Microbial ‘Parasites’ meet Climate Change: Marine Cyanobacteria as Viral Puppets.

Tamsin Redgwell

A thesis submitted for the degree of Doctor of Philosophy in Biological Sciences

School of Life Sciences,
University of Warwick
July 2019

Table of Contents

Acknowledgements.

I would firstly like to thank my supervisors Prof. David Scanlan and Dr. Andrew Millard whose unique blend of 'tireless optimism' and 'dense questioning' have helped me through my PhD. Your guidance, advice, and intolerance of one-word answers has made me a better scientist.

I would also like to thank all members past and present of the Scanlan and Millard groups, who have made my PhD a much more enjoyable, coffee, and cake-filled experience. Particularly to Branko and Blanca for showing me the basics when I first arrived. To Fran, whose optimism and interest in so many things will always continue to be an inspiration. To Aurelija, for always being ready for coffee, and the endless supply of cute animal videos. Finally to Rich, the most elderly (and wisest) thirty-something I know, thank you for knowing how to do everything I needed to know your help was/still is greatly appreciated.

To my friends outside of Warwick, Ash & Alex, House J, and Josie for reminding me of 'real life', and putting up with long periods of silence.

To my family, thank you – I promise I'll get a real job now.

Whilst this work may be my own, I have come to realise that a PhD is not something someone does by themselves. So to anyone who has helped me in the last three and a half years, thank you.

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 submitted in any previous application for any degree.

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

- The *psbAII* construct illustrated in Figure 5.21 Panel B was designed and ordered by a previous PhD student, Hayley Preston.
- Figure 3.13 was plotted by Dr Andrew Millard, University of Leicester.
- The script in Appendix 1 was written by Dr Andrew Millard, University of Leicester.

Summary.

Cyanophages play a significant role in oceanic biogeochemistry through lysis of their host and possession of auxiliary metabolic genes (AMGs). AMGs are homologues of host genes that can manipulate the metabolism of their cyanobacterial host during infection, and are widespread in nature. Many cyanophages have been found to contain genes that play roles in photosynthesis, including one of the most well studied AMGs, *psbA*. PsbA forms part of the core photosystem II complex and is thought to maintain host photosynthesis during infection to optimise phage infection.

A major aim of my thesis was to understand the function of the viral *psbA* and its associated regulatory elements. Components of the cyanophage S-PM2d *psbA* region were heterologously expressed in a *Synechocystis* sp. PCC6803 strain lacking its own *psbA* copies. I showed successful expression of a cyanophage *psbA* in *Synechocystis*; however, photophysiological measurements suggest it is not capable of maintaining photosynthesis. The reasons behind this require further investigation.

The identity of other AMGs in cyanophages is limited by the lack of a genetic system. Another aim was to create a random chemical mutagenesis system to create mutants of cyanophages and identify novel AMGs. I successfully used hydroxylamine mutagenesis to create mutants of S-PM2d with SNPs identified using high throughput sequencing. I characterised these mutants phenotypically, leading to the identification of novel gene functions, including a SNP in the S-PM2d *psbA* gene causing an amino acid change. This mutant shows markedly different infection kinetics compared to wild type.

This thesis makes an important contribution towards our understanding of how cyanophages manipulate host metabolism, through analysing the function of the cyanophage encoded *psbA* gene. Additionally, this work has led to the identification of novel gene functions through chemical mutagenesis. Through this we have increased our understanding of how cyanophages interact with their cyanobacterial hosts.

Terms Used:

ADP	Adenosine DiPhosphate
AMG	Auxiliary Metabolic Gene
asRNA	Antisense RNA
ASW	Artificial Sea Water
ATP	Adenosine TriPhosphate
BAM	Binary Alignment Map format
BLAST	Basic Local Alignment Search Tool
cDNA	Complementary DNA
CFU	Colony Forming Units
Cyt b6f	Cytochrome b6f complex
DCMU	3-(3,4-dichlorophenyl)-1,1-dimethylurea
DNA	DeoxyRibonucleic Acid
GT	Glucose Tolerant
F’	Instantaneous fluorescence
FD	Ferredoxin
Fm’	Maximal fluorescence after light exposure
Fq	Difference in fluorescence between Fm’ and F’
Fq/Fm	PSII operating efficiency
HE	Homing Endonuclease
HMM	Hidden Markov Model
HP	Hypothetical Protein
MOI	Multiplicity Of Infection
MWCO	Molecular Weight Cut Off
NADP	Nicotinamide Adenine Dinucleotide Phosphate

NCBI	National Centre for Biotechnology Information
ncRNA	Non-Coding RNA
OD	Optical Density
OEC	Oxygen Evolving Complex
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PET	Photosynthetic Electron Transport
PFU	Plaque Forming Units
PMF	Proton Motive Force
PPP	Pentose Phosphate Pathway
PQ	PlastoQuinone
PSI	Photosystem I
PSII	Photosystem II
PTOX	Plastoquinol Terminal Oxidase
RCF	Relative Centrifugal Force
rETR	Relative Electron Transport Rate
RNA	Ribonucleic Acid
RPM	Revolutions Per Minute
RT-PCR	Reverse Transcriptase PCR
SAM	Sequence Alignment Map Format
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SNP	Single Nucleotide Polymorphism
TEM	Transmission Electron Microscopy
WT	Wild Type

List of Figures

List of Tables

List of Appendices

Chapter 1: Introduction

1.1.0 Cyanobacteria and their global importance

Cyanobacteria are known to be amongst the oldest living organisms on the planet; their existence is documented as far back as 3.4 billion years ago(Schopf 1993). These ancient organisms are credited with driving the atmospheric composition we see today through the advent of oxygenic photosynthesis, and continue to play a vital role in oceanic photosynthetic processes now(Kirschvink *et al.* 2000). It is currently estimated that half of all global primary production occurs in the oceans photic zone, with the cyanobacterial contribution ~25% (Flombaum *et al.* 2013).

In addition to this, cyanobacteria also play important roles in other global biogeochemical cycles, such as the nitrogen cycle. Dinitrogen fixation by marine cyanobacteria is thought to be one of the most prevailing sources of 'new' nitrogen in the oceans(Martínez-Pérez *et al.* 2016). The colony forming N_2 fixing cyanobacterium *Trichodesmium* is among one of the most important contributors, along with the unicellular cyanobacterium group A (UCYN-A). Some estimates report the former to be responsible for as much as half of the total N_2 fixation in subtropical gyre biomes of the oceans (Hutchins *et al.* 2015) and the latter to account for up to 20% of N_2 fixation in other regions such as the tropical North Atlantic (Martínez-Pérez *et al.* 2016). Due to the predicted warming of the oceans and consequent increases in the areal extent of nutrient impoverished gyre regions (Polovina *et al.* 2008) the abundance of the two major marine cyanobacterial genera, *Synechococcus* and *Prochlorococcus*, is set to increase in the future (Flombaum *et al.* 2013). Similarly, levels of the N_2 fixing cyanobacterium *Trichodesmium*, which has been shown to display increased growth rates and metabolic activity under elevated CO_2 levels (Hutchins *et al.* 2015), are also predicted to increase. Due to their widespread abundance cyanobacteria have been shown to be vital in a number of important global cycles (Flombaum *et al.* 2013; Martínez-Pérez *et al.* 2016; Fike *et al.* 2008; Sundby *et al.* 1992; Cottingham *et al.* 2015). This contribution is thought to become more important with the changing conditions predicted in the future (Polovina *et al.* 2008).

Cyanobacteria occupy a wide range of ecological niches and can be found in a variety of environments, such as hot springs (Castenholz 1978; Alcamán *et al.* 2017), microbial mats (Stal 1995; Pessi *et al.* 2018), deserts (Lacap-Bugler *et al.* 2017), freshwater (Frangeul *et al.* 2008) and marine ecosystems (Flombaum *et al.* 2013). In oceanic waters the picocyanobacteria *Prochlorococcus* and *Synechococcus* dominate, but with *Synechococcus* inhabiting a much wider geographical range than *Prochlorococcus*. The latter genus is found in surface waters down to the base of the photic zone (ca. 150m depth), and dominates waters between 40°N and 40°S, with numbers up to 1-2x10^5 cells per ml in some regions (Johnson *et al.* 2006) making it the most abundant phototroph on Earth (Partensky *et al.* 1999b). Marine *Synechococcus* extend much further N and S into polar regions, being found in waters up to 82.5°N (Paulsen *et al.* 2016) as well as in more eutrophic coastal regions (Scanlan 2012; Scanlan et al., 2009; Partensky *et al.* 1999a). The large geographic extent occupied by *Prochlorococcus* and *Synechococcus* is mirrored by the high genetic diversity found in each genus. This includes high light- and low light-adapted *Prochlorococcus* ecotypes (Johnson *et al.* 2006; Chisholm *et al.* 2006) as well as a number of *Synechococcus* clades (Scanlan *et al*., 2009; Dufresne *et al.* 2008; Scanlan and West 2002) that have recently been assigned into ecological significant taxonomic units dictated by their ecological distribution (Farrant *et al.* 2016). A range of both environmental and biological factors determine the extent of these distributions.

Thus, the ocean is not a homogenous entity, but rather there are distinct ecological differences between oceanic provinces (Longhurst 2007), and conditions can vary throughout the water column. As such, the distribution of cyanobacteria can be influenced by a wide range of factors that can largely be divided into biotic and abiotic influences. Abiotic determinants include nutrient availability (Zwirglmaier *et al.* 2008), light levels (Zinser *et al.* 2007; Chisholm *et al.* 2006), and temperature (Flombaum *et al.* 2013; Johnson *et al.* 2006) amongst others. Biotic factors include grazing by heterotrophic protists (Christaki *et al.* 1999) and viral lysis (Proctor and Fuhrman 1990). The role of viruses in particular is poorly understood, but studies have shown

that they may be of great importance in shaping bacterial communities and thereby influencing global biogeochemical cycles (Fuhrman 1999; Weitz and Wilhelm 2012; Roux *et al.* 2016). Indeed, it has been noted that within communities of closely related cyanobacterial strains, the areas of the genome that do vary are often those that encode extracellularly exposed products that could serve as viral receptors (Avrani *et al.* 2011).

1.1.1 Ecology of Marine Viruses

Bacteriophages are viruses that specifically infect bacteria and are the most abundant biological entity on the planet (Suttle 2005). In every millilitre of seawater there are millions of viral particles, so it is unsurprising that through lysis of their host they can play a significant role in shaping assemblage diversity (Suttle 2007). One method by which phage maintain high levels of host diversity is through 'Kill-the-winner' dynamics (Thingstad and Lignell 1997). An increase in any one host lineage would lead to an increase in phage that targets this lineages receptor. This causes selection against this variant and provides an opportunity for concurrent lineages with differing receptor variants to be maintained (Rodriguez-Valera *et al.* 2009). For example, Yoshida *et al* (2008) observed an increase in cyanophage abundance correlated with an increase in host *Microcystis* numbers. Alternatively, cyanobacteria and their respective phage can co-occur, as is commonly observed with marine *Synechococcus* and *Prochlorococcus* species (Waterbury 1993; Suttle 2005; Millard and Mann 2006; Sandaa and Larsen 2006; Avrani *et al.* 2011; Clasen *et al.* 2013). This is usually explained by an evolutionary 'arms race' between phage and their respective host (Breitbart *et al.* 2018). The way in which viruses interact with their host is highly variable (Breitbart *et al.* 2018)and likely depends on a range of factors including the phage, the host, and the ecosystem in question.

1.1.2 Cyanophage Diversity

Marine cyanophages were first isolated ~25 years ago in the early 1990's (Wilson *et al.* 1993; Suttle and Chan 1993). The cyanophages isolated since

can generally be divided into three broad families: Myoviridae, Siphoviridae, and Podoviridae (Figure 1.1). These differ morphologically with Myoviruses possessing contractile tails, Siphoviruses having long, non-contractile tails, and the short non-contractile tails of Podoviruses (Ackermann 2003). As well as differing morphologically these families have distinct genetic identities.

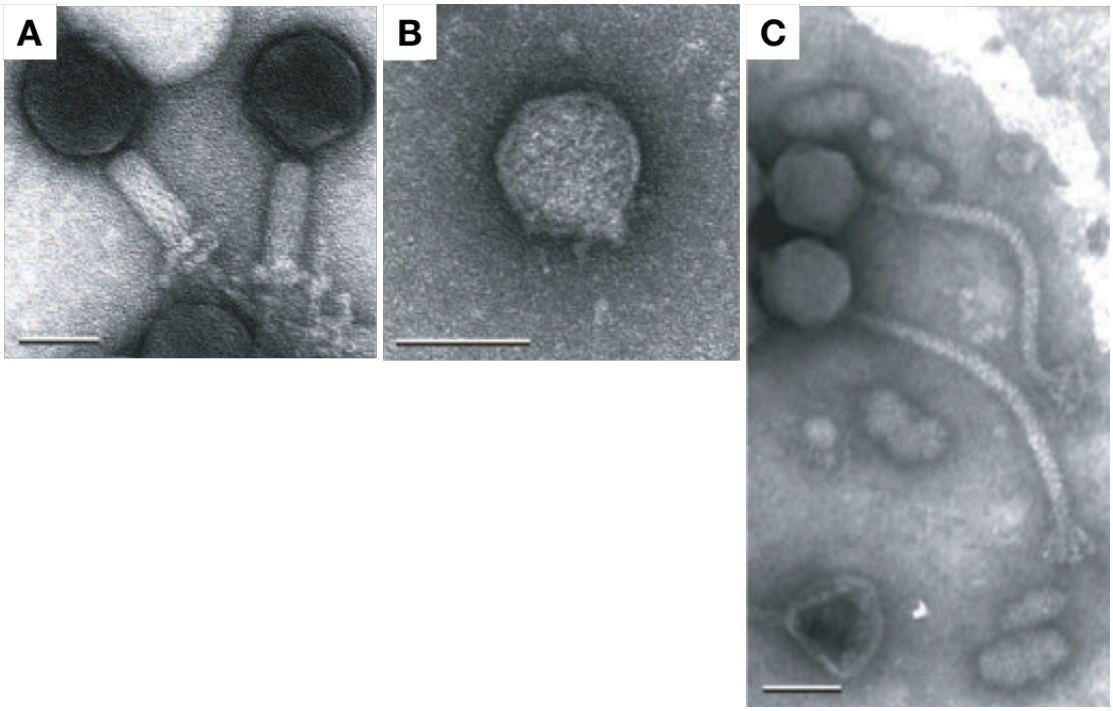


Figure 1.1. TEM of the three families of bacteriophages. A) Myovirus with a contractile tail. **B)** Podoviruses with short non-contractile tails. **C)** Siphoviruses with long non-contractile tails. Scale bar 50 nm. Taken from Suttle (2005).

Myoviruses typically have larger genomes and lower %GC content (Puxty *et al.* 2015), combined with the ability to infect a much broader host range then podo or siphoviruses (Enav *et al.* 2012; Dekel-Bird *et al.* 2015). Cyanopodoviruses are highly host specific and found in a range of marine habitats (Sullivan *et al.* 2003). Gene content studies have found that all known Cyanopodoviruses are genetically similar to the archetypal coliphage T7, but on average have a conserved genome size larger than that of T7 (42-47kbp as opposed to 37-39kbp) (Puxty *et al.* 2015). In contrast, Siphoviruses have a relatively broad host range and much more variable %GC content (Puxty, Millard, *et al.* 2015). Together, these three families encompass the enormous genetic and biological diversity that has been

discovered in cyanophages through genome sequencing and more traditional culture-based methods.

1.1.3 Viral shunt

As well as affecting host community diversity there are also significant ways in which phage can affect the wider ecosystem. Viral lysis of microbial hosts causes the release of previously trapped carbon and other nutrients back into the environment in a mechanism termed the 'viral shunt' (Wilhelm and Suttle 1999; Suttle 2005, 2007; Breitbart *et al.* 2018). This prevents these cells from being targeted by grazers and other organisms higher up the food chain. The released nutrients can then be utilised by microorganisms for other metabolic processes (Figure 1.2). Overall, this diversion of carbon and other resources leads to an increase in community respiration and a decrease in the amount of carbon transferred to higher trophic levels (Wilhelm and Suttle 1999; Suttle 2005). It also reduces the rate at which carbon sinks to the deep ocean where it can stay trapped for millennia (Suttle 2005; Breitbart *et al.* 2018). In this way we know that cyanophage can directly influence biogeochemical cycles but there are alternative ways in which they can do this, one of which is through the possession of auxiliary metabolic genes.

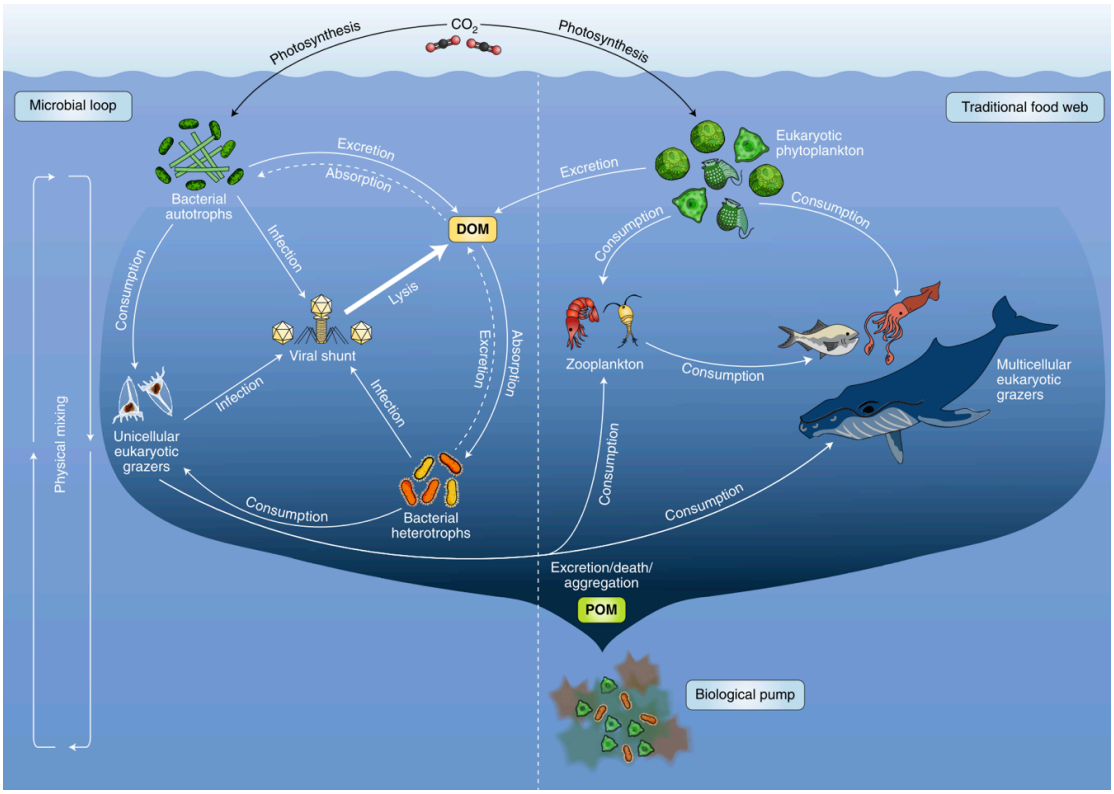


Figure 1.2. Overview of the marine food web. Highlights the role of the viral shunt in recycling dissolved organic matter (DOM). When bacteria are lysed by phages their carbon and organic matter is released into the ocean and flows through the viral shunt. Taken from Breitbart *et al.* (2018).

1.1.4 Auxiliary Metabolic Genes

In addition to diverting nutrient flow through lysis, phages can also affect global biogeochemical cycles through the expression of auxiliary metabolic genes (AMGs) (Breitbart *et al.* 2007). During phage infection the complement of 'accessory' genes present in the phage may affect the behaviour of the virus in question. Some of these genes are homologues of bacterial metabolic genes known as AMGs (Breitbart *et al.* 2007). AMGs often encode proteins involved in rate-limiting steps of host metabolism during infection, enabling viruses to boost these previously limiting steps and promoting viral propagation (Lindell *et al.* 2005). Table 1.1 shows a list of AMGs found in cyanophage, and their putative functions.

Table 1.1. List of auxiliary metabolic genes and proteins that have been identified in cyanophages, through both genomic and metagenomic analysis.

Gene/Protein Name	Predicted Function	Source	Reference
	Photosynthesis		
Ho1	Heme oxygenase	Genome	(Dammeyer *et al.* 2008)
pcyA	Phycocyanobilin:ferredoxin oxidoreductase	Genome	(Dammeyer *et al.* 2008)
pebS	Phycoerythrobilin synthase	Genome	(Dammeyer *et al.* 2008)
cpeT	Phycobiliprotein lyase	Genome	(Mann *et al.* 2005)
psbA	PSII component	Genome	(Mann *et al.* 2003)
psbD	PSII component	Genome	(Lindell *et al.* 2004)
speD	Polyamine biosynthesis	Genome	(Sullivan *et al.* 2005)
petE	Plastocyanin	Genome	(Lindell *et al.* 2004)
petF	Ferredoxin	Genome	(Lindell *et al.* 2004)
ptox	Plastoquinol terminal oxidase	Genome	(Weigele *et al.* 2007)
hli	Thylakoid associated protein	Genome	(Lindell *et al.* 2004)
psaA	PSI component	Metagenome	(Sharon *et al.* 2009)
psaB	PSI component	Metagenome	(Sharon *et al.* 2009)
psaC	PSI component	Metagenome	(Sharon *et al.* 2009)
psaD	PSI component	Metagenome	(Sharon *et al.* 2009)
psaE	PSI component	Metagenome	(Sharon *et al.* 2009)
psaJF	PSI component	Metagenome	(Sharon *et al.* 2009)
psaK	PSI component	Metagenome	(Sharon *et al.* 2009)
psaJ	PSI component	Metagenome	(Sharon *et al.* 2011)
	Carbon metabolism		
zwf	Glucose 6-phosphate dehydrogenase	Genome	(Millard *et al.* 2009)

gnd	6-phosphogluconate dehydrogenase	Genome	(Millard *et al.* 2009)
cp12	Calvin cycle inhibitor	Genome	(Sullivan *et al.* 2010a)
talC	Transaldolase	Genome	(Sullivan *et al.* 2010a)
	Nucleotide metabolism		
purH	Purine biosynthesis		
purN	Purine biosynthesis		
purL	Purine biosynthesis		
purS	Purine biosynthesis		
purM	Purine biosynthesis	Genome	(Sullivan *et al.* 2005)
noI	carbomoyltransferase		
nrdA	RNR domain		
nrdB	RNR domain		
nrdJ	RNR domain		
pyrE	Pyramidine biosynthesis		
nrdC	Aerobic thioredoxin		
	Phosphate stress		
phoA	Alkaline phosphatase	Genome	(Sullivan *et al.* 2010a)
pstS	Phosphate uptake	Genome	(Sullivan *et al.* 2005)
phoH	Phosphate stress-induced	Genome	(Rohwer *et al.* 2000)
	Cell protection		
prnA	Bacterial tryptophan halogenase	Genome	(Sullivan *et al.* 2005)
hsp	Heat shock protein chaperonin	Genome	(Sullivan *et al.* 2005)
nblA	Phycobilisome degradation	Genome	(Zhang *et al.* 2015)
	Other cell metabolism		
mazG	Pyrophosphohydrolase/ pyrophosphatase	Genome	(Sullivan *et al.* 2005)
cAMP phosphodiesterase		Genome	(Crummett *et al.* 2016)
carboxylesterase		Genome	(Sullivan *et al.* 2010a)
spoT	ppGpp synthetase/hydrolase	Genome	(Dreher *et al.* 2011)
nusG	Transcription anti-termination factor	Genome	(Dreher *et al.* 2011)
cobS	Vitamin B12 synthesis	Genome	(Sullivan *et al.* 2005)

ferrochetalase domain		Genome	(Sullivan *et al.* 2010a)
2OG-Fe(II) oxygenase		Genome	(Sullivan *et al.* 2010a)
Haemaglutanin neuraminidase	Viral attachment	Genome	(Sullivan *et al.* 2005)
Fatty Acid Desaturase			
sufA	Fe-S cluster assembly genes		
iscU	Fe-S cluster assembly genes		
glutaredoxin	Fe-S cluster assembly genes		
clpP	Fe-S cluster assembly genes		
moaA	Radical SAM superfamily genes		
nrdG	Radical SAM superfamily genes		
aslB	Radical SAM superfamily genes		
paaD	Radical SAM superfamily genes	Metagenome	(Hurwitz *et al.* 2015)
glgA	Glycogen synthase carbohydrate		
autotrans_barl	Ion transport and metabolism		
cysK/M	Fe-S cluster assembly genes		
iscA	Fe-S cluster assembly genes		
sensory box	Signal transduction		
rfoB			
galE	Carbohydrate transport/metabolism		
cyt_trans	Cell wall/membrane biogenesis		
bclB	Exosporium		
queA	Queosine biosynthesis protein		
ydeH	Signal transduction		
succinate semialdehyde dehydrogenase	Energy production /conversion		
fructose bisphosphate aldolase		Metagenome	(Sharon *et al.* 2011)
type I NAD(P)H dehydrogenase		Metagenome	(Sharon *et al.* 2011)

PDF	Peptide deformylase	Genome	(Frank *et al.* 2013)
ndhI	NAD(P)H dehydrogenase (NDH-1)	Metagenome	(Alperovitch-Lavy *et al.* 2011)
ndhD	NAD(P)H dehydrogenase (NDH-1)	Metagenome	(Sharon *et al.* 2011)
ndhP	NAD(P)H dehydrogenase (NDH-1)	Genome	(Nowaczyk *et al.* 2011)
denV	N-glycosidase	Genome	(Sullivan *et al.* 2005)
MsmA	Methanesulfonate monooxygenase	Genome	(Kang *et al.* 2013)
RNR	Ribonucloetide reductase	Genome	(Rohwer *et al.* 2000)
Thy1	Thymidylate synthase	Genome	(Rohwer *et al.* 2000)
RP (Roseophage) endodeoxyribonuclease I	endodeoxyribonuclease I	Genome	(Rohwer *et al.* 2000)
FlhA	Motility	Metagenome	(Dinsdale *et al.* 2008)
CheA	Chemotaxis	Metagenome	(Dinsdale *et al.* 2008)
CheB	Chemotaxis	Metagenome	(Dinsdale *et al.* 2008)

These genes are widespread in nature and metagenomic studies have shown metabolic genes of viral origin to be abundant (Sharon *et al.* 2007, 2009, 2011; Thompson *et al.* 2011; Hurwitz *et al.* 2015). This suggests that viral communities can act as reservoirs of this genetic material that might be advantageous for infection. Cyanophage contain AMGs that they may have acquired vertically from their host through transduction (Sullivan *et al.* 2006a). Once the gene has been acquired there is potential for it to continue to evolve within the virus, potentially developing modified or improved functions compared to the original. The *hli* multigene family in *Prochlorococcus* is thought to be an example of this. Multiple instances of re-acquisition of the *hli* gene from phage may have led to the multiple copies of the gene found in *Prochlorococcus* (Lindell *et al.* 2004). This gives it an advantage in particular environmental conditions and has led to niche differentiation and the formation of distinct high-light ecotypes (Lindell *et al.* 2004). Alternatively, AMGs can recombine in both directions between phage and their host. Zeidner *et al* found evidence of this type of intragenic recombination segments of the *psbA* gene, leading to novel arrangements of the sequence (Zeidner *et al.* 2005).

