrams can be seen in Appendix 2.

Temporal clustering analysis revealed that the choice of sampling points for the infection experiment caused the omission of an 'early' S-PM2 gene expression cluster which probably reached maximum expression between time points 0-3 hours. Due to this omission, the clustering profile for +P infected samples does not resemble the typical clustering profile which includes samples from the early infection period (see Appendix 2). Instead, the 'early' gene expression cluster appears to capture a decline in expression levels, levels which then appear to remain constant and high over the remaining time points (Figure 6.6). In contrast, the S-PM2 'late' gene cluster only reaches a maximum at the 6-hour time point, with expression remaining high until the end of the experiment.
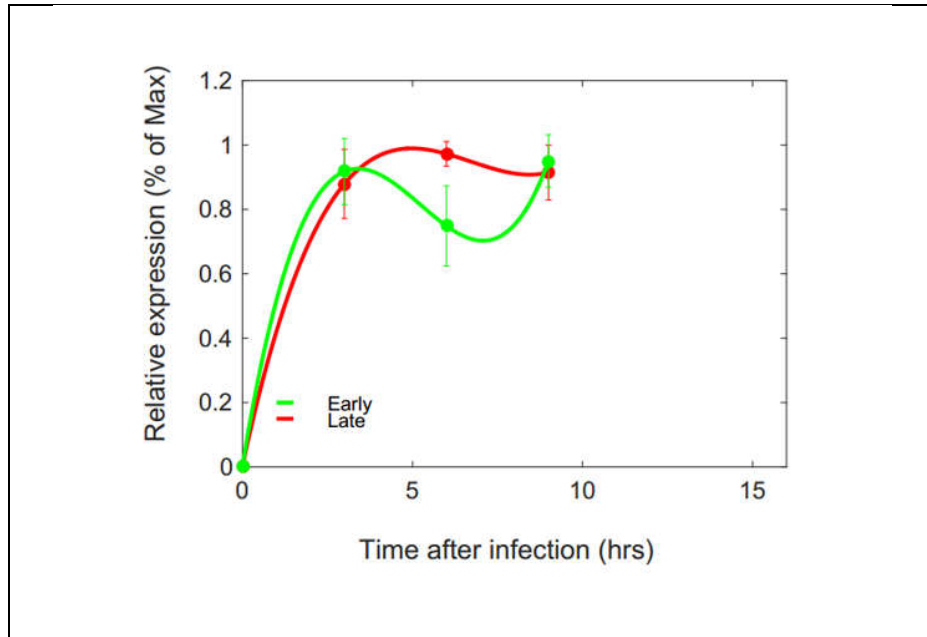
Figure 6.6: Average relative expression of cyanophage S-PM2 'early' and 'late' gene clusters during infection of *Synechococcus* sp. WH7803 under P-replete conditions.

Cyanophage S-PM2 gene expression, during infection of a *Synechococcus* host under P-deplete conditions, shows, in agreement with data showing DNA replication to be prolonged over the course of the 18-hour one-step growth experiment (see Fig. 6.2A), that maximum expression level of the 'early' gene cluster occurs around 3 hours post-infection, while 'late' genes reach their maximum expression 6 hours post-infection (Figure 6.7).
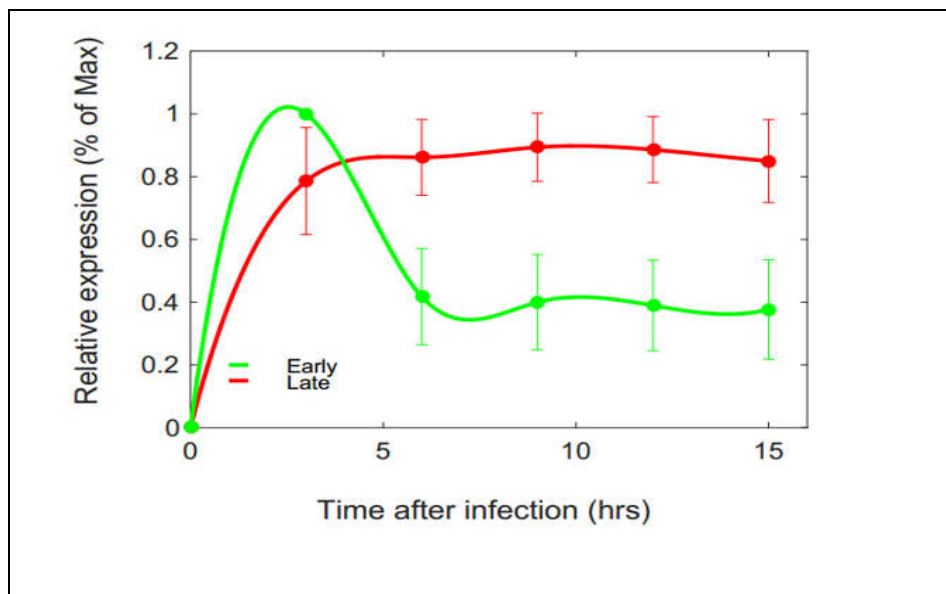
Figure 6.7: Average relative expression of cyanophage S-PM2 'early' and 'late' gene clusters during infection of *Synechococcus* sp. WH7803 under P-deplete conditions.

Interestingly, the high level of expression of the 'late' gene cluster continues throughout the experiment, in contrast to previous expression data where the expression of late genes has been observed to decrease 9 hours post-infection, with the initiation of virion assembly and cell lysis (Clokie, *et al.*, 2006).

In order to assess the robustness of the above described clustering, k-means clustering with silhouette analysis was performed (see section 6.3.6.1). In agreement with the above observation, P-replete infected samples do not reliably cluster into any attempted number of clusters, while the P-deplete infected samples show most reliable clustering into two temporal clusters (Figure 6.8A and B).

If we perform the clustering analysis across all time points (0 – 15 hours) for the -P infected samples, the analysis shows that still, the only clustering model that has a relatively high silhouette score is the two-cluster model (Figure 6.8C).
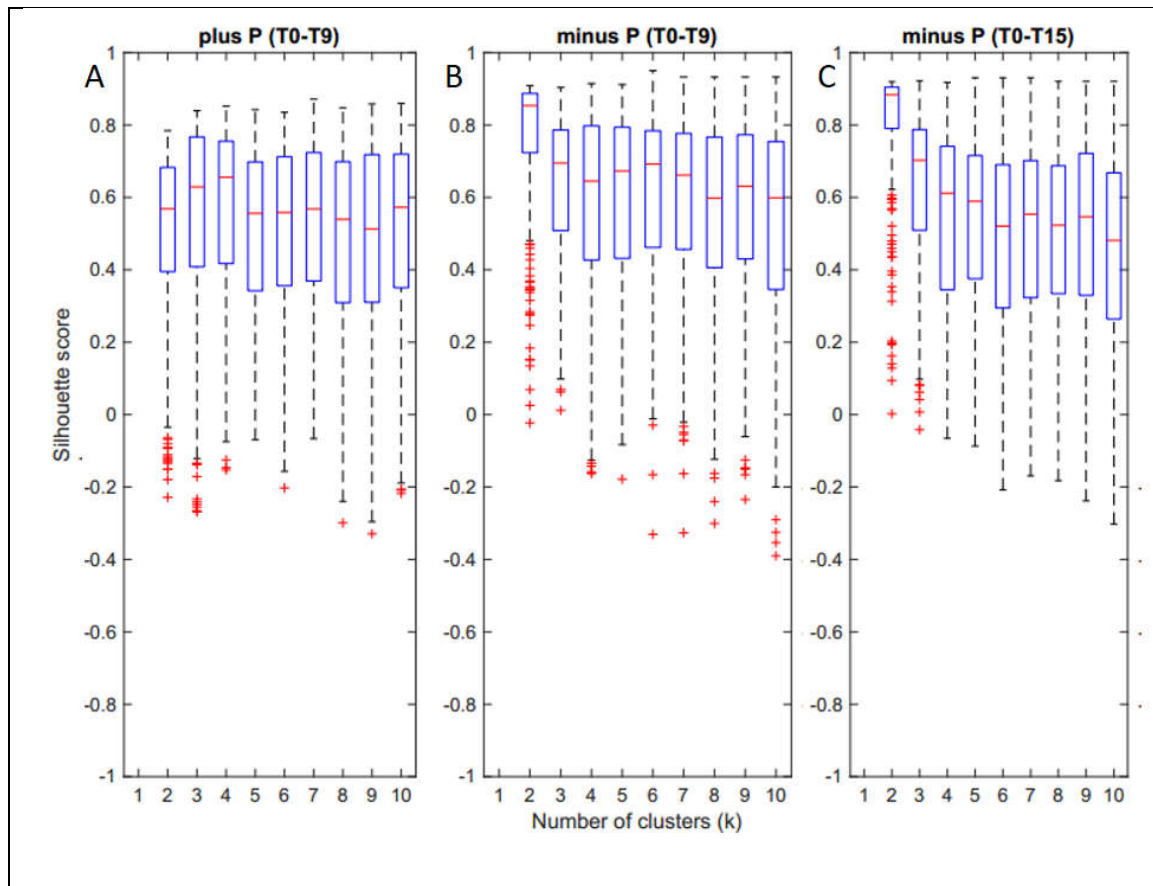
Figure 6.8: Silhouette score analysis of clustering models of temporal gene expression. A – Analysis of the +P infected samples; B – analysis of the -P infected samples, when only taking time points T0-T9 into account; C – analysis of the -P infected samples with all of the time points (T0-T15) taken into account.

6.4.2.3 Differential gene expression

In order to investigate whether there is a difference in S-PM2 gene expression levels between the equivalent time points in P-replete and P-deplete infected cultures, differential gene expression analysis was performed using the DESeq2 R package. The only fold change values that were taken into account were those which received a False Discovery Rate amended p-value <0.05.

After comparing samples from the same time point during infection of P-replete/P-deplete cultures, the only significant differences in gene expression levels were 17 genes which had higher relative expression under P-deplete conditions 3 hours post-infection, compared to their expression at the same time point under P-replete conditions (see Table 6.3). The RPKM value change of each of the identified differentially expressed genes over time can be seen in Figure 6.9.

Figure 6.9 – RPKM values of cyanophage S-PM2 genes found to be differentially expressed 3 hours post-infection under P-deplete conditions, compared to expression levels under P-replete conditions at the same time point.
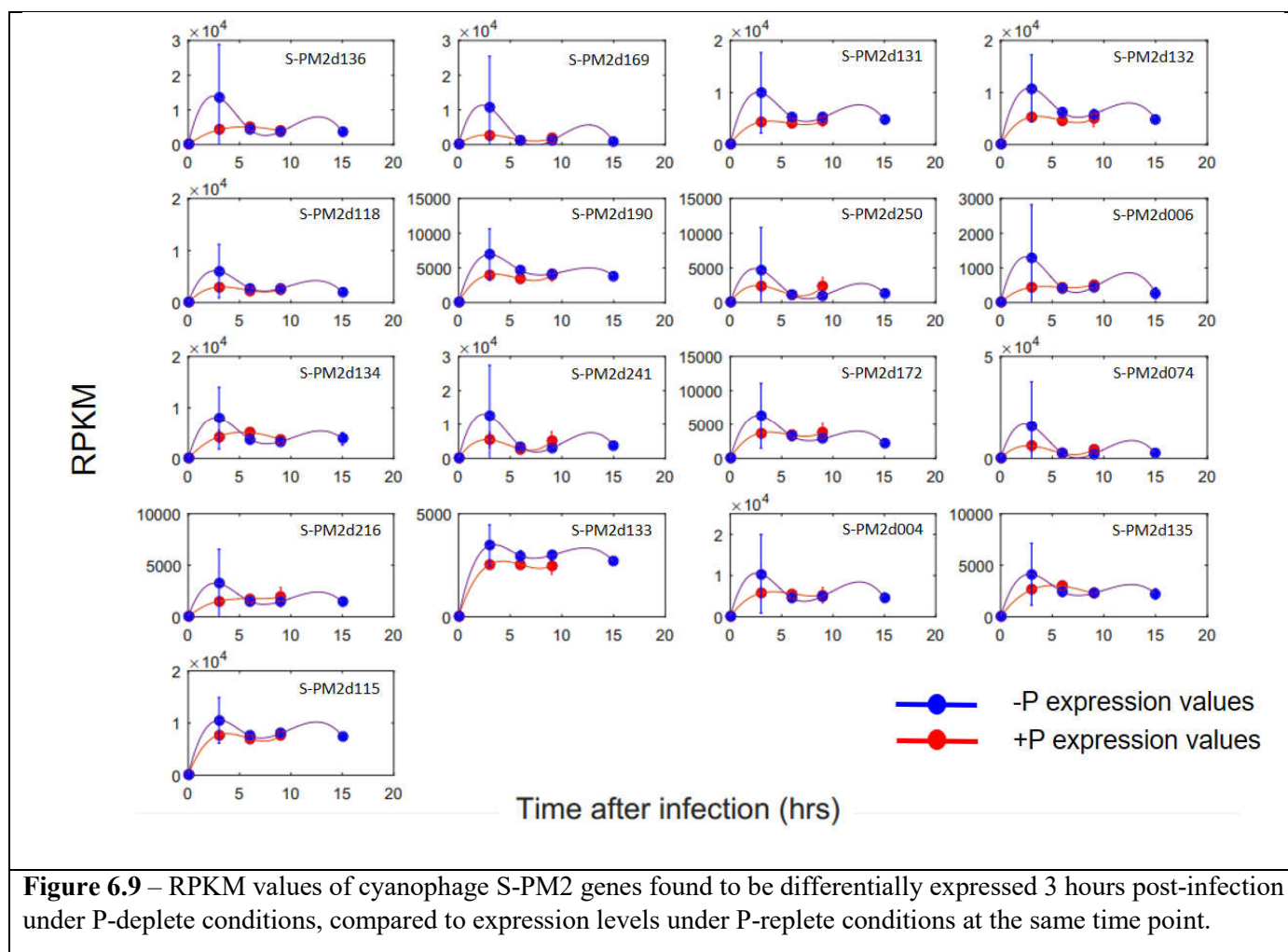
CD	Product	Gene	Start codon position	Stop codon position	Strand	log2 Fold Change	p-value	FDR-adjusted p-value
S-PM2d169	hypothetical protein		114418	114768	+	1.42	1.84E-05	2.11E-03
S-PM2d136	Pyrophosphatase	*mazG*	92757	93164	+	1.33	1.11E-05	2.11E-03
S-PM2d131	hypothetical protein		87134	87601	+	1.12	5.76E-05	4.40E-03
S-PM2d006	hypothetical protein		1913	2116	+	1.12	6.28E-04	1.60E-02
S-PM2d132	hypothetical protein		87598	87849	+	1.04	1.18E-04	5.86E-03
S-PM2d074	hypothetical protein		14226	14432	-	1.03	1.74E-03	3.06E-02
S-PM2d241	hypothetical protein		14805	15005	+	0.99	1.51E-03	2.88E-02
S-PM2d118	recombination endonuclease subunit	gp46	77014	78744	+	0.99	1.61E-04	6.15E-03
S-PM2d216	hypothetical protein		155436	155594	+	0.91	2.01E-03	3.29E-02
S-PM2d190	hypothetical protein		132902	133072	+	0.87	3.10E-04	1.01E-02
S-PM2d134	UvsX RecA-like recombination protein	UvsX RecA-like	90352	91383	+	0.84	1.21E-03	2.78E-02
S-PM2d004	hypothetical protein		1255	1560	+	0.79	3.06E-03	4.36E-02
S-PM2d172	High Light Inducible Proteins	hli2	115295	115414	+	0.76	1.41E-03	2.88E-02
S-PM2d250	putative ATPase		155233	155376	+	0.71	3.19E-02	1.61E-01
S-PM2d135	DNA primase-helicase	gp41	91343	92755	+	0.67	3.24E-03	4.36E-02
S-PM2d133	DNA polymerase	gp43	87849	90341	+	0.63	2.62E-03	3.99E-02
S-PM2d115	hypothetical protein		75481	75699	+	0.63	4.18E-03	5.03E-02

Table 6.3 – Cyanophage S-PM2 genes differentially expressed 3 hours post-infection of *Synechococcus* sp. WH7803 under P-deplete conditions compared to the gene expression levels under +P conditions.

There are several interesting observations that can be made based on the identity of the S-PM2 genes that were found to be differentially expressed under P-deplete conditions compared to P-replete conditions 3 hours post-infection. Thus, genes S-PM2d131-S-PM2d136 appear to be organised in an operon, and with some of the genes in this operon possessing bacterial homologues involved in DNA replication e.g. gp43 – encoding viral DNA polymerase (Petrov and Karam, 2004), gp41 – encoding a putative DNA primase-helicase involved in DNA replication initiation (Dong, *et al.*, 1996), UvsX-RecA like recombination protein – possibly involved in the repair of errors created during DNA replication (Ando and Morrical, 1998), and several hypothetical proteins of unknown function.

Furthermore, at the end of this operon is the cyanophage-encoded *mazG* gene. As described extensively in Chapter 5 of this thesis, *mazG* encodes an active pyrophosphatase with potential activity in aberrant nucleotide removal (Lu, *et al.*, 2010). This is consistent with a potential role for MazG in viral DNA replication and with *mazG* being up-regulated following cyanophage infection of a P-deplete host.

6.4.2.4 PhoB regulation of the cyanophage S-PM2d131-d136 gene cluster?

In previous work, cyanophage genes that were found to be differentially expressed under P-stress conditions, were shown to possess upstream located binding sites for PhoB, the transcriptional regulator involved in the P-stress cellular response in cyanobacteria (Lin, *et al.*, 2016). In some cases, these Pho-boxes were shown to be bound by the host-encoded PhoB protein (Zeng and Chisholm, 2012) and the expression level of genes found downstream from the binding site was shown to be elevated during cyanophage infection under P deplete conditions (Lin, *et al.*, 2016). In order to examine whether such putative PhoB binding sites exist in the promoter regions of those genes found to be differentially expressed under P-deplete conditions in this work, upstream sequences were examined using the Pattern Locator script (Mrazek and Xie, 2006). The consensus sequence previously identified as a putative Pho box binding site in cyanobacteria (Su, *et al.*, 2007) was used to scan the upstream region of these S-PM2 genes and several potential binding sites were found (see Table 6.4). The list of genes examined for the presence of putative Pho boxes included S-PM2d130. This

cyanophage gene was not found to be differentially expressed during infection under P-deplete conditions. However, S-PM2d130 is found directly upstream of the potentially up-regulated operon S-PM2d131-d136. Moreover, this gene contains a putative Pho box at a relevant distance (31bp upstream) of the start of translation of the protein, hinting that this gene is also transcribed as part of the up-regulated S-PM2d131-d136 operon.

ORF	Gene annotation	Putative Pho Box	Position
S-PM2d136	Pyrophosphatase (m*azG*)	**CTTTTGAA***GAC***CTTTCAAC***TCG***CCAAAA**GC	-63
S-PM2d169	hypothetical protein	NA	
S-PM2d131	hypothetical protein	**TTTGAAGAT***GGT***CTTTTGAA***GA*T**CTCACTGG**	-138
S-PM2d132	hypothetical protein	**CTAATCCA***AAC***CTTGAT**GC	-91
S-PM2d118	Recombination endonuclease subunit (gp46)	NA	NA
S-PM2d190	hypothetical protein	NA	NA
S-PM2d250	putative ATPase	**CCTATACT***GAT***CTCAGTTC**	-30
S-PM2d006	hypothetical protein	NA	
S-PM2d134	UvsX RecA-like	NA	
S-PM2d241	hypothetical protein	**CTTCCCCG***CCG***CTGAGGTG**	-105
S-PM2d172	hli2	NA	NA
S-PM2d074	hypothetical protein	NA	NA
S-PM2d216	hypothetical protein	NA	NA
S-PM2d133	DNA polymerase (gp43)	**GTTAAAAT***ATC***CATATGCA**	-162
S-PM2d004	hypothetical protein	**CTTTCCTC***TCG***TTTAACTG** **ATTATACA**ATT**CTTGGAGA**	-62 and -25
S-PM2d135	DNA primase-helicase (gp41)	NA	NA
S-PM2d115	hypothetical protein	NA	NA
S-PM2d130	Heat shock protein (Hsp20)	**A***TT*C**AAAC**TCG**CTTAAATA**	-31

Table 6.4 –Sequence and position of putative Pho box motifs found upstream of cyanophage S-PM2 genes differentially expressed during infection of a P-deplete **Synechococcus** host. Bold nucleotides represent residues comprising part of the pho box; underlined nucleotides correspond to the recognised consensus sequence of the Pho box binding site (5'-PyTTAAPyPyT/A-3') (Su*, et al.*, 2007). NA – no putative Pho box found. Position is expressed as the number of nucleotides upstream of the putative start of translation for each protein.

6.5 Discussion

As discussed in Chapter 3 of this thesis, one of the main hallmarks of S-PM2 infection of a P-deplete *Synechococcus* host is the process of "pseudolysogeny" – in which the infection cycle extends beyond the normal timescale of the lytic cycle and without any apparent ability of the phage to integrate into the host genome. Previous work has shown that the delay in lysis and reduction in burst size associated with pseudolysogeny are not due to a reduced ability of the infecting phage to adhere to the host, nor to the host entering the stationary phase of growth. Rather, at least in the case of S-PM2 infection of *Synechococcus* sp. WH7803, this process appears directly linked to the lack of available phosphorus (Wilson*, et al.*, 1996). Indeed, this 'phosphorus effect' on viral replication kinetics has now been documented both in other cyanobacteria, specifically *Prochlorococcus* (Zeng and Chisholm, 2012; Lin*, et al.*, 2016) as well as in the eukaryotic phytoplankter *Micromonas pusilla* (Maat*, et al.*, 2014; Maat and Brussaard, 2016; Maat*, et al.*, 2016). However, no studies have thus far been performed to investigate the molecular mechanism(s) regulating this lysis delay.

As was shown in Chapter 3, the possession of host orthologues of P-stress related genes may provide cyanophage with a competitive advantage in P-deplete environments. The dramatic impact on viral replication kinetics of a lack of phosphorus in the medium, on those cyanophage that do not possess P-stress related genes, points towards a distinct molecular mechanism, either controlled by the host or the phage, by which the infection process is delayed and the burst size reduced. From the host perspective, this mechanism would potentially make a P-deplete host less "attractive" to the phosphorus-demanding phage allowing persistence of the host population throughout periods of P-scarcity in the dynamic marine environment. From the cyanophage perspective, the pseudolysogeny process would be an alternative to undertaking an abortive infection pathway– and even though the infection is significantly affected and sub-optimal, the low number of viral progeny produced would still allow phage persistence, facilitating future host infection once the depleted nutrient(s), in this case P, became available again.

In order to decipher the molecular mechanism underlying the pseudolysogeny process, in this chapter an RNAseq experiment was performed using cyanophage S-PM2 (which lacks known P-stress related genes) and *Synechococcus* sp. WH7803 as the model system. One of the ways to interpret the phage RNA expression profiles produced is to divide them into temporal clusters, since phages have a well-defined and extensively studied temporal program of gene expression (Luke*, et al.*, 2002; Mosig and Eiserling, 2006; Clokie*, et al.*, 2006; Doron*, et al.*, 2016). Accordingly, genes can be roughly grouped into two clusters – 'early' and 'late' genes, where early genes usually encode for proteins involved in viral DNA replication and the

interference of host metabolism, while the 'late' genes encode proteins that are necessary for virion assembly and lysis of the infected host cell.

In the RNAseq experiment performed here (see section 6.4.2), sampling points were chosen based on the need to achieve a compromise between extensive sampling (due to cost) and the fact that in order to ensure infection of the majority of cells in each sample, the infection needed to occur at a high MOI. This in turn required accumulation of a large stock of purified phage prior to initiation of the experiment. Furthermore, in order to accurately compare the molecular processes occurring in P-deplete and P-replete infected cultures, the same time points needed to be used when sampling both cultures. In order to achieve this compromise, preliminary qPCR and flow cytometry data were used to choose the most suitable time points for sampling (see Figs 6.2 and 6.3). This resulted in choosing the 3 hours post-infection sample as the first sampling point. While this allowed extensive sampling of the infection cycle under P-deplete conditions, it meant that an earlier time point (i.e. 1 hour post-infection) was not taken. Unfortunately this led us to 'miss' the maximum expression levels of the cyanophage 'early' genes and caused problems with clustering of the infected P-replete culture. In order to examine the congruence of clustering for both the P-replete and P-deplete samples with previously described work, a comparison between genes clustering into each of the 'early' and 'late' groups was performed. Currently, there are only two studies which predict the assignment of S-PM2 genes into temporal clusters. The first one was based on a bioinformatic prediction of promoters of S-PM2 genes (Mann, *et al.*, 2005) and the other was based on an RNAseq experiment during S-PM2 infection of *Synechococcus* sp. WH7803 under high- and low-light conditions (Puxty, 2014). A comparison of these approaches with the RNAseq data produced in this study, can be seen in Table 6.5.

	+P	-P, T0-T9	-P, T0-T15, 2 clusters	-P, T0-T15, 3 clusters
Promoter prediction – Early[1]	0.20000	0.25000	0.25000	0.20000
Promoter prediction – Late[1]	0.80645	0.83871	0.83871	0.64516
Promoter prediction – Total[1]	0.56863	0.607843	0.60784	0.47059
RNAseq - Early[2]	0.24138	0.58621	0.28736	0.36782
RNAseq- Late[2]	0.68085	0.92553	0.94681	0.85106
RNAseq - Total[2]	0.57333	0.78667	0.76000	0.46698
RNAseq -T1 Early[3]	0.235955	0.573034	0.28090	0.35955
RNAseq -T1 Late[3]	0.695652	0.945652	0.96740	0.86957
RNAseq -T1 Total[3]	0.66667	0.81333	0.78667	0.49057

Table 6.5: Comparison of cluster prediction success compared to previously published predictions. [1] – comparison to clustering based on promoter motif prediction (Mann, *et al.*, 2005) [2] – comparison to clustering predicted in (Puxty, 2014), including all the time points. [3] – same data as in [2], with the omission of 1-hour post-infection time point. Values are between 0 (no congruency) and 1 (identical clusters).

As can be seen from this comparison, the omission of a sampling time point prior to 3 hours greatly affects the ability of the data produced in this experiment to accurately predict the 'early' cluster genes. In all of the comparisons, 'late' gene prediction is much more congruent with previous predictions than the 'early' cluster prediction. Furthermore, removing the 1 hour post-infection time point from the data set produced in (Puxty, 2014), does not seem to improve the ability to correctly classify genes as 'early'. However, it does improve the classification of genes into the 'late' cluster, thus generally improving the clustering accuracy.

An additional observation from this comparison is that the prediction of both 'early' and 'late' clusters is significantly better for P-deplete samples than for P-replete samples (57.6%-66.7% congruency for P-deplete samples, compared to 78.67%-81.33% congruency for P-deplete (T0-T9) samples). Additionally, when analysing the clustering congruency for the complete set of P-deplete samples (T0-T15), the 'late' cluster prediction further improves (92.55%-94.57% for -P, T0-T9 compared to 94.68%-96.74% for -P, T0-T9 2 clusters). Forcing the third cluster model onto the P-deplete samples makes the clustering prediction less congruent with the previous models.

These observations point towards S-PM2 infection under P-deplete conditions occurring with a similar temporal pattern to the P-replete infection process, but with this temporal gene expression program delayed, presumably due to a lack of P required for transcription. Better 'early' gene cluster prediction in the P-deplete samples, suggests that the 'early' genes are transcribed around 3-hours post infection, as opposed to ca. 1-hour post infection in the P-replete samples. An additional observation that supports this explanation is the relatively lower abundance of viral RNA transcripts under P-deplete conditions compared to P-replete conditions (Figure 6.4). It seems likely that RNA transcription is both delayed and reduced in magnitude during P stress conditions, possibly to favour an alternative 'molecular process' that optimises phage yield under nutrient limitation conditions.

The differential gene expression data (Fig. 6.10; Table 6.3) provides further hints of the molecular processes occurring in the initial window post infection. Comparison of equivalent time points between infection under P-replete and P-deplete conditions, showed the only statistically significant differences in gene expression occurred 3 hours post infection. This finding lends itself to the hypothesis that the RNA transcription program during infection under P-deplete conditions is indeed changed, but only at the initial stages. The identity of genes that are differentially expressed during infection under P-deplete conditions 3 hours post-infection (Table 6.3) hints at the prioritisation of transcription of those genes involved in

the DNA replication process. Further support for this conclusion can be found in the qPCR data. Thus, while the general yield of viral DNA under P-deplete conditions is lower than that produced under P-replete conditions (see Figure 6.2A, Figure 3.3; Wilson*, et al.*, 1996), the rate of DNA replication seems to be similar between both conditions (Figure 6.2B). It is possible that by maintaining the same cyanophage DNA replication rate under both conditions, that during P stress cyanophage compensate for a lack of P resources by up-regulating the enzymatic machinery involved in DNA replication. Interestingly, no other genes seem to be significantly differentially expressed at other time points, suggesting that the main difference during infection of a P-deplete host is in the DNA replication process.

Given that *Synechococcus* host RNA levels are greatly and significantly reduced in both P-replete and P-deplete infected samples (see Figure 6.5; Table 6.2), with no significant differences between both conditions, suggests irrespective of nutrient conditions cyanophages attempt to redirect all available resources, in the shape of inorganic phosphate, nucleotides, energy, etc., preferentially towards viral transcription and/or DNA replication.

Previously, regulation of some of the P-stress related cyanophage AMGs was shown to be controlled by the host transcription factor PhoB (Zeng and Chisholm, 2012). In order to investigate whether this scenario could also be involved in the increased transcription levels of genes identified in this study, a bioinformatics analysis of promoter regions of differentially expressed ORFs was performed. Several of the genes found to be over-expressed seem to possess a putative Pho box in their promoter region (Table 6.4), with varying degree of conformity to the consensus sequence as identified in (Su*, et al.*, 2007). The variability in predicted Pho boxes stems both from the level to which the putative binding sites adhere to the predicted consensus at the nucleotide sequence level, as well as in the number of repeated consensus motifs predicted in the promoter area. ORFs S-PM2d136 and S-PM2d131 have 3 predicted repeats of the consensus binding motif, while the remaining ORFs possessing a predicted Pho box in their promoter, have only 2 repeats. While promoters with both variations of the number of binding sites were identified as active previously (Zeng and Chisholm, 2012), this variability could be a sign of differential regulation, or merely a result of false positive motif identification, an issue which should be resolved in future experiments.

Interestingly, one of the cyanophage S-PM2 genes that was up-regulated during infection of *Synechococcus* sp. WH7803 under P-deplete conditions was *mazG*. As was extensively discussed in Chapter 5, *mazG* in bacteria encodes a pyrophosphatase which plays a regulatory role in the SOS response (Gross*, et al.*, 2006; Lee*, et al.*, 2008). There are several potential roles that the cyanophage-encoded *mazG* could be playing during viral adaptation to infection under P-deplete conditions as follows:

- **The stringent response** – as discussed in Chapter 5, *mazG* plays a role in the stringent response reducing cellular levels of the alarmone (p)ppGpp. P-stress can induce the stringent response in bacteria (Braeken*, et al.*, 2006) and (p)ppGpp accumulation has been linked to the ability to form intracellular poly-P bodies (Rao*, et al.*, 1998). Upon infection under stringent response conditions, bacteriophage DNA replication has been shown to be disproportionately negatively affected (Nowicki*, et al.*, 2013b) and intracellular levels of (p)ppGpp have been shown to be significantly reduced following phage infection (Borbély*, et al.*, 1980).

- **Aberrant nucleotide removal** – previous work has shown that the MazG protein could have a potential role in removing non-canonical nucleotides (Lu*, et al.*, 2010) thus preventing oxidative stress-induced mutations. It is possible that the increase in transcription levels of genes involved in viral DNA replication observed during phage infection of a P-deplete *Synechococcus* host in work described in this chapter, requires an increased response rate of mutation-preventing mechanisms in which viral MazG could play an active role.

- **Phosphate supply** – work outlined in chapter 5, has shown that the cyanophage MazG possesses pyrophosphatase activity capable of hydrolysing orthophosphate groups from (d)NTPs and making them available for other metabolic reactions. Increased expression of the cyanophage *mazG* gene could be an attempt by the phage to provide additional phosphate for its needs. This possibility seems counterintuitive to what is currently known about the metabolic requirements of infecting cyanophages and the ways in which they alter infected host metabolism (Puxty*, et al.*, 2015a; Thompson*, et al.*, 2011). Since the carbon metabolism reactions of infected hosts seem to be redirected towards nucleotide and energy production, it seems counterproductive for the phage to produce a pyrophosphatase which will reduce the pool of available nucleotides in the cell. However, as was shown in Chapter 5, viral MazG has a wide range of affinities towards different nucleotides, in a way which can diverge from the affinity patterns of the host MazG orthologue (Figure 5.12) and these differences in affinity for the various substrates could serve as a regulator of the hydrolytic activity of the cyanophage MazG.