AMGs are associated with a range of functions including photosynthesis (Mann *et al.* 2003; Lindell *et al.* 2004; Dammeyer *et al.* 2008; Thompson *et al.* 2011), phosphate acquisition (Sullivan *et al.* 2010a; Zeng and Chisholm 2012; Martiny *et al.* 2009), carbon (Sullivan *et al.* 2010a; Thompson *et al.* 2011; Puxty *et al.* 2016) and sulphur metabolism (Anantharaman *et al.* 2014), stress tolerance(Lindell *et al.* 2004; Lindell *et al.* 2005; Kelly *et al.* 2013), rhodopsin (Philosof and Béjà 2013), and DNA/RNA processing (Sullivan *et al.* 2005; Pandolfini *et al.* 2013). The abundance of these AMGs can vary between phage isolates, but those that are common encode metabolic functions that are essential under a wide range of conditions. For example, *psbA*, encoding the D1 polypeptide of photosystem II (PSII) in cyanobacteria, is thought of as a 'nearly core' gene, having been found in almost all of the cyanomyoviruses (Crummett *et al.* 2016).

Work by Fridman *et al* (2017) previously identified a cyanophage that contains a PSI gene cassette in addition to the PSII genes *psbA* and *psbD*. Both sets were found

to be transcribed at the same time during infection; this could provide an important supplementary source of PSI proteins during this period (Fridman *et al.* 2017). This work also showed that whilst the rate of PSII activity remained constant during phage infection, the rate of PSI reduction increased. These results point towards a PSII-independent electron flow to PSI. This increased cyclic electron flow could provide the additional ATP required for phage DNA replication and nucleotide biosynthesis (Fridman *et al.* 2017).

It has also recently been found that cyanophage are capable of directly affecting CO_2 fixation in their *Synechococcus* host (Puxty *et al.* 2016). By comparing two cyanophage with differing carbon metabolism and photosynthetic electron transport gene complements, Puxty *et al* (2016) found that cyanophage-dependent inhibition of CO_2 fixation was evident early in the infection period. This is indicative of phage redirection of host photosynthesis for their own development. It is suggested that anywhere between 0.02 and 5.39 Pg of carbon could be 'lost' to this viral inhibition of CO_2 fixation per year (Puxty *et al.* 2016). This strategy of inhibition of host CO_2 fixation has important implications for our estimates of global primary productivity, and our understanding of the role of phage in this process. With an increase in cyanobacterial abundance thought likely to occur due to global climate change (Flombaum *et al.* 2013), greater appreciation of these interactions between host and virus will be of increasing importance in the future.

1.2.0 Oxygenic Photosynthesis

Oxygenic photosynthesis arose ~2.5 Ga ago and without this oxidation of the biosphere we would not have the complex multicellular life that exists today (Buick 1992; Payne *et al.* 2011). Oxygenic photosynthesis is the conversion of CO_2 and water to sugars using light energy (Vermaas 2001). This conversion is made up of two processes: the light reactions (photophosphorylation) and the dark reactions (the Calvin cycle). In cyanobacteria the light reactions occur in the thylakoid membrane, the internal membrane system that separates the cytoplasm

from the lumen (Vermaas 2001). A schematic representation of the key protein complexes involved can be found in Figure 1.3.

Primarily, photons are harvested by the chromophore molecules in the antennae complexes of PSII and PSI (Fassioli *et al.* 2014). This energy is then passed down the electron transport chain and used to reduce the special chlorophyll pair of P680 in PSII. P680 then donates its electron to pheophytin, the primary electron acceptor in the chain and then on to Q_A, the quinone binding residue. This oxidized $P680^+$ must be rapidly reduced and so H_2O is split to be used as an electron donor, a reaction that is extremely unfavourable (Lodish *et al.* 2000). This oxidation of water not only allows the reduction of $P680^+$, but also causes the acidification of the lumen through the release of H^+. This acidification is necessary for the maintenance of the proton motive force across the membrane. Plastoquinone (PQ) is then reduced to Plastoquinol (PQH2), a reaction that requires two electrons from PSII and $2H^+$ from the cytoplasm. This two-electron cost causes this step to become rate limiting due to the slow kinetics of the reduction. When it is reduced PQH2 can donate its electrons to cytochrome b6f, an electron transporter complex that in turn donates electrons to Plastocyanin (PC) – a luminal electron transporter of cyanobacteria. PSI uses the electrons it receives from PC to fill the electron hole created by the light – mediated reduction of the chlorophyll P700 reaction centre at its core. P700+ then donates electrons to ferredoxin. The resulting Ferredoxin-NADP+ reductase then catalyses the reduction of $NADP^+$ to NADPH, a process that requires one H^+ from the cytoplasm (Lodish *et al.* 2000).

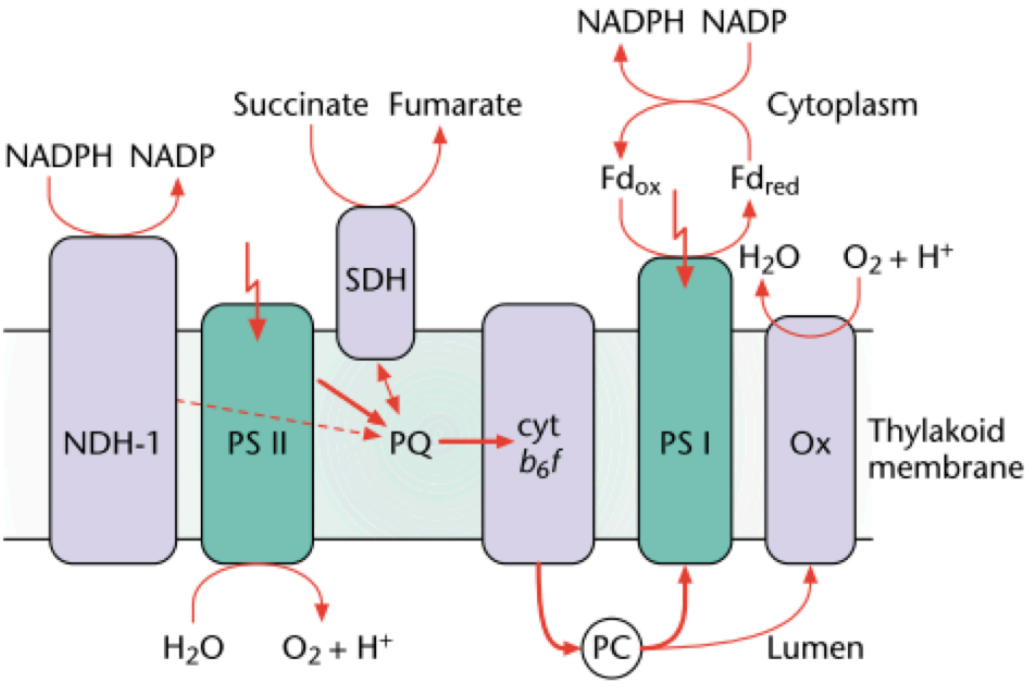


Figure 1.3. Representation of the key complexes involved in photosynthesis and the intersecting respiration pathway in *Synechocystis* PCC6803. Complexes specifically involved in photosynthesis are photosystems I & II (PSI & PSII). Those for respiration are NDH-1, SDH, and the terminal oxidase. PQ, cyt b_6f and PC are shared by both pathways. Taken from Vermaas (2001).

Throughout this process the translocation of protons maintains a proton motive force (PMF) that is sustained by both the proton concentration, and an electrochemical proton gradient (Vermaas 2001). The ATP synthase complex facilitates the movement of H^+ through its internal H^+ channel. This movement is a vital method of equilibrating the transmembrane potential required for the PMF. The net products of the light reactions of photosynthesis are NADPH and ATP. These are produced at a ratio of 2:3 – the exact ratio required by the Calvin Cycle (Allen 2003).

1.2.1 Carbon Fixation

The Calvin Cycle is the process of carbon fixation to produce carbohydrate. The process is fueled by ATP and NADPH from the light reactions (Berg *et al.* 2002). The transformations involved in this cycle can be divided into three steps:

1. **Carbon Fixation.** CO_2 combines with ribulose 1, 5–bisphosphate (RuBP) to produce a 6 carbon compound that splits into two molecules of 3 – phosphoglyceric acid (3-PGA), a three carbon molecule. This step is catalyzed by the enzyme RuBP Carboxylase/Oxygenase (RuBisCO).
2. **Reduction.** In this stage ATP and NADPH convert 3-PGA into glyceraldehyde – 3 – phosphate (G3P).
3. **Regeneration.** The final step of the Calvin Cycle is the regeneration of RuBP (Berg *et al.* 2002).

Cyanophages that infect the environmentally important photosynthetic cyanobacteria *Synechococcus* and *Prochlorococcus* are known to encode multiple AMGs. Many of these are involved in the light and dark reactions of photosynthesis (Crummett *et al.* 2016). This includes genes involved in electron transfer (Philosof *et al.* 2011), genes encoding reaction centre proteins (Mann *et al.* 2003), and those involved in carbon metabolism (Thompson *et al.* 2011). Details of these and further cyanophage auxiliary metabolic genes can be found in Table 1.1.

1.3.0 Cyanobacterial *psbA*

PSII is a large membrane-bound complex present in cyanobacteria, plants, and green algae, that is vital for oxygenic photosynthesis (Shen 2015). Its role involves mediating the light-induced, highly oxidative chemistry of water splitting into molecular oxygen. The D1 monomer, encoded by the *psbA* gene, provides cofactors and ligands required for this process (Shen 2015). Due to the oxidative nature of the reaction the D1 protein is subject to constant damage and needs to be replaced to prevent photo inhibition (Figure 1.4). In cyanobacteria under low light conditions this happens every 5 hours, but can occur up to every 20 minutes under intense illumination (Mulo *et al*. 2009).

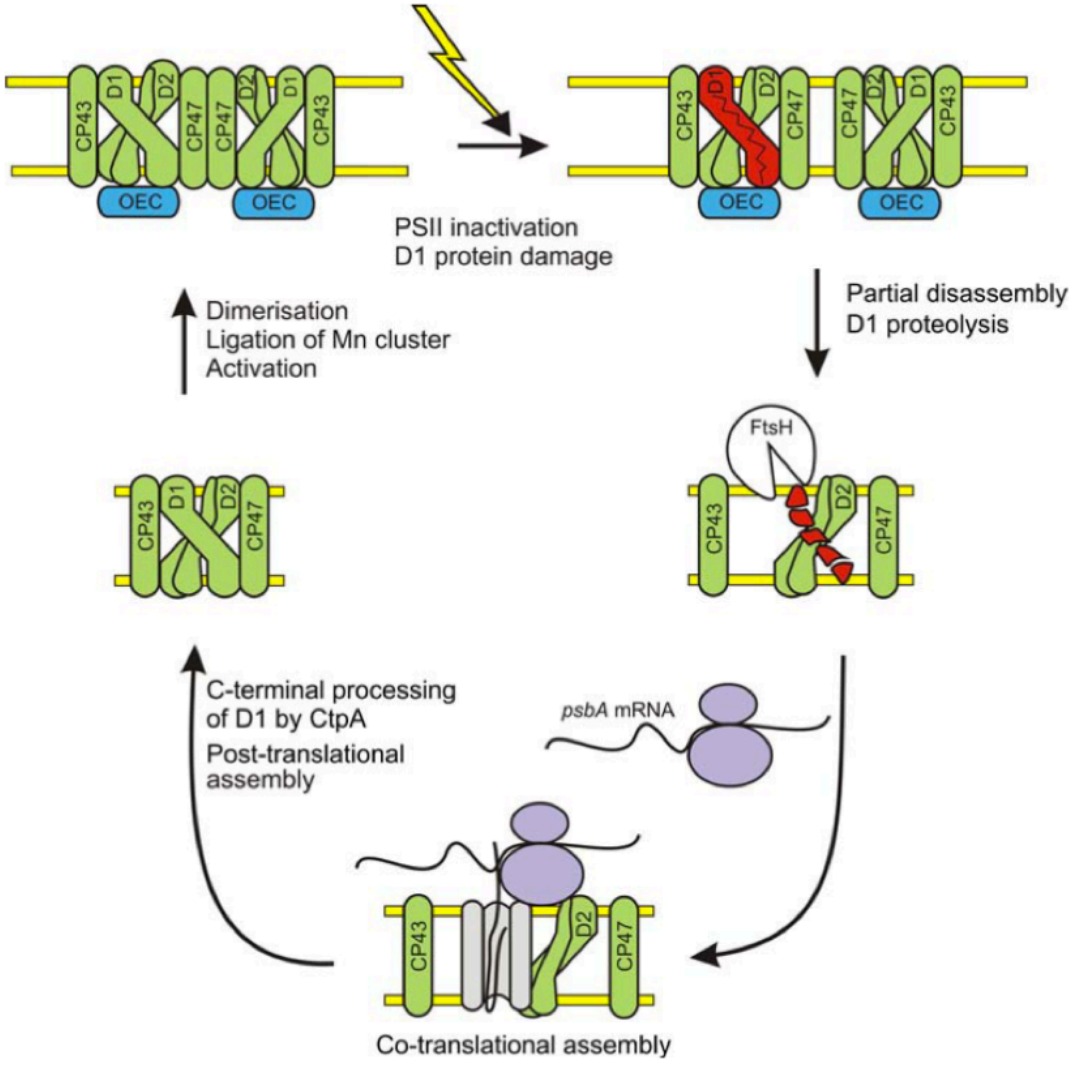


Figure 1.4 Photosystem II repair cycle. Light causes the inactivation of PSII dimers, and the D1 protein is damaged. The FtsH protease targets this for degradation and the subsequent ribosome – nascent D1 chain complex is targeted to the thylakoid membrane. Here the D1 protein is inserted into the PSII complex in the membrane. The C-terminus of the D1 protein is processed post-translationally and PSII is re-assembled into functional dimers. Taken from Mulo *et al*. (2009).

Unlike plants, all cyanobacterial strains have been found to encode a small family of *psbA* genes. The model species *Synechocystis* sp. PCC6803 has 3 copies of *psbA* that are differentially expressed under different conditions (Mohamed *et al.* 1993). *psbAII* and *psbAIII* encode forms of D1 that are identical at the amino acid level. 90% of transcripts are produced by *psbAII* in standard growth conditions, with the remaining 3-10% accounted for by *psbAIII* (Mohamed *et al.* 1993). However, in strains devoid of *psbAII* it has been shown that *psbAIII* transcript

levels increase to make up the difference so ultimately transcript levels remain the same (Huang *et al.* 2002). Previously thought to be a silent copy, *psbAI* is now known to be induced under microaerophilic conditions (Sicora *et al.* 2009a; Summerfield *et al.* 2008). These alternate genes presumably allow PSII to remain functional across a range of environmental conditions, a necessity for such a crucial complex.

In order to maintain functionality of the photosynthetic machinery over a range of conditions, expression of the *psbA* gene(s) is under strict control. This control has been demonstrated to occur at both the transcriptional and to a lesser extent, the translational level (Mulo *et al*. 2009). There are at least two distinct mechanisms for regulation of the *psbA* gene family under stress conditions that have been shown in multiple cyanobacterial species. The first of these is to replace the D1 protein present in the PSII reaction centres under non-stressed conditions with a different form when stress is detected (Clarke *et al.* 1993; Sane *et al.* 2002; Kulkarni and Golden 1994). One of the best examples of this behavior is the regulation of high light and low light isoforms. In the model organism *Synechococcus* PCC7942 three *psbA* genes encode the two D1 isoforms. In low light conditions *psbAI* encodes the D1:1 isoform that accounts for >80% of transcripts. However, in high light conditions there is a rapid change of D1:1 to D1:2, encoded by the *psbAII* and *III* genes. These genes have been shown to produce functionally different isoforms that allow the adaptation of *Synechococcus* cells to changing environmental conditions (Clarke *et al.* 1993; Kulkarni and Golden 1994). Alternatively, upon the induction of stress the turnover rate of the same D1 produced under basic growth conditions can be increased (Mohamed *et al.* 1993).

In addition to the regulation of *psbA* expression at the transcriptional level there are multiple other mechanisms that act only during translation. One such mechanism is the elongation of translation of *psbA* transcripts. These transcripts can be attached to ribosomes even in the dark and only become targeted to the thylakoid membrane when illuminated. Therefore, the completion of D1 synthesis can only be completed in the light, and as such the elongation of translation acts as a regulatory step (Tyystjarvi *et al.* 2001).

Another way in which cyanobacteria are able to regulate *psbA* expression is through the expression of antisense RNAs (asRNAs). Sakurai *et al* (2012) have demonstrated the presence of two asRNAs (PsbA2R & PsbA3R) of the *psbAII* and *psbAIII* genes from *Synechocystis* sp. PCC6803, located in the 5' untranslated regions. Whilst found in a low steady state abundance, their expression becomes up-regulated by light and down-regulated by darkness just as is witnessed for their target mRNAs. The creation of a PsbA2R(-) suppressor strain showed that the mRNA degraded at a much faster rate in the absence of the asRNA than in the wild type strain (Sakurai *et al.* 2012). This provides strong evidence for a role in the maintenance of mRNA stability and expression regulation for the asRNA PsbA2R. asRNAs have now been identified in a number of cyanobacteria (Georg and Hess 2011; Georg *et al.* 2009), including marine *Synechococcus* and *Prochlorococcus* (Gierga *et al.* 2012; Axmann *et al.* 2005), yet the role of most of these, with a few exceptions, remains unclear.

One of the few exceptions involves the cyanobacterial response to iron limitation. Cyanobacteria respond to the stress of iron limitation by expressing iron stress-induced protein A (IsiA) (Dühring *et al.* 2006). This protein forms additional antenna structures around PSI; in this way it enhances light adsorption and helps to relieve the stress from the reduction in PSI complexes caused by iron depletion. Once transcription of *isiA* is activated it is modulated by the cis-encoded antisense RNA IsrR (Iron stress-repressed RNA). The sequence of IsrR is complementary to the central third of the *isiA* mRNA. This leads to the transient formation of IsrR-*isiA* RNA duplexes that target them for degradation (Dühring *et al.* 2006). In this way IsrR acts as a method of posttranscriptional control, providing a reversible switch to respond to changes in environmental iron levels.

Additionally, in cyanobacteria an asRNA RblR has been found to positively regulate RuBisCO (Hu *et al.* 2017).This is the enzyme that catalyses carbon fixation and, in this way, RblR can potentially exert a level of control of cyanobacterial photosynthesis. RblR is completely complimentary to the target gene *rbcL*, that encodes the large chain of RuBisCO. Under various stress conditions, including

high light levels, the complementary base pairing of RblR to *rbcL* promotes the translation of *rbcL* mRNA into the RbcL protein(Hu *et al.* 2017).

1. 3.1 Cyanophage *psbA*

Cyanophages contain a diverse array of AMGs (see Table 1.1), a large proportion of which are associated with crucial steps and components of photosynthesis. *psbA,* the gene encoding the D1 protein of PSII was the first AMG to be identified (Mann *et al.* 2003) and through genome sequencing and metagenomic surveys has been shown to be widespread in nature (Crummett *et al.* 2016; Sharon *et al.* 2011; Angly *et al.* 2006; Roux *et al.* 2016; Sullivan *et al.* 2010a). Currently, the gene has been found in all but a few cyanomyoviruses (Crummett *et al.* 2016). In Podoviruses, metagenomic data shows that *psbA* is more abundant with 89% containing the gene. What's more, in this family the possession of the gene appears to be clade dependent; nearly all of clade B contain a copy, whereas no clade A podoviruses have yet been found to possess one (Zheng *et al.* 2013). However, the *psbA* gene has only been identified in 1/5 of the sequenced siphoviruses, P-SS1 (Puxty *et al.* 2015).

Whilst the function of the cyanophage encoded *psbA* has yet to be proven, the conservation in sequence between the cyanobacterial and cyanophage PsbA proteins suggests they likely perform the same function. Interestingly there are a few regions of the cyanophage encoded *psbA* that differ from that of the traditional cyanobacterial sequences (Sharon, Tzahor, Williamson, Shmoish, Man-Aharonovich, Rusch, Yooseph, Zeidner, Golden, Mackey, *et al.* 2007a). The first region resides within the PEST-like domain, located in the loop between transmembrane helices D and E. The second area is a stretch of variable residues after helix E but before the C-terminal residues of the D1 protein. These regions have both been implicated as having a potential effect on the stability of PSII complexes. The PEST domain is thought to be the initial site of cleavage to initiate degradation of D1. The second variable region likely comes into contact with PsbEF (cyt. B559), PsbJ and PsbV (cyt C-550) and could therefore play a vital role

in the correct assembly of PSII complexes (Sharon, Tzahor, Williamson, Shmoish, Man-Aharonovich, Rusch, Yooseph, Zeidner, Golden, Mackey, *et al.* 2007a).

Activity of the cyanophage *psbA* gene is supported by the ability to detect cyanophage *psbA* transcripts throughout the infection cycle, and that the phage D1 polypeptide abundance increases throughout infection (Clokie *et al.* 2006). As shown by Lindell *et al* (2005) phage-encoded *psbA* mRNA is also translated into protein. This evidence strongly suggests an active role for the cyanophage *psbA* during infection of the host.

The current view of phage *psbA* function is that it acts as a kind of viral 'life support', facilitating more successful infections for cyanophage. It does this through preventing photo-inhibition of the host; this occurs when the rate of D1 polypeptide damage exceeds the rate of synthesis to replace it and maintain photosynthetic capacity. With the lytic cycle of many cyanophage lasting longer than the D1 damage rate an auxiliary source of D1 would be required to maintain photosynthesis (Blot *et al.* 2011; Mella-Flores *et al.* 2012). The theory that this is provided by the phage encoded copy of the gene is supported by the high-light fitness advantage afforded to cyanophage with the *psbA* gene (Bragg and Chisholm 2008; Hellweger 2009). Sullivan *et al* (2006) have proposed that viral D1 sequences belonging to phage infecting *Synechococcus* strains are more similar to D1 Form II; in freshwater strains this is known to be the high-light and stress induced variant of the protein, as opposed to the 'housekeeping' Form I encoded by the host *Synechococcus* species (Sullivan *et al.* 2006b). Recent work by Puxty *et al* (2018) has shown that the cyanophage S-PM2 actively upregulates its copy of *psbA* in response to a higher light intensity, and that PSII photochemistry of the host remains stable in these conditions (Puxty *et al.* 2018). This would suggest that more energy would be available through photophosphorylation for phage development.

1.3.2 Genetic organisation of the cyanophage S-PM2 *psbA* region.

One of the most interesting things to note about the cyanophage S-PM2 *psbA* region is the genetic organization of this module (Figure 1.5). This *psbA* gene has

been found to contain a self-splicing group 1 intron (Mann *et al.* 2005), that is spliced throughout the infection cycle (Millard *et al.* 2010). These *psbA* introns are widespread in marine metagenomes with the *psbA* gene commonly occurring next to a gene encoding a homing endonuclease (HE) (Millard *et al.* 2010). Characterisation of a HE adjacent to *psbA* in cyanophage S-PM2 suggests that the former mediates the mobility of the intron. It has recently been discovered that F-CphI does not resemble any previously characterized HE and instead, more closely resembles an endonuclease. F-CphI is the first representative of this new family of homing endonucleases that uses the Endo VII catalytic motif (Fang *et al.* 2018). The fact that the region joining the 3' end of *psbA* and the 5′ end of F-CphI (HE) is occupied by an antisense non-coding RNA, CfrI, implies that the rearrangement of the HE gene into the *psbA* intron would lead to CfrI disruption and a lack of function (Fig. 1.5) (Millard *et al.* 2010). This would explain why a mobile group 1 intron has not been formed yet. It has been suggested that the regulation of gene expression through cis-encoded asRNAs constitutes a distinct layer of regulation in bacteria, through mechanisms such as imperfect base pairing with their target (Georg and Hess 2011). We know that the antisense ncRNA CFrI in cyanophage S-PM2 overlaps considerably with the 5' end of *psbA* (Millard *et al.* 2010). As such this could provide an elegant mechanism for the regulation of the cyanophage *psbA* gene. It is known that the *isiA* gene in *Synechocystis* sp. PCC6803 is regulated by the asRNA IsrR (Duhring *et al.* 2006), so this type of regulation is not uncommon and could be the mechanism of action in cyanophage S-PM2.

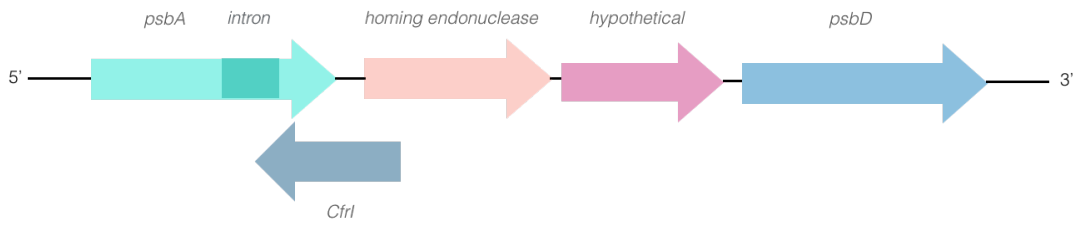


Figure 1.5 Genetic organisation of the cyanophage S-PM2 *psbA* region. The *psbA* gene is interrupted by a group 1 intron. Immediately downstream of *psbA* is a homing endonuclease. The antisense RNA CfrI overlaps the 3' end of the intron. Adapted from Millard et al. (2010).