Another cyanophage gene that was found to be up-regulated under P-deplete condition was GP46 (S-PM2d118). This gene is putatively annotated as a recombination endonuclease subunit and as such could have a potential role in DNA metabolism. However, examination of the structural domain similarity between the protein sequence encoded by this gene and PstB – a component of the ABC transporter responsible for import of inorganic phosphate, shows similarity in the structural domain content, in spite of amino acid sequence divergence between the two proteins (Figure 6.10).

Figure 6.10. Putative active domains in the cyanophage S-PM2 GP46 and cyanobacterial PstB proteins, the latter encoding part of the ABC phosphate transporter.
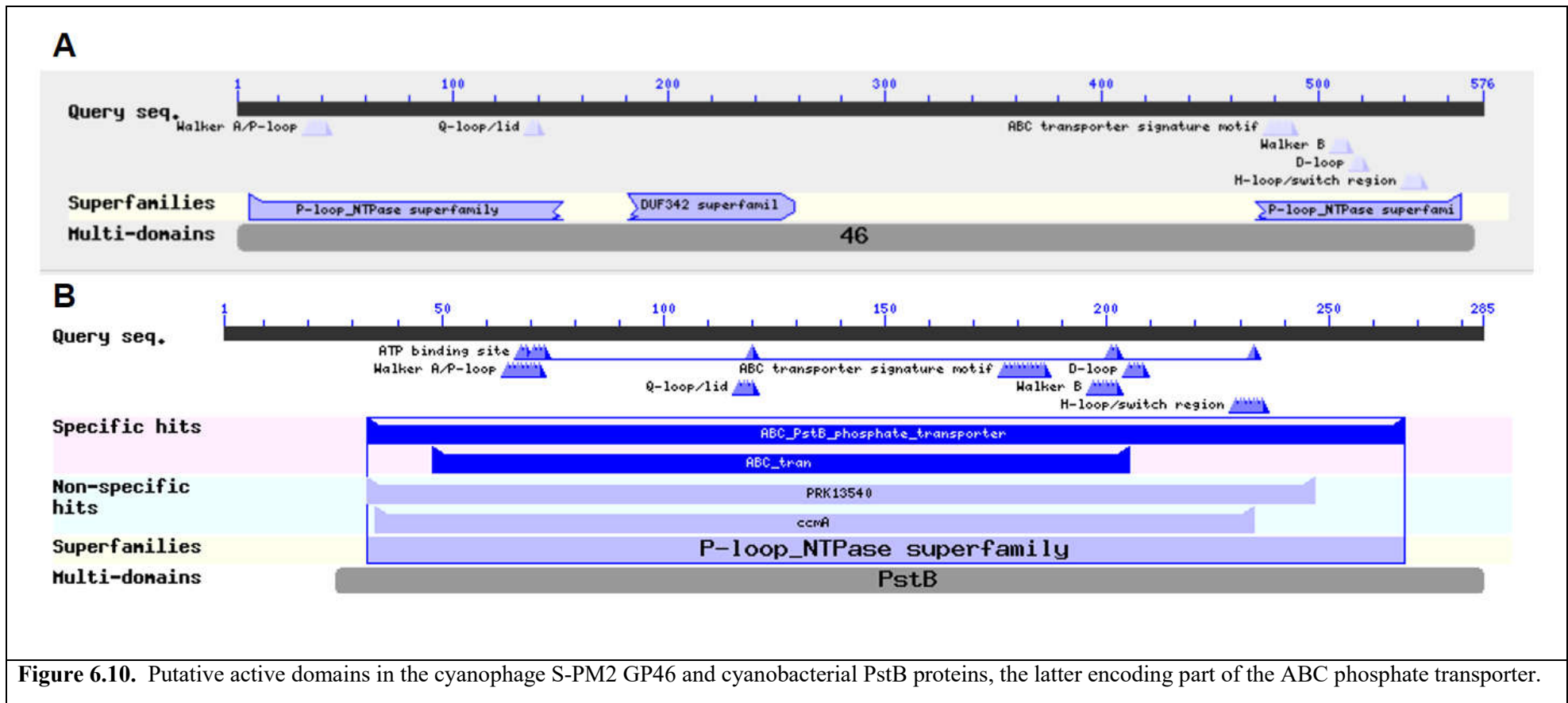
Both proteins contain a putative P-loop NTPase domain, an ABC transporter signature motif, a D0-loop, Walker -B, N-loop switch region, as well as Q-loop/lid and Walker A/P loop domains. Since the function of GP46 is not known, it is possible that the protein has been mis-annotated based on sequence similarity and its cellular function should be further examined.

Finally, there are obvious differences in host transcription levels between infected and uninfected samples under both P-replete and P-deplete conditions. It has previously been shown that upon infection, cyanophage-induced RNase E activity selectively degrades host RNA (Stazic, *et al.*, 2016). The levels of host RNA in this study were shown to be substantially decreased, compared to uninfected samples (Figure 6.5). However, further work is required to investigate differential expression of host genes during the infection process under both P-replete and P-deplete conditions. Of particular interest would be changes to host gene expression levels that are a reflection of the interference of cyanophage gene products (such as MazG) in host metabolism and the transcriptional landscape.

Chapter 7:

Conclusions and future directions

Work in this thesis addressed two main aims:

1. To understand the molecular processes that underlie the observed phenomenon of 'pseudolysogeny', a process which occurs during infection of *Synechococcus* sp. WH7803 by cyanomyovirus S-PM2 under P-deplete conditions.

2. To investigate the activity and potential physiological function of both the cyanophage S-PM2 and *Synechococcus* sp. WH7803 host orthologues of the pyrophosphatase MazG.

Previous work investigating the dynamics of infection of cyanophage S-PM2 under P-deplete conditions has shown that P-stress exerts a significant effect on the infection process. It was also apparent that this effect was not universal, with some cyanophage not being 'affected' by P-deficiency (Wilson*, et al.*, 1996). After confirming this observation using several independent methods (Chapter 3, Figure 3.3, Figure 3.4 and Chapter 6, Figure 6.2 and Figure 6.3), my first goal was to try and understand whether genetic differences between individual cyanophages was responsible for this differential response to P-depletion.

Previously published work has shown that some phages possess host-like genes, dubbed auxiliary metabolic genes (AMGs), which help the infecting phage 'take over' host metabolism and harness it towards maximising viral production during the infection cycle (Lindell*, et al.*, 2007; Millard*, et al.*, 2009). Some of these cyanophage AMGs may have a potential role in phosphate metabolism, and I speculated that possession of these genes could be correlated with the ability of the phage to successfully infect its host under P-deplete conditions. Out of three potential P-metabolism related orthologues (*phoH, phoA* and *pstS)*, I chose to focus on *pstS* due to its high similarity to the corresponding host orthologue and its relatively well characterised expression patterns and function (Scanlan*, et al.*, 1993). Additional studies have shown that the presence and expression levels of both the viral and bacterial orthologue of *pstS* are well correlated with the phosphorus status of the oceanic niche from which they were isolated (Fuller*, et al.*, 2005; Kelly*, et al.*, 2013). Furthermore, RNA sequencing following cyanophage infection of a closely related cyanobacterium *Prochlorococcus*, under P-deplete conditions, revealed that the phage *pstS* orthologue was one of the only genes to be differentially expressed, compared to P-replete conditions (Lin *et al*., 2016).

In agreement with these previous findings, my experiments have shown that there is a correlation between the presence of *pstS* in the cyanophage genome and its ability to avoid P-

stress induced 'pseudolysogeny' during the course of infection (Chapter 3). All the tested cyanophages that were found to lack *pstS* were subjected to significant latent phase prolongation and burst size reduction. These findings may have significant implications for estimates of marine primary production and the effect that viral infection plays on the biogeochemical cycling of carbon and other nutrients, particularly in large nutrient depleted areas of the world's oceans. In light of the results presented in this thesis, the presence of viral *pstS* orthologues in environmental metagenomes could be used as a way to amend estimates of primary production in P-depleted areas of the ocean in the infected microbial population fraction.

In light of these results, I hypothesised that if we isolated cyanophages from P-deplete oceanic environments a significant fraction of these isolates would contain P-stress related AMGs in their genomes. This prediction was supported via the previous isolation of *Prochlorococcus* cyanophage P-RSM4, containing *pstS*, which was isolated in September after months of stratification from the Gulf of Aqaba, Red Sea (Sullivan*, et al.*, 2010). This specific marine environment undergoes seasonal cycles of summer water column stratification and deep winter mixing, resulting in nutrient depletion of the photic layer during the summer stratified months (Lindell and Post, 1995). Due to the above, I examined the collection of cyanomyoviruses that was previously isolated in our lab (Millard and Mann, 2006) over the course of a whole year from the Gulf of Aqaba. In order to confirm this hypothesis, I focused on cyanophage isolated on dates in which water column stratification was thought to be the most pronounced i.e. during a period when the chances of isolating cyanophage which have adapted to infecting a nutrient-stressed host would be the highest. Surprisingly, none of the tested phages possessed *pstS* in their genomes, nor were they capable of evading the P-depletion induced pseudolysogeny response (Chapter 3, Table 3.5). A possible reason for this result is that the host (*Synechococcus* sp. WH7803) on which these cyanophages were isolated is not a member of the dominant *Synechococcus* clade found in the Gulf of Aqaba during the period of isolation. Therefore, the profile of cyanophages isolated may not be representative of the dominant cyanophage community during these months. Indeed, previous investigation of *Synechococcus* ecotypes dominating the photic zone over different seasons in the Gulf of Aqaba, shows that clade V, to which *Synechococcus* WH7803 belongs, is found in minimal, almost un-detectable numbers over the summer months of water column stratification (Post*, et al.*, 2011). In order to really test the idea that the presence of P-related AMGs in cyanophage genomes correlates with the prevailing nutrient conditions in the water column at the time of isolation, cyanophages should be isolated on a host belonging to the dominant clade of *Synechococcus* occurring at that time. Previous work shows that during the summer stratified months, there is a dramatic increase in clade III, and less so of clade II ecotypes (Post*, et al.*,

2011). Therefore, more appropriate hosts for cyanophage isolation would be *Synechococcus* sp. WH8102 (clade III) or *Synechococcus* sp. RS9902 (clade II).

Whole genome sequencing of some of the cyanophage that were used in the above experiments (Chapter 4), confirmed the presence of P-stress related genes in their genomes but also pointed out another interesting finding that needs to be taken into consideration when performing environmental phage isolation. This was that the presence of a specific gene in a phage lysate does not mean that the lysate is monoclonal, i.e. contains only a single phage. Thus, PCR using degenerate primers suggested that cyanophage S-BM3 possesses *pstS* in its genome (Figure 3.2), whilst one-step experiments showed this cyanophage could evade pseudolysogeny under P-deplete conditions (Table 3.5). However, subsequent whole genome sequencing showed that S-BM3 does not contain any of the P-stress related genes in its genome. One possible explanation to this discrepancy is that the lysate from which it was produced contained more than a single phage, perhaps many. Whilst PCR confirmed the presence of *pstS* from one cyanophage, possibly the same one that successfully infected *Synechococcus* under P-deplete conditions and evaded pseudolysogeny, it is possible that the lysate used for whole genome sequencing (propagated following infection of a P-replete host) selected for production of a different phage, one that is not adapted to infection under P-deplete conditions.

This finding suggests that cyanophage isolation from environmental samples is more complex that is assumed using current isolation methods – i.e. that three times single plaque purification may be insufficient to ensure a clonal phage population is produced and there is the need for examining additional ways of ensuring isolated phage contain a single 'strain'. One example of such method improvement could be to infect several potential hosts with the same isolated phage strain, under a number of environmental conditions and perform whole genome sequencing of each isolate produced under different conditions and on different hosts.

Clearly, chapters 3 and 4 of this thesis extend previous work describing the changing kinetic parameters of cyanophage infection under P-deplete conditions, and begin to link this with the possession of specific AMGs in cyanophage that either can, or cannot, overcome the pseudolysogeny process. However, such work is really focused from the cyanophage perspective and fails to address host responses. This is important since recent work has shown that cyanophage infection has a significant effect on host *Synechococcus* CO_2 fixation capacity during infection under nutrient replete conditions (Puxty*, et al.*, 2016). However, nothing is known about how this CO_2 fixation capacity changes following cyanophage infection during nutrient-deplete growth. Such work would potentially bring even greater accuracy to current models of marine CO_2 fixation, as well as the general biogeochemical cycling of carbon in the ocean.

In order to understand the transcriptional landscape of the observed phenomenon of 'pseudolysogeny', occurring during cyanophage infection under P-deplete conditions, I carried out RNA sequencing of total cyanophage and host mRNAs over the whole infection cycle. The results of this experiment (chapter 6) showed no significant changes to the cyanophage transcriptional program that result in a prolonged infection cycle or reduced burst size. However, during the initial stage of infection, several cyanophage genes, e.g. gp41, gp43 and gp46 potentially involved in DNA replication, are over-expressed under P-deplete conditions, compared to the P-replete control. DNA replication has previously been identified as the bottleneck stage in the phage infection process, one that appears to be most affected during infection of 'stationary phase' cultures (Bryan, *et al.*, 2016), so the findings I report here are entirely consistent within what was previously known about phage infection under nutrient stress conditions. Subsequently however, there does not seem to be any statistically significant differentially expressed cyanophage genes between P-replete/deplete infected cultures at later time points. It seems likely that the delay in the latent phase and the observed reduction in burst size is not a transcriptionally controlled event, but rather a product of the simple lack of building blocks necessary for the production of all the required phage components to produce an optimal viral progeny in the next generation.

An additional observation supporting the delay in the latent period is the strengthening of the significance in temporal clustering of the P-deplete samples, in spite of the omission of an early time point covering the early gene cluster for the P-replete control. However, ideally it would be important to repeat the RNAseq experiment ensuring that a full range of time points is taken, covering all of the temporal clusters and possibly discovering new expression patterns. Future analysis should also investigate whether there are any significant differences in *Synechococcus* host gene expression under the P-replete/deplete experimental conditions described here.

Another interesting observation that arises from the RNAseq data presented here is that some of the cyanophage genes that were found to be over-expressed under P-deplete conditions (in a cyanophage undergoing pseudolysogeny) possess a putative PhoB-binding motif in their promoters. Previously, only cyanophage overcoming pseudolysogeny have been analysed and here those genes that were found to be over-expressed under P-deplete conditions i.e. *pstS* and *phoA* were also found to have Pho boxes in their promoters. Moreover, these binding motifs were shown to be bound by the host transcriptional regulator PhoB (Zeng and Chisholm, 2012). It would therefore be very interesting to perform gel-shift experiments with the promoters identified in this study and over-expressed *Synechococcus* PhoB protein, in order to confirm binding by this host-encoded transcriptional regulator.

In the latter part of my thesis (Chapter 5), I attempted to investigate the possible differences in enzymatic activity of the pyrophosphatase MazG, present in both cyanophage and *Synechococcus* sp. WH7803 host genomes. Comparison of the relative activities of the over-expressed host and cyanophage MazG proteins showed that the cyanophage protein possessed a significantly higher level of activity when incubated with dGTP compared to any other substrate (Figs. 5.10 and 5.11). This activity could stem from the structural similarity between dGTP and (p)ppGpp which is the predicted substrate of the host MazG and a potential signalling molecule in the *Synechococcus* stringent response. In addition to dGTP, the cyanophage orthologue also showed a higher activity with dCTP as a substrate, both compared to the host and to itself using other nucleotides. This difference could result from the relatively low GC content of cyanophage genomes, so that cyanophage would preferentially hydrolyse those nucleotides that it needs less off whilst at the same time providing phosphate for its energy needs. These hypotheses could be examined using radioactively labelled nucleotides as a sole phosphate source to the phage infected host and following the incorporation of labelled phosphate into the viral progeny produced in the infection.

During the course of my thesis I also examined the role of the *Synechococcus* MazG protein by constructing a *Synechococcus* sp. WH7803 mazG mutant strain using interposon mutagenesis (Fig. 5.4). While no significant difference in growth rates between the WT and *Synechococcus mazG* mutant were observed (Figure 5.5), there were differences in cyanophage infection dynamics under nutrient deplete growth, most noticeably during P-depletion where infection of the *mazG* mutant was slower and less productive (Figure. 5.7). This potentially suggests that the pyrophosphatase activity of *Synechococcus* MazG could provide the cyanophage with phosphate necessary for transcription and replication under –P-deplete conditions. Thus, it would be interesting to examine the potential changes in transcription levels of *Synechococcus mazG,* infected by S-PM2 under -P conditions Interestingly, one of the few S-PM2 genes that were over-expressed under P-deplete conditions, compared to the P-replete control, was the cyanophage *mazG* orthologue (S-PM2d136), providing further weight to this hypothesis.

In conclusion, this thesis provides a greatly improved mechanistic understanding of cyanophage infection dynamics during sub-optimal host growth conditions. Particularly, it highlights how the presence of specific AMGs relevant to a specific 'limiting condition' provides a significant advantage to the cyanophage which might otherwise be severely 'limited' by nutrient depletion. Understanding the ways in which phages hijack the metabolic processes of their host during the infection process, both via sequencing of novel phages and examining their genetic "arsenal" to identify those components that might aid a successfully

infection, as well as via molecular and biochemical characterisation of AMGs, is crucial to more completely understand how cyanophage affect the flow of the major macro- and micro-elements in marine systems and in turn how this affects global biogeochemical cycling, a critical facet given current global climate change.

References

Albi-Rodríguez, T. and Serrano, A. (2008). Identification of genes and functional characterization of proteins involved in polyphosphate metabolism in photosynthetic organisms. In: *Photosynthesis. Energy from the Sun. Proceedings of the 14th International Congress on Photosynthesis*, pp. 957-963: Springer, Heidelberg.

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L. and Beatson, S.A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *Bmc Genomics* 12pp.

Ammerman, J.W., Hood, R.R., Case, D.A. and Cotner, J.B. (2003). Phosphorus deficiency in the Atlantic: An emerging paradigm in oceanography. *Eos, Transactions American Geophysical Union* **84**(18), pp. 165-170.

Ando, R.A. and Morrical, S.W. (1998). Single-stranded DNA binding properties of the UvsX recombinase of bacteriophage T4: Binding parameters and effects of nucleotides. *Journal of Molecular Biology* **283**(4), pp. 785-796.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J.M., Mueller, J.E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C.A. and Rohwer, F. (2006). The marine viromes of four oceanic regions. *Plos Biology* **4**(11), pp. 2121-2131.

Avarbock, D., Avarbock, A. and Rubin, H. (2000). Differential regulation of opposing Rel(Mtb) activities by the aminoacylation state of a tRNA center ribosome center · mRNA center · Rel(Mtb) complex. *Biochemistry* **39**(38), pp. 11640-11648.

Bachhawat, P., Swapna, G. and Stock, A.M. (2005). Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active States. *Structure* **13**(9), pp. 1353-1363.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. and Pevzner, P.A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**(5), pp. 455-477.

Benitez-Nelson, C.R. (2000). The biogeochemical cycling of phosphorus in marine systems. *Earth-Science Reviews* **51**(1-4), pp. 109-135.

Bennett, E.M., Carpenter, S.R. and Caraco, N.F. (2001). Human impact on erodable phosphorus and eutrophication: A global perspective. *Bioscience* **51**(3), pp. 227-234.

Bertilsson, S., Berglund, O., Karl, D.M. and Chisholm, S.W. (2003). Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. *Limnology and Oceanography* **48**(5), pp. 1721-1731.

Bessey, O.A., Lowry, O.H. and Brock, M.J. (1946). A method for the rapid determination of alkaline phosphatase with five cubic millimeters of serum. *The Journal of Biological Chemistry* **164**, pp. 321-329.

Bohme, H. (1998). Regulation of nitrogen fixation in heterocyst-forming cyanobacteria. *Trends in Plant Science* **3**(9), pp. 346-351.

Borbély, G., Kari, C., Gulyás, A. and Farkas, G.L. (1980). Bacteriophage infection interferes with guanosine 3'-diphosphate-5'-diphosphate accumulation induced by energy and nitrogen starvation in the cyanobacterium *Anacystis nidulans*. *Journal of Bacteriology* 144(3), pp. 859–864.

Bouman, H.A., Ulloa, O., Scanlan, D.J., Zwirglmaier, K., Li, W.K.W., Platt, T., Stuart, V., Barlow, R., Leth, O., Clementson, L., Lutz, V., Fukasawa, M., Watanabe, S. and Sathyendranath, S. (2006). Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**(5775), pp. 918-921.

Braeken, K., Moris, M., Daniels, R., Vanderleyden, J. and Michiels, J. (2006). New horizons for (p)ppGpp in bacterial and plant physiology. *Trends in Microbiology* **14**(1), pp. 45-54.

Brahamsha, B. (1996). A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Applied and Environmental Microbiology* **62**(5), pp. 1747-1751.

Breitbart, M., Thompson, L.R., Suttle, C.A. and Sullivan, M.B. (2007). Exploring the vast diversity of marine viruses. *Oceanography* **20**(2), pp. 135-139.

Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J.M., Gorsky, G., Gregory, A.C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B.T., Schwenck, S.M., Speich, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Bork, P., Bowler, C., Sunagawa, S., Wincker, P., Karsenti, E., Sullivan, M.B. and Tara Oceans, C. (2015). Patterns and ecological drivers of ocean viral communities. *Science* **348**(6237), pp. 1261498

Bryan, D., El-Shibiny, A., Hobbs, Z., Porter, J. and Kutter, E.M. (2016). Bacteriophage T4 Infection of Stationary Phase *E. coli* : Life after Log from a Phage Perspective. *Frontiers in Microbiology* **7**pp.

Burkill, P.H., Leakey, R.J.G., Owens, N.J.P. and Mantoura, R.F.C. (1993). *Synechococcus* and its importance to the microbial foodweb of the Northwestern Indian Ocean. *Deep-Sea Research Part II-Topical Studies in Oceanography* **40**(3), pp. 773-782.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L. and Brussow, H. (2003). Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* **6**(4), pp. 417-424.

Capella-Gutierrez, S., Silla-Martinez, J.M. and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**(15), pp. 1972-1973.

Capone, D.G., Burns, J.A., Montoya, J.P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A.F. and Carpenter, E.J. (2005). Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles* **19**(2), pp.

Capone, D.G., Zehr, J.P., Paerl, H.W., Bergman, B. and Carpenter, E.J. (1997). *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**(5316), pp. 1221-1229.

Carini, P., Steindler, L., Beszteri, S. and Giovannoni, S.J. (2013). Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME Journal* **7**(3), pp. 592-602.

Carini, P., Van Mooy, B.A.S., Thrash, J.C., White, A., Zhao, Y., Campbell, E.O., Fredricks, H.F. and Giovannoni, S.J. (2015). SAR11 lipid renovation in response to phosphate starvation. *Proceedings of the National Academy of Sciences of the United States of America* **112**(25), pp. 7767-7772.

Carini, P., White, A.E., Campbell, E.O. and Giovannoni, S.J. (2014). Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. *Nature Communications* **5**, pp.

Carmichael, W.W. (2001). Health effects of toxin-producing cyanobacteria: "The CyanoHABs". *Human and Ecological Risk Assessment* **7**(5), pp. 1393-1407.

Casadaban, M.J. and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. *Journal of Molecular Biology* **138**(2), pp. 179-207.

Chan, J.Z.M., Millard, A.D., Mann, N.H. and Schafer, H. (2014). Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like Roseophage specific genes. *Frontiers in Microbiology* **5**, pp.

Chan, R.K., Botstein, D., Watanabe, T. and Ogata, Y. (1972). Specialized transduction of tetracycline resistance by phage P22 in *Salmonella typhimurium*. II. Properties of a high-frequency-transducing lysate. *Virology* **50**(3), pp. 883-898.

Chandler, J.W., Lin, Y.J., Gainer, P.J., Post, A.F., Johnson, Z.I. and Zinser, E.R. (2016). Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer. *Environmental Microbiology Reports* **8**(2), pp. 272-284.

Chatterji, D. and Ojha, A.K. (2001). Revisiting the stringent response, ppGpp and starvation signaling. *Current Opinion in Microbiology* **4**(2), pp. 160-165.

Chen, P.S., Toribara, T.Y. and Warner, H. (1956). Microdetermination of phosphorus. *Analytical Chemistry* **28**(11), pp. 1756–1758.

Chenard, C. and Suttle, C.A. (2008). Phylogenetic diversity of sequences of cyanophage photosynthetic gene *psbA* in marine and freshwaters. *Applied and Environmental Microbiology* **74**(17), pp. 5317-5324.

Chisholm, S.W., Frankel, S.L., Goericke, R., Olson, R.J., Palenik, B., Waterbury, J.B., Westjohnsrud, L. and Zettler, E.R. (1992). *Prochlorococcus marinus* Nov. Gen. Nov. Sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Archives of Microbiology* **157**(3), pp. 297-300.

Clokie, M.R.J., Shan, J.Y., Bailey, S., Jia, Y., Krisch, H.M., West, S. and Mann, N.H. (2006). Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environmental Microbiology* **8**(5), pp. 827-835.

Coale, K.H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S., Chavez, F.P., Ferioli, L., Sakamoto, C., Rogers, P., Millero, F., Steinberg, P., Nightingale, P., Cooper, D., Cochlan, W.P., Landry, M.R., Constantinou, J., Rollwagen, G., Trasvina, A. and Kudela, R. (1996). A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**(6600), pp. 495-501.

Costanzo, A. and Ades, S.E. (2006). Growth phase-dependent regulation of the extracytoplasmic stress factor, sigma(E) by guanosine 3 ',5 '-bispyrophosphate (ppGpp). *Journal of Bacteriology* **188**(13), pp. 4627-4634.

Cushing, D.H. (1989). A difference in structure between ecosystems in strongly stratified waters and in those that are only weakly stratified. *Journal of Plankton Research* **11**(1), pp. 1-13.

Dalmasso, M., Hill, C. and Ross, R.P. (2014). Exploiting gut bacteriophages for human health. *Trends in Microbiology* **22**(7), pp. 399-405.

De Smet, J., Zimmermann, M., Kogadeeva, M., Ceyssens, P.-J., Vermaelen, W., Blasdel, B., Jang, H.B., Sauer, U. and Lavigne, R. (2016). High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. *ISME Journal* pp.

Dekel-Bird, N.P., Sabehi, G., Mosevitzky, B. and Lindell, D. (2015). Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. *Environmental Microbiology* **17**(4), pp. 1286-1299.

Desplats, C., Dez, C., Tetart, F., Eleaume, H. and Krisch, H.M. (2002). Snapshot of the genome of the pseudo-T-even bacteriophage RB49. *Journal of Bacteriology* **184**(10), pp. 2789-2804.

Dong, F., Weitzel, S.E. and vonHippel, P.H. (1996). A coupled complex of T4 DNA replication helicase (gp41) and polymerase (gp43) can perform rapid and processive DNA strand-displacement synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **93**(25), pp. 14456-14461.

Doron, S., Fedida, A., Hernandez-Prieto, M.A., Sabehi, G., Karunker, I., Stazic, D., Feingersch, R., Steglich, C., Futschik, M., Lindell, D. and Sorek, R. (2016). Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME Journal* **10**(6), pp. 1437-1455.

Dufresne, A., Garczarek, L. and Partensky, F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biology* **6**(2), pp. 1

Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., Paulsen, I.T., de Marsac, N.T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A.F., Hess, W.R. and Partensky, F. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biology* **9**(5), pp. 1

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., Duprat, S., Galperin, M.Y., Koonin, E.V., Le Gall, F., Makarova, K.S., Ostrowski, M., Oztas, S., Robert, C., Rogozin, I.B., Scanlan, D.J., de Marsac, N.T., Weissenbach, J., Wincker, P., Wolf, Y.I. and Hess, W.R. (2003). Genome sequence of the cyanobacterium

Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences of the United States of America* **100**(17), pp. 10020-10025.

Duhaime, M.B., Wichels, A., Waldmann, J., Teeling, H. and Gloeckner, F.O. (2011). Ecogenomics and genome landscapes of marine Pseudoalteromonas phage H105/1. *ISME Journal* **5**(1), pp. 107-121.

Dyhrman, S.T., Chappell, P.D., Haley, S.T., Moffett, J.W., Orchard, E.D., Waterbury, J.B. and Webb, E.A. (2006). Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* **439**(7072), pp. 68-71.

d'Herelle, F.H. (1917). Sur une microbe invisible antagoniste des bacilles dysentérique. p. 373–375, vol. **165**, *C. R. Acad. Sci.* (Paris).

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, pp. 1-19.

Elhai, J. and Wolk, C.P. (1988). Conjugal transfer of DNA to cyanobacteria. *Methods in Enzymology* **167**, pp. 747-754.

Enav, H., Beja, O. and Mandel-Gutfreund, Y. (2012). Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts. *ISME Journal* **6**(3), pp.

English, B.P., Hauryliuk, V., Sanamrad, A., Tankov, S., Dekker, N.H. and Elf, J. (2011). Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proceedings of the National Academy of Sciences of the United States of America* **108**(31), pp. E365-E373.

Falkowski, P.G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. *Nature* 387(6630), pp. 272-275.

Farrant, G.K., Dore, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M., Pitt, F.D., Wincker, P., Scanlan, D.J., Iudicone, D., Acinas, S.G. and Garczarek, L. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* **113**(24), pp. E3365-E3374.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K., Lomas, M.W., Veneziano, D., Vera, C.S., Vrugt, J.A. and Martiny, A.C. (2013). Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America* (**24**), pp. 9824-9829.

Flores, E. and Herrero, A. (1994). Assimilatory nitrogen metabolism and its regulation. In: Bryant, D.A. *The Molecular Biology of Cyanobacteria*, pp. 487-517: Springer Netherlands.

Flores, E., López-Lozano, A. and Herrero, A. (2015). Nitrogen fixation in the oxygenic phototrophic prokaryotes (Cyanobacteria): the fight against oxygen. In: Bruijn, F.J.d. *Biological Nitrogen Fixation*, pp. 879-890. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Freifeld, D. and Meselson, M. (1970). Topological relationship of prophage lambda to the bacterial chromosome in lysogenic cells. *Proceedings of the National Academy of Sciences of the United States of America* **65**(1), pp. 200.

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, **399** (6736), pp. 541-548

Fuller, N.J., Wilson, W.H., Joint, I.R. and Mann, N.H. (1998). Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. *Applied Environmental Microbiology* **64**(6), pp. 2051-2060.

Fuller, N.J., West, N.J., Marie, D., Yallop, M., Rivlin, T., Post, A.F. and Scanlan, D.J. (2005). Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea*. Limnology and Oceanography* **50**(1), pp.

Galperin, M.Y., Moroz, O.V., Wilson, K.S. and Murzin, A.G. (2006). House cleaning, a part of good housekeeping. *Molecular Microbiology* **59**(1), pp. 5-19.

Garcia-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Gotz, S., Tarazona, S., Dopazo, J., Meyer, T.F. and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**(20), pp. 2678-2679.

Garcia-Fernandez, J.M., de Marsac, N.T. and Diez, J. (2004). Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. *Microbiology and Molecular Biology Reviews* **68**(4), pp. 630.

Garczarek, L., Dufresne, A., Blot, N., Cockshutt, A.M., Peyrat, A., Campbell, D.A., Joubin, L. and Six, C. (2008). Function and evolution of the *psbA* gene family in marine *Synechococcus*: *Synechococcus* sp. WH7803 as a case study. *ISME Journal* **2**(9), pp. 937-953.

Gentry, D.R., Hernandez, V.J., Nguyen, L.H., Jensen, D.B. and Cashel, M. (1993). Synthesis of the stationary-phase sigma-factor σS is positively regulated by ppGppSynthesis of the stationary-phase sigma-factor σS is positively regulated by ppGpp. *Journal of Bacteriology* **175**(24), pp. 7982-7989.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, M., Rappe, M.S., Short, J.M., Carrington, J.C.

and Mathur, E.J. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**(5738), pp. 1242-1245.

Golden, S.S. (1995). Light-responsive gene expression in cyanobacteria. *Journal of Bacteriology* **177**(7), pp. 1651-1654.

Goldsmith, D.B., Crosti, G., Dwivedi, B., McDaniel, L.D., Varsani, A., Suttle, C.A., Weinbauer, M.G., Sandaa, R.-A. and Breitbart, M. (2011). Development of *phoH* as a novel signature gene for assessing marine phage diversity. *Applied and Environmental Microbiology* **77**(21), pp.

Gomez-Garcia, M.R., Losada, M. and Serrano, A. (2003). Concurrent transcriptional activation of *ppa* and *ppx* genes by phosphate deprivation in the cyanobacterium Synechocystis sp strain PCC 6803. *Biochemical and Biophysical Research Communications* **302**(3), pp. 601-609.

Griese, M., Lange, C. and Soppa, J. (2011). Ploidy in cyanobacteria. *FEMS Microbiology Letters* **323**(2), pp. 124-131.

Gross, M., Marianovsky, I. and Glaser, G. (2006). MazG - a regulator of programmed cell death in Escherichia coli. *Molecular Microbiology* **59**(2), pp. 590-601.