1.4. Bacteriophages as model organisms

Bacteriophages have long served as model systems for use in genetic studies. It was phages that were used to unambiguously show that DNA, rather than protein is the hereditary material of life (Hershey and Chase 1952). The first gene (Jou *et al.* 1972), RNA genome (Fiers *et al.* 1976), and DNA genome (Sanger *et al.* 1977) to be fully sequenced were all phage–based. More recently, studies of phage resistance mechanisms in bacteria led to the discovery of the CRISPR/Cas system (Cong *et al.* 2013; Mali *et al.* 2013); This system is now at the forefront of targeted mutagenesis methods. Phages continue to be of importance many years after their vital contribution to basic biological understanding.

1.4.1 Bacteriophage Mutagenesis.

In the past, a wide range of mutagenesis methods have been used to determine the function of genes in bacteriophages. These can be split into two main categories: random mutagenesis and targeted mutagenesis.

1.4.2 Random Mutagenesis.

Random mutagenesis is the process of introducing changes into a DNA sequence without conscious design of these changes. Random mutagenesis occurs spontaneously in nature through processes such as erroneous DNA replication, but mutations can also be induced through the use of many physical or chemical mutagens. Traditional physical modes of mutagenesis include ionization, ultraviolet (UV) radiation, and heat treatment. In the cases of ionization and heat treatment the processes cause physical breaks in the DNA which can lead to changes in the sequence when repaired. In the case of ionizing radiation, the most common type of mutations caused are single base pair substitutions, although a lower frequency of frameshifts have also been recorded (Breimer 1988). The work of Tindall *et al* in the late 1980's showed the effects of a large number of lambda *cl* gene mutations caused by irradiation of lambda as a prophage. As well as identifying common mutations caused by irradiation such as double-stranded breaks, larger rearrangements of the DNA were also found (Tindall *et al.* 1988).

Heat mutagenesis has been thoroughly characterized previously by mutating bacteriophage T4 (Kricker and Drake 1990; Bingham *et al.* 1976; Baltz *et al.* 1976), but has more recently been used to isolate mutants of phage φX174 that can overcome heat inhibition due to the occurrence of point mutations in the genome (Bull *et al.* 2000).

UV radiation is strongly absorbed by bases that can cause pyrimidine dimers that leads to erroneous DNA replication. UV radiation has been widely used in the past to understand more about the basic biology of bacteriophages. It has previously been used to identify the function of, amongst others, the *dinB* and the *lex* genes in phage lambda (Brotcorne-Lannoye and Maenhaut-Michel 1986; Defais *et al.* 1971). More recently, UV mutagenesis has been applied to lactococcal phage TP901-1 to construct clear plaque mutants, therefore allowing the detection of genes involved in the lysogeny decision, such as the *cl* gene (Kot *et al.* 2016).

Chemical mutagens that have been widely used in mutagenesis include acridine dyes, base analogues such as 5-bromouracil, nitrous acid, and hydroxylamine amongst others. Acridine dyes are flat molecules that intercalate themselves between bases to create frameshift mutations. As well as being used as a mutagen itself, acridine derivatives have previously been used to sensitize organisms to increase the action of UV mutagenesis. The use of proflavine and UV was common and has been used on phage T4 to increase the production of 'rapid-lysis' mutants (Ritchie 1964). The mutagen 5-bromouracil substitutes itself for thymine in DNA, altering the physical structure. This has been shown to be capable of mutating bacteriophage DNA, and Benzer and Freese identified mutagenic 'hot spots', where the production of mutations was 10^4 times higher than would arise by spontaneous mutation alone (Benzer and Freese 1958). In contrast, nitrous acid and hydroxylamine cause single base pair changes causing the deamination of bases and transition mutations, respectively. The mutagenic potential of nitrous acid has previously been used on both the single stranded phage φX174 and the double stranded phage T4, in order to show that only a single base change was necessary for creation of a mutant (Tessman 1959).

Random mutagenesis requires no previous knowledge of the targeted genome. The basic premise is the construction of random mutations by use of chemical of physical methods and the identification of a phenotype. An association between the displayed phenotype and any changes in genotype can then be used to identify the functions of genes, as has been shown in the examples above.

1.4.3 Targeted Mutagenesis.

Targeted mutagenesis, alternatively known as site-directed mutagenesis, is the introduction of pre-designed changes to a specific site in the genome (Hodyra and Dąbrowska 2015). Arguably the most well-established method for achieving this is homologous recombination. Whilst this phenomenon occurs spontaneously in nature, it can be commandeered to incorporate foreign DNA into a genome, by designing the desired construct flanked by regions of homology (Lodish *et al.* 2000). This leads to a crossover event in these regions thereby constructing the desired gene insertions, deletions, or replacements. Homologous recombination can occur between regions of homology as short as 23 base pairs (Pires *et al.* 2016). However, when used in phage usually only a small percentage of the progeny phage will be recombinant, with rates of recombination often between 10^{-10} and 10^{-4} depending on the phage and gene being manipulated (Pires *et al.* 2016; Loessner *et al.* 1996).

There are a number of examples of the successful use of homologous recombination to create recombinant phages. One of these is the creation of a diagnostic bioluminescent phage that can be used for the detection of *Yersinia pestis*, by conferring a bioluminescent phenotype to the cells (Schofield *et al.* 2009). Homologous recombination was used to integrate luxAB ('light')-reporter genes into a non-coding region of the CDC plague-diagnostic phage φA1122. This recombinant reporter phage then transferred the bioluminescence phenotype to *Y. pestis* within 12 minutes of infection (Schofield *et al.* 2009). This is just one of the ways in which phages can be manipulated via targeted mutagenesis methods.

Alternatively, another frequently used method for engineering lytic phages is bacteriophage recombineering of electroporated DNA (BRED) (Marinelli *et al.* 2008). This method consists of electroporating the recombineering substrates (phage DNA and construct DNA) into electrocompetent cells that contain a plasmid that promotes high levels of homologous recombination (Marinelli *et al.* 2008). Recombinant phages are then identified through selective plating methods. As well as creating deletions, insertions, and gene replacements, BRED can also be used to create specific point mutations. Whilst originally constructed using mycobacteriophages (Marinelli *et al.* 2008), BRED has since been utilized to construct mutations in a range of bacteriophage systems.

One of the more recent examples of BRED is the use of the technique to engineer phage T7 for improved pH tolerance (Nobrega *et al.* 2016). The *E. coli* outer membrane phosphoporin protein E (PhoE) signal peptide was electroporated into phage T7, with regions of homology leading to fusion with the capsid. These recombinant phages showed increase tolerance to acidic pH levels and recombineering provided an alternative method for phage engineering.

The use of CRISPR-Cas-Mediated genome editing has become a much more popular method over the last decade (Cong *et al.* 2013; Mali *et al.* 2013). Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated components (Cas) form an adaptive immune system that protects microbial cells from DNA invasion (Barrangou *et al.* 2007; Deveau *et al.* 2010). Unlike other methods it can be utilized to cleave any desired DNA sequence, so any DNA and any part of the genome can be used as a counter selection marker (Kiro *et al.* 2014). CRISPR-Cas systems can be divided into three main categories (I, II, and III), with multiple sub-groups in each (Martel and Moineau 2014). The CRISPR-Cas9 system is one of the most widely utilized and has successfully been applied to a range of organisms including cells from humans (Cong *et al.* 2013; Mali *et al.* 2013), plants (Jiang *et al.* 2013), yeast (DiCarlo *et al.* 2013), bacteria (Jiang *et al.* 2013), and latent eukaryotic viruses (Ebina *et al.* 2013), as well as bacteriophages (Martel and Moineau 2014). CRISPR is based on a natural bacterial method of protection against viral infection. CRISPR loci are made up of 'repeats' and

variable regions of sequence, 'spacers'. When spacers are transcribed they guide nucleases (Cas proteins) to the target DNA by sequence homology. This is then cut, degraded, and repaired (Barrangou *et al.* 2007; Deveau *et al.* 2010). But by introducing specific 'guide' sequences to match the desired sequence, the targeted recognition of these sites can and has been utilized for mutagenic purposes in bacteriophages, providing a novel method for identifying the functions of uncharacterized genes amongst other things.

For example, the type II CRISPR-Cas system of *Streptococcus thermophilus* DGCC7710 has been used to edit the genomes of virulent phages, such as phage 2972, at a very high efficiency (Martel and Moineau 2014). Using point mutations, small frameshift deletions, and larger deletions, Martel *et al* were able to identify genes essential for phage 2972 propagation. They were also able to demonstrate complete gene replacement in 2972, by replacing a complete gene region with a methyltransferase gene from *L. lactis* that was shown to be functional in the recombinant phage (Martel and Moineau 2014).

The use of CRISPR-Cas in genome editing is not limited to one system. Kiro *et al* recently demonstrated the use of the type I-E system to engineer the *E. coli* phage T7. Using this system they were able to introduce scarless deletions of the 1.7 gene, encoding a nucleotide kinase, a non-essential gene for T7 propagation (Kiro *et al.* 2014).

Targeted methods of mutagenesis offer greater control over changes in DNA sequence. However, they require more prior knowledge of the bacteriophage sequence in order to enact this change. Due to this, it could be argued that targeted mutagenesis is more restrictive, and limits the identification of novel gene functions provided more readily in random mutagenesis approaches. When combined with a screen for specific phenotypes, random mutagenesis could allow a wider range of mutations to be identified and better understood.

1.4.4 Cyanophage Mutagenesis.

Unlike bacteriophages, cyanophages currently lack a reliable genetic system. The mutagenesis methods known to work in bacteriophage have not been successfully

translated to use in cyanophages. To date, cyanophages have been found to share 64-65 core genes, representing only 19-32% of total gene content. The remaining 68-89% is made up of accessory and unique genes, many of which are genes of unknown function (Figure 1.6). The development of a mutagenesis system for use in cyanophage could illuminate the possible functions of these uncharacterized genes, by associating a change in phenotype to a particular genotype.

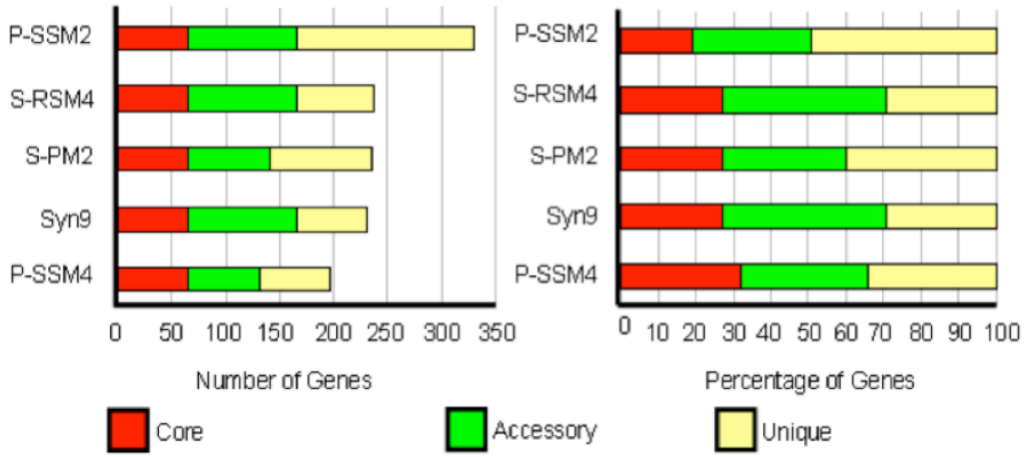


Figure 1.6 The core and accessory genomes of marine cyanomyoviruses. Core genes are present in all, accessory are present in more than one, but not all cyanomyoviruses. Unique genes are only present in one cyanomyovirus. Taken from Millard et al. (2009).

1.5 PhD Aims

a) To develop a random chemical mutagenesis system for use in cyanophage, to help illuminate the role of potential AMGs. By testing any cyanophage mutants for phenotypes we can associate these with any identified changes in the genome (Chapters 3 & 4).

b) To understand the role of the viral S-PM2 *psbA* during infection of its *Synechococcus* host. It is thought that the viral PsbA is used to maintain host photosynthesis whilst it is infected. We will therefore test to see if the viral copy is capable of maintaining photosynthesis in the host. (Chapter 4).

c) To understand the role of the associated regulatory elements of the cyanophage S-PM2 *psbA* region. This region contains the *psbA* gene containing an intron, along with an antisense RNA and homing endonuclease. The function of these elements is unknown. Using a heterologous system *Synechocystis* sp. PCC6803, and by replacing the wild type *psbA* genes with variations of this region we sought to identify the possible functions of these elements. (Chapter 5).

Chapter 2: General Methods

Methods described in this chapter are referred to multiple times throughout this work. More specific methods are explained within the relevant individual results chapter for reasons of clarity and fluidity.

2.1 Strains and plasmids.

Table 2.1 List of general strains and bacteriophages used in this work

Strain/Bacteriophage	Genotype	Comments	Reference
Strain			
Synechococcus sp. WH7803	WT	Host for S-PM2d and S-RSM4 cyanophages.	Waterbury *et al,* (1986)
Synechocystis sp. PCC6803 WT	WT	WT homologous system for recombination. Naturally transformable.	(Stanier *et al.* 1971)
Synechocystis sp. PCC6803 GT	Glucose Tolerant	GT homologous system for recombination. Naturally transformable.	
Synechocystis nWT	(DA1:DA2Em:DA3-H6) x (pRD1031) Gm^r, Km^r	'WT' control strain with a triple *psbA* knockout that has been transformed with plasmid pRD1031 to reconstruct the *psbA2* locus.	Nagarajan *et al* (2011)
Synechocystis DA1:DA3	DA1:DA3-H6 Gm^r	Double knockout strain, *psbA1* & *psbA3* deletion. PSII complex is his-tagged at the c-terminus.	Nagarajan *et al* (2011)
Synechocystis 'Triple deletion' (*psbA2*NS)	DA1:DA2Em:DA3-H6 Em^r, Gm^r	Triple *psbA* deletion strain where the *psbA2* locus is interrupted with an erythromycin cassette. PSII complex is his-tagged at the c-terminus.	Nagarajan *et al* (2011)
E. coli DH5α (K-12 deriv.)	F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15	General cloning strain.	Meselson & Yuan (1968)

E.coli MG1655 K12	Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ– F-, lambda-, rph-1	Host for SLUR29	(Blattner *et al.* 1997)
Bacteriophage			
SLUR29	WT	Control bacteriophage.	
S-PM2d	Deletion strain of S-PM2^{WT}	Deletion strain of S-PM2, contains deletion of ORFs p017-050.	Puxty *et al* (2015)
S-RSM4	WT		Millard & Mann (2006)

2.2 Growth media and conditions.

Synechococcus sp. WH7803 was grown in Artificial Seawater (ASW) medium (Tables 2.1 & 2.2) at 23°C, under illumination at 10-20 µmol photons m^{-2} s^{-1} with constant agitation. Cultures were routinely maintained in 100 ml volumes in 250 ml Erlenmeyer flasks, but larger volumes were grown up to 500 ml in 1 L flasks.

Table 2.2 Composition of Artificial Seawater medium, values for 1L.

Compound	***Mass (gL^{-1})***	***Final Concentration (mM)***
NaCl	25	428
$NaNO_3$	0.75	8.8
$MgCl_2$	2	10
KCl	0.5	6.7
$CaCl_2$	0.5	4.5
$MgSO_4$	3.5	11
Tris	1.1	9
K_2HPO_4	0.03	0.15
	Volume (ml)	
Trace metals solution	1	-

Table 2.3 Composition of ASW trace metals stock solution, values for 1 L

Compound	*Mass (gL^{-1})*	*Final Concentration (µM)*
H_3BO_3	2.86	46.26
$MnCl_2 – 4H_2O$	1.81	9.15
$ZnSO_4 – H_2O$	0.222	0.77
$Na_2MoO_4 – 2H_2O$	0.390	1.61
$CuSO_4 – 5H_2O$	0.008	0.03
$Co(NO_3)_2 – 6H_2O$	0.0494	0.02
$FeCl_3 – 6H_20$	3.0	11.10
EDTA (NaMg)	0.5	1.71

Synechocystis sp. PCC6803 was grown in BG11 medium (Tables 2.3 & 2.4) at 30°C under illumination at 25 µmol photons m^{-2} s^{-1} with constant agitation. Cultures were maintained in volumes of 100 ml in 250 ml flasks, but smaller volumes of 50 ml in 100 ml, and larger volumes of 500 ml in 1 L flasks were also grown. *Synechocystis* were also grown on solid BG11 medium with the addition of 1.5 % (w/v) Bacto Agar (Formedium). For strains where antibiotics were required, concentrations of the appropriate antibiotic were added as described in table 2.6.

Table 2.4 Composition of BG11 medium.

Compound	*Conc. (gL^{-1})*	*Vol (L^{-1})*
$NaNO_3$	1.5	
$K_2HPO_4 – 3H_2O$	0.04	
$MgSO_4 – 7H_2O$	0.075	
$CaCl_2 – 2H_2O$	0.036	
Group A solution		1
Trace metals solution		1

Table 2.5 Composition of the group A solution and trace metals solution required for BG11 medium.

Group A solution		***Trace metals Solution***	
Compound	***1000x Stock (gL^{-1})***	***Compound***	***Stock (gL^{-1})***
Citric acid	6	H_3BO_3	2.86
Ferric NH_4 citrate	6	$MnCl_2 – 4H_2O$	1.81
EDTA Na_2Mg	1	$ZnSo_4 – H_2O$	0.222
Na_2CO_3	20	$Na_2MoO_4 – 2H_2O$	0.390
		$CuSO_4 – 5H_2O$	0.08
		$Co(NO_3) – 6H_2O$	0.0494

Table 2.6 Antibiotic concentrations used to treat *Synechocystis* strains used in this study

Antibiotic	***Final Concentration***	***Strains treated***
Gentamicin	100μg/ml	Mutants constructed in the *Synechocystis sp*. PCC6803 GT & WT backgrounds. Also *Synechocystis* nWT and DA1:DA3-H6
Apramycin	50μg/ml	Mutants constructed in the DA1:DA2Em:DA3-H6 background
Erythromycin	1μg/ml	DA1:DA2Em:DA3-H6

Escherichia coli strains were maintained in Luria broth (LB). Growth on solid media contained 1.5 % (w/v) Bacto Agar. Expression of transformed strains happened in Super

Optimal Broth with Catabolite repression (SOC) media, with antibiotics added in the appropriate concentration as described in table 7. This media is used for obtaining higher transformation efficiencies of *E. coli.*

Table 2.7 Antibiotic concentrations used with *E. coli* strains

Antibiotic	***Final Concentration*** **(µg/ml)**
Gentamicin	10
Kanamycin	50
Spectinomycin	50
Ampiciliin	100
Apramycin	50

2.3 Propagation and concentration of cyanophages.

Cyanophages were regularly purified from liquid lysates of host *Synechococcus* sp. WH7803. Smaller volumes (100-200 ml) were concentrated using centrifugation with Vivaspin-15 ultrafiltration centrifugal concentrator (Sartorius) at 4000 rpm for 15 min at room temperature (Eppendorf Centrifuge 5810R). Centrifugation was repeated if necessary and concentrated phage were then resuspended off the filter with a small volume of fresh ASW medium. For larger volumes (1 L) cyanophages were concentrated with precipitation using polyethylene glycol (PEG). Lysate was spun down using the Avanti J-25/JLA 10.5 rotor at 8000 rpm for 20 min to pellet any cell debris. PEG 6000 was added to a final concentration of 10 % (w/v) and left overnight at 4°C in the dark with stirring to dissolve. The PEG mixture was precipitated by centrifugation at 8000 rpm for 30 min at 4°C

(Avanti J-25/JLA 10.5 rotor). Supernatant was discarded and the pellet resuspended in as small a volume of ASW possible by gentle pipetting. The resuspended pellet was treated with an equal volume of chloroform to remove any remaining debris and PEG, and the top layer was removed and stored at 4°C.

Lysates were further concentrated to a higher titre using CsCl ultracentrifugation. CsCl stock solutions were made with ASW to densities p=1.45, 1.5, and 1.7. Gradients of CsCl were made in ultracentrifugation tubes using 2 ml p=1.7, 3 ml p=1.5 and 3 ml p=1.45. 6 ml of Vivaspin concentrated or PEG precipitated phage was layered on top of the step gradient, and the tubes were ultracentrifuged at 35, 000 rpm in a SW40Ti rotor for 2 hours at 4°C. The band of concentrated phage was removed using aspiration, and then dialysed overnight at 4°C in 2L ASW to remove any CsCl. The membranes used for dialysis had a 12-14 kDa molecular weight cut-off and were activated according to the manufacturer's instructions (Visking-Medicell). Purified, concentrated lysate was then stored at 4°C in the dark.

2.4 Cyanophage enumeration.

Cyanophages were enumerated using variations on a plaque assay as described in Millard (2009). *Synechococcus* sp. WH7803 was grown to an OD_{750nm} of 0.35-0.45, concentrated ten-fold by centrifugation at 8000 g for 20 min at 23°C and resuspended in fresh ASW media. 0.375 ml of concentrated *Synechococcus* was pipetted into the wells of a 6-well plate. 10-fold serial dilutions of the cyanophage was made in ASW, and 25 µl were added to the concentrated cells. This was left to

adsorb for 1 hour at 23°C under illumination at 10-15 µmol photons m^{-2} s^{-1}. After this time 1.6 ml of molten 0.4 % (w/v) agarose was added to the wells and mixed. Plates were incubated at 23°C under constant illumination of 10-15 µmol photons m^{-2} s^{-1} until plaques had formed (~10 days). Plaques were counted using an appropriate dilution and the final titre was calculated as a mean of 3 replicates.

Alternatively, full plate plaque assays were also used to determine cyanophage titres according to Millard (2009). Again, *Synechococcus* sp WH7803 was grown to an OD_{750nm} of 0.35-0.45, concentrated by centrifugation and resuspended in fresh ASW. 25 µl of each 10-fold serial phage dilution was added to a glass screw top vial containing 0.5 ml concentrated cells. This was then left in the light at 23°C for at ~1 hour. 2.5 ml of warm 0.4 % (w/v) agar was added to each vial and gently mixed before pouring on top of a solid 1% (w/v) ASW agarose plate (made by adding 10 g agarose to 500 ml milliQ water and 500 ml of 2 X ASW). This was allowed to set for 1 hour before inverting and incubating at 23°C in constant illumination at 10-15 µmol photons m^{-2} s^{-1} until plaques had formed.

2.5 Cyanophage DNA extraction.

DNA was routinely extracted from different cyanophages using concentrated lysate. An equal volume of Tris-saturated phenol (pH 8) was added to the concentrated lysate, vortexed and left at room temperature for 5 min. The mixture was then centrifuged at 13, 000 rpm (Eppendorf Centrifuge 5424) for 5 min at room temperature. The top layer was removed and this initial step repeated. An equal volume of phenol:chloroform (25:24) was added to the tube, vortexed and

left at room temperature for 5 min. The tube was then centrifuged using the same conditions as previously stated and the resulting top layer was removed into a fresh tube. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the tube vortexed and left at room temperature for 5 min. The tube was centrifuged again and the top layer was added to an excess of ice cold 100 % (w/v) ethanol, with $1/10^{th}$ volume 3 M sodium acetate (pH 8) and left for between 1 and 12 hours at -20°C. Tubes were centrifuged for 40 min at 16, 000 rpm at 4°C (Eppendorf Centrifuge 5415R). The supernatant was removed and the pellet washed with 70 % (v/v) ethanol and centrifuged as before but for 15 min. The supernatant was removed and the pellet allowed to air dry before being resuspended in nuclease free H_2O. The extracted DNA was then cleaned and concentrated using the Zymo DNA clean and concentrator column (Zymo Research, Cambridge Biosciences) according to the manufacturer's instructions. DNA was stored at -20°C if not being used immediately.

2.6 RNA extraction.

Total RNA was extracted from *Synechocystis* sp. PCC6803 strains using an adaptation of the hot phenol method as described by (Scanlan *et al.* 1993). Briefly, 20 ml of *Synechocystis* cells were centrifuged at 4000 rpm at 23°C for 15 min before the supernatant was removed and the pellet resuspended in 20 ml wash buffer (0.05 M EDTA, 0.12 M NaCl). The resuspended cells were centrifuged again in the same way and the pellet resuspended in 0.8 ml extraction buffer (100 mM LiCl, 50 mM Tris-HCl, 30 mM EGTA, 1 % (w/v) SDS (pH 7.5). An equal volume of phenol pre-warmed to 65°C was added and the mixture incubated at 65°C for 5

min. The tube was centrifuged at 13, 000 rpm for 4 min at room temperature and the top layer extracted. The acid phenol step was repeated and the top layer again extracted to a fresh tube. An equal volume of phenol:chloroform:isoamyl alcohol was added and the tube vortexed and left at room temperature for 5 min. The mixture was then centrifuged again and the final aqueous layer was precipitated in 0.6 volumes of 100 % (v/v) ethanol with 1\10th volume 3 M sodium acetate (pH 8), and left at -20°C for 1-12 hours. The tubes were centrifuged at 16, 000 rpm for 40 min at 4°C (Eppendorf Centrifuge 5415R). The supernatant was removed and the pellet washed with 70 % (v/v) ethanol before being centrifuged again as before but for 15 min. The supernatant was removed and the pellet allowed to air dry before being resuspended in nuclease free H_2O. The extracted RNA was DNAse treated using DNase I according to manufacturer's instructions (New England Biolabs). It was then further cleaned and concentrated using the zymo RNA clean and concentrator column (Zymo Research, Cambridge Biosciences) according to the manufacturer's instructions, being eluted in a final volume of 25 µl. gDNA contamination was tested by PCR using primers 6803M_P2, the sequences of which are below, and target downstream of the *psbA* region in *Synechocystis* sp. PCC6803. Any samples that resulted in a PCR product were subject to a further DNAse treatment.

6803M_P2 F: 5' – GATATCGACCCAAGTACCGC – 3'

6803M_P2 R: 5' – GGATAAACCGCCTGATAGGTG – 3'

2.7 Preparation of chemically competent *E. coli* cells.

5 ml of LB was inoculated with a single colony of the relevant *E. coli* strain and incubated at 37°C overnight. 1 ml of this culture was then used to inoculate a 100 ml flask of LB and grown to an OD_{600} of 0.4-0.6. The cells were chilled on ice for 5 min before being centrifuged for 5 min at 2500 rpm at 4°C (Eppendorf Centrifuge 5415R). The resultant pellet was resuspended with 50 ml of cold filter-sterilised 0.1M $CaCl_2$. The cells were then incubated on ice for 30 min, then centrifuged as before for 15 min. The pellet was resuspended in 5 ml cold filter-sterilised $CaCl_2$:Glycerol (70:30). 30 µl aliquots were snap frozen using liquid N_2 and stored at -80°C until needed.

2.8 Chemical transformation of *E. coli*.

A 30 µl aliquot of competent cells was thawed on ice before 270 µl of ice-cold sterile $CaCl_2$ 0.1 M was added. 1-50 ng DNA was added to the cells which were then incubated on ice for 30 min. The cells were heat shocked at 42°C for 45 sec and then put back on ice for 5 min. 1 ml of pre-warmed SOC medium was added to the cells and they were incubated at 37°C for 1-2 hours. The tubes were centrifuged for 3 min at 3000 rpm at room temperature (Eppendorf centrifuge 5424). The SOC was discarded by inversion of the tube and cells resuspended using the residual 100 µl of medium. The 100 µl of culture was spread on an LB agar plate (1.5 % (w/v)) with any appropriate antibiotic included (see Table 7). Plates were incubated at 37°C overnight until colonies were visible.

2.9 Routine freezing of stocks.

100 ml of culture was centrifuged for 20 min at 4000 rpm at room temperature (Eppendorf Centrifuge 5810R) and resuspended in 6 ml of appropriate growth medium. 1.5 ml of cells were added to cryovials containing 75 µl of DMSO and snap frozen in liquid nitrogen before being stored at -80°C until needed.

2.10 Polymerase Chain Reaction (PCR) and Gel Electrophoresis.

PCR reactions were performed with GoTaq Green Master Mix 2X (Promega) according to the manufacturer's instructions, for 28 cycles at an annealing temperature appropriate for the primer pair being used. PCR products were run on ethidium bromide stained 1% (w/v) agarose (Sigma) gels, at 100V until the bands were visible three quarters of the way down the gel. Gels were imaged using the Syngene U:Genius 3 machine.

2.11 Flow Cytometry.

Synechococcus sp. WH7803 cell concentration was calculated by measuring the number of cells using a CytoFLEX flow cytometer (Beckman Coulter) and CytExpert v2.0 software. Events were counted for a fixed amount of time (180 secs), whilst maintaining an event rate of less than 500. The parameters counted against were PEA at a bandpass filter of 585/42nm, and PC7A at a bandpass filter of 780/60nm. Cells were then diluted to a concentration of $1x10^8$ cells per ml.

Chapter 3: Development of a random chemical mutagenesis system for use in cyanophages

3.1 Introduction.

The genomes of cyanophages contain a high proportion of accessory and unique genes (Millard *et al.* 2009). A clear roadblock in our understanding of the function of these genes is the lack of a genetic system for use in cyanophages. A commonly used method to identify the function of a gene is to mutate it and relate any corresponding change in phenotype to a difference in genotype. Ideally, a single nucleotide polymorphism (SNP) in a genome would allow a higher degree of certainty as to the cause of any phenotypic difference. There are a number of chemical mutagens that have been used historically to mutate bacterial DNA (Kada *et al.* 1972), plasmids (Humphreys *et al.* 1976), and bacteriophage (Tessman 1968). Some of these mutagens have a high degree of specificity, mutating only one of the bases in DNA, such as hydroxylamine.

Hydroxylamine is a chemical mutagen that has been characterised as making SNPs in DNA, thought to be limited specifically to causing C-T transition mutations (Tessman *et al.* 1964). Its mechanism of action is known to be split into two parts *in vitro*. First is a rapid action that is purely lethal and non-mutagenic, followed by a slower but highly mutagenic stage (Ernst Freese *et al.* 1961). Incubation with hydroxylamine allows the phage genome to be mutated whilst DNA is still packaged in the virion, before replication in the host cell (Robins *et al.* 2013). One benefit of this method is that due to the recovery of mutagenized phage particles on their host to form plaques, only those mutations that give rise to viable, infective phage particles will be maintained.

A major setback in random mutagenesis is often the devising of a technique for rapid screening of putative mutants. However, with hydroxylamine the probability of mutating a single gene can be made so close to 1.0 that this should no longer be an issue (Hall and Tessman 1966). Hydroxylamine has previously been used to mutate a number of bacteriophages such as T4 (Tessman 1968)and P22 (Venza Colon *et al.* 2004), and has more recently been coupled with deep sequencing methods to probe essential residues in T7 and virus JSF7 of *Vibrio cholerae* (Robins *et al.* 2013). As such, the method shows promise for application in the mutagenesis of cyanophages.

The aims of this chapter are:

a) To experimentally confirm that a 3-log drop in titre correlates to 1 SNP per genome by sequencing the genomes of putative mutant phages.
b) To develop a random chemical mutagenesis system for use in cyanophages, capable of creating single nucleotide changes.
c) To compare mutagenesis of S-PM2d to a different cyanophage, and identify if the mutagenesis behaviour is unique or more widely representative of cyanophages in general.

3.2 Methods.

3.2.1 Hydroxylamine Mutagenesis.

Phage mutagenesis was conducted according to Davis *et al* (1980). 0.4 ml of phosphate EDTA buffer (0.5 M KH_2PO_4 adjusted to pH 6 using 1M KOH, and 0.5 M EDTA) was mixed with 0.5 ml sterile water, 0.8 ml of freshly prepared hydroxylamine solution (1 M NH_2OH at pH 6 made fresh by adding 0.56 ml of 4 M NaOH to 0.35g of NH_2OH and adding sterile water to 5 ml), 0.1 ml of sterile 0.2 M $MgSO_4$ and 0.2 ml phage stock (between $1x10^4$ – $1x10^{13}$ depending on the phage). This gave a final hydroxylamine concentration of 0.4M. Three replicates were used for each phage and three replicates containing no hydroxylamine were used as a negative control. Samples were incubated for 24 hours at 23°C under illumination of 35 µmol photons m^{-2} s^{-1}, with samples taken at hours 0, 1, 6, 12, 18, and 24 by diluting into fresh medium. The mutagen was removed immediately by dialysing the sample through a Vivaspin centrifugal column (Sartorius) by centrifugation at 4000 g for 15 minutes in the Eppendorf centrifuge 5810R, using fresh medium. Phage were resuspended off the filter into 2 ml of fresh media by pipetting. A dilution series of each sample was then titered for plaques on the appropriate host and incubated until plaques appeared, as described in section 2.4.

3.2.2. Phage DNA preparation and extraction.

Due to the volume of phages being sequenced DNA was prepared from a single plaque. Using a sterile toothpick or pipette tip a plaque was picked and added to 200 µl ASW medium for cyanophages and LB for coliphages, and this was left overnight to form a lysate. 20 µl of this lysate was added to a 2 ml culture of

exponential phase host in order to propagate the phage. DNA was then extracted and concentrated following the method previously described in section 2.5.

3.2.3. Illumina Nextera XT library preparation and Sequencing.

Phages were sequenced using both the traditional method of Nextera XT library preparation (following manufacturer's instructions) and an adapted 2/5 reagent volume method described below. In both methods the protocol is the same, only the volume of reagents is different as can be seen in Table 3.1. In order to optimise the PCR ligator attachment step, different cycling conditions to the manufacturer's instructions were used. These were an increase in denaturation temperature to 97°C, and an increase in extension time to 1 minute at the lower temperature of 65°C, for a total of 16 cycles.

Table 3.1 Differences in reagent volumes between standard Illumina Nextera XT library preparation and the adapted 2/5 volume protocol.

Reagent	***Standard Illumina Protocol (µl)***	***Adapted 2/5 volume protocol (µl)***
DNA	5 µl (0.2ng/µl)	2 µl (0.7ng/µl)
Tagment DNA buffer (TD)	10	4
Amplicon Tagment Mix (ATM)	5	2
Index 1 (i7)	5	2
Index 2 (i5)	5	2
Nextera PCR Master Mix (NPM)	15	6
H2O (Library amplification step)	-	30
AMPure XP beads	30	25

3.2.4 Bioinformatic analysis of Illumina sequencing data.

Reads were trimmed with sickle (v1.33) using the paired end and Sanger options, with a quality cut off -30 and length cut off -30. Resulting reads were then mapped against the indexed reference (S-PM2d accession LN828717, S-RSM4 accession NC_013085, Slur29 accession ERS3527802) using bwa mem (v0.7.12-r1039) and default parameters. Samtools view was then used to make a bam file using the -b -S flags, and sorted using -sort. From the sorted bam file samtools(v1.6) was used to produce an mpileup file outputted as a text file, using -B -f arguments. VarScan (v2.3.6) was then used to call SNPs with the mpileup2snp option. Arguments for -min-coverage 10 -min-var-freq 0.90 -min-avg-qual 30 and -p-value 0.05 were used. The genomes of the mutants were also assembled, this was done using SPAdes (v3.12.0) with the paired end option and the '- - only- assembler' option.

3.2.5 Confirmation of predicted SNPs.

Predicted SNPs in the phages were confirmed using PCR (section 2.10) and primers that targeted a ~600 bp region containing the SNP of interest (Tables 3.2 & 3.3). PCR products were cleaned using the Wizard SV clean up kit (Promega) according to manufacturer's instructions and sent for Sanger sequencing using 5 µl (5 pmol/µl) of the same forward or reverse primer used in the PCR reaction, and 5 µl of the PCR product at a concentration between 20-80 ng/µl. This was done for all predicted SNPs in both directions in the coliphage and most of the SNPs predicted in cyanophage S-PM2d that were later sequenced with the MinION (Figure 3.2).

Table 3.2 List of Primer sequences used to confirm SNPs identified in Slur29.

Primer Name	Sequence 5'-3'
SL29_46000_F	GCGTCCCGTACACTTATCGT

SL29_46000_R	**GCGAGTACAAAACGCAAAGC**
SL29_2000_F	**TAAGACCCCAGCAATCGAAC**
SL29_2000_R	**CGCCCAAGCCAGTACAATAG**
SL29_7500_F	**CTAACAAGTTTGCCCCAGGA**
SL29_7500_R	**TGATGGCGAACGTTATGTTG**
SL29_13242F	**AACCAACCACAACCTTTTCG**
SL29_13242R	**TTCGCGGAATGAGAGATAGC**
SL29_39966F	**ACTGACCCGATCACAGCAAT**
SL29_39966R	**GCTGCGTTTGGCGTTATACT**
SL29_29921F	**CACTTACTGCCAATGCTGG**
SL29_29921R	**AAGCGCGCATATCTGAATAAA**
SL29_12724F	**ATTGCTCGCCATCTTCTTTG**
SL29_12724R	**TTGTGGTTGGTTCTTGTGCT**
SL29_21344F	**ACTTCATGCCCTGCCTTAAC**
SL29_21344R	**GAGCATGTGGAGCTTGTG**
SL29_36854F	**ACCACGATCGGAAAGTATGC**
SL29_36854R	**GTCTTTACCTCGCGTTC**
SL29_10752F	**CTGCCGGATACTCTTTCTCG**
SL29_10752R	**ATGATGAAGCGGCGTTATCT**
SL29_HA1_F	**ATCCCCTCTAGTCCGGTAGC**
SL29_HA1_R	**TCTTTCGCGGAATGAGTTTT**
SL29_HA2_F	**TGCCATCTACTTTGGTTTGC**
SL29_HA2_R	**CTTACCGGAACCTCTACGCC**
SL29_HA3_F	**GTTGGTCAGGCTGGCGTAAT**
SL29_HA3_R	**GAATTTACGGCGCAGTGCT**

Table 3.3 List of primers used to confirm the SNPs identified in the S-PM2d mutants that were later sequenced with the MinION.

Primer Name	Sequence 5'-3'
10473_SPM2dd_F	**AGCAAATTGGGGCAATTCTT**
10473_SPM2dd_R	**GCGAAGTACAATGGACCAAAA**
100793_SPM2dd_F	**TTCGGTGATGGATCAAAACA**
100793_SPM2dd_R	**TACTGCCCATTGTCAAGACG**
115714_SPM2dd_F	**TGGATCTGGTGGTGATGATG**
115714_SPM2dd_R	**AGATTTCACACTGGGGTTCG**
15951_SPM2dd_F	**GCCGATTTAATGCAGGATGT**
15951_SPM2dd_R	**CCTTCACACGACTCGTAGCA**
126728_F	**CTACATGCTTGTTTTCCAGGC**
126728#2_R	**CGTGCTAAGTTCGTGTAGAC**

3.2.6 Oxford Nanopore MinION sequencing.

Due to the higher concentration of DNA required for MinION sequencing (1 μg as opposed to 1 ng for Illumina Nextera XT) and the practical difficulty this presented using cyanophages, a much smaller number of S-PM2d mutant phages were sequenced using the MinION. DNA was quantified using the Qubit fluorometer (Life Technologies, Paisley, UK) using the High sensitivity ds DNA reagents according to the manufacturer's instructions. 1 μg of DNA per phage was used as input for the library preparation. Library preparation was done according to the manufacturers protocol for 1D Native barcoding for genomic DNA, using a FLO MIN 106 R9 flow cell and the native barcoding expansion EXP-NBD103 in conjunction with the SQK-LSK 109 ligation sequencing kit.

3.2.7 Bioinformatic analysis of MinION sequencing data.

Long reads from MinION sequencing were base called using Albacore (v1.2.1) and converted to FASTQ format using Poretools (v0.6.0). Genomes were assembled using Unicycler (v0.4.3) and default parameters). Hybrid assemblies using both short reads from Illumina sequencing and long reads from MinION sequencing were also performed. In this instance short reads were trimmed as described in section 3.2.4 and long reads were base called as mentioned above. Hybrid de novo assemblies were created using Unicycler (v0.4.3) and default parameters. Assemblies were viewed in Tablet Genome Browser (v1. 17.18.17).

3.2.8 Subsampling experiment analysis.

Multiple sets of primers were used to target 'deletions' identified in S-PM2d mutants (listed below). However, all of these proved unsuccessful in confirming that these were real. In order to test whether the deletions detected in Illumina sequencing were true or a result of problems with sequencing chemistry, a replicate sequencing experiment was carried out. 3 library preparations of the same Slur29 DNA extraction were prepared for Illumina Nextera XT, according to the manufacturer's instructions. Two preparations of a selected S-PM2d mutant were also used, as well as two preparations of the wild type S-PM2d DNA extraction. Afterwards, sequencing reads were mapped against reference genomes (S-PM2d accession LN828717) with bwa mem using default parameters. Deletions were identified as regions of no coverage using a custom script (appendix 1) on the assemblies from each library. The resulting replicate libraries were compared to see whether the same 'deletions' were found in all replicates of the same DNA extraction which would suggest that they were real. In addition to this, the reads from each replicate library of each phage were pooled together and the reads were subsampled using Seqtk (v1.0-r31) to different coverage levels, to identify if the appearance of 'deletions' was coverage dependent.

Table 3.4 List of primers used to try and confirm a subset of identified deletions in S-PM2d mutants.

Primer Name	Sequence 5'-3'
HAdel_171492_F	**GATACCAAGCACAGATCCAA**
Hadel_172024_R	**CCAGAACCATATCTCCAACC**
Del_HA_#1_F	**CTCCAGACACATATATATTCTC**
Del_HA_#1_R	**CTGAATAAGATAACTCAATTG**
Del_HA_#2_F	**GATGAATTATCTATTTCCAC**
Del_HA_#2_R	**GTATGAATAGGTGGATTTG**
HA_del_V2_F	**TCCCCTGGAATAACTTTGGA**
HA_del_V2_R	**AGGTGGTTCATCTGGACCTG**
HA_del_V2#1_F	**TTGCTGAAATTGAATCTGTTGA**
HA_del_V2#1_R	**TACCGCTGCTTGAGGTCTTT**

HA_del_V3#1_F	**CTGTTTCTTCAGAAAATGCAG**
HA_del_V3#1_R	**TTCTACTACCGCTGCTTGAGG**
HA_del_V3#2_F	**CTCACGCATTCAAGCAAATG**
HA_del_V3#2_R	**AAAGTTTTTCGCACTTTGGAA**
HA_del_V3#3_F	**CCAAGCACAGATCCAATTTTT**
HA_del_V3#3_R	**TAGTTTGCCCTTCTGCTACG**

3.2.9 Distribution of SNPs throughout the S-PM2d genome and in nature.

In order to check whether the SNPs present in the genome of S-PM2d were normally distributed, a Shapiro-Wilk test was conducted on the distance between each SNP, to check for normality. After this I wanted to see if the SNPs accumulated in particular regions of the genome, such as those parts described as 'hypervariable' or as 'core'. To do this I determined the hypervariable regions of the S-PM2d genome using the 91 viral metagenomes from the TARA dataset (accessions in appendix 2). Hypervariable regions have previously been described as regions that are under-represented from metagenomic datasets, specifically over 500 bp long and with coverage less than 20% of the mean (Mizuno *et al.* 2014; Enav *et al.* 2018). The reads from the metagenomes were mapped against the S-PM2d reference genome (accession LN828717) at an identity level of 70%, 80%, 90%, and 95% using bbmap (v37.54) and altering the 'minid' option as appropriate e.g 'minid=0.7' for 70%, the 'covstats' option was also used. I then determined regions of the genome of any length that had coverage of less than 20% of the mean, using a custom script (appendix 3) in R studio (v1.1.468). I also wanted to see if the SNPs created by mutagenesis were mutations that were present in nature. To do this I called SNPs on the reads that mapped to the S-PM2d genome at 90% identity from the 91 viral metagenomes used previously. I used

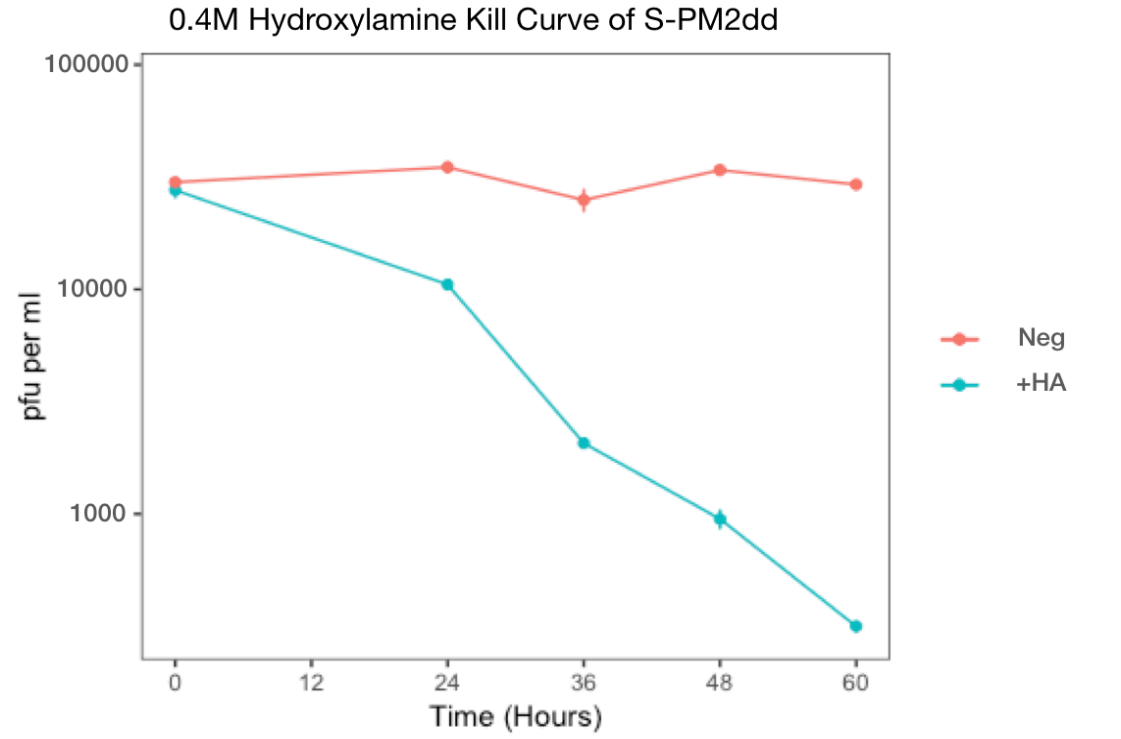
samtools (v1.6) to create a bam file of the mapped reads which I then converted to a sorted sam file using samtools sort. From this I made an mpileup file which I could read into varscan (v2.3.6) to call SNPs. SNPs were called using the following flags '-min-coverage 5' '-p-value 0.01' '-min-var-freq 0.1'.

3.3 Results.

3.3.1 Identifying an appropriate concentration of hydroxylamine.

Hydroxylamine mutagenesis relies on the concentration of the chemical mutagen being sufficient to incur the desired 3-log drop in titre, thought to correlate with one SNP per genome. The range of concentrations tested here showed no significant differences in the time taken to produce a 3-log drop in titre, as shown in Figure 3.1. For this reason, the highest concentration (0.8M) of hydroxylamine was used hereafter in all mutagenesis experiments, as higher concentrations of hydroxylamine have been shown to inhibit the initial non-mutagenic mechanism of the mutagen(Freese *et al.* 1961; Freese *et al.* 1961).

A



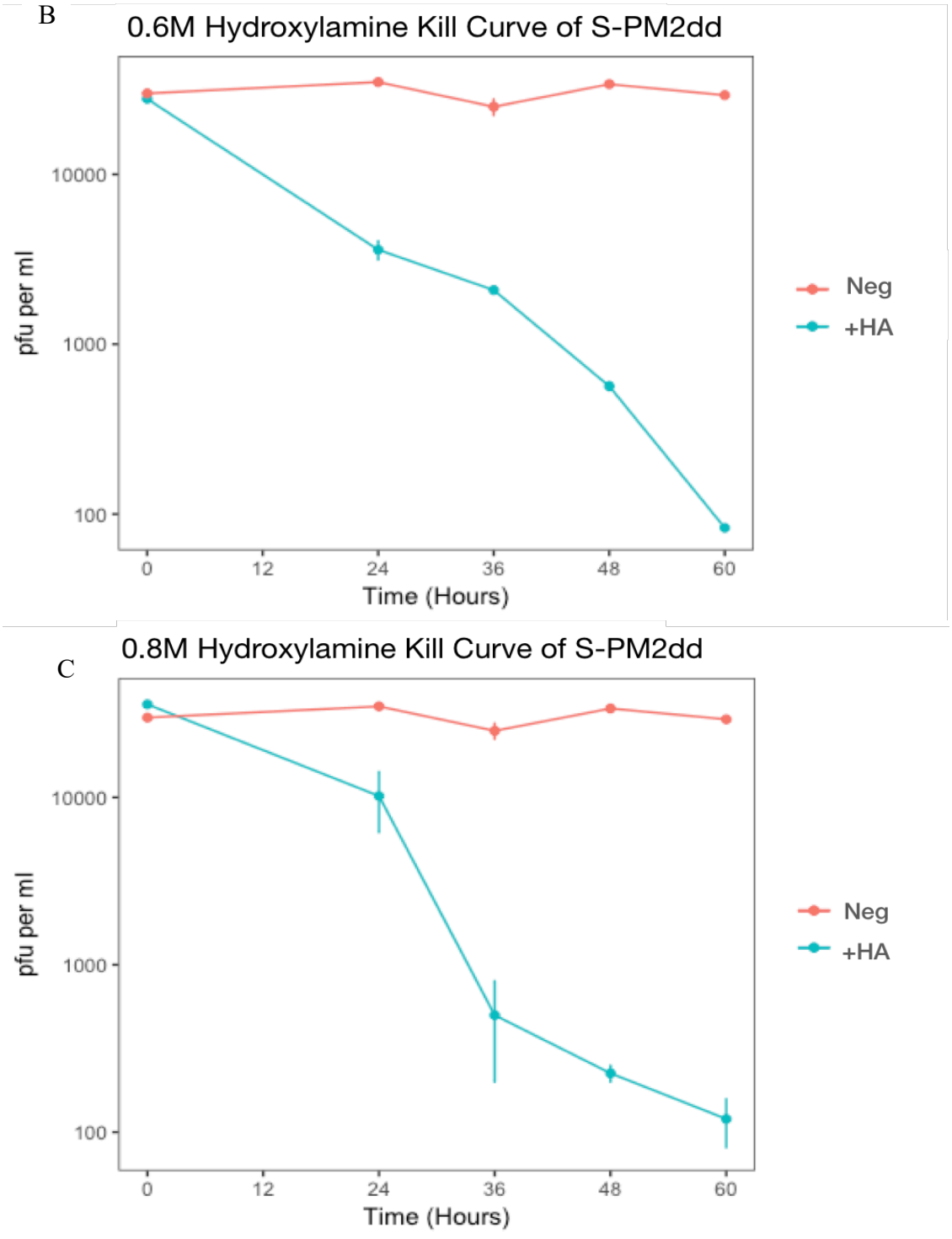


Figure 3.1 Kill curve of S-PM2dd when incubated with 0.4 M hydroxylamine (A), 0.6 M hydroxylamine (B) or 0.8 M hydroxylamine (C). In all graphs the red line represents the negative control of phage without any mutagen added. The green lines represent S-PM2dd when incubated with hydroxylamine. Error bars show standard deviation of the mean.

3.3.2 Confirming 3-log drop in titre correlates to 1 SNP per genome in a control coliphage.

Previously, the correlation between 1 SNP per genome and a 3-log drop in titre was established using coliphage (Freese *et al.* 1961). However, this work was done before the advent of next generation sequencing and so has never been confirmed using sequencing of the putative mutant coliphage. The correlation between a three-log drop in titre and one SNP per genome was tested by the sequencing of ~30 individual coliphage plaques from each timepoint in the mutagenesis. This allowed the identification of the time at which the average number of SNPs per genome equalled one, and whether that corresponded to the point at which the titre is three logs lower than the starting titre. The SNPs identified were confirmed by Sanger sequencing; an example gel of these results is shown in Figure 3.2. The average number of SNPs per genome in the coliphage reaches a single SNP after 18 hours (Figure 3.4). By comparing this to the kill curve for slur29 (Figure 3.3), this shows that this timepoint does indeed correlate with the time at which the titre has dropped by three logs. Therefore, a three-log drop in phage titre resulting from incubation with hydroxylamine, can be used as a proxy for the time at which the average number of SNPs per genome is one in a coliphage.