Hagemann, M., Henneberg, M., Felde, V.J.M.N.L., Drahorad, S.L., Berkowicz, S.M., Felix-Henningsen, P. and Kaplan, A. (2015). Cyanobacterial diversity in biological soil crusts along a precipitation gradient, Northwest Negev desert, Israel. *Microbial Ecology* **70**(1), pp. 219-230.

Haseltin, W.A. and Block, R. (1973). Synthesis of guanosine tetraphosphate and pentaphosphate requires presence of a codon-specific, uncharged transfer ribonucleic-acid in acceptor site of ribosomes. *Proceedings of the National Academy of Sciences of the United States of America* **70**(5), pp. 1564-1568.

Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C. and Gourse, R.L. (2006). rRNA promoter regulation by nonoptimal binding of a region 1.2: An additional recognition element for RNA polymerase. *Cell* **125**(6), pp. 1069-1082.

Hauryliuk, V., Atkinson, G.C., Murakami, K.S., Tenson, T. and Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nature Reviews Microbiology* **13**(5), pp. 298-309.

Heldal, M., Scanlan, D.J., Norland, S., Thingstad, F. and Mann, N.H. (2003). Elemental composition of single cells of various strains of marine *Prochlorococcus* and *Synechococcus* using X-ray microanalysis. *Limnology and Ocean*ography **48**(5), pp. 1732-1743.

Henn, M.R., Sullivan, M.B., Stange-Thomann, N., Osburne, M.S., Berlin, A.M., Kelly, L., Yandava, C., Kodira, C., Zeng, Q., Weiand, M., Sparrow, T., Saif, S., Giannoukos, G., Young, S.K., Nusbaum, C., Birren, B.W. and Chisholm, S.W. (2010). Analysis of high-throughput sequencing and annotation strategies for phage genomes. *PLOS One* **5**(2), pp.

Hurwitz, B.L. and Sullivan, M.B. (2013). The Pacific Ocean Virome (POV): A Marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLOS One* **8**(2), pp.

Hurwitz, B.L. and U'Ren, J.M. (2016). Viral metabolic reprogramming in marine ecosystems. *Current Opinion in Microbiology* **31**, pp. 161–168.

Jardillier, L., Zubkov, M.V., Pearman, J. and Scanlan, D.J. (2010). Significant CO_2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME Journal* **4**(9), pp. 1180-1192.

Jiang, M., Sullivan, S.M., Wout, P.K. and Maddock, J.R. (2007). G-protein control of the ribosome-associated stress response protein SpoT. *Journal of Bacteriology* **189**(17), pp. 6140-6147.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S. and Chisholm, S.W. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**(5768), pp. 1737-1740.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**(9), pp. 1236-1240.

Jungblut, A.D., Lovejoy, C. and Vincent, W.F. (2010). Global distribution of cyanobacterial ecotypes in the cold biosphere. *ISME Journal* **4**(2), pp. 191-202.

Kang, I., Oh, H.-M., Kang, D. and Cho, J.-C. (2013). Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proceedings of the National Academy of Sciences of the United States of America* **110**(30), pp. 12343-12348.

Karl, D., Letelier, R., Tupas, L., Dore, J., Christian, J. and Hebel, D. (1997). The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**(6642), pp. 533-538.

Karl, D.M., Letelier, R., Hebel, D.V., Bird, D.F. and Winn, C.D. (1992). *Trichodesmium* blooms and new nitrogen in the North Pacific gyre. *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs* **362**, pp. 219-237.

Karpinets, T.V., Greenwood, D.J., Sams, C.E. and Ammons, J.T. (2006). RNA: protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biology* **4**, pp. 1-10

Kelly, L., Ding, H., Huang, K.H., Osburne, M.S. and Chisholm, S.W. (2013). Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME Journal* **7**(9), pp. 1827-1841.

Keren, N., Berg, A., VanKan, P.J.M., Levanon, H. and Ohad, I. (1997). Mechanism of photosystem II photoinactivation and D1 protein degradation at low light: The role of back electron flow. *Proceedings of the National Academy of Sciences of the United States of America* **94**(4), pp. 1579-1584.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G.M., Richardson, P. and Chisholm, S.W. (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLOS Genetics* **3**(12), pp. 2515-2528.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K. and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**(17), pp. 2283-2285.

Kokjohn, T.A., Sayler, G.S. and Miller, R.V. (1991). Attachment and replication of *Pseudomonas aeruginosa* bacteriophages under conditions simulating aquatic environments. *Journal of General Microbiology* **137**, pp. 661-666.

Koskella, B. and Brockhurst, M.A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews* **38**(5), pp. 916-931.

Krasny, L. and Gourse, R.L. (2004). An alternative strategy for bacterial ribosome synthesis: Bacillus subtilis rRNA transcription regulation. *EMBO Journal* **23**(22), pp. 4473-4483.

Krom, M.D., Herut, B. and Mantoura, R.F.C. (2004). Nutrient budget for the Eastern Mediterranean: Implications for phosphorus limitation. *Limnology and Oceanography* **49**(5), pp. 1582-1592.

Kromkamp, J., Vandenheuvel, A. and Mur, L.R. (1989). Phosphorus uptake and photosynthesis by phosphate-limited cultures of the cyanobacterium *Microcystis aeruginosa*. *British Phycological Journal* **24**(4), pp. 347-355.

Labry, C., Herbland, A. and Delmas, D. (2002). The role of phosphorus on planktonic production of the Gironde plume waters in the Bay of Biscay. *Journal of Plankton Research* **24**(2), pp. 97-117

Larkin, A.A., Blinebry, S.K., Howes, C., Lin, Y.J., Loftus, S.E., Schmaus, C.A., Zinser, E.R. and Johnson, Z.I. (2016). Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME Journal* **10**(7), pp. 1555-1567.

Laurie, A.D., Bernardo, L.M.D., Sze, C.C., Skarfstad, E., Szalewska-Palasz, A., Nystrom, T. and Shingler, V. (2003). The role of the alarmone (p)ppGpp in sigma(N) competition for core RNA polymerase. *Journal of Biological Chemistry* **278**(3), pp. 1494-1503.

Lee, S., Kim, M.H., Kang, B.S., Kim, J.-S., Kim, G.-H., Kim, Y.-G. and Kim, K.J. (2008). Crystal structure of *Escherichia coli* MazG, the regulator of nutritional stress response. *Journal of Biological Chemistry* **283**(22), pp. 15232-15240.

Lennon, J.T., Khatana, S.A.M., Marston, M.F. and Martiny, J.B.H. (2007). Is there a cost of virus resistance in marine cyanobacteria? *ISME Journal* **1**(4), pp. 300-312.

Levine, A., Vannier, F., Dehbi, M., Henckes, G. and Seror, S.J. (1991). The stringent response blocks DNA replication outside the ori region in Bacillus subtilis and at the origin in Escherichia coli. *Journal of Molecular Biology* **219**(4), pp. 605-613.

Lewis, M.R., Harrison, W.G., Oakey, N.S., Hebert, D. and Platt, T. (1986). Vertical nitrate fluxes in the oligotrophic ocean. *Science* **234**(4778), pp. 870-873.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. http://arxiv.org/pdf/1303.3997pp.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14), pp. 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data, P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16), pp. 2078-2079.

Li, W.K.W. (1998). Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnology and Oceanography* **43**(7), pp. 1746-1753.

Limor-Waisberg, K., Carmi, A., Scherz, A., Pilpel, Y. and Furman, I. (2013). The Role of tRNAs in Cyanophages. **15th International Conference on Photosynthesis**.

Lin, X., Ding, H. and Zeng, Q. (2016). Transcriptomic response during phage infection of a marine cyanobacterium under phosphorus-limited conditions. *Environmental Microbiology* (**18**), pp. 450-460.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., Hess, W.R., Church, G.M. and Chisholm, S.W. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**(7158), pp. 83-86.

Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M. and Chisholm, S.W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**(7064), pp. 86-89.

Lindell, D. and Post, A.F. (1995). Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat), Red Sea. *Limnology and Oceanography* **40**(6), pp. 1130-1141.

Linhart, C. and Shamir, R. (2005). The degenerate primer design problem: Theory and applications. *Journal of Computational Biology* **12**(4), pp. 431-456.

Love, M.I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**(12), pp.

Lu, L.-d., Sun, Q., Fan, X.-y., Zhong, Y., Yao, Y.-f. and Zhao, G.-P. (2010). Mycobacterial MazG is a novel NTP pyrophosphohydrolase involved in oxidative stress response. *Journal of Biological Chemistry* **285**(36), pp. 28076-28085.

Luke, K., Radek, A., Liu, X.P., Campbell, J., Uzan, M., Haselkorn, R. and Kogan, Y. (2002). Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* **299**(2), pp. 182-191.

Luo, Y.W., Doney, S.C., Anderson, L.A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Bostrom, K.H., Bottjer, D., Capone, D.G., Carpenter, E.J., Chen, Y.L., Church, M.J., Dore, J.E., Falcon, L.I., Fernandez, A., Foster, R.A., Furuya, K., Gomez, F., Gundersen, K., Hynes, A.M., Karl, D.M., Kitajima, S., Langlois, R.J., LaRoche, J., Letelier, R.M., Maranon, E., McGillicuddy, D.J., Moisander, P.H., Moore, C.M., Mourino-Carballido, B., Mulholland, M.R., Needoba, J.A., Orcutt, K.M., Poulton, A.J., Rahav, E., Raimbault, P., Rees, A.P., Riemann, L., Shiozaki, T., Subramaniam, A., Tyrrell, T., Turk-Kubo, K.A., Varela, M., Villareal, T.A., Webb, E.A., White, A.E., Wu, J. and Zehr, J.P. (2012). Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth System Science Data* **4**(1), pp. 47-73.

Maat, D.S. and Brussaard, C.P.D. (2016). Both phosphorus- and nitrogen limitation constrain viral proliferation in marine phytoplankton. *Aquatic Microbial Ecology* **77**(2), pp. 87-97.

Maat, D.S., Crawfurd, K.J., Timmermans, K.R. and Brussaard, C.P.D. (2014). Elevated CO_2 and phosphate limitation favor *Micromonas pusilla* through stimulated growth and reduced viral impact. *Applied and Environmental Microbiology* **80**(10), pp. 3119-3127.

Maat, D.S., van Bleijswijk, J.D.L., Witte, H.J. and Brussaard, C.P.D. (2016). Virus production in phosphorus-limited *Micromonas pusilla* stimulated by a supply of naturally low concentrations of different phosphorus sources, far into the lytic cycle. *FEMS Microbiology Ecology* **92**(9), pp.

Mandeville, R., Griffiths, M., Goodridge, L., McIntyre, L. and Ilenchuk, T.T. (2003). Diagnostic and therapeutic applications of lytic phages. *Analytical Letters* **36**(15), pp. 3241-3259.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989). *Molecular Cloning: a Laboratory Manual*. 2nd ed.: Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., USA.

Mann, K.H. and Lazier, J.R.N. (1996). Dynamics of marine ecosystems; Biological-physical interactions in the oceans. 2nd ed.: Blackwell, Cambridge, USA.

Mann, N.H., Clokie, M.R.J., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A. and Krisch, H.M. (2005). The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *Journal of Bacteriology* **187**(9), pp. 3188-3200.

Mann, N.H., Cook, A., Millard, A., Bailey, S. and Clokie, M. (2003). Marine ecosystems: Bacterial photosynthesis genes in a virus. *Nature* **424**(6950), pp. 741.

Manuel Santorum, J., Darriba, D., Taboada, G.L. and Posada, D. (2014). jmodeltest.org: selection of nucleotide substitution models on the cloud. *Bioinformatics* **30**(9), pp. 1310-1311.

Martiny, A.C., Coleman, M.L. and Chisholm, S.W. (2006b). Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **103**(33), pp. 12552-12557.

Martiny, A.C., Huang, Y. and Li, W.Z. (2009). Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environmental Microbiology* **11**(6), pp. 1340-1347.

Martínez-Pérez, C., Mohr, W., Löscher, C.R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B.M., Lavik, G., Schmitz, R.A., LaRoche, J. and Kuypers,

M.M.M. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, doi:10.1038/nmicrobiol.2016.163pp.

Mazard, S., Wilson, W. H. and Scanlan, D.J. (2012). Dissecting the physiological response to phophorus stress in marine *Synechococcus* isolates (Cyanophyceae). *Journal of Phycology* **48,** pp. 94-105.

Millard, A., Clokie, M.R.J., Shub, D.A. and Mann, N.H. (2004). Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proceedings of the National Academy of Sciences of the United States of America* **101**(30), pp.

Millard, A.D. (2009). Isolation of Cyanophages from Aquatic Environments. *Methods in Molecular Biology* 501pp. 33-42.

Millard, A.D. and Mann, N.H. (2006). A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba, Red Sea. *Journal of the Marine Biological Association of the United Kingdom* 86(3), pp. 507-515.

Millard, A.D., Zwirglmaier, K., Downey, M.J., Mann, N.H. and Scanlan, D.J. (2009). Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environmental Microbiology* **11**(9), pp. 2370-2387.

Miller, E.S., Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Durkin, A.S., Ciecko, A., Feldblyum, T.V., White, O., Paulsen, I.T., Nierman, W.C., Lee, J., Szczypinski, B. and Fraser, C.M. (2003a). Complete genome sequence of the broad-host-range vibriophage KVP40: Comparative genomics of a T4-related bacteriophage. *Journal of Bacteriology* **185**(17), pp. 5220-5233.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. and Ruger, W. (2003b). Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews* **67**(1), pp. 86.

Miller, S.R., Castenholz, R.W. and Pedersen, D. (2007). Phylogeography of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Applied and Environmental Microbiology* **73**(15), pp. 4751-4759.

Mills, M.M., Ridame, C., Davey, M., La Roche, J. and Geider, R.J. (2004). Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* **429**(6989), pp. 292-294.

Milon, P., Tischenko, E., Tomsic, J., Caserta, E., Folkers, G., La Teana, A., Rodnina, M.V., Pon, C.L., Boelens, R. and Gualerzi, C.O. (2006). The nucleotide-binding site of

bacterial translation initiation factor 2 (IF2) as a metabolic sensor. *Proceedings of the National Academy of Sciences of the United States of America* **103**(38), pp. 13962-13967.

Mohr, W., Vagner, T., Kuypers, M.M.M., Ackermann, M. and LaRoche, J. (2013). Resolution of conflicting signals at the single-cell level in the regulation of cyanobacterial photosynthesis and nitrogen fixation. *PLOS One* **8**(6), pp.

Moore, C.M., Mills, M.M., Arrigo, K.R., Berman-Frank, I., Bopp, L., Boyd, P.W., Galbraith, E.D., Geider, R.J., Guieu, C., Jaccard, S.L., Jickells, T.D., La Roche, J., Lenton, T.M., Mahowald, N.M., Maranon, E., Marinov, I., Moore, J.K., Nakatsuka, T., Oschlies, A., Saito, M.A., Thingstad, T.F., Tsuda, A. and Ulloa, O. (2013). Processes and patterns of oceanic nutrient limitation. *Nature Geoscience* **6**(9), pp. 701-710.

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

Moore, L.R., Post, A.F., Rocap, G. and Chisholm, S.W. (2002). Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnology and Oceanography* **47**(4), pp. 989-996.

Moore, L.R., Rocap, G. and Chisholm, S.W. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**(6684), pp. 464-467.

Moroz, O.V., Murzin, A.G., Makarova, K.S., Koonin, E.V., Wilson, K.S. and Galperin, M.Y. (2005). Dimeric dUTPases, HisE, and MazG belong to a new superfamily of all-alpha NTP pyrophosphohydrolases with potential "house-cleaning" functions. *Journal of Molecular Biology* **347**(2), pp. 243-255.

Mosig, G. and Eiserling, F. (2006). T4 and related phages: structure and development. In: *Calendar, R. The Bacteriophages*, pp. 225–267: Oxford University Press, New York, NY, USA.

Mrazek, J. and Xie, S.H. (2006). Pattern locator: a new tool for finding local sequence patterns in genomic DNA sequences. *Bioinformatics* **22**(24), pp. 3099-3100.

Nowicki, D., Kobiela, W., Wegrzyn, A., Wegrzyn, G. and Szalewska-Palasz, A. (2013b). ppGpp-dependent negative control of DNA replication of Shiga toxin-converting bacteriophages in *Escherichia coli*. *Journal of Bacteriology* **195**(22), pp. 5007-5015.

Ochoa de Alda, J.A.G., Esteban, R., Luz Diago, M. and Houmard, J. (2014). The plastid ancestor originated among one of the major cyanobacterial lineages. *Nature Communications* **5**, pp.

Ogunseitan, O.A., Sayler, G.S. and Miller, R.V. (1990). Dynamic interactions of *Pseudomonas aeruginosa* and bacteriophages in lake water. *Microbial Ecology* **19**(2), pp. 171-186.