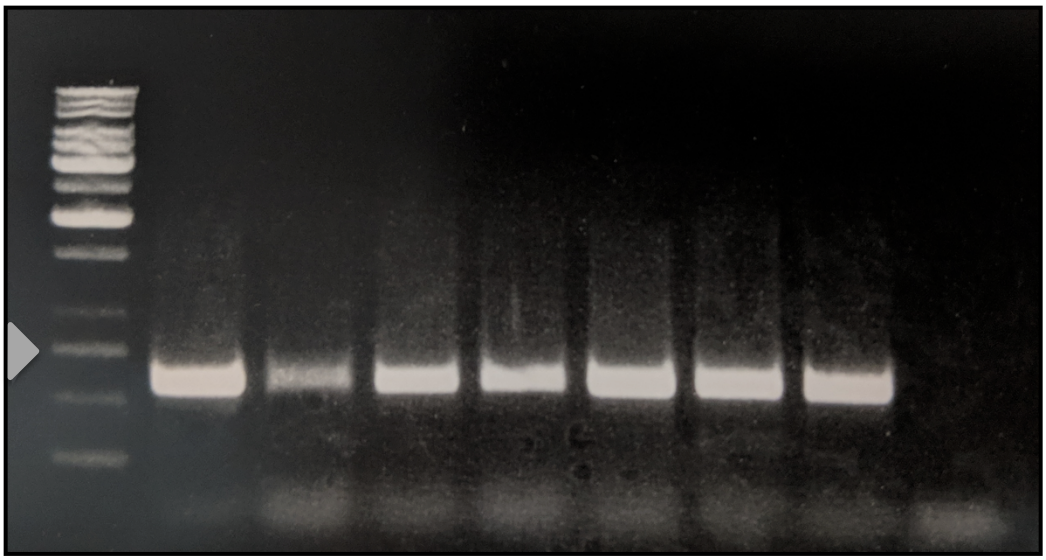


Figure 3.2 Agarose gel electrophoresis for the confirmation of predicted SNPs in Slur29 and S-PM2d mutants. Primers target a ~600 bp region encapsulating the SNP of interest using the primers in Tables 3.2 and 3.3. PCR products were run on a gel (section 2.10) with a 1 Kb ladder.

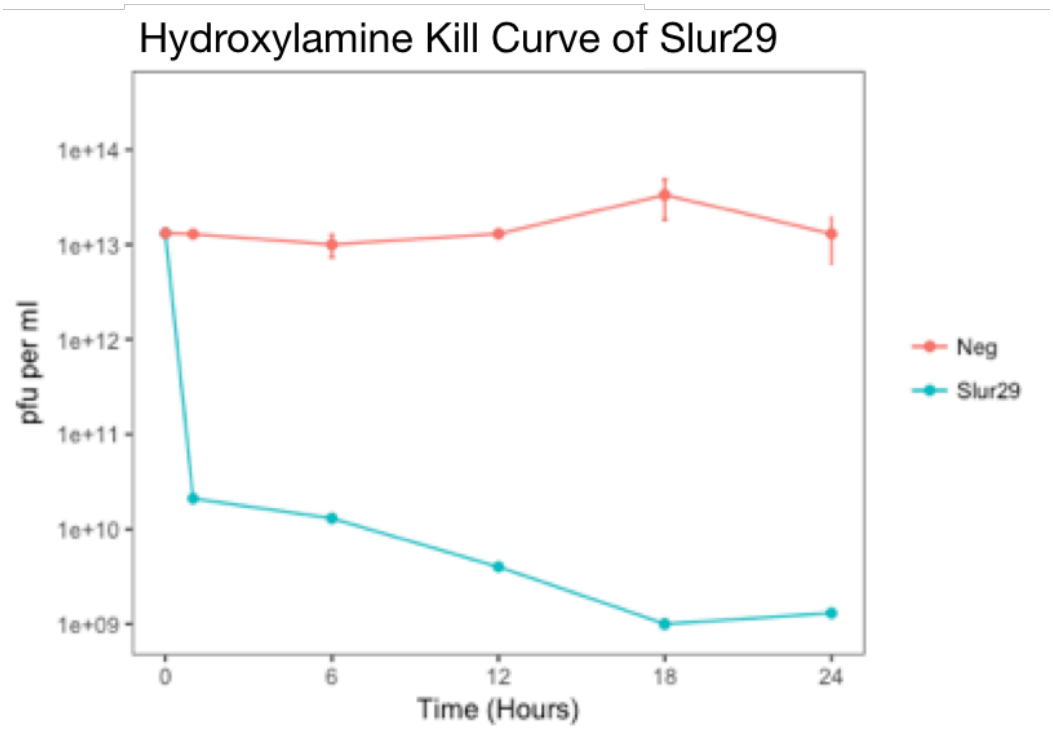


Figure 3.3 Kill curve of coliphage Slur29 when incubated with 0.8 M hydroxylamine for 24 hours. The red line is the negative control of phage without mutagen. The blue is the phage incubated with mutagen. Error bars are standard deviation of the mean

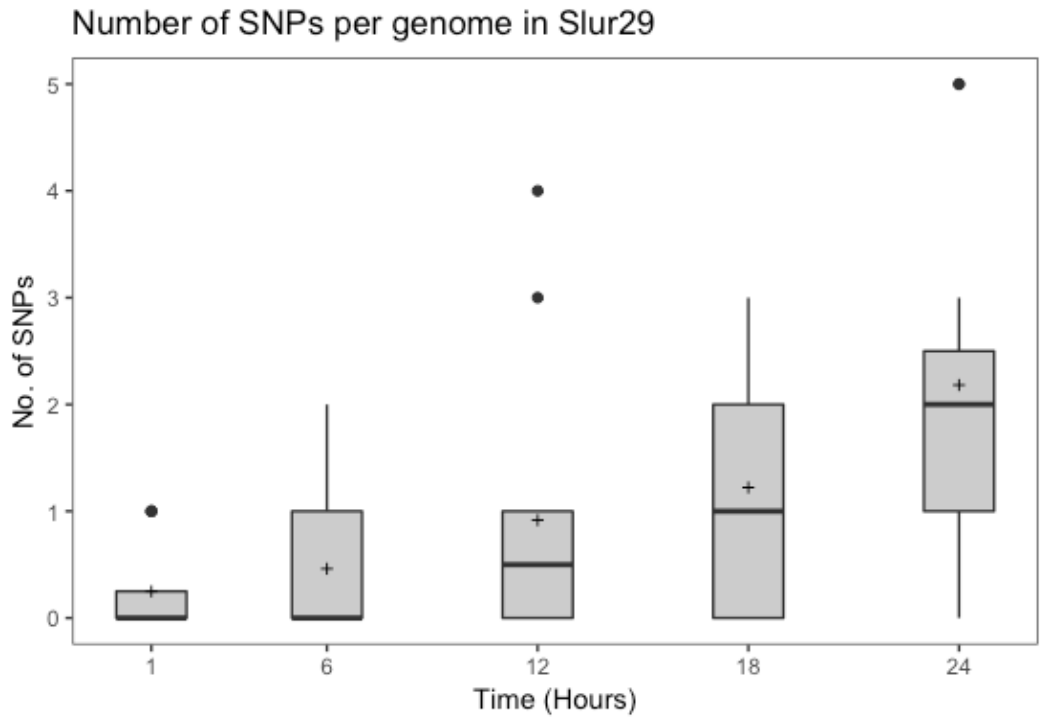


Figure 3.4 Number of SNPs for each group of Slur29 phages at each timepoint. The '+' represents the mean number of SNPs per genome at that time.

However, upon more in depth analysis of the SNPs present in the coliphage it was found that they were not all C – T transition mutations as expected. The majority of SNPs that were identified were indeed transition mutations (Figure 3.5), but these did not all target cytosine as has been previously described as the mutagenic action of hydroxylamine. Additionally, 16% of the SNPs identified were transversion mutations, something that has not previously been described for hydroxylamine induced mutations. When plotting the position of these SNPs along the Slur29 genome it becomes apparent that they are not normally distributed; the type of distribution you would expect if mutagenesis was random. If the SNPs were normally distributed then the distance from one SNP to the next would be equal. Therefore. The distance of each SNP relative to every other was calculated and tested for normality using a Shapiro-Wilk test. The results of this confirm that the SNPs are not normally distributed, where $W=0.91646$, $p = <0.001$. There are a

few regions in the genome that appear to be resistant to the appearance of SNPs (Figure 3.6), suggesting that SNPs in this region could lead to mutations that prove to be fatal to the bacteriophage.

Table 3.5 Complete list of Slur29 mutants containing SNPs, showing the position and type of nucleotide change, as well as the change in amino acid. HP: Hypothetical protein, S/NS: Synonymous/Non-Synonymous, INT: Intergenic.

Isolate Name	No. SNPs	Position	Gene Product	Ref	Mut	S/NS	AA Ref	AA Mut
S31	1	40683	HP	C	A	S	A	A
S33	1	46602	HP	C	T	S	Q	Q
S36	1	41696		G	A	INT		
S47	2	2217	Tail assembly protein	G	A	NS	P	S
		21482	HP	T	C	NS	I	V
S48	1	7775	Major tail protein	G	A	NS	L	F
S49	1	7775	Major tail protein	G	A	NS	L	F
S50	2	46344	HP	G	A	S	Y	Y
		47749	HP	G	A	NS	A	V
S54	1	19980	HP	C	T	NS	G	S
S61	1	46728	HP	C	A	S	V	V
S62	1	4865	Tape measure protein	G	A	NS	A	V
S64	3	5070	Tape measure protein	G	A	NS	L	F
		9719		G	A	INT		
		42184		G	T	INT		
S66	1	3626	HP	C	T	S	K	K
S70	4	5848	Tape measure protein	C	T	S	Q	Q
		6147	Tape measure protein	C	T	NS	E	K
		8522	HP	C	T	NS	S	N
		21658		C	A	INT		
S71	2	19950	HP	G	A	NS	R	C
		27419	HP	C	A	NS	L	I
S72	2	42613	DNA Binding protein	C	T	S	Q	Q
		43230	HP	C	T	NS	D	N
S73	3	25667	HP	G	A	S	K	K
		27302	HP	G	A	NS	G	S
		30448	HP	G	A	S	P	P
S74	1	14769	Head Morphogenesis protein	G	A	NS	R	W
S76	1	2104	Tail assembly protein	A	C	S	G	G
S78	3	3193	HP	A	G	NS	V	A
		25898	HP	C	T	S	L	L

		27244	HP	C	T	S	D	D
S80	2	35810	Helicase	C	T	NS	G	R
		38872	DNA Primase	C	T	S	G	G
S81	1	46728	HP	C	A	S	V	V
S82	1	25732	HP	G	A	NS	G	D
S83	1	41745	DNA Binding protein	C	T	S	Q	Q
S85	5	7201	Major tail protein	G	A	S	G	G
		10917	HP	G	A	NS	T	I
		17190	Terminase large subunit	G	A	NS	P	S
		23610		G	T	INT		
		35809	Helicase	C	T	NS	G	E
S86	3	23552		G	A	INT		
		24666	HP	G	A	NS	V	I
		39876	DNA Primase	C	A	NS	P	H
S87	2	22105	HP	C	T	S	E	E
		46602	HP	C	T	S	Q	Q
S89	5	2641	HP	C	T	NS	R	K
		13555	Head Morphogenesis protein	C	T	NS	V	I
		13714	Head Morphogenesis protein	C	T	NS	A	T
		13715	Head Morphogenesis protein	C	T	NS	M	I
		35787	Helicase	C	T	S	Q	Q
S90	1	12426	HP	C	T	NS	A	T
S91	2	28084	HP	G	A	NS	E	K
		46721	HP	G	A	NS	L	F
S93	2	10989	HP	C	T	NS	W	*
		30673	HP	C	A	S	V	V
S94	2	13857	Head Morphogenesis protein	C	T	NS	S	N
		27743	HP	C	T	NS	H	Y

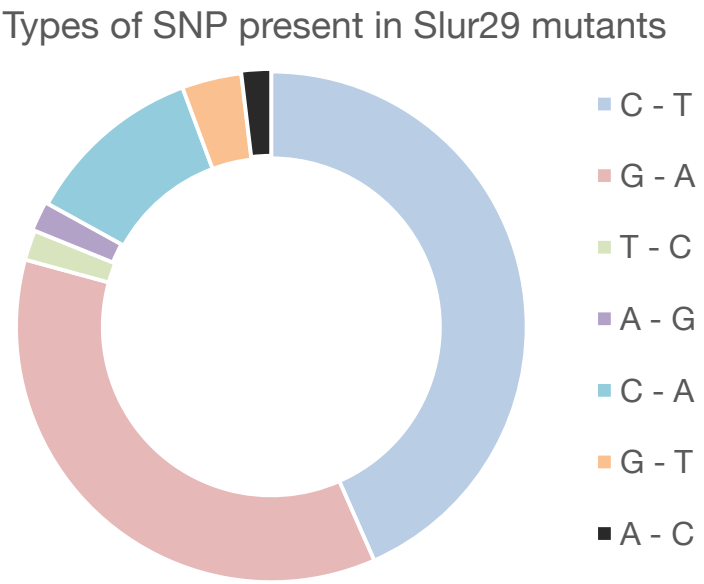


Figure 3.5 The types of single nucleotide polymorphisms identified in the control coliphage Slur29. Sections that are shaded blue are transition mutations, those in grey are transversion mutations.

3.3.3 Hydroxylamine mutagenesis applied to cyanophage S-PM2d.

Once it had been established genetically that 1 SNP per genome did correlate to a 3-log drop in coliphage titre, this method was applied to the model cyanophage S-PM2d. After incubation with hydroxylamine for 24 hours, titering the putative mutant phage from each timepoint showed that S-PM2d reached a three-log drop by between hours 18 and 24 (Figure 3.7). This is similar to that achieved by the coliphage slur29. However, after sequencing of these putative mutants with Illumina it was found that this does not correlate to finding one SNP per genome. At none of the time-points was an average of one SNP observed the average being consistently much higher (Figure 3.8).

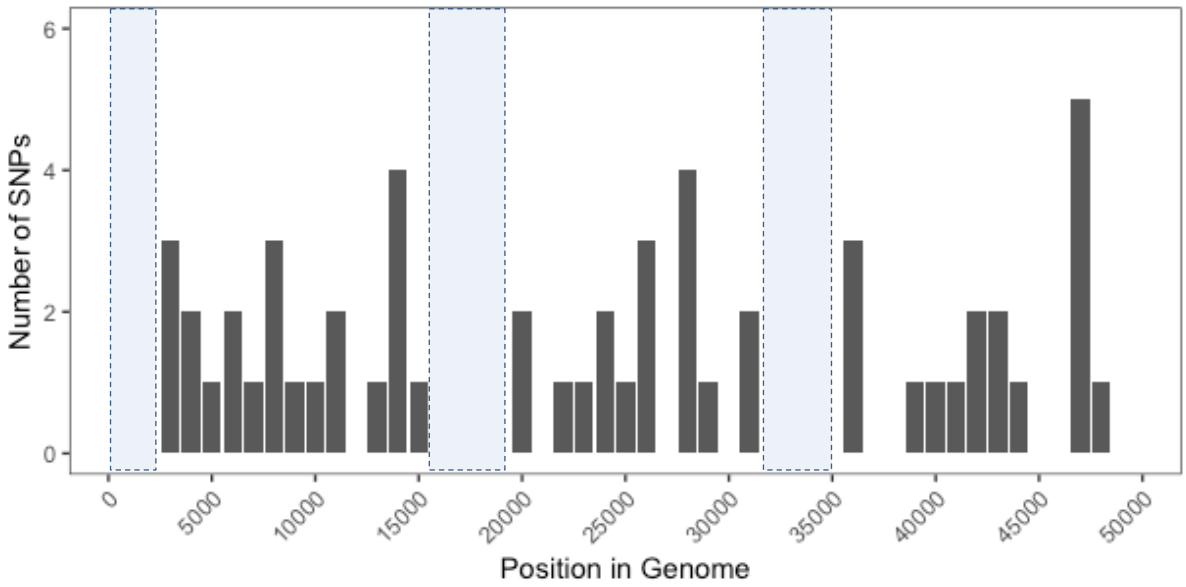


Figure 3.6 Distribution of the total number of SNPs in 1000bp windows along the Slur29 genome. Regions highlighted in blue are those that appear to be 'resistant' to SNPs.

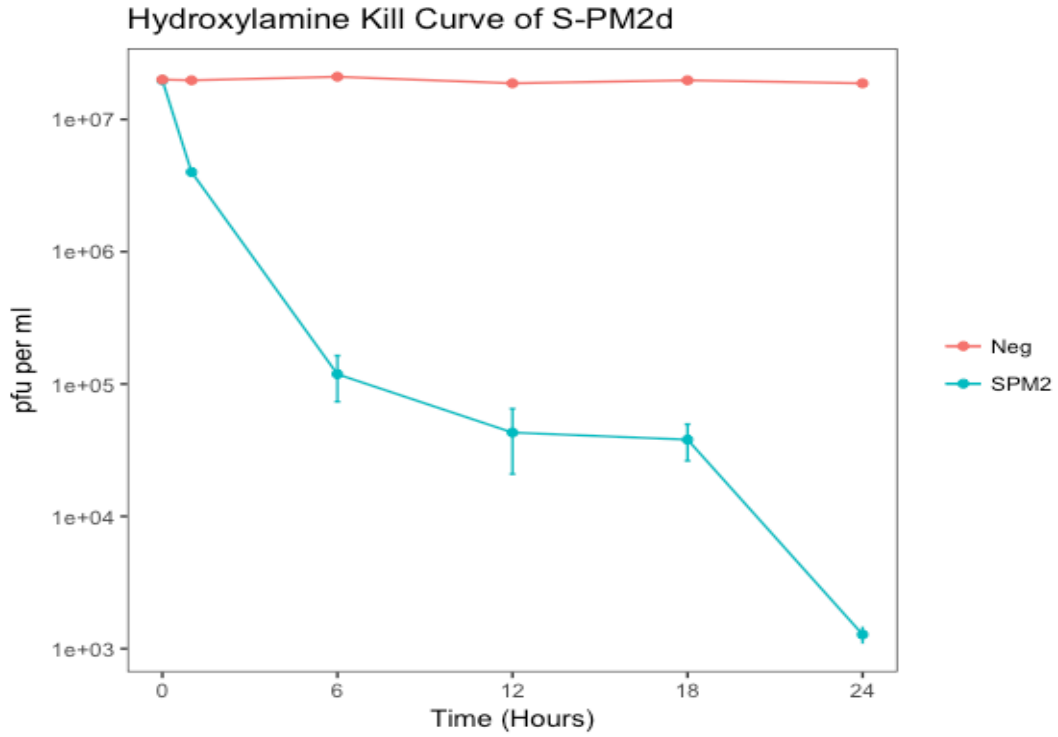


Figure 3.7 Kill curve of cyanophage S-PM2d over a 24-hour period using 0.8M hydroxylamine. The red line is phage without any mutagen added, the blue line is phage with mutagen. Error bars are standard deviation of the mean.

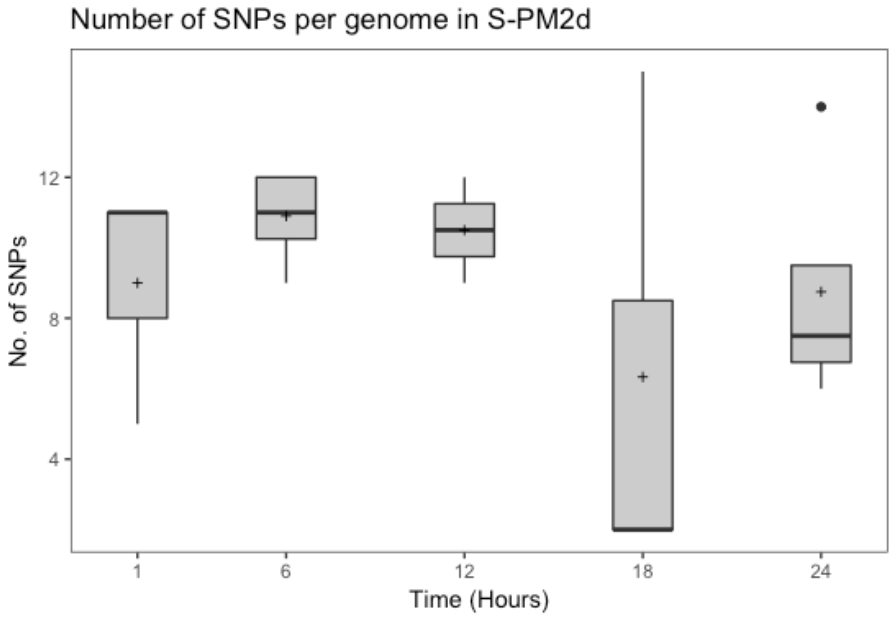


Figure 3.8 Number of SNPs present in the genomes of mutant S-PM2d phages at each time point of the mutagenesis experiment. '+' represents the mean value of SNPs per genome.

In addition to identifying SNPs, mapping the assembled mutant genomes against the reference S-PM2d genome showed that there were larger regions of the genome that lacked any coverage. Therefore, there appeared to be larger deletions as well as single nucleotide polymorphisms. However, these deletions were unable to be confirmed with PCR targeting these regions, using the primers described in Table 3.4. This led to the suggestion that these were artefacts of sequencing, something that was confirmed later using a library replicate experiment, and is explained in more detail in the next section.

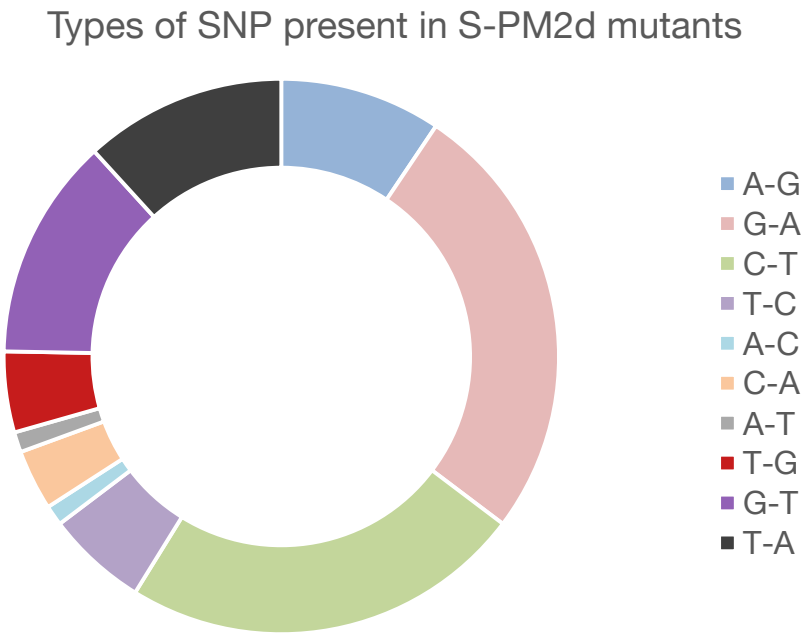


Figure 3.9 The types of single nucleotide polymorphisms identified in cyanophage S-PM2d. Sections that are shaded blue are transition mutations, those in grey are transversion mutations.

3.3.4 Coverage dependent 'deletions' in cyanophage S-PM2d.

No regions of no coverage were identified in the 3 replicates of Slur29. However, replicate libraries of the same S-PM2d DNA extraction, in both the wild type and the mutant, gave differing complements of deletions. This would suggest that these 'deletions' are just an artefact of sequencing, and not real deletions in the phages. The libraries pooled together and subsampled to differing degrees showed that the size of these deletions depended on the level of coverage of the phage as shown in the selected S-PM2d mutant in Figure 3.10.

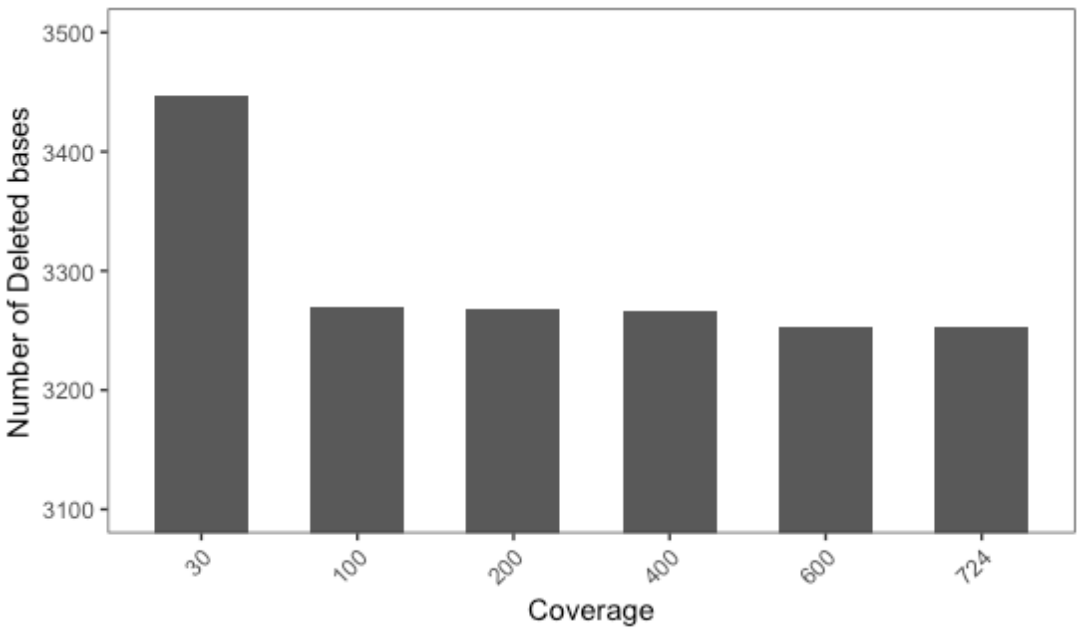


Figure 3.10 Decrease in number of deleted bases in pooled libraries of an S-PM2d mutant, as total coverage increases.

3.3.5 Hybrid genome assemblies using long-read Nanopore and short read Illumina sequencing.

Combining the long reads of Nanopore sequencing with the shorter reads of Illumina sequencing allowed full coverage of the genome, overcoming the issues of using only short read technologies for sequencing S-PM2d. All apart from one deletion were shown to be artefacts of sequencing. Long Nanopore reads allowed the identification of a novel deletion in the S-PM2d genome, not caused by the mutagenesis (Figure 3.11). This deletion spans positions 12071 to 13441, and is in the same 'ORFanage' region in which S-PM2 previously lost ~10 kb of its genome (Puxty, Perez-Sepulveda, *et al.* 2015). This results in the loss of 4 ORFs that encode hypothetical proteins (S-PM2d071, S-PM2d250, S-PM2d072, and -S-PM2d073). As the deletion is present in the original 'wild type' used in the

mutagenesis experiments it was determined that the deletion was a spontaneous event, and not a result of the hydroxylamine mutagenesis. This new variant, which we name S-PM2dd, had been used as the wild type in all experiments up to this point and so I continued to use this as the wild type in all further experiments.

The hybrid assemblies of long and short read sequencing technologies were used to confirm SNPs of a small pool of S-PM2dd mutants. These mutants were characterised in more detail in Chapter 4 but can be found shaded in grey in Table 3.5. The complement of SNPs in each phage matched that predicted by the Illumina sequencing and PCR confirmation previously (PCR product sequences can be found in appendix 4). These SNPs are discussed in Chapter 4.

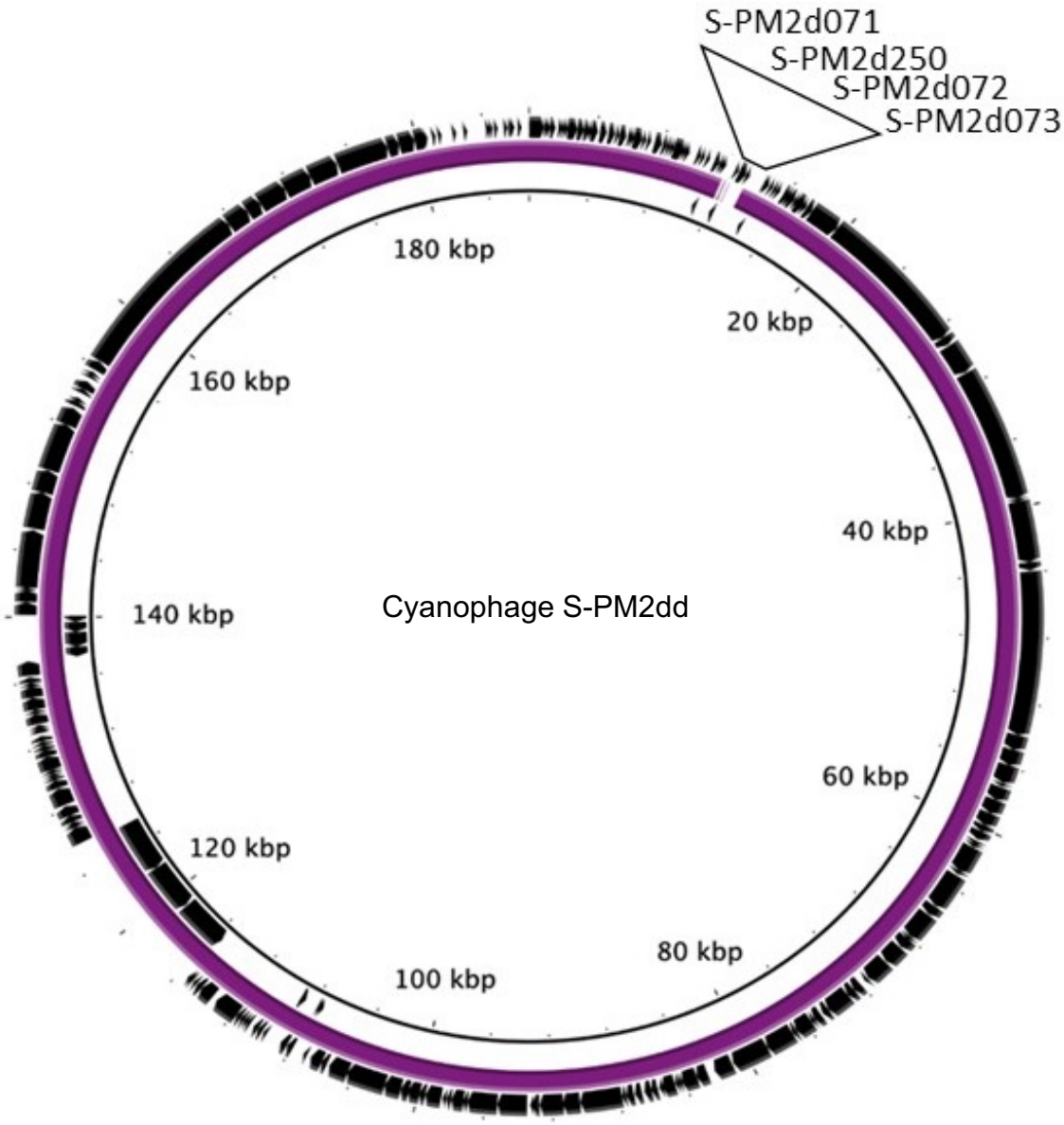


Figure 3.11 Genome visualisation of the 'new' S-PM2dd cyanophage with an additional deletion causing the loss of 4 ORFs. The purple ring is the genome of S-PM2dd compared to that of S-PM2d. Forward and reverse genes are shown in black, and the ORF numbers have been added in the region that the deletion occurs, and the areas immediately adjacent.

Table 3.6 List of SNPs present in all S-PM2dd mutants sequenced that fully assembled. The Table shows the nucleotide and amino acid changes where appropriate, and whether the SNP was found in the TARA dataset when mapped at both 90 and 70% identity. Phages shaded in grey are those that are further characterised in Chapter 4. HP: Hypothetical protein.

Isolate Name	No. SNPs	Position	Gene Product	Ref	Mut	S/NS	AA Ref	AA Mut	InTARA90	InTARA70
20	5	10506	HP	A	G	NS	Y	C	no	no
		49638	HP	T	G	NS	V	G	no	no
		99358	Structural protein	C	A	NS	Q	K	no	no
		100717	HP	G	A	NS	E	K	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
21	5	100793	HP	T	A	NS	L	H	no	no
		1151	Structural protein	C	T	NS	G	D	no	no
		175049	HP	T	C	S	N	N	no	no
		175050	HP	A	G	NS	N	D	no	no
		175068	HP	G	T	NS	A	S	no	no
30	7	4695	Intergenic	C	T	INT			no	no
		37678	Virulence Associated	G	A	NS	A	T	no	no
		71729	Intergenic	C	T	INT			no	no
		100717	HP	G	A	NS	E	K	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
		125516	Intergenic	G	A	INT			no	no
		148153	HP	A	T	NS	Q	L	no	no
31	10	7534	HP	G	T	NS	R	L	no	no
		33383	Virulence Associated	C	T	NS	A	V	no	no
		44990	HP	G	A	NS	G	S	no	no
		44991	HP	G	A	NS	G	D	no	no
		49590	HP	C	T	NS	T	I	no	no
		100717	HP	G	A	NS	E	K	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
		175049	HP	T	C	S	N	N	no	no
		175050	HP	A	G	NS	N	D	no	no
		175068	HP	G	T	NS	A	S	no	no
32	3	48666	HP	C	A	NS	A	D	no	yes

		100793	HP	T	A	NS	L	H	no	no
		185120	HP	G	A	NS	W	Stop	no	no
33	2	49638	HP	T	G	NS	V	G	no	no
		100717	HP	G	A	NS	E	K	no	no
37	5	44990	HP	G	A	NS	G	S	no	no
		44991	HP	G	A	NS	G	D	no	no
		100793	HP	T	A	NS	L	H	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
		145039	HP	C	T	S	L	L	no	no
39	1	115714	Structural protein	C	T	NS	G	D	no	no
60	2	100793	HP	T	A	NS	L	H	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
65	3	100793	HP	T	A	NS	L	H	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
		158028	HP	G	A	NS	V	I	no	no
80	2	15951	HP	G	T	NS	A	S	no	no
		100793	HP	T	A	NS	L	H	no	no
83	4	10506	HP	A	G	NS	Y	C	no	no
		49638	HP	T	G	NS	V	G	no	no
		100717	HP	G	A	NS	E	K	no	no
		119890	HP	G	T	NS	A	D	no	no
84	3	10473	HP	G	T	NS	S	I	no	no
		100793	HP	T	A	NS	L	H	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
85	1	100793	HP	T	A	NS	L	H	no	no
86	5	100717	HP	G	A	NS	E	K	no	no
		115714	Structural protein	C	T	NS	G	D	no	no
		175049	HP	T	C	S	N	N	no	no
		175050	HP	A	G	NS	N	D	no	no
		175068	HP	G	T	NS	A	S	no	no
87	3	44990	HP	G	A	NS	G	S	no	no
		44991	HP	G	A	NS	G	D	no	no
		175068	HP	G	T	NS	A	S	no	no
91	4	100793	HP	T	A	NS	L	H	no	no
		175049	HP	T	C	S	N	N	no	no
		175050	HP	A	G	NS	N	D	no	no
		175068	HP	G	T	NS	A	S	no	no
92	6	100793	HP	T	A	NS	L	H	no	no

		115714	Structural protein	C	T	NS	G	D	no	no
		164632	HP	C	A	NS	A	D	no	no
		175049	HP	T	C	S	N	N	no	no
		175050	HP	A	G	NS	N	D	no	no
		175068	HP	G	T	NS	A	S	no	no
94	1	115714	Structural protein	C	T	NS	G	D	no	no
95	4	44990	HP	G	A	NS	G	S	no	no
		44991	HP	G	A	NS	G	D	no	no
		49590	HP	C	T	NS	T	I	no	no
		100717	HP	G	A	NS	E	K	no	no
96	4	10506	HP	A	G	NS	Y	C	no	no
		100717	HP	G	A	NS	E	K	no	no
		118592	Structural protein	C	T	NS	D	N	no	no
		119890	HP	G	T	NS	A	D	no	no
B1	1	100717	HP	G	A	NS	E	K	no	no
psbASNP	4	12031	Intergenic	A	C	INT			no	no
		49638	HP	T	G	NS	V	G	no	no
		100717	HP	G	A	NS	E	K	no	no
		126728	psbA (PSII)	C	T	NS	T	I	no	no

3.3.6 Distribution of SNPs across the genome of S-PM2dd, and in nature.

If the SNPs in S-PM2dd were created at random, as would be expected in chemical mutagenesis, then the presence of SNPs would be normally distributed across the genome. However, when plotting the appearance of SNPs across the S-PM2dd genome there appear to be areas where SNPs occur more often than others (Figure 3.15), and visually they do not seem to be normally distributed. This is confirmed by a Shapiro-Wilk test for normality where $W=0.944$, $p = <0.001$, meaning that the null hypothesis that SNPs are normally distributed across the genome can be rejected. If this mutagen specifically targets cytosine, as previously suggested, then the regions containing more SNPs could be expected to have a higher %GC content than the rest of the genome; the average %GC content for S-

PM2dd is 37.8%. Whilst some mutagenic regions do have a much higher %GC content, such as between 10-15,000bp, this is not a consistent pattern. The areas around 95-100,000 bp and 115-120,000 bp show no significant increase in %GC content yet have some of the highest levels of SNPs. A complete list of all the SNPs identified is shown in Table 3.5.

As the identified SNPs were not normally distributed I checked to see if they occurred in hypervariable regions more often than those that are conserved. I mapped the TARA ocean viral metagenome reads (appendix 2) against the S-PM2dd genome at differing levels of minimum percent identity, from 70% to 90%. Previously, hypervariable regions have been identified using a cut-off as low as 70% (Enav *et al.* 2018), but this is low enough that reads could be mapping against a phage of a different species, so I used the higher minimum identity of 90% for more certainty. Figure 3.12 shows the difference in the reads mapping against S-PM2dd at different levels of % identity. The average level of coverage of the genome at 90% identity was 12X, so anything with less than 20% of this was classified as hypervariable, and above 20% was considered core. By comparing the positions of SNPs identified in the hydroxylamine mutants with the regions found to be conserved or hypervariable, it became evident that the SNPs appear more often in the hypervariable regions. Only two of the SNPs found appeared in conserved regions, at positions 126728 in *psbA*, and position 119890 in a hypothetical protein.

To see if the SNPs identified in mutagenesis appeared in nature I compared their position against lists of SNPs called using varscan on the reads that mapped to the S-PM2dd genome at 70% and 90% identity. From the 336 variable positions

identified (appendix 5), none of these are found in the mutants created in this work (Table 3.5), apart from one SNP (position 48666) although this is only when using the reads mapped at 70% identity and not 90%.

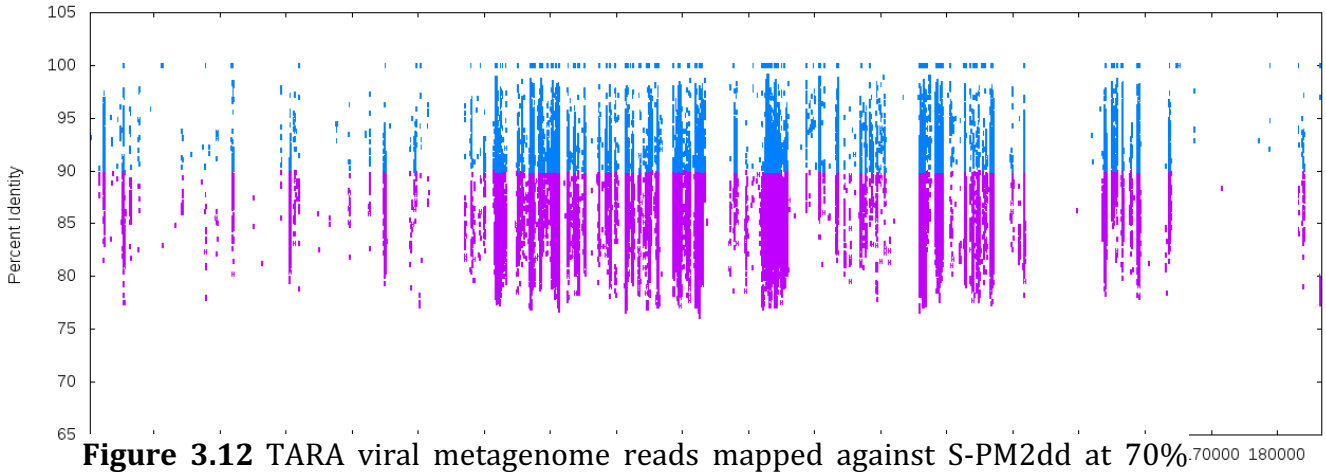


Figure 3.12 TARA viral metagenome reads mapped against S-PM2dd at 70% identity (purple) and 90% identity (blue).

3.3.7 Comparing mutagenic action of hydroxylamine on S-PM2dd to that of another cyanophage, S-RSM4.

In order to determine whether the results obtained for S-PM2dd are representative more widely of cyanophages, or somewhat more limited to itself, a second cyanophage was subjected to hydroxylamine mutagenesis. In comparison to both previously tested phage S-RSM4 showed a significantly faster drop in titre, reaching the required 3-log drop in titre in only a single hour (Figure 3.13). This pattern was again confirmed using a spot-assay (appendix 6).

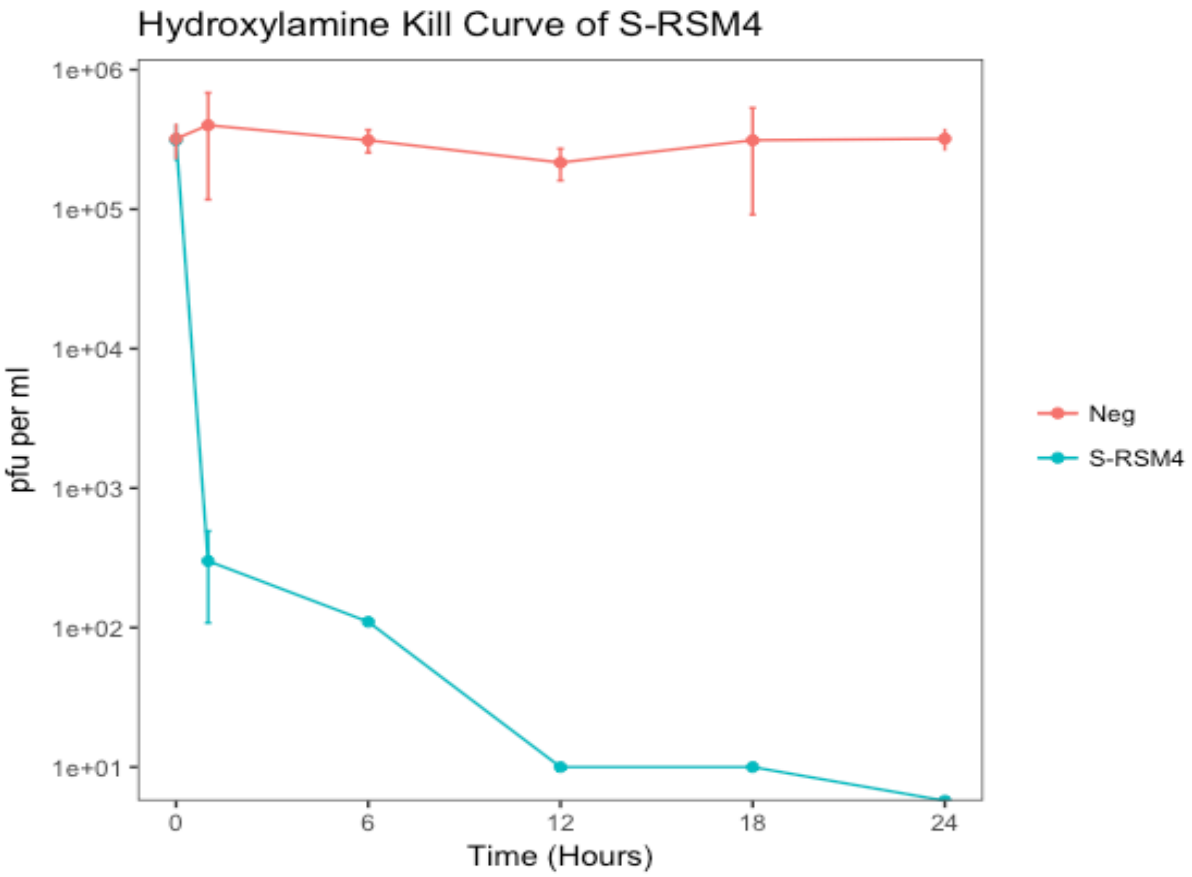


Figure 3.13 Kill curve of cyanophage S-RSM4 over a 24-hour period using 0.8M hydroxylamine. The red line is S-RSM4 without any hydroxylamine added, the blue line is phage incubated with the mutagen. Error bars are standard deviation of the mean.

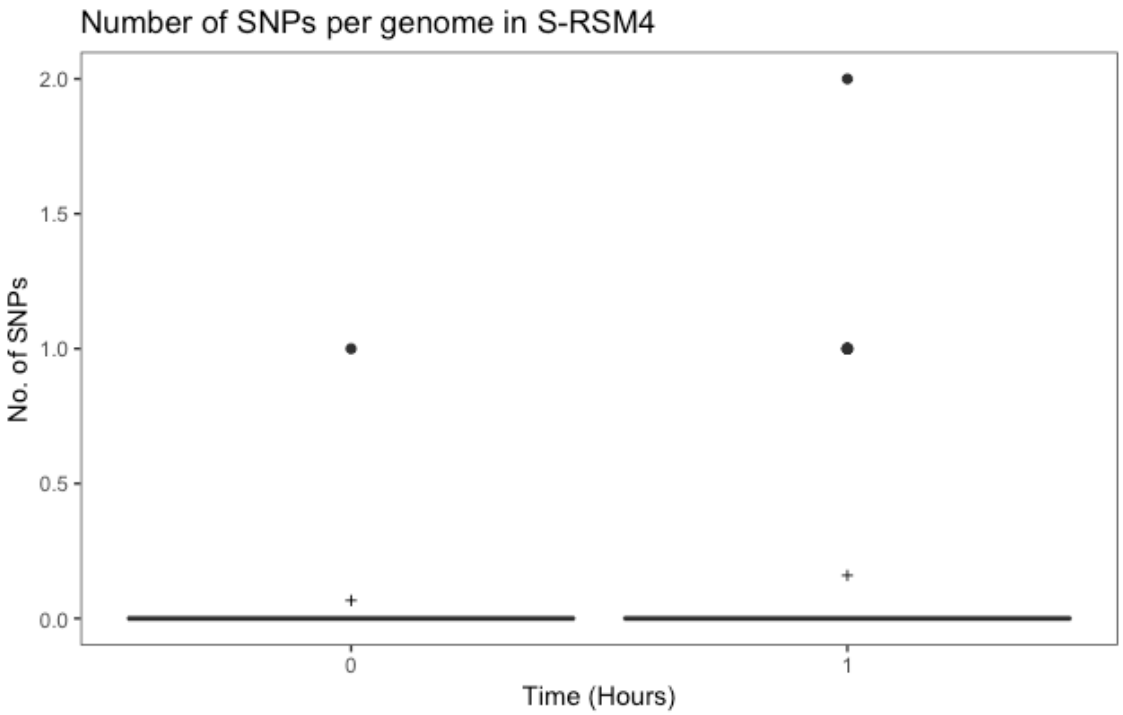


Figure 3.14 Number of SNPs present per genome in mutant S-RSM4 phages from each timepoint. '+' represents the mean number of SNPs per genome.

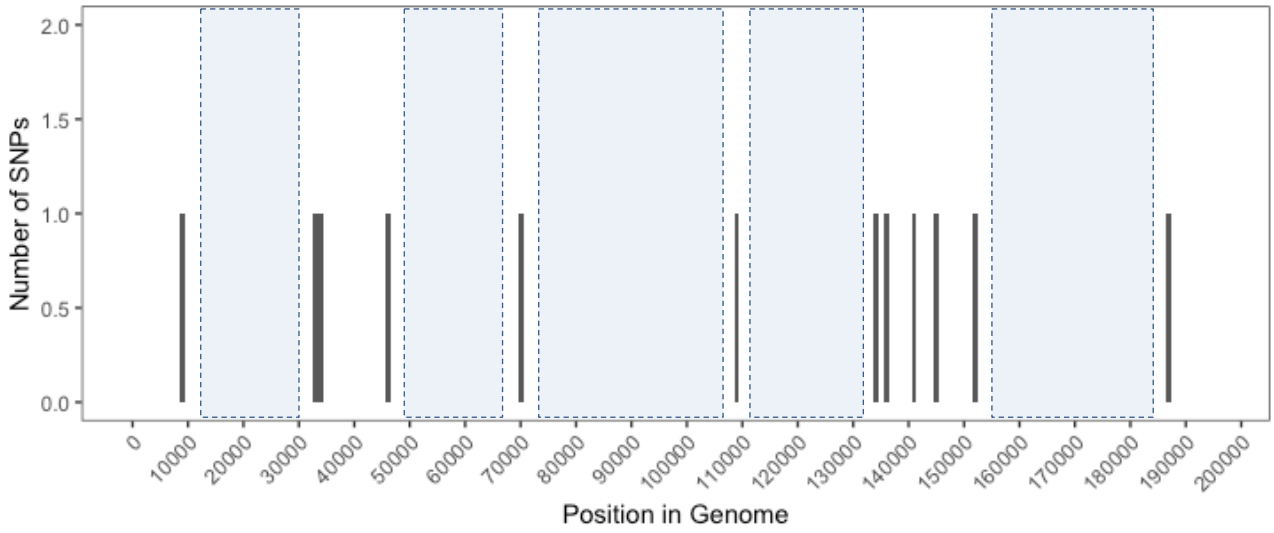


Figure 3.15 Distribution of total number of SNPs in 1000bp windows along the S-RSM4 genome. Those regions that appear to be devoid of SNPs are highlighted in blue.

Sequencing phage from timepoints 0 and 1 hour when the 3-log drop in titre occurred showed different results to both Slur29 and S-PM2dd. At hour one there was an average of 0.14 SNPs per genome (Figure 3.14), with only 11 SNPs present in the 75 phages sequenced from that timepoint. When plotting these along the genome they also do not appear to be normally distributed (Figure 3.16), something that is shared between all phage tested. Results of a Shapiro-Wilk test confirm this with W=0.946, p = <0.001. Examining the types of SNPs identified shows that whilst the C-T transition mutations known to be caused by hydroxylamine are the most prevalent, there are a range of different SNPs present (Figure 3.17). This is also a result that is common between both cyanophages, and the control coliphage Slur29.

Table 3.7 List of all SNPs identified in S-RSM4 mutants showing the nucleotide and amino acid changes where appropriate.

Isolate Name	No. SNPs	Position	Gene Product	Ref	Mut	S/NS	AA Ref	AA Mut
SR7	1	140867		C	T	INT		
SR13	1	33063		C	A	INT		
SR27	1	32007		A	G	INT		
SR30	1	8829		T	A	INT		
SR34	1	144680		C	T	INT		
SR42	2	108807	HP	A	G	NS	Y	H
		135479		C	T	INT		
SR45	1	45169	MazG	G	A	S	F	F
SR55	1	69636	Prohead Scaffold protein	C	A	NS	S	A
SR76	1	186216	HP	C	T	NS	S	G
SR77	1	133568		T	G	INT		
SR83	1	151370		G	A	INT		

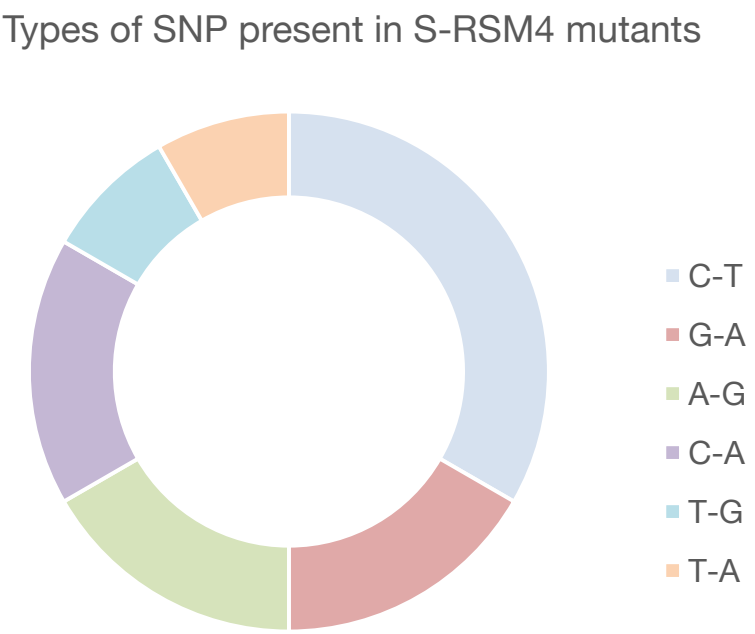


Figure 3.16 The types of single nucleotide polymorphisms identified in cyanophage S-RSM4. Sections that are shaded blue are transition mutations, those in grey are transversion mutations.