Ostrowski, M., Mazard, S., Tetu, S.G., Phillippy, K., Johnson, A., Palenik, B., Paulsen, I.T. and Scanlan, D.J. (2010). PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine *Synechococcus*. *ISME Journal* **4**(7), pp. 908-921.

Paerl, H.W. and Otten, T.G. (2013). Harmful cyanobacterial blooms: causes, consequences, and controls. *Microbial Ecology* **65**(4), pp. 995-1010.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A. and Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**(22), pp. 3691-3693.

Pal, C., Macia, M.D., Oliver, A., Schachar, I. and Buckling, A. (2007). Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**(7172), pp. 1079-1081.

Palenik, B. (2001). Chromatic adaptation in marine *Synechococcus* strains. *Applied and Environmental Microbiology* **67**(2), pp. 991-994.

Palenik, B., Brahamsha, B., Larimer, F.W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E.E., McCarren, J., Paulsen, I., Dufresne, A., Partensky, F., Webb, E.A. and Waterbury, J. (2003). The genome of a motile marine *Synechococcus*. *Nature* **424**(6952), pp. 1037-1042.

Palenik, B., Ren, Q., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., Madupu, R., Nelson, W.C., Brinkac, L.M., Dodson, R.J., Durkin, A.S., Daugherty, S.C., Sullivan, S.A., Khouri, H., Mohamoud, Y., Halpin, R. and Paulsen, I.T. (2006). Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proceedings of the National Academy of Sciences of the United States of America* **103**(36), pp. 13555-13559.

Partensky, F. and Garczarek, L. (2010). *Prochlorococcus*: Advantages and Limits of Minimalism. *Annual Review of Marine Science* **2**, pp. 305-331.

Paul, B.J., Berkmen, M.B. and Gourse, R.L. (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. *Proceedings of the National Academy of Sciences of the United States of America* **102**(22), pp. 7823-7828.

Paytan, A., Cade-Menun, B.J., McLaughlin, K. and Faul, K.L. (2003). Selective phosphorus regeneration of sinking marine particles: evidence from P-31-NMR. *Marine Chemistry* **82**(1-2), pp. 55-70.

Perez-Sepulveda, B., Redgwell, T., Rihtman, B., Pitt, F., Scanlan, D.J. and Millard, A. (2016). Marine phage genomics: the tip of the iceberg. *FEMS Microbiology Letters* **363**(15).

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**(10), pp. 785-786.

Petrov, V.M. and Karam, J.D. (2004). Diversity of structure and function of DNA polymerase (gp43) of T4-related bacteriophages. *Biochemistry-Moscow* **69**(11), pp. 1213-+.

Philosof, A., Battchikova, N., Aro, E.-M. and Beja, O. (2011). Marine cyanophages: tinkering with the electron transport chain. *ISME Journal* **5**(10), pp. 1568-1570.

Pierce, J. and Omata, T. (1988). Uptake and utilization of inorganic carbon by cyanobacteria. *Photosynthesis Research* **16**(1-2), pp. 141-154.

Post, A.F., Penno, S., Zandbank, K., Paytan, A., Huse, S.M. and Welch, D.M. (2011). Long term seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, Northern Red Sea. *Frontiers in Microbiology* **2** (131).

Poulton, S.W. and Raiswell, R. (2002). The low-temperature geochemical cycle of iron: From continental fluxes to marine sediment deposition. *American Journal of Science* **302**(9), pp. 774-805.

Puxty, R.J. (2014). The role of light in photosynthetic cyanophages: from physiology to gene expression. Doctoral dissertation, University of Warwick.

Puxty, R.J., Millard, A.D., Evans, D.J. and Scanlan, D.J. (2016). Viruses inhibit CO_2 fixation in the most abundant phototrophs on Earth. *Current Biology* **26**(12), pp. 1585-1589.

Puxty, R.J., Perez-Sepulveda, B., Rihtman, B., Evans, D.J., Millard, A.D. and Scanlan, D.J. (2015b). Spontaneous Deletion of an "ORFanage" Region Facilitates Host Adaptation in a "Photosynthetic" Cyanophage. *PLOS One* **10**(7), pp.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6), pp. 841-842.

Rakhuba, D.V., Kolomiets, E.I., Dey, E.S. and Novik, G.I. (2010). Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Polish Journal of Microbiology* **59**(3), pp. 145-155.

Rao, N.N., Liu, S.J. and Kornberg, A. (1998). Inorganic polyphosphate in *Escherichia coli*: the phosphate regulon and the stringent response. *Journal of Bacteriology* **180**(8), pp. 2186-2193.

Ribalet, F., Swalwell, J., Clayton, S., Jimenez, V., Sudek, S., Lin, Y.J., Johnson, Z.I., Worden, A.Z. and Armbrust, E.V. (2015). Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proceedings of the National Academy of Sciences of the United States of America* **112**(26), pp. 8008-8012.

Rihtman, B., Meaden, S., Clokie, M.R.J., Koskella, B. and Millard, A.D. (2016). Assessing Illumina technology for the high-throughput sequencing of bacteriophage genomes. *PeerJ* **4**, pp. e2055-e2055.

Ripp, S. and Miller, R.V. (1997). The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology* **143**, pp. 2065-2070

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., Johnson, Z.I., Land, M., Lindell, D., Post, A.F., Regala, W., Shah, M., Shaw, S.L., Steglich, C., Sullivan, M.B., Ting, C.S., Tolonen, A., Webb, E.A., Zinser, E.R. and Chisholm, S.W. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**(6952), pp. 1042-1047.

Rodriguez-Garcia, A., Sola-Landa, A., Apel, K., Santos-Beneit, F. and Martin, J.F. (2009). Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: direct and indirect negative control of *glnR*, *glnA*, *glnII* and *amtB* expression by the response regulator PhoP. *Nucleic Acids Research* **37**(10), pp. 3230-3242.

Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pasic, L., Thingstad, T.F., Rohwer, F. and Mira, A. (2009). Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7**(11), pp. 828-836.

Rohwer, F., Segall, A., Steward, G., Seguritan, V., Breitbart, M., Wolven, F. and Azam, F. (2000). The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnology and Oceanography* **45**(2), pp. 408-418.

Roitman, S., Flores-Uribe, J., Philosof, A., Knowles, B., Rohwer, F., Ignacio-Espinoza, J.C., Sullivan, M.B., Cornejo-Castillo, F.M., Sánchez, P., Acinas, S.G., Dupont, C.L. and Béjà, O. (2015). Closing the gaps on the viral photosystem-I *psaDCAB* gene organization. *Environmental Microbiology* **17**(12), pp. 5100-5108

Roszak, D.B. and Colwell, R.R. (1987). Survival strategies of bacteria in the natural environment. *Microbiological Reviews* **51**(3), pp. 365-379.

Rousseeuw, P.J. (1987). Silhouettes - A graphical aid to the interpretation and validation of cluster-analysis. *Journal of Computational and Applied Mathematics* **20,** pp. 53-65.

Rusch, D.B., Martiny, A.C., Dupont, C.L., Halpern, A.L. and Venter, J.C. (2010). Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proceedings of the National Academy of Sciences of the United States of America* **107**(37), pp. 16184-16189.

Sabehi, G., Shaulov, L., Silver, D.H., Yanai, I., Harel, A. and Lindell, D. (2012). A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proceedings of the National Academy of Sciences of the United States of America* **109**(6), pp. 2037-2042.

Sachs, J.P. and Repeta, D.J. (1999). Oligotrophy and nitrogen fixation during eastern Mediterranean sapropel events. *Science* **286**(5449), pp. 2485-2488.

Sañudo-Wilhelmy, S.A., Kustka, A.B., Gobler, C.J., Hutchins, D.A., Yang, M., Lwiza, K., Burns, J., Capone, D.G., Raven, J.A. and Carpenter, E.J. (2001). Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* **411**(6833), pp. 66-69.

Scanlan, D. (2012). Marine picocyanobacteria. In: Whitton, B.A. *Ecology of Cyanobacteria II*, pp. 503–533. Dordrecht: Springer.

Scanlan, D.J. (2003). Physiological diversity and niche adaptation in marine *Synechococcus*. *Advances in Microbial Physiology* **47**, pp. 1-64.

Scanlan, D.J., Mann, N.H. and Carr, N.G. (1993). The response of the picoplanktonic marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate-binding protein of *Escherichia coli*. Molecular *Microbiology* **10**(1), pp. 181-191

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., Post, A.F., Hagemann, M., Paulsen, I. and Partensky, F. (2009). Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews* **73**(2), pp. 249-299

Scanlan, D.J. and West, N.J. (2002). Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiology Ecology* **40**(1), pp. 1-12.

Schopf, J.W. (1993). Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. *Science* 260(5108), pp. 640-646.

Schreiber, G., Ron, E.Z. and Glaser, G. (1995). ppGpp-mediated regulation of DNA-replication and cell-division in *Escherichia coli*. *Current Microbiology* **30**(1), pp. 27-32.

Schwarz, R. and Forchhammer, K. (2005). Acclimation of unicellular cyanobacteria to macronutrient deficiency: emergence of a complex network of cellular responses. *Microbiology* **151**, pp. 2503-2514.

Sebastian, M., Smith, A.F., Gonzalez, J.M., Fredricks, H.F., Van Mooy, B., Koblizek, M., Brandsma, J., Koster, G., Mestre, M., Mostajir, B., Pitta, P., Postle, A.D., Sanchez, P., Gasol, J.M., Scanlan, D.J. and Chen, Y. (2016). Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. *ISME Journal* **10**(4), pp. 968-978.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**(14), pp. 2068-2069.

Silva, P., Thompson, E., Bailey, S., Kruse, O., Mullineaux, C.W., Robinson, C., Mann, N.H. and Nixon, P.J. (2003). FtsH is involved in the early stages of repair of photosystem II in Synechocystis sp PCC 6803. *Plant Cell* **15**(9), pp. 2152-2164.

Six, C., Thomas, J.-C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., Scanlan, D.J. and Partensky, F. (2007). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biology* **8**(12), pp. 1

Sohm, J.A., Ahlgren, N.A., Thomson, Z.J., Williams, C., Moffett, J.W., Saito, M.A., Webb, E.A. and Rocap, G. (2016). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME Journal* **10**(2), pp. 333-345.

Srivatsan, A. and Wang, J.D. (2008). Control of bacterial transcription, translation and replication by (p)ppGpp. *Current Opinion in Microbiology* **11**(2), pp. 100-105.

Stazic, D., Pekarski, I., Kopf, M., Lindell, D. and Steglich, C. (2016). A novel strategy for exploitation of host RNase E activity by a marine cyanophage. *Genetics* **204**(1), pp. 115

Steglich, C., Mullineaux, C.W., Teuchner, K., Hess, W.R. and Lokstein, H. (2003). Photophysical properties of *Prochlorococcus marinus* SS120 divinyl chlorophylls and phycoerythrin in vitro and in vivo. *FEBS Letters* **553**(1-2), pp. 79-84.

Su, Z.C., Olman, V. and Xu, Y. (2007). Computational prediction of Pho regulons in cyanobacteria. *BMC Genomics* **8**(8), pp. 1

Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P.R., DeFrancesco, A.S., Kern, S.E., Thompson, L.R., Young, S., Yandava, C., Fu, R., Krastins, B., Chase, M., Sarracino, D., Osburne, M.S., Henn, M.R. and Chisholm, S.W. (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology* **12**(11), pp. 3035-3056.

Sullivan, M.J., Petty, N.K. and Beatson, S.A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* **27**(7), pp. 1009-1010.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., Cornejo-Castillo, F.M., Costea, P.I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J.M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B.T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M.B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S.G., Bork, P. and Tara Oceans, C. (2015). Structure and function of the global ocean microbiome. *Science* **348**(6237), pp. 1261359

Suttle, C.A. (2005). Viruses in the sea. *Nature* **437**(7057), pp. 356-361.

Suttle, C.A. (2007). Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology* **5**(10), pp. 801-812.

Tai, V. and Palenik, B. (2009). Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME Journal* **3**(8), pp. 903-915.

Tamoi, M., Miyazaki, T., Fukamizo, T. and Shigeoka, S. (2005). The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant Journal* **42**(4), pp. 504-513.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**(12), pp. 2725-2729.

Taylor, R.G., Walker, D.C. and McInnes, R.R. (1993). *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research* **21**(7), pp. 1677-1678.

Tetu, S.G., Brahamsha, B., Johnson, D.A., Tai, V., Phillippy, K., Palenik, B. and Paulsen, I.T. (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp WH8102. *ISME Journal* **3**(7), pp. 835-849.

Thingstad, T. F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, **45**(6), pp. 1320-1328.

Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J. and Chisholm, S.W. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism *Proceedings of the National Academy of Sciences of the United States of America* **108**(39), pp. E757-764.

Trautinger, B.W., Jaktaji, R.P., Rusakova, E. and Lloyd, R.G. (2005). RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Molecular Cell* **19**(2), pp. 247-258.

Twort, F.W. (1915). An investigation on the nature of ultra-microscopic viruses. *The Lancet* **186**(4814), pp. 1241-1243.

Van Mooy, B.A.S., Rocap, G., Fredricks, H.F., Evans, C.T. and Devol, A.H. (2006). Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences of the United States of America* **103**(23), pp. 8607-8612.

Van Wambeke, F., Christaki, U., Giannokourou, A., Moutin, T. and Souvemerzoglou, K. (2002). Longitudinal and vertical trends of bacterial limitation by phosphorus and carbon in the Mediterranean Sea. *Microbial Ecology* **43**(1), pp. 119-133.

Vaulot, D., LeBot, N., Marie, D. and Fukai, E. (1996). Effect of phosphorus on the *Synechococcus* cell cycle in surface Mediterranean waters during summer. *Applied Environmental Microbiology* **62**(7), pp. 2527-2533.

Veldhuis, M.J.W., Kraay, G.W., VanBleijswijk, J.D.L. and Baars, M.A. (1997). Seasonal and spatial variability in phytoplankton biomass, productivity and growth in the northwestern Indian Ocean: The southwest and northeast monsoon, 1992-1993. *Deep-Sea Research Part I-Oceanographic Research Papers* **44**(3), pp. 425-449.

Wanner, B.L. (1990a). Phosphorus assimilation and its control of gene expression in Escherichia Coli. p. 152-163. In Hauska, G.a.R.T. (ed.), Colloquium Der Gesellschaft Fuer Biologische Chemie Mosbach; 41. *The Molecular Basis of Bacterial Metabolism*. Springer-Verlag: Berlin, Germany, Mosbach, Germany.

Waterbury, J.B., Watson, S.W., Valois, F.W. and Franks, D.G. (1986). Biological and ecological characterisation of the marine unicellular cyanobacterium *Synechococcus* sp. In Platt, T. and Li, W. (ed.), *Canadian Bulletin of Fisheries and Aquatic Sciences* vol. **214**. p. 71–120

Weigele, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., King, J., Hatfull, G.F., Lawrence, J.G. and Hendrix, R.W. (2007). Genomic and structural

analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environmental Microbiology* **9**(7), pp. 1675-1695.

Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28**(2), pp. 127-181.

Wendrich, T.M., Blaha, G., Wilson, D.N., Marahiel, M.A. and Nierhaus, K.H. (2002). Dissection of the mechanism for the stringent factor RelA. *Molecular Cell* **10**(4), pp. 779-788.

West, N.J. and Scanlan, D.J. (1999). Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Applied and Environmental Microbiology* **65**(6), pp. 2585-2591.

Whelan, S. and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18**(5), pp. 691-699.

Wilhelm, S. W., & Suttle, C. A. (1999). Viruses and nutrient cycles in the sea viruses play critical roles in the structure and function of aquatic food webs. *Bioscience*, **49** (10), 781-788.

Williams, H.T.P. (2013). Phage-induced diversification improves host evolvability. *BMC Evolutionary Biology* **13** (1), pp. 1

Willner, D., Thurber, R.V. and Rohwer, F. (2009). Metagenomic signatures of 86 microbial and viral metagenomes. *Environmental Microbiology* **11**(7), pp. 1752-1766.

Wilson, W.H., Carr, N.G. and Mann, N.H. (1996). The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp WH7803. *Journal of Phycology* **32**(4), pp. 506–516.

Wilson, W.H., Joint, I.R., Carr, N.G. and Mann, N.H. (1993). Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Applied and Environmental Microbiology* **59**(11), pp. 3736-3743.

Wolf-Vecht, A., Paldor, N. and Brenner, S. (1992). Hydrographic indications of advection/convection effects in the Gulf of Eilat. *Deep Sea Research Part A. Oceanographic Research Papers* **39**(7-8), pp. 1393-1401.

Wu, J., Sunda, W., Boyle, E.A. and Karl, D.M. (2000). Phosphate depletion in the western north Atlantic Ocean. *Science* **289**(5480), pp. 759-762.

Xiao, H., Kalman, M., Ikehara, K., Zemel, S., Glaser, G. and Cashel, M. (1991). Residual guanosine 3',5'-bispyrophosphate synthetic activity of relA null mutants can be eliminated by spoT null mutations. *Journal of Biological Chemistry* **266**(9), pp. 5980-5990.

Young, C.L. and Ingall, E.D. (2010). Marine dissolved organic phosphorus composition: insights from samples recovered using combined electrodialysis/reverse osmosis. *Aquatic Geochemistry* **16**(4), pp. 563-574.

Zeng, Q. and Chisholm, S.W. (2012). Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Current Biology* **22**(2), pp. 124–128.

Zhang, J.J. and Inouye, M. (2002). MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with era, an essential GTPase in *Escherichia coli*. *Journal of Bacteriology* **184**(19), pp. 5323-5329.

Zhang, J.J., Zhang, Y.L. and Inouye, M. (2003). *Thermotoga maritima* MazG protein has both nucleoside triphosphate pyrophosphohydrolase and pyrophosphatase activities. *Journal of Biological Chemistry* **278**(24), pp. 21408-21414.

Zhao, Y.L., Temperton, B., Thrash, J.C., Schwalbach, M.S., Vergin, K.L., Landry, Z.C., Ellisman, M., Deerinck, T., Sullivan, M.B. and Giovannoni, S.J. (2013). Abundant SAR11 viruses in the ocean. *Nature* **494**(7437), pp. 357-360.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., Not, F., Massana, R., Ulloa, O. and Scanlan, D.J. (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology* **10**(1), pp. 147-161.

Zwirglmaier, K., Spence, E.M., Zubkov, M. V., Scanlan, D.J., and Mann, N.H. (2009). Differential grazing of two heterotrophic nanoflagellates on marine *Synechococcus* strains. *Environmental Microbiology* **11**, 1767–1776

Appendices

Appendix 1 - Sequences of the *Synechococcus* sp. WH7803 and S-PM2 *mazG* genes after optimization for *E. coli* codon usage. The changed codons in the optimized sequence are underlined.

```
Synechococcus WH7803   ATGGCCCAGCACGCCACGGTTCCTGGTGATCAGGATCCCCTGCGCTACCTCGAAAGCGTG
Protein                 M  A  Q  H  A  T  V  P  G  D  Q  D  P  L  R  Y  L  E  S  V
E. coli                ATGGCACAGCATGCAACCGTTCCGGGTGATCAGGATCCGCTGCGTTATCTGGAAAGCGTT

Synechococcus WH7803   GTGGCACGCCTGCGGGATCCAGTGAACGGGTGCCCCTGGGACCTGGAGCAGACCCATGCC
Protein                 V  A  R  L  R  D  P  V  N  G  C  P  W  D  L  E  Q  T  H  A
E. coli                GTTGCACGTCTGCGTGATCCGGGTTAATGGTTGTCCGTGGGATCTGGAACAGACCCATGCA

Synechococcus WH7803   TCCCTGGTGCCCTACGTCCTGGAGGAAGCCCATGAGGTGGCTGACGCCATCCGCCATGGC
Protein                 S  L  V  P  Y  V  L  E  E  A  H  E  V  A  D  A  I  R  H  G
E. coli                AGCCTGGTTCCGTATGTTCTGGAAGAAGCACACGAAGTTGCAGATGCAATTCGTCATGGT

Synechococcus WH7803   GACGACCGGCATCTGAAAGAAGAACTTGGCGATCTGCTGCTGCAGGTGGTGCTGCACGCC
Protein                 D  D  R  H  L  K  E  E  L  G  D  L  L  L  Q  V  V  L  H  A
E. coli                GATGATCGTCATCTGAAAGAAGAACTGGGCGATCTGCTGCTGCAGGTTGTTCTGCATGCA

Synechococcus WH7803   CGCATCGGCGCAGAAAACAACCGCTTTGATCTCGATGCCATCGCCACGACCATCAGCGAC
Protein                 R  I  G  A  E  N  N  R  F  D  L  D  A  I  A  T  T  I  S  D
E. coli                CGTATTGGTGCAGAAAATAATCGTTTTGATCTGGATGCAATTGCCACCACCATTAGCGAT

Synechococcus WH7803   AAACTGATCCGCCGCCATCCCCACGTGTTCGGCGAGGCCCGAGCGGAAAACACTGAGGCC
Protein                 K  L  I  R  R  H  P  H  V  F  G  E  A  R  A  E  N  T  E  A
E. coli                AAACTGATTCGTCGTCATCCGCATGTTTTTGGTGAAGCACGTGCCGAAAATACCGAAGCA

Synechococcus WH7803   GTGCGCTTGAGCTGGGAGGCCATCAAGGCCGCCGAACGGGCGGAACAATCGGGAGGGGAG
Protein                 V  R  L  S  W  E  A  I  K  A  A  E  R  A  E  Q  S  G  G  E
E. coli                GTTCGTCTGAGCTGGGAAGCAATTAAAGCAGCAGAACGTGCAGAACAGAGCGGTGGTGAA

Synechococcus WH7803   CAATCCTCCAGCCCGCTGAGCGATCAACTGGCCGGCAAGGTGCGAGGGCAACCGGCCCTG
Protein                 Q  S  S  S  P  L  S  D  Q  L  A  G  K  V  R  G  Q  P  A  L
E. coli                CAGAGCAGCAGTCCGCTGAGCGATCAGCTGGCAGGTAAAGTTCGTGGTCAGCCTGCACTG

Synechococcus WH7803   GCCGCTGCCATGACCATCTCGCGCAAGGCCGCCAAGGCCGGTTTCGAGTGGGATGCCATC
Protein                 A  A  A  M  T  I  S  R  K  A  A  K  A  G  F  E  W  D  A  I
E. coli                GCAGCAGCAATGACCATTAGCCGTAAAGCAGCCAAAGCAGGTTTTGAATGGGATGCCATT

Synechococcus WH7803   GACGGCGTGTGGGGGAAGGTGCAGGAGGAGCTCGATGAGCTCAAGGAGGCCATCGCCTCG
Protein                 D  G  V  W  G  K  V  Q  E  E  L  D  E  L  K  E  A  I  A  S
E. coli                GATGGTGTTTGGGGTAAAGTGCAAGAGGAACTGGATGAACTGAAAGAGGCCATTGCATCA

Synechococcus WH7803   GGGGACCGCCGCCATGCTCAAGATGAGCTCGGCGATGTGCTGTTCACCCTTGTGAATGTG
Protein                 G  D  R  R  H  A  Q  D  E  L  G  D  V  L  F  T  L  V  N  V
E. coli                GGCGATCGTCGCCATGCACAGGATGAGCTGGGTGATGTTCTGTTTACCCTGGTTAATGTT

Synechococcus WH7803   GCCCGCTGGTGCGGCCTTGATCCTGAGGAGGGGCTGGCCGCAACCAATCAGCGCTTCCTG
Protein                 A  R  W  C  G  L  D  P  E  E  G  L  A  A  T  N  Q  R  F  L
E. coli                GCCCGTTGGTGTGGTCTGGATCCGGAAGAAGGTCTGGCAGCAACCAATCAGCGTTTTCTG

Synechococcus WH7803   GATCGTTTCTCCCGGGTTGAGAGCGCCCTGAATGGAGATCTGCAAGGACGGAGCATCCAG
Protein                 D  R  F  S  R  V  E  S  A  L  N  G  D  L  Q  G  R  S  I  Q
E. coli                GATCGTTTTAGCCGTGTTGAAAGCGCACTGAATGGTGATCTGCAGGGTCGTAGCATTCAA

Synechococcus WH7803   GAGCTCGAGGCTCTGTGGCAGCAAGCGAAGGCTGCGATCCGCGCTGAAAACACCCAGTCG
Protein                 E  L  E  A  L  W  Q  Q  A  K  A  A  I  R  A  E  N  T  Q  S
E. coli                GAACTGGAAGCACTGTGGCAGCAGGCAAAAGCAGCAATTCGTGCGGAAAACACCCAGAGC

Synechococcus WH7803   TCTTGA
Protein                 S  *
E. coli                AGCTAG
```

```
S-PM2      ATGTCTAAAGTAAACTTTGAACGCTATCAAGAATTTGTGTCGGAAGTTACTTCCGATGCT
Protein     M  S  K  V  N  F  E  R  Y  Q  E  F  V  S  E  V  T  S  D  A
E. coli    ATGAGCAAAGTGAACTTTGAGCGCTATCAAGAATTTGTTAGCGAAGTTACCAGTGATGCC

S-PM2      TCTACAAACTTCGTTGACTTCGCTGATCGTATTGGCGAGTTGGATCGTGAAGGTGCCAAT
Protein     S  T  N  F  V  D  F  A  D  R  I  G  E  L  D  R  E  G  A  N
E. coli    AGCACCAATTTTGTTGATTTTGCAGATCGTATTGGTGAACTGGATCGTGAAGGTGCAAAT

S-PM2      ATTGAACGACTTCTTACTGCTGGTGTTGGCATCAATGCTGAGGGTGGTGAGTTTCTTGAG
Protein     I  E  R  L  L  T  A  G  V  G  I  N  A  E  G  G  E  F  L  E
E. coli    ATTGAACGTCTGCTGACCGCAGGCGTTGGTATTAATGCCGAAGGTGGTGAATTTCTGGAA

S-PM2      ATCATTAAGAAGATGGTATTCCAAGGTAAGCCTTGGAACCGCGATAATCGAGAACATCTT
Protein     I  I  K  K  M  V  F  Q  G  K  P  W  N  R  D  N  R  E  H  L
E. coli    ATCATCAAAAAAATGGTGTTCCAGGGTAAACCGTGGAATCGTGATAATCGTGAACATCTG

S-PM2      ATTATTGAGTTGGGTGACATTATGTGGTATGTGGCACAAGCATGTATTGCGCTAGGTGTT
Protein     I  I  E  L  G  D  I  M  W  Y  V  A  Q  A  C  I  A  L  G  V
E. coli    ATTATTGAGCTGGGCGATATTATGTGGTATGTTGCACAGGCATGTATTGCACTGGGTGTT

S-PM2      TCTTTTGATGATGTCATTTCTGGCAACGTCAAGAAACTTGAAAAACGTTATCCAGGAGGA
Protein     S  F  D  D  V  I  S  G  N  V  K  K  L  E  K  R  Y  P  G  G
E. coli    AGCTTTGATGATGTGATTAGCGGCAATGTGAAAAAACTGGAAAAACGTTATCCGGGTGGC

S-PM2      GAATTTGATGTCTTCTATTCCGAAAATAGATCAGCAGACGACCGATAA
Protein     E  F  D  V  F  Y  S  E  N  R  S  A  D  D  R  *
E. coli    GAATTTGATGTGTTTTATAGCGAAAATCGTAGCGCAGATGATCGCTAG
```

Appendix 2: Clustergrams showing hierarchical clustering of relative normalised expression estimates for each S-PM2 gene. Left - +P infected sample; Middle - -P infected sample T0-T9; Right - -P infected sample T0-T15
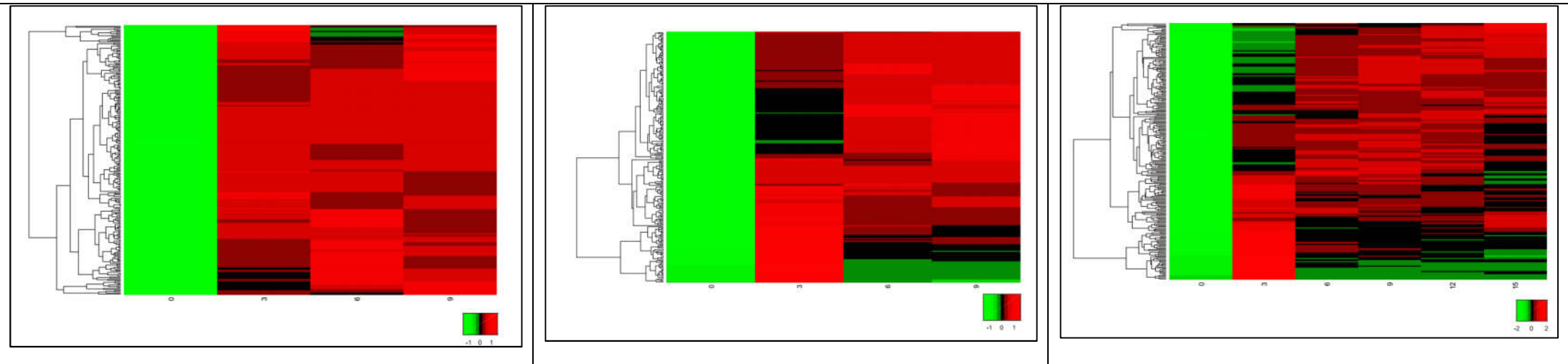
Appendix 3: Temporal cluster grouping for all the S-PM2 ORFs. ORF numbers and coordinates are as reported in the S-PM2d genome submitted in Genbank accession LN828717.1. Strand designation – (+) – Watson strand, (-) – Crick strand. Clustering methods used are described in Chapter 6

ORF	Gene annotation	Gene product	Strand	Start codon position	Stop codon position	Experimental conditions			
						+P (T0-T9)	-P (T0-T9, 2 clusters)	-P (T0-T15, 2 clusters)	-P (T0-T15, 3 clusters)
S-PM2d001	DNA adenine Methylase	Dam	+	1	888	Late	Late	Late	Late
S-PM2d002	hypothetical protein		+	925	1059	Late	Early	Early	Early
S-PM2d003	hypothetical protein		+	1049	1258	Late	Early	Early	Early
S-PM2d004	hypothetical protein		+	1255	1560	Late	Early	Early	Early
S-PM2d005	hypothetical protein		+	1729	1932	Early	Late	Late	Middle
S-PM2d006	hypothetical protein		+	1913	2116	Early	Early	Early	Early
S-PM2d007	hypothetical protein		+	2113	2292	Early	Early	Early	Early
S-PM2d008	hypothetical protein		+	2289	2852	Late	Late	Late	Middle
S-PM2d009	hypothetical protein		+	2856	3251	Late	Late	Late	Late
S-PM2d010	hypothetical protein		+	3248	3556	Late	Late	Late	Late
S-PM2d011	hypothetical protein		+	3550	3678	Late	Late	Late	Middle
S-PM2d012	hypothetical protein		+	3675	4079	Late	Early	Early	Middle
S-PM2d013	hypothetical protein		+	4201	4605	Late	Late	Late	Middle
S-PM2d014	hypothetical protein		+	4710	5090	Late	Early	Early	Middle
S-PM2d015	hypothetical protein		+	5160	5330	Late	Late	Late	Late
S-PM2d016	hypothetical protein		+	5370	5780	Late	Late	Late	Late
S-PM2d051	hypothetical protein		+	5906	6055	Late	Late	Late	Late
S-PM2d052	hypothetical protein		+	6052	6234	Late	Early	Late	Middle
S-PM2d053	hypothetical protein		+	6227	6739	Early	Early	Early	Early
S-PM2d054	hypothetical protein		+	6768	6971	Late	Early	Late	Middle
S-PM2d055	hypothetical protein		+	7044	7268	Late	Late	Late	Middle
S-PM2d056	hypothetical protein		+	7488	7922	Late	Early	Late	Middle
S-PM2d057	hypothetical protein		+	7984	8253	Late	Early	Late	Middle