3.4 Discussion

3.4.1 A range of mutations are present in phages treated with hydroxylamine.

The development of a random chemical mutagenesis system for use in cyanophage would be of significant use to the field. It would help in the identification of the functions of hypothetical proteins as well as illuminating the roles of identified auxiliary metabolic proteins such as PsbA. Hydroxylamine was chosen as the chemical mutagen due to its long history of use in bacteriophages, and its reported specificity. Hydroxylamine is thought to specifically cause C-T transition mutations at a rate of 1 SNP per genome when a corresponding 3-log drop in phage titre is observed (Tessman *et al.* 1964; Hall and Tessman 1966; Villafane 2009).

However, within the phage sequenced here this did not prove to be the case apart from when used on the control coliphage Slur29. Here, a three log drop in titre correlated to a single SNP per genome. Even when there were multiple SNPs identified it was more often the case that multiples were synonymous and as such would not have any effect on the amino acid. These patterns were not true of the two cyanophages used. In S-PM2dd there were consistently multiple SNPs per genome, the majority of which were non-synonymous, whereas S-RSM4 showed opposing results of an average of less than one SNP per genome. Historically, the correlation between drop in titre and number of SNPs has been well documented (Hall and Tessman 1966; Tessman 1968). However, the majority of these experiments were carried out before the widespread use of whole genome sequencing. When this method has previously been coupled with whole genome

sequencing this correlation has not been observed, and an average of 3.8 SNPs per genome was identified when using phage T7 (Robins *et al.* 2013).In this work Robins et al., discount this observation as mutations that were not caused by hydroxylamine, but give no further explanation. On rare occasions where hydroxylamine was found to cause more than a single base change in other work this has been explained by the presence of 'hot spots' that are both very rare, and not very hot (Tessman *et al.* 1964). This is something I looked into in the exploration of hypervariable regions in the cyanophage S-PM2d and will talk about in further detail later.

It has previously been noted that the presence of mutants containing more mutations than expected occurs more frequently than expected (Drake *et al.* 2005). A wide variety of organisms have clearly shown that instances of multiple mutation is significantly higher than would be predicted from known organism mutation frequencies and random distribution of mutations (Buettner *et al.* 2000; Bebenek *et al.* 2001; Colgin *et al.* 2002). Two causes of this phenomenon have been widely considered; firstly, the creation of 'mutator mutations,' such that the mutation created increases the genetic instability of the DNA and leads to a higher than normal rate of mutation (Drake *et al.* 2005). Secondly, a condition known as 'transient hypermutability', a general mutational mechanism found in a range of organisms, with the ability to generate multiple synchronous mutations (Drake *et al.* 2005; Chen *et al.* 2009). These might be caused by errors in transcription/translation that lead to the production of proteins with alterations in their primary sequence, errors in folding that cause dominant-inactive proteins, or regulatory errors that cause the over or underproduction of proteins. This

could also affect proteins involved in the fidelity of replication, and through this hypermutability could be caused through a number of mechanisms (Drake *et al.* 2005; Chen *et al.* 2009). From the limited knowledge of the functions of the proteins with SNPs in phage, it is unclear whether any of them have a role in transcription/translation and whether an alteration in their structure could cause this state of transient hypermutability. However, it is a possibility and would explain the presence of multiple mutations in the phages sequenced here.

As well as varying number of SNPs identified per genome, different to that expected, both the bacteriophage and cyanophages all showed a range of mutations in addition to the expected C-T transition mutations. This result was surprising as the mechanism of action for hydroxylamine has been characterised to attack only the cytosine residue in genetic material (Tessman *et al.* 1964). Previous work that has coupled hydroxylamine mutagenesis and deep sequencing also noted this result (Robins *et al.* 2013). Although in this case this was explained by the significant depth of sequencing coverage (200,000 – 500, 000 per nucleotide) that allowed the detection of extremely rare mutations introduced by PCR, flow cell clustering, or inaccuracies in phage replication. However, with the levels of sequencing depth achieved here (10s-100s per nucleotide), whilst more than sufficient to allow the assembly of genomes (Rihtman *et al.* 2016) and to call SNPs (Song *et al.* 2016), would not be as likely to detect these rare events, and as such is an unlikely explanation. One type of mutation specifically over-represented in all three phages sequenced was a G-A change, also noted by Robins in their work. This was explained as a result of strand-specific DNA replication, whereby only one strand of DNA is copied into the template used for replication

during infection (Robins *et al.* 2013). This could also be the reason for the over-representation of these mutations in this work.

3.4.2 Distribution of SNPs throughout the genomes and in nature.

If hydroxylamine was not specifically targeting cytosine bases as previously thought, then it may be that the mutagen is acting in an entirely random way. If this was the case then mutations caused would be normally distributed across the genome, such as those created by random evolutionary processes (Koenig 2002). However, in this work this is also not the case and SNPs appear more frequently in some regions than others. Tests of normality (Shapiro-Wilk) on the distances between SNPs show that they are not normally distributed in any of the phages tested. Through the identification of the hypervariable regions in cyanophage S-PM2 I was able to test whether the SNPs appeared more frequently in these variable regions or the 'core' genes. My analysis showed that the SNPs were more often found in these variable regions, which may not be surprising due to the conserved nature of 'core' genes. SNPs may not be as well tolerated in 'core' genes that are required for the correct functioning of the phage, and so phage with SNPs in these regions will not be propagated.

One interesting observation is that the SNPs created by hydroxylamine mutagenesis in S-PM2dd are not SNPs that are identified in nature. One reason for this could be that the SNPs created by mutagenesis are deleterious, reducing 'fitness' and as such you would not expect them to be propagated by phages in nature. This could also help explain the higher than expected number of mutations in the S-PM2dd phages. It has previously been noted in bacteriophages that those

phages that obtain a mutation that proves to be deleterious are more likely to obtain a further mutation to compensate for the effects, rather than a reversion of the original change (Poon and Chao 2005). It could be that only some of the mutations in S-PM2dd were caused by the mutagenesis process, and others are compensatory mutations in an attempt to silence the prior mutation.

3.4.3 Mutations obtained in each phage.

A small pool of S-PM2d mutants were chosen to further characterise in the following chapter. These mutants were selected for a number of reasons, each of which contained interesting mutations or a single SNP that would be easy to associate with any phenotype that was presented. Mutant d39 contains a single SNP when comparing to the new 'WT' identified in this work. This SNP causes an amino acid change in what is annotated as a virion structural protein. Due to the single genetic change it will be easy to associate any phenotypic change with this. The mutants d85 and d80 were chosen due to their complement of SNPs that would jointly allow the characterisation of associated phenotypes. d85 contains a single SNP at position 100793, and mutant d80 contains the same mutation along with an additional SNP at position 15951. Finally, the psbASNP mutant was selected due to its SNP causing an amino acid change in the *psbA* gene, an auxiliary metabolic gene that is of particular interest in this work. This SNP is in a position that has not been previously investigated, and as such could provide interesting information on the functioning of the protein when observing any phenotypic change.

3.4.4 Identifying a new deletion mutation in cyanophage S-PM2d.

The primary sequencing of S-PM2dd mutants was done using only Illumina Nextera sequencing. This appeared to show the introduction of multiple 'deletions' in the genomes; 'deletions' that were later shown to be false – apart from one that is discussed later on. Illumina Nextera is a transposon-based library preparation method, designed for quick production of libraries from low DNA yields. However, it has previously been noted that the method has an innate bias due to the consensus sequence of the Tn5 targeting AT-rich regions (Marine *et al.* 2011). The PCR amplification step has also been shown to drastically affect the coverage of genomes causing low coverage in AT rich regions (Marine *et al.* 2011). The GC% content of S-PM2 is 37.8%, making the genome AT rich. It could be that the high levels of AT rich sequences in S-PM2d caused the transposon to cut more frequently, over-fragmenting the genome and making it impossible to sequence. Combined with the effects of the amplification bias due to PCR, this could have caused any regions of the genome that were rich in AT to have a reduced coverage. This could have caused the numerous regions of no coverage that was seen when sequencing the S-PM2dd mutants. The control coliphage has a higher GC% of 44.7% and this may help explain why it didn't show the regions of no coverage found in S-PM2dd. In order to overcome this a combination of Illumina and Nanopore sequencing was used. Nanopore sequencing is not a transposon-based approach and so does not show the same biases that Illumina sequencing does. However, the error rate of this method is still higher (5-15% compared to >0.1% of Illumina Nextera XT) (Rang *et al.* 2018; Glenn 2011),and as such a combination of the two was used to ensure full coverage of the genome, and the accuracy in

base calling required to identify SNPs. This approach led to the discovery of a novel deletion of the wild type S-PM2d phage.

Cyanophage S-PM2 has previously spontaneously deleted a ~10 kb fragment of its genome, removing 33 ORFs all of which had no significant similarity following BLAST analysis (Puxty, Perez-Sepulveda, *et al.* 2015). It was this deletion mutant that was used as the original wild type in this work. It was identified here through a combination of short and long read data that S-PM2d had spontaneously deleted another set of ORFs from its genome. A total of three ORFs in the same region as those previously deleted were lost, specifically a deletion from bp 12071 to 13441. The tendency for S-PM2(d) to lose parts of its genome in this region could suggest that these are genes involved in adaptation to particular environmental conditions, that the phage does not encounter when propagated under 'ideal' conditions in the lab and so can get rid of these without any noticeable fitness cost. This raises an interesting question as to how much of the genome is accessory to the core genes that have been identified in cyanophages in general, as well as those specifically identified in S-PM2d.

This result also means that the mutants created, were created in a background that is not S-PM2d, but is the new mutant S-PM2dd that has been used as the WT for all experiments in this chapter, and those following.

Overall, this chapter has shown that whilst in coliphages hydroxylamine mutagenesis does result in 1 SNP per genome when a corresponding 3-log drop in titre is observed, this is not true of the two cyanophages tested here. Additionally, the SNPs that are present are not limited to the transition mutations expected, and in S-PM2dd are different from SNPs found in nature. The reasons for this are

unclear and would require further work to illuminate any possible explanations. Whilst this method may not have worked optimally it has led to the production of mutants that could increase our understanding of proteins of unknown function, something that is explored in the next chapter. The model cyanophage S-PM2(d) has shown new levels of genetic instability, not expected in this work. This is shown by the identification of a new deletion in the phage, leading to a loss of ~1 kb DNA. This could suggest that this region is only required for adaptation to different environmental conditions, but does pose an interesting question of how much of this cyanophages genome is really 'core' and how much is 'accessory'

Chapter 4: Characterisation of novel cyanophage S-PM2dd mutants

4.1 Introduction

Cyanophage S-PM2 is often used as a model phage in conjunction with its host *Synechococcus sp*. WH7803. Much characterisation has been done on the wild type phage; it was the first cyanophage to have its genome fully sequenced (Mann *et al.* 2005) and the presence of AMGs were first identified in this phage (Mann *et al.* 2003). Recently, a spontaneous deletion mutant of cyanophage S-PM2 was described (called S-PM2d), which lacks ~10 kb of the genome and possesses a fitness advantage compared to the wild type phage under standard laboratory conditions (Puxty *et al.* 2015). Interestingly, of 238 coding sequences in S-PM2d, 121 encode hypothetical proteins. The creation of mutants in S-PM2d (see Chapter 3) should greatly increase our understanding of the function of some of these hypothetical proteins. It was identified in the previous chapter that the S-PM2d WT phage used here has incurred a second spontaneous deletion to create a new mutant called S-PM2dd in this work. This is the WT phage that has been used in all experiments in the previous chapter, and in this one. Therefore, the S-PM2dd mutants that were fully sequenced with both long and short read technologies (see Chapter 3) were further characterised to identify possible phenotypic changes in these mutant phages, and hence link genotype with a phenotype.

4.2 Methods

4.2.1 Transmission Electron Microscopy

TEM images were taken at the University of Leicester Microscopy facility. The five phages were imaged using a JEOL JEM-1400 Transmission electron microscope (TEM) as follows: 10 µl of a $1x10^8$-$1x10^{10}$ phage stock was added to a glow-discharged formvar copper grid (200 µm mesh). This was left for 2 mins before being wicked off and the grid washed using 10 µl of water that was also wicked off using filter paper. 10 µl 2% (w/v) uranyl acetate stain was added to the grid and left for 30 secs before its removal. The grid was air-dried and then imaged using a JEOL JEM-1400 TEM with an accelerating voltage of 100kV. Digital images were collected with a Megaview III digital camera using iTEM software. Images were processed in ImageJ (Schindelin *et al.* 2015) using the measure tool and the scale bar present on each image to obtain phage particle size. When measuring the capsid size of each mutant, values are the average of 50 particles.

4.2.3 Plaque size determination

Phages were plated for plaques using the 6-well method previously described (see section 2.4). Plaque size was measured manually for at least 30 plaques for each phage. Mean values are an average of 30 size measurements.

4.2.4 Phage one-step infection experiments

Phage infection parameters were determined using a 96 well plate one step method. Host *Synechococcus* sp. WH7803 cell concentration was measured by flow cytometry (Section 2.11) and cultures were diluted such that they were $1x10^8$ cells per ml. Each phage was added to culture at an MOI of 0.01, in triplicate. A sample

was taken immediately and diluted 1/4000 into fresh ASW medium (see section 2.2). 1.5 ml of infected culture was transferred into a 2 ml deep 96 well plate and covered with sealing lids. The plate was then incubated on a thermomixer set at 23°C and 800 rpm, under constant illumination at 10 µmol photons m^{-2} s^{-1}. 50 µl samples were taken every hour for 24 hours. To sample, a clear flat-bottomed 96 well plate was used as a collection plate in the Vac-Man 96 Vacuum manifold in conjunction with a Millipore vacuum motor. A 0.45 µm pore size 96 well filtration plate was placed on top of the divider. Once the 50 µl sample was loaded into the wells using a multi-channel pipette the vacuum pump was switched on and run for ~30 secs so the sample was filtered through the wells onto the collection plate below. Phages were titered using 10 µl of sample and the 6-well method described previously (see section 2.4).

4.3 Results

4.2.1 Selection of cyanophage S-PM2dd mutants for phenotypic characterisation

The cyanophage S-PM2dd mutant containing a SNP that causes an amino acid change in the *psbA* gene was an obvious choice for phenotypic characterisation, given a major aim of my thesis was determining the functional role of this viral PsbA protein. Cyanophage S-PM2dd mutants d39, d80, and d85 were also chosen for phenotypic characterisation based on their complement of SNPs (see Table 4.1) and particularly that these cyanophage shared SNPs that meant these then differed only by a single SNP. In this way any phenotype present could potentially be associated with a particular SNP. As mentioned in Chapter 3, the combination of long and short read sequencing had allowed for the genomes to be sequenced in full. Apart from the novel deletion identified in the original stock, and therefore present in all mutants, there were no other mutations (insertions/deletions) found in these phages.

Table 4.1 Position of SNPs in the genomes of mutant cyanophage S-PM2dd chosen for further characterisation, with the corresponding protein product in S-PM2d shown under the position of the SNP. Shown on the first line of each box is the nucleotide change, followed by the type of mutation where 'NS' stands for non-synonymous, and in brackets is the amino acid change.

Phage	***12031 INT***	***15951 p075***	***49638 p088***	***100717 p148***	***100793 p148***	***115714 p173***	***126728 p176***
WT							
psbASNP	A-C INT		T-G NS (V-G)	G-A NS (E-K)			C-T NS (T-I)
d39						C-T NS (G-D)	
d80		G-T NS (A-S)			T-A NS (L-H)		
d85					T-A NS (L-H)		

4.3.1 Morphology of S-PM2dd mutants

To determine whether the presence of a particular SNP affected an uncharacterised cyanophage structural gene I performed transmission electron microscopy of each of the mutant cyanophage shown in Table 4.1 to assess whether there were any visible changes in morphology of these mutant phages.

Inspection of the TEM images (Figure 4.1) showed that one of the mutants, d85, possesses an obvious difference in morphology (panel C), with the capsid of this phage not being as clearly defined as those of the wild type S-PM2dd, and indeed the other mutant phages. Thus, the icosahedral shape of the capsid is much more evident in the wild type cyanophage S-PM2dd (A) when compared to d85 (C). This phenotype was present in all the d85 isolates imaged. Evidence of this can be seen in Figure 4.2.

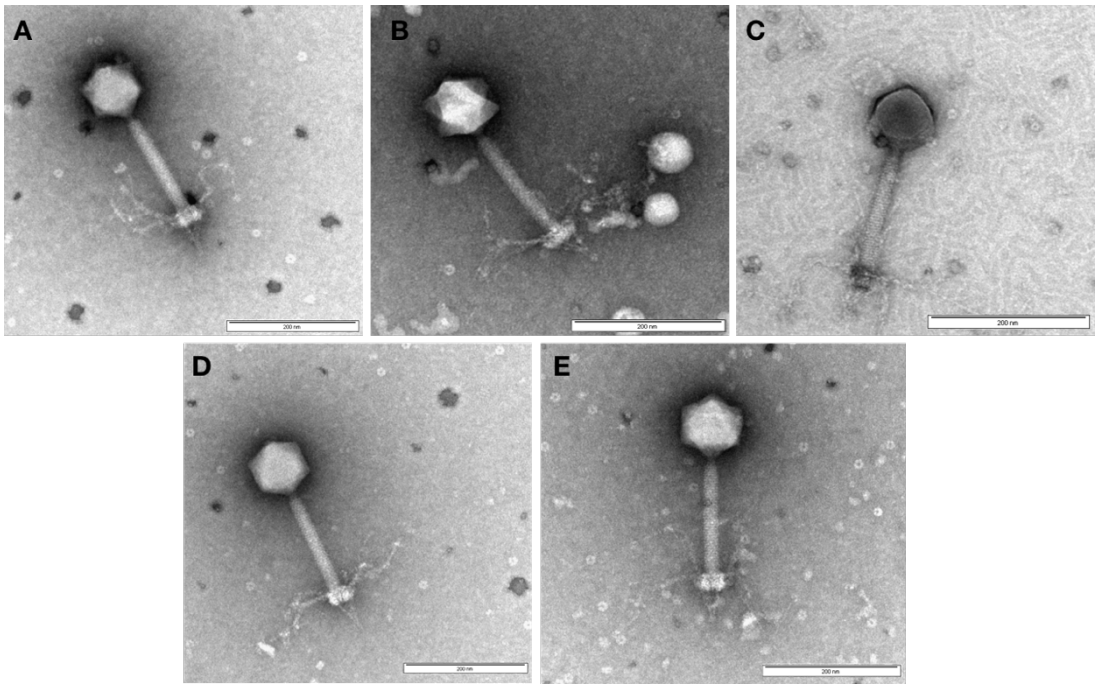


Figure 4.1 TEM images of the S-PM2dd mutant phages. Panel C highlights the deformed head structure of the d85 mutant phage. **A)** S-PM2dd 'WT' **B)** psbASNP **C)** d85 **D)** d80 **E)** d39. The scale bar in each individual image represents 200 nm.

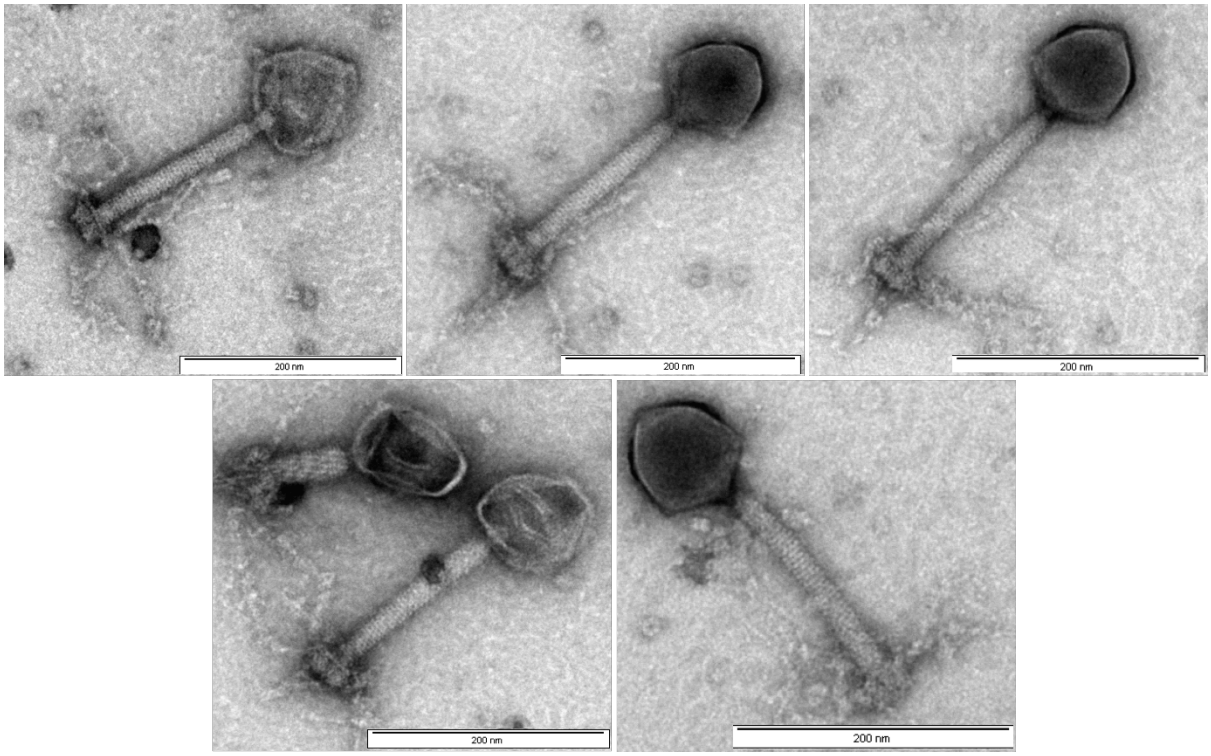


Figure 4.2 TEM images of the S-PM2dd mutant d85, clearly showing the deformity in head structure in all images. The scale bar in each image represents 200 nm.

By measuring the diameter of the capsid, clear differences between the mutant phages were observed (Figure 4.3). There was a statistically significant difference between the capsid diameters of the phages, as determined by ANOVA (F (4, 223) = 30.726, p= <0.001). A post hoc Tukey HSD test showed that the only mutant phage to differ in capsid size significantly from the wild type was d85 (p = <0.001). A further summary of the statistical results can be found in the Appendix (see appendix 7). The mean capsid size of d85 is ~9 nm smaller than that of the wild type. There was much variation between the mutant phages, but this was within the normal range of capsid diameter for the wild type. The significant differences between capsid diameters of the phages are highlighted in red in Figure 4.4.

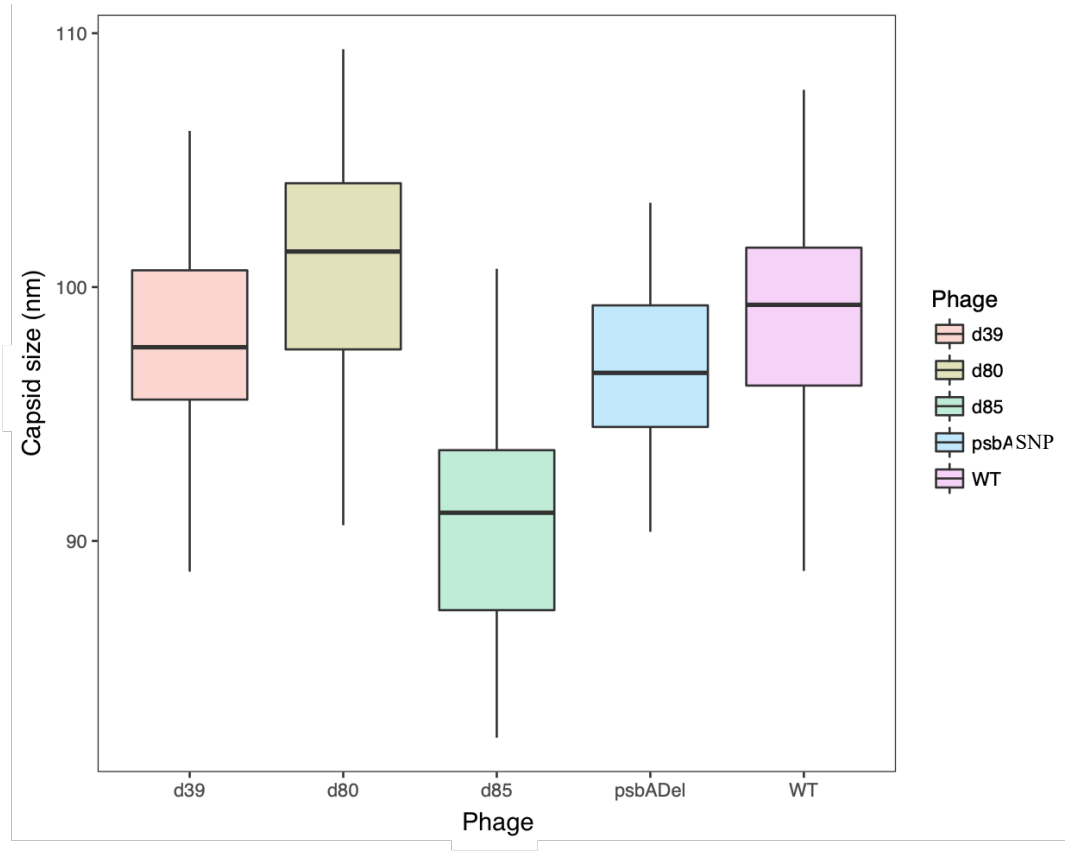


Figure 4.3 Differences in capsid diameter of S-PM2dd mutant phages. Phages from left to right: d39, d80, d85, psbASNP, and WT.