S-PM2d058	hypothetical protein	+	8250	8477	Late	Late	Late	Middle
S-PM2d059	hypothetical protein	+	8474	8677	Early	Early	Early	Early
S-PM2d060	2OG-FeII oxygenase	+	8674	9276	Late	Late	Late	Late
S-PM2d061	hypothetical protein	+	9249	9440	Late	Late	Late	Late
S-PM2d062	hypothetical protein	+	9421	9654	Late	Late	Late	Late
S-PM2d063	hypothetical protein	+	10095	10346	Early	Early	Early	Early
S-PM2d064	hypothetical protein	+	10400	10564	Early	Early	Early	Early
S-PM2d065	hypothetical protein	+	10740	10859	Early	Late	Late	Middle
S-PM2d066	hypothetical protein	-	10913	11158	Late	Early	Late	Middle
S-PM2d068	hypothetical protein	+	11196	11525	Late	Early	Late	Middle
S-PM2d069	hypothetical protein	+	11564	11737	Late	Late	Late	Late
S-PM2d070	hypothetical protein	+	11734	11934	Late	Early	Late	Middle
S-PM2d071	hypothetical protein	-	12114	12386	Early	Early	Early	Early
S-PM2d250	hypothetical protei	+	12555	12722	Early	Early	Early	Early
S-PM2d072	hypothetical protein	+	12798	13232	Early	Early	Early	Early
S-PM2d073	hypothetical protein	+	13286	13432	Early	Early	Early	Early
S-PM2d074	hypothetical protein	-	14226	14432	Early	Early	Early	Early
S-PM2d239	hypothetical protein	+	14459	14560	Early	Early	Early	Early
S-PM2d240	hypothetical protein	+	14538	14717	Early	Early	Early	Early
S-PM2d241	hypothetical protein	+	14805	15005	Early	Early	Early	Early
S-PM2d242	hypothetical protein	+	15075	15260	Early	Early	Early	Middle
S-PM2d243	hypothetical protein	+	15286	15486	Early	Late	Late	Late
S-PM2d075	hypothetical protein	+	15726	16055	Late	Late	Late	Late
S-PM2d244	hypothetical protein	+	15982	16245	Late	Late	Late	Late
S-PM2d245	hypothetical protein	+	16173	16496	Late	Late	Late	Late
S-PM2d076	hypothetical protein	+	16536	16679	Early	Late	Late	Late
S-PM2d077	hypothetical protein	+	16663	17127	Late	Late	Late	Late
S-PM2d078	hypothetical protein	+	17096	17290	Early	Late	Late	Late

S-PM2d079	base plate wedge subunit	gp25	+	17324	17716	Late	Late	Late	Late
S-PM2d080	baseplate wedge	gp6	+	17716	19524	Late	Late	Late	Late
S-PM2d081	hypothetical protein		+	19530	28676	Late	Late	Late	Late
S-PM2d082	virion structural protein		+	28677	29204	Early	Late	Late	Late
S-PM2d083	baseplate wedge	gp8	+	29267	31171	Late	Late	Late	Late
S-PM2d084	Virulence Associated protein		+	31216	39459	Late	Late	Late	Late
S-PM2d085	hypothetical protein		+	39456	39659	Early	Late	Late	Late
S-PM2d086	virion structural protein		+	39637	43392	Early	Late	Late	Late
S-PM2d087	virion structural protein		+	43394	43900	Early	Late	Late	Late
S-PM2d088	hypothetical protein		+	43949	53965	Late	Late	Late	Late
S-PM2d089	virion structural protein		+	53975	54895	Late	Late	Late	Late
S-PM2d090	virion structural protein		+	54897	55880	Early	Late	Late	Late
S-PM2d091	virion structural protein		+	55882	57021	Early	Late	Late	Late
S-PM2d092	hypothetical protein		+	57053	57238	Early	Late	Late	Middle
S-PM2d093	neck protein	gp13	+	57291	58121	Late	Late	Late	Late
S-PM2d094	neck protein	gp14	+	58123	59001	Late	Late	Late	Late
S-PM2d095	proximal tail sheath stabilization	gp15	+	59001	59801	Late	Late	Late	Late
S-PM2d096	terminase DNA packaging enzyme - small subunit	gp16	+	59798	60217	Late	Late	Late	Late
S-PM2d097	hypothetical protein		+	60217	60759	Late	Late	Late	Late
S-PM2d098	hypothetical protein		+	60862	61092	Early	Late	Late	Late
S-PM2d099	hypothetical protein		+	61070	61288	Early	Late	Late	Late
S-PM2d100	terminase DNA packaging enzyme large subunit	gp17	+	61314	62960	Late	Late	Late	Late
S-PM2d101	T4-like Endonuclease VII		+	62948	63190	Late	Late	Late	Late
S-PM2d102	tail sheath monomer	gp18	+	63275	65506	Late	Late	Late	Late
S-PM2d103	tail tube monomer	gp19	+	65545	66159	Late	Late	Late	Late

S-PM2d104	portal vertex protein of head	gp20	+	66195	67889	Late	Late	Late	Late
S-PM2d105	hypothetical protein		+	67912	68088	Early	Early	Early	Early
S-PM2d106	prohead core scaffold and protease	gp21	+	68085	68729	Early	Late	Late	Late
S-PM2d107	scaffoldprohead core protein	gp22	+	68786	69964	Late	Late	Late	Middle
S-PM2d108	precursor of major head subunit	gp23	+	70016	71422	Late	Early	Late	Middle
S-PM2d109	hypothetical protein		+	71498	71662	Late	Late	Late	Late
S-PM2d110	head-proximal tip of tail tube tail completion - sheath stabilizer protein	gp3	+	72022	72531	Late	Late	Late	Late
S-PM2d111	UvsY	UvsY	+	72528	73001	Late	Late	Late	Late
S-PM2d112	RNA-DNA/DNA-DNA helicase	UvsW	+	73001	74464	Late	Late	Late	Late
S-PM2d113	Methylamine util.		+	74464	74862	Early	Late	Late	Late
S-PM2d114	Sigma factor for late transcription	gp55	+	74990	75484	Early	Late	Late	Middle
S-PM2d115	hypothetical protein		+	75481	75699	Late	Late	Late	Middle
S-PM2d116	recombination endonuclease subunit	gp47	+	75696	76745	Early	Late	Late	Middle
S-PM2d117	hypothetical protein		+	76742	77017	Early	Early	Early	Middle
S-PM2d118	recombination endonuclease subunit	gp46	+	77014	78744	Early	Early	Early	Early
S-PM2d119	Peptidase		+	78747	80969	Late	Late	Late	Late
S-PM2d120	porphyrin biosynthetic protein	CobS	+	81043	82161	Late	Late	Late	Middle
S-PM2d121	sliding clamp DNA polymerase accessory protein	gp45	+	82610	83275	Late	Late	Late	Middle
S-PM2d122	clamp loader subunit	gp44	+	83275	84219	Late	Late	Late	Middle

S-PM2d123	hypothetical protein		+	84216	84410	Late	Early	Early	Middle
S-PM2d124	hypothetical protein		+	84403	84603	Late	Late	Late	Late
S-PM2d125	possible cytosine-specific methyl transferase		+	84635	85414	Late	Late	Late	Late
S-PM2d127	hypothetical protein		+	85404	85547	Late	Late	Late	Late
S-PM2d128	clamp loader subunit	gp62	+	85695	86081	Late	Late	Late	Middle
S-PM2d129	RegA	RegA	+	86078	86503	Late	Late	Late	Late
S-PM2d130	heat shock protein	Hsp20	+	86654	87064	Late	Late	Late	Late
S-PM2d131	hypothetical protein		+	87134	87601	Late	Early	Early	Early
S-PM2d132	hypothetical protein		+	87598	87849	Early	Early	Early	Early
S-PM2d133	DNA polymerase	gp43	+	87849	90341	Late	Late	Late	Middle
S-PM2d134	UvsX RecA-like	UvsX RecA-like	+	90352	91383	Late	Early	Early	Early
S-PM2d135	DNA primase-helicase	gp41	+	91343	92755	Late	Early	Early	Middle
S-PM2d136	Pyrophosphatase	MazG	+	92757	93164	Late	Early	Early	Early
S-PM2d137	hypothetical protein		+	93124	93300	Late	Early	Early	Early
S-PM2d138	hypothetical protein		+	93485	95260	Late	Late	Late	Late
S-PM2d140	hypothetical protein		+	95257	97017	Late	Late	Late	Late
S-PM2d141	possible cytitidyltransferase		+	97014	97289	Late	Late	Late	Late
S-PM2d142	hypothetical protein		+	97264	97989	Late	Late	Late	Late
S-PM2d143	hypothetical protein		+	97998	98273	Late	Late	Late	Middle
S-PM2d144	hypothetical protein		+	98302	98496	Late	Late	Late	Late
S-PM2d145	virion structural protein		+	98590	99540	Late	Late	Late	Late
S-PM2d146	hypothetical protein		+	99547	99903	Late	Late	Late	Late
S-PM2d147	hypothetical protein		+	99876	100589	Early	Late	Late	Late
S-PM2d148	hypothetical protein		+	100540	101034	Late	Late	Late	Late
S-PM2d149	DNA primase subunit	gp61	+	101031	102020	Late	Late	Late	Late

S-PM2d150	ribonucleotide reductase A subunit	NrdA	+	102005	104335	Late	Late	Late	Middle
S-PM2d151	ribonucleotide reductase A subunit	NrdB	+	104316	105500	Late	Early	Late	Middle
S-PM2d152	hypothetical protein		+	105656	106003	Late	Late	Late	Late
S-PM2d153	hypothetical protein		+	105990	106685	Late	Late	Late	Late
S-PM2d154	putative lysin		+	106955	107173	Late	Late	Late	Late
S-PM2d156	hypothetical protein		-	107396	107566	Early	Late	Late	Middle
S-PM2d157	hypothetical protein		-	107571	107726	Late	Late	Late	Late
S-PM2d158	hypothetical protein		+	107894	108256	Late	Late	Late	Late
S-PM2d159	hypothetical protein		+	108287	108685	Late	Late	Late	Late
S-PM2d160	hypothetical protein		-	108647	109003	Late	Late	Late	Late
S-PM2d247	tRNA1-Arg		+	108699	108773	Late	Late	Late	Late
S-PM2d248	tRNA2-Lys		+	109108	109181	Late	Late	Late	Late
S-PM2d161	tRNA3-Tyr		+	109220	109305	Late	Early	Late	Middle
S-PM2d162	tRNA4-Asp		+	109306	109380	Late	Late	Late	Middle
S-PM2d163	tRNA5-Glu		+	109437	109512	Late	Early	Early	Middle
S-PM2d164	tRNA6-Ile		+	109513	109586	Early	Late	Late	Middle
S-PM2d165	tRNA7-Ser		+	109589	109681	Early	Late	Late	Middle
S-PM2d167	tRNA8-Ser		+	109682	109767	Early	Late	Late	Middle
S-PM2d168	hypothetical protein		+	109704	109853	Late	Early	Early	Early
S-PM2d169	tRNA9-Met		+	109770	109842	Early	Early	Early	Early
S-PM2d170	tRNA10-His		+	109843	109917	Early	Early	Early	Early
S-PM2d171	tRNA11-Gln		+	109918	109993	Early	Early	Early	Early
S-PM2d172	hypothetical protein		+	109966	110265	Late	Early	Early	Early
S-PM2d173	tRNA12-Trp		+	109993	110068	Late	Late	Late	Late
S-PM2d174	tRNA13-Arg		+	110067	110141	Late	Late	Late	Late
S-PM2d175	tRNA14-Val		+	110182	110254	Late	Late	Late	Late
S-PM2d176	hypothetical protein		+	110271	110555	Late	Late	Late	Late

S-PM2d176	tRNA15-Pro		+	110695	110769	Late	Late	Late	Late
PsbA intron	glutaredoxin	NrdC	+	110804	111049	Late	Late	Late	Late
S-PM2d177	PhDYefM tox-ant domain		+	111034	111249	Late	Late	Late	Late
S-PM2d178	hypothetical protein		+	111246	111485	Early	Late	Late	Late
S-PM2d179	hypothetical protein		+	111491	111775	Late	Late	Late	Late
S-PM2d180	Tryptophan Halognenase		+	111772	113247	Late	Late	Late	Late
S-PM2d181	tRNA16-Gly		+	113455	113527	Early	Late	Late	Late
S-PM2d182	putative S-layer protein		+	113572	114411	Early	Late	Late	Late
S-PM2d184	hypothetical protein		+	114418	114768	Late	Late	Late	Late
S-PM2d185	High Light Inducible Proteins	hli1	+	114818	115012	Late	Late	Late	Late
S-PM2d186	hypothetical protein		+	115098	115298	Late	Late	Late	Late
S-PM2d187	High Light Inducible Proteins	hli2	+	115295	115414	Late	Late	Late	Late
S-PM2d188	virion structural protein		-	115485	118772	Late	Late	Late	Late
S-PM2d189	hypothetical protein		-	118830	122006	Late	Late	Late	Late
S-PM2d190	virion structural protein		-	122017	125550	Early	Early	Early	Middle
S-PM2d191	photosystem II D1 protein	PsbA	+	125782	126783	Early	Late	Late	Middle
S-PM2d192	photosystem II D1 protein	PsbA	+	126996	127073	Late	Late	Late	Late
S-PM2d193	experimentally confirmed Group I intron		+	126784	126996	Late	Late	Late	Middle
S-PM2d194	homing endonuclease	F-Cphl	+	127140	127568	Late	Late	Late	Middle
S-PM2d195	hypothetical protein		+	127565	128194	Late	Late	Late	Late
S-PM2d196	photosystem II D2 protein	PsbD	+	128276	129337	Late	Late	Late	Late
S-PM2d197	hypothetical protein		+	129370	129882	Early	Late	Late	Middle
S-PM2d198	glutaredoxin	NrdC	+	129886	130134	Early	Early	Early	Middle
S-PM2d199	hypothetical protein		+	130131	130310	Late	Early	Early	Early
S-PM2d200	hypothetical protein		+	130326	130562	Early	Late	Late	Late
S-PM2d201	rnh RNaseH		+	130549	131412	Late	Late	Late	Late

S-PM2d202	hypothetical protein		+	131409	131843	Late	Late	Late	Late
S-PM2d203	hypothetical protein		+	131825	132112	Early	Late	Late	Late
S-PM2d204	hypothetical protein		+	132109	132372	Late	Late	Late	Late
S-PM2d205	SpeD	SpeD	+	132379	132711	Early	Late	Late	Late
S-PM2d206	hypothetical protein		+	132902	133072	Early	Late	Late	Late
S-PM2d207	hypothetical protein		+	133065	133178	Late	Late	Late	Late
S-PM2d208	hypothetical protein		+	133214	133369	Late	Late	Late	Late
S-PM2d209	Td thymidylate synthetase	Td	+	133451	134086	Late	Late	Late	Late
S-PM2d210	hypothetical protein		+	134089	134328	Late	Late	Late	Late
S-PM2d211	P-Starvation inducible protein	PhoH	+	134335	135087	Early	Late	Late	Late
S-PM2d212	exonuclease		+	135087	135776	Late	Late	Late	Late
S-PM2d213	late promoter transcription accessory protein	gp33	+	135754	136002	Late	Late	Late	Middle
S-PM2d214	DUF1825	DUF1825	+	135999	136346	Late	Late	Late	Late
S-PM2d215	ssDNA binding protein	gp32	+	136425	137312	Late	Late	Late	Late
S-PM2d250	base plate wedge component	gp53	-	137337	137999	Early	Early	Early	Early
S-PM2d216	baseplate tail tube cap	gp48	-	137999	138997	Late	Early	Early	Early
S-PM2d217	gp2	gp2	-	138997	139659	Late	Late	Late	Late
S-PM2d218	head completion protein	gp4	-	139662	140099	Late	Late	Late	Late
S-PM2d219	hypothetical protein		+	140133	140921	Late	Late	Late	Late
S-PM2d220	baseplate hub subunit	gp26	+	140927	141643	Late	Late	Late	Late
S-PM2d221	gp51	gp51	+	141646	141819	Late	Late	Late	Late
S-PM2d222	hypothetical protein		+	141812	145315	Late	Late	Late	Late
S-PM2d223	hypothetical protein		+	145408	147657	Late	Late	Late	Late
S-PM2d224	hypothetical protein		+	147657	149051	Late	Late	Late	Late
S-PM2d225	baseplate hub - tail lysozyme	gp5	+	149081	152026	Early	Late	Late	Late

S-PM2d226	base plate lysozyme		+	152026	153120	Early	Late	Late	Late
S-PM2d227	hypothetical protein		+	153281	153625	Early	Late	Late	Late
S-PM2d228	hypothetical protein		+	153647	153778	Early	Late	Late	Late
S-PM2d229	tRNA17-Ile		+	153837	153912	Early	Late	Late	Late
S-PM2d230	tRNA18-Met		+	153915	153988	Early	Late	Late	Late
S-PM2d231	tRNA19-Leu		+	154033	154117	Late	Late	Late	Late
S-PM2d232	hypothetical protein		+	154143	154415	Late	Late	Late	Middle
S-PM2d251	antenna proteins	cpeT	+	154408	154935	Late	Late	Late	Late
S-PM2d233	tRNA20-Thr		+	154929	155003	Late	Late	Late	Late
S-PM2d234	tRNA21-Ala		+	155188	155262	Early	Late	Late	Late
S-PM2d235	putative ATPase		+	155233	155376	Late	Late	Late	Late
S-PM2d236	tRNA22-Met		+	155279	155356	Late	Late	Late	Late
S-PM2d237	hypothetical protein		+	155436	155594	Early	Late	Late	Late
S-PM2d238	tRNA23-Asn		+	155600	155673	Early	Late	Late	Late