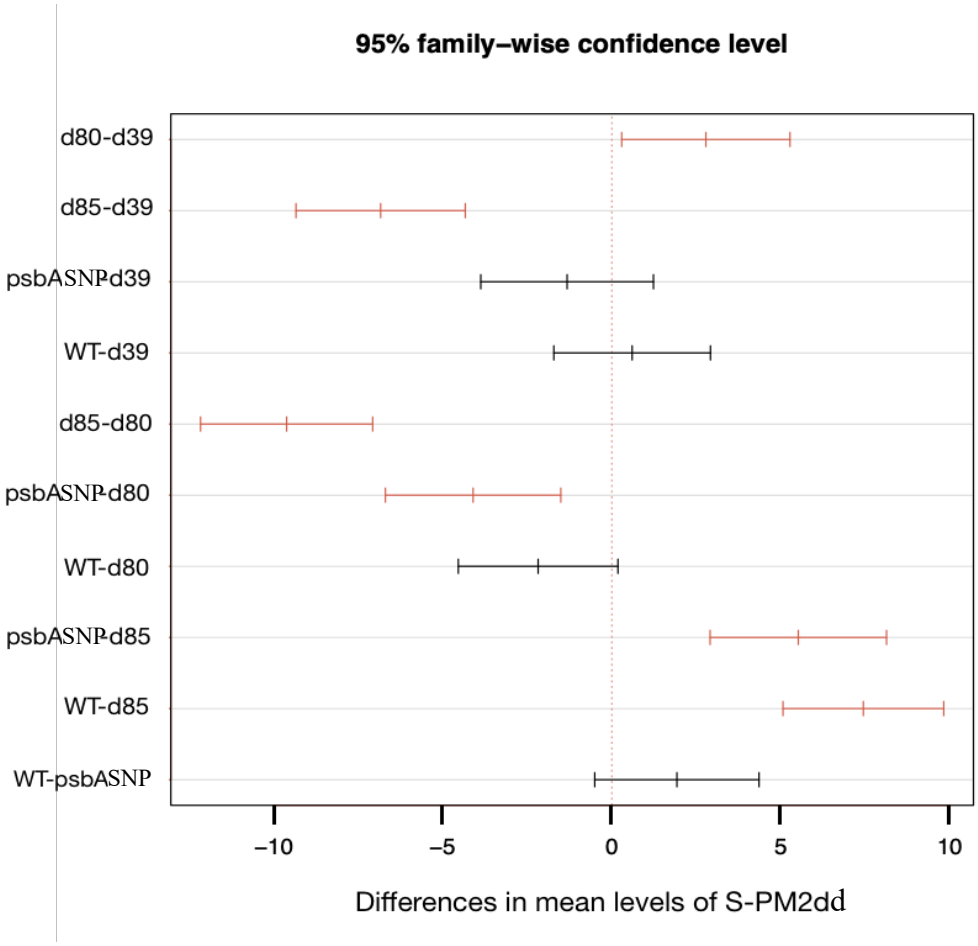


Figure 4.4 Differences in capsid size between cyanophage S-PM2dd mutants are statistically significant. Pairwise comparisons are named on the y axis, significant differences are represented by those lines that do not cross zero. These are highlighted in red.

4.3.2 Differences in plaque size between mutants

The plaque size of phages is often used as a proxy for general phage fitness. Indeed, when cyanophage S-PM2 lost 10 kb of its genome to become S-PM2d used here, one obvious phenotype that conveyed an increase in phage fitness was an increase in plaque size (Puxty *et al.* 2015). Therefore, a noticeable phenotype that may be present is a change in plaque size.

Plaque size determined manually (see section 4.2.3) showed that there were significant differences in phenotype between both the mutants and the wild type, and between individual mutants (Figures 4.5 and 4.6). This was confirmed by ANOVA (F (4, 195) = 12.734, p = <0.001). A post hoc Tukey HSD test confirmed that mutant d80 showed the greatest difference in plaque size compared to wild type (p = <0.001), but phages d85 and psbASNP also showed statistically significant differences compared to wild type (p= 0.002, p= 0.04 respectively). A further summary of the statistical results can be found in the Appendix (see appendix 8) The significantly smaller phenotype shown by d80, likely represents a decrease in fitness caused by the mutations present in this phage (see Table 4.1). Whilst d39 showed a range of plaque sizes bigger than the wild type, this difference was not statistically significant (p= 0.92) and so none of the phages showed a phenotype that may represent an increase in phage fitness.

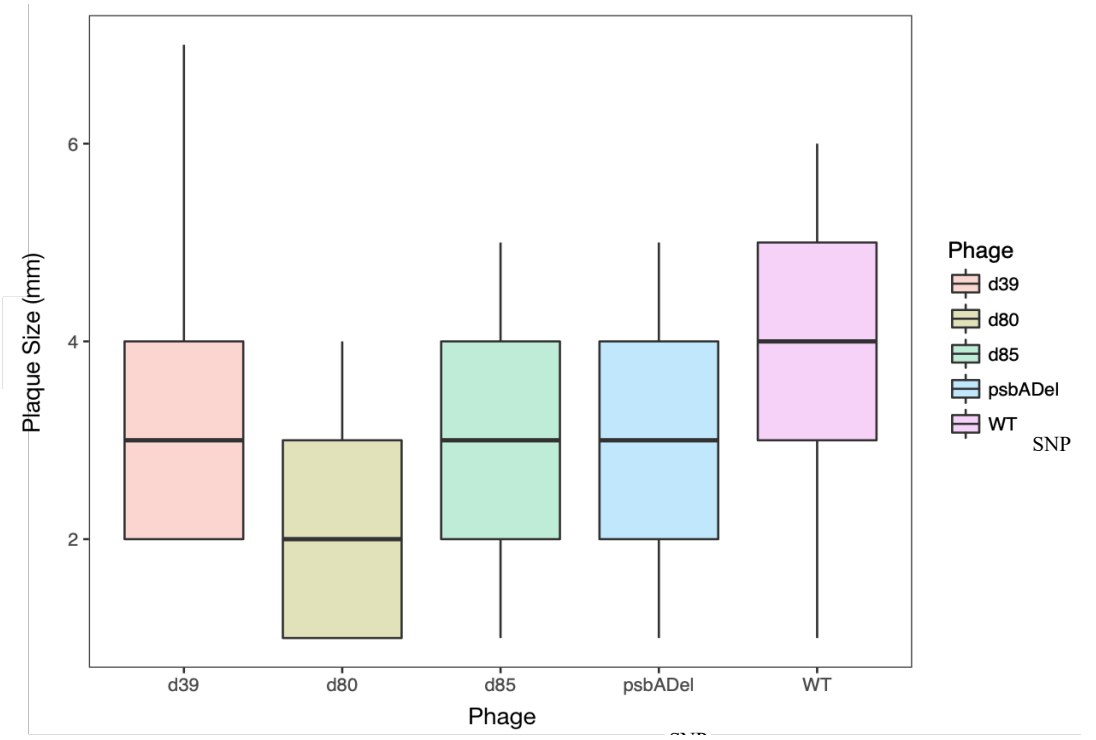


Figure 4.5 Difference in plaque size of S-PM2dd mutant phages. Phages from left to right: d39, d80, d85, psbASNP, and WT.

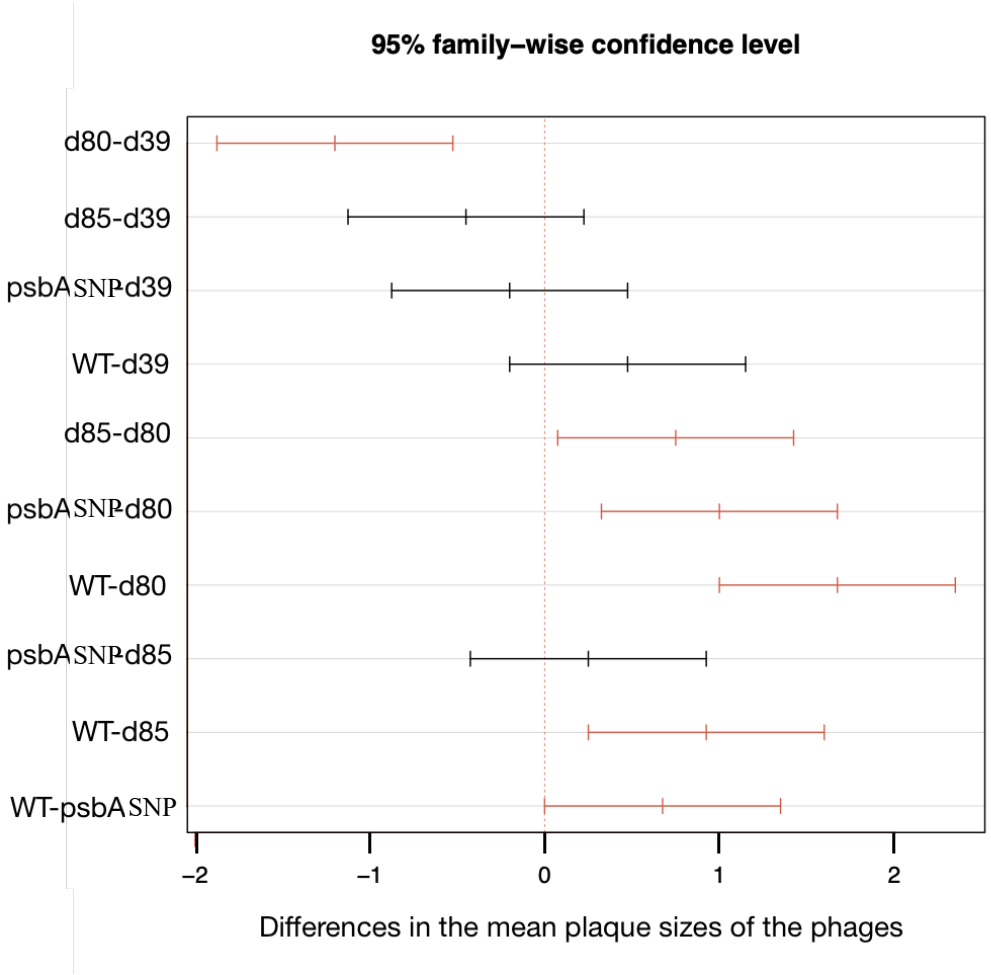


Figure 4.6 Differences in plaque size between cyanophage S-PM2dd mutants are statistically significant. Pairwise comparisons are named on the y axis, significant differences are represented by those lines that do not cross zero. These are highlighted in red.

4.3.3 One step infection analysis of mutant S-PM2dd phages.

All S-PM2dd mutants showed an extended latent period compared to that of the wild type S-PM2dd (Table 4.2). All except one mutant had a reduced burst size, whereas the d80 mutant showed an increased burst size (47, compared to 33 for the wild type). Mutant d39 had a burst size of 1. However, it is clear from Figure 4.7 that the obvious burst and plateau steps shown in the other phages were not present here, making it difficult to accurately assess the burst size and latent period of this mutant. The SNP in mutant d39 is in a gene annotated as encoding a

virion structural protein, causing a glycine to aspartic acid amino acid change. This results in a change in charge of the amino acid from a hydrophobic side chain to an acidic charge. This is the only SNP identified in this mutant, and as such the change in infection dynamics can be associated with this change.

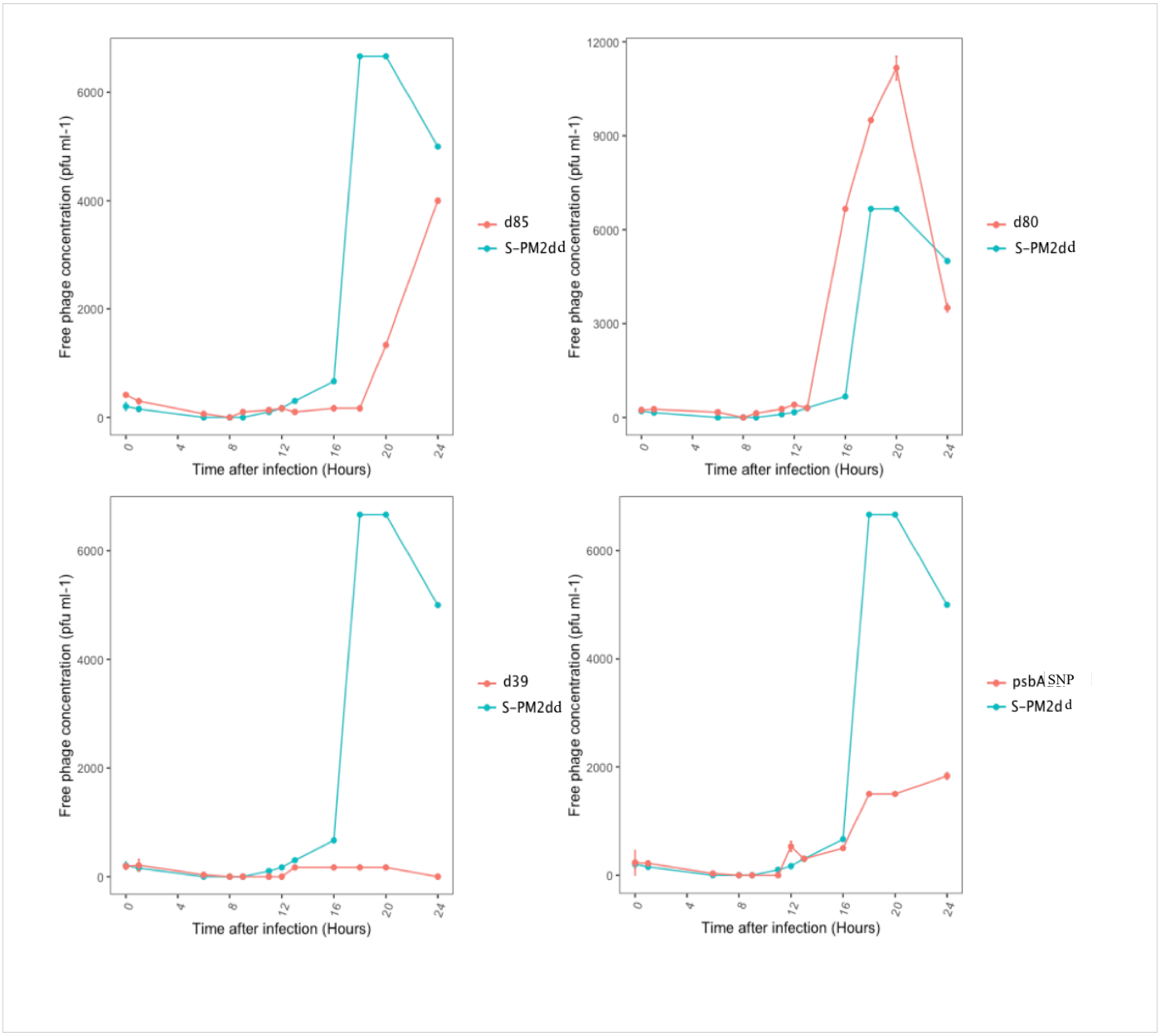


Figure 4.7 One step growth curves of the S-PM2dd mutants compared to the wild type S-PM2dd phage. Error bars show the standard deviation of the mean. Mutant d85: top left panel; mutant d80: top right panel; mutant d39: bottom left panel; mutant psbASNP: bottom right panel.

Table 4.2 Estimation of the burst size and latent period of the S-PM2dd mutants compared to that of the wild type S-PM2dd.

Phage	Burst Size	Latent Period (Hrs)
S-PM2dd WT	33	13
d85	12	18
d80	47	13
d39	1	N/A
psbASNP	8	16

4.3.4 Identification of a previously unknown cyanophage S-PM2dd head structural protein

The S-PM2dd mutant d85 contains a SNP at position 100793 causing a T-A change in gene S-PM2d148, a gene that is annotated to encode a hypothetical protein of 164 amino acids in length. This SNP is the only difference across the entire genome between the S-PM2dd wild type and the d85 mutant and causes a leucine (aliphatic) to histidine (basic) amino acid change in S-PM2d148 (Figure 4.8). Therefore, the presence of a distinct morphological change in the phage head (Figure 4.1) can likely be put down to this change. This open reading frame was not previously identified as a structural protein through other proteomic predictions that have been carried out on S-PM2 (Clokie *et al.* 2008). However, by performing an HMM search (HMMer) using the S-PM2d148 protein amino acid sequence it is obvious that this protein is widespread in cyanophages (Table 4.3) as an 'uncharacterised protein'. My work thus suggests that this protein is in fact a structural protein that plays a role in forming the shape of the capsid head. When this protein is mutated, as in mutant d85, the phage capsid appears to lose its structural definition as can be seen in the TEM images (Figure 4.1 & Figure 4.2). The capsid diameter of this mutant is also significantly smaller than the other

mutants as well as the wild type (Figures 4.3 and 4.4). Together these results suggest that this hypothetical protein is in fact involved in the correct formation of the phage capsid.

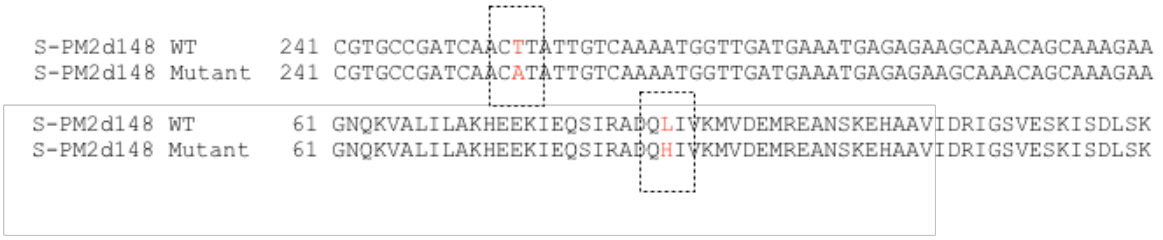


Figure 4.8 Alignment of the nucleotide (above) and amino acid (below) sequences of open reading frame S-PM2d148 in wild type cyanophage S-PM2dd and the d85 mutant cyanophage with a SNP at position 100793. The changes in sequence are highlighted in red inside the dotted boxes.

Table 4.3 Results of a protein hidden markov model search using PHMMER with an E-value cut-off of 0.01, showing that the protein encoded by S-PM2d148 is widespread in cyanophages.

Species	***Description***	***E-value***
Synechococcus phage S-PM2	Hypothetical-Protein / belonging to T4-LIKE GC: 201	6.20E-100
Synechococcus phage S-WAM1	Uncharacterized protein	3.90E-29
Synechococcus phage S-RIM2 R1_1999	Uncharacterized protein	8.80E-28
Synechococcus phage S-PRM1	Uncharacterized protein	4.10E-27
Synechococcus phage S-RSM4	Hypothetical cyanophage protein	6.40E-27
Synechococcus phage ACG-2014a	Uncharacterized protein	1.20E-26
Prochlorococcus phage Syn1	Uncharacterized protein	2.00E-26
Synechococcus phage S-ShM2	Uncharacterized protein	2.40E-26
Synechococcus phage S-CAM1	Uncharacterized protein	2.50E-26
Synechococcus phage S-CAM4	Uncharacterized protein	2.50E-26
Synechococcus phage ACG-2014e	Uncharacterized protein	2.60E-26
Synechococcus phage ACG-2014c	Uncharacterized protein	5.40E-26
Synechococcus phage S-WAM2	Uncharacterized protein	1.50E-25
Synechococcus phage ACG-2014h	Uncharacterized protein	1.60E-25
Synechococcus phage Syn19	Uncharacterized protein	3.40E-25
Synechococcus phage S-SM1	Uncharacterized protein	3.40E-25
Synechococcus phage S-CAM9	Uncharacterized protein	4.00E-25
Synechococcus phage S-IOM18	Uncharacterized protein	8.80E-25
Synechococcus phage S-CAM3	Uncharacterized protein	3.80E-24
Synechococcus phage S-H35	Uncharacterized protein	7.20E-23
Synechococcus phage syn9	Gp191	1.00E-22
Synechococcus phage ACG-2014d	Uncharacterized protein	3.60E-21
Synechococcus phage S-RIM8	Uncharacterized protein	7.20E-21
Cyanophage S-RIM32	Uncharacterized protein	1.90E-18
Cyanophage S-TIM5	Uncharacterized protein	5.10E-11
Prochlorococcus phage P-TIM68	Uncharacterized protein	2.10E-08
Synechococcus phage Bellamy	Uncharacterized protein	3.50E-07
Synechococcus phage S-SM2	Uncharacterized protein	1.50E-06
Synechococcus phage S-CAM22	Uncharacterized protein	9.80E-05
Prochlorococcus phage P-SSM7	Uncharacterized protein	0.00017
Synechococcus phage S-CAM8	Uncharacterized protein	0.00028
Synechococcus phage S-SSM7	Uncharacterized protein	0.00045
Prochlorococcus phage P-SSM4	Uncharacterized protein	0.00049
Cyanophage P-TIM40	Uncharacterized protein	0.00062
Prochlorococcus phage P-HM2	Uncharacterized protein	0.00069
Prochlorococcus phage P-RSM4	Uncharacterized protein	0.001

Prochlorococcus phage Syn33	Uncharacterized protein	0.0011
Synechococcus phage S-CAM7	Uncharacterized protein	0.0021
Synechococcus phage ACG-2014b	Uncharacterized protein	0.0029
Synechococcus phage metaG-MbCM1	Uncharacterized protein	0.003
Cyanophage P-RSM1	Uncharacterized protein	0.0062
Synechococcus phage S-SSM5	Uncharacterized protein	0.0062
Prochlorococcus phage P-HM1	Uncharacterized protein	0.0091

4.3.5 Smaller plaque size may be indicative of a decrease in phage fitness in cyanophage mutant d80.

Mutant d80 contains the same SNP at position 100793 as mutant d85 but also contains a second non-synonymous SNP occurring at position 15951 causing a G-T change which results in an alanine (aliphatic, non-polar side chain polarity) to serine (hydroxyl-containing, polar side-chain polarity) amino acid alteration in S-PM2d075 (Figure 4.9). However, whilst the d80 mutant does not display the deformed capsid head phenotype nor a smaller capsid size like the d85 mutant, it does show a smaller plaque size that could be indicative of a decrease in phage fitness. The S-PM2d075 gene that this SNP appears in is annotated as encoding a hypothetical protein of 109 amino acids in length. The results of a protein HMM search using PHMMER indicate that this protein is not widely found, with homologues only identified in two other cyanophages. It is unclear why the d80 mutant does not exhibit the deformed head phenotype of mutant d85, even though it shares the same SNP at position 100793. It may be that the decrease in plaque size of the d80 mutant causes a general decrease in phage fitness which indirectly over-rides the phenotype caused by mutation in the S-PM2d148 gene or that the mutation in S- PM2d075 directly suppresses the S-PM2d148 mutation.

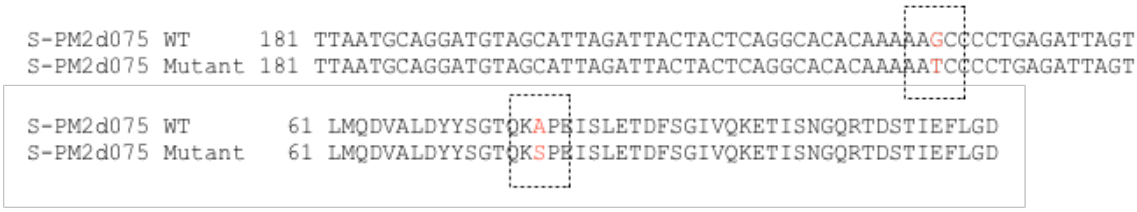


Figure 4.9 Alignment of the nucleotide (above) and amino acid (below) sequences of open reading frame S-PM2d75 in wild type cyanophage S-PM2dd and the d80 mutant cyanophage mutant with a SNP at position 15951. The changes in sequence are highlighted in red inside the dotted boxes.

Table 4.4 Significant hits of a protein hidden markov model search using the S-PM2d75 protein and an E-value cut-off of 0.01.

Species	***Description***	***E-value***
Cyanophage S-RIM32	Uncharacterized protein	1.40E-24
Prochlorococcus phage Syn1	Uncharacterized protein	4.40E-22

4.3.6 Assessing the effect of the identified SNP in the cyanophage S-PM2d *psbA* gene.

The cyanophage S-PM2dd mutant psbASNP possesses a SNP that causes a change in amino acid sequence of the S-PM2dd encoded *psbA* gene from a threonine (hydroxyl-containing, polar side chain polarity) to isoleucine (aliphatic, non-polar side chain polarity) at residue 316 of the protein (see Figure 4.10). After aligning the mutant sequence against all *psbA* sequences identified in cyanophages using MUSCLE, it is evident that this is a change that does not appear in the natural cyanophage population. In the 303 sequences that were identified as cyanophage *psbA* genes, all of the amino acids at this residue were threonine. By using SNAP2 (software that uses neural networks to predict the functional effects of mutations (Hecht *et al.* 2015)) on the amino acid sequence of the wild type S-PM2dd PsbA

sequence I was able to predict the effect that mutations would have on each residue (Figure 4.11). It appears that the T-I amino acid change caused by the SNP identified is likely to have an effect on the structure of the protein, and therefore potentially its function.

```
psbA WT     901 AACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGATGTG
psbA mutant 901 AACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACATCTGGGCAGATGTG

psbA WT      301 NFNQSIVSSEGRVLNTWADVLNRAGLGMEVMHERNAHNFPLDLAAAEATPVALTAPAIG
psbA mutant  301 NFNQSIVSSEGRVLNIWADVLNRAGLGMEVMHERNAHNFPLDLAAAEATPVALTAPAIG
```

Figure 4.10 Alignment of the wild type S-PM2dd and the sequence of the mutant containing a SNP at position 126728, in the *psbA* gene. The change in nucleotide (above) and amino acid (below) is highlighted in red inside the dotted boxes.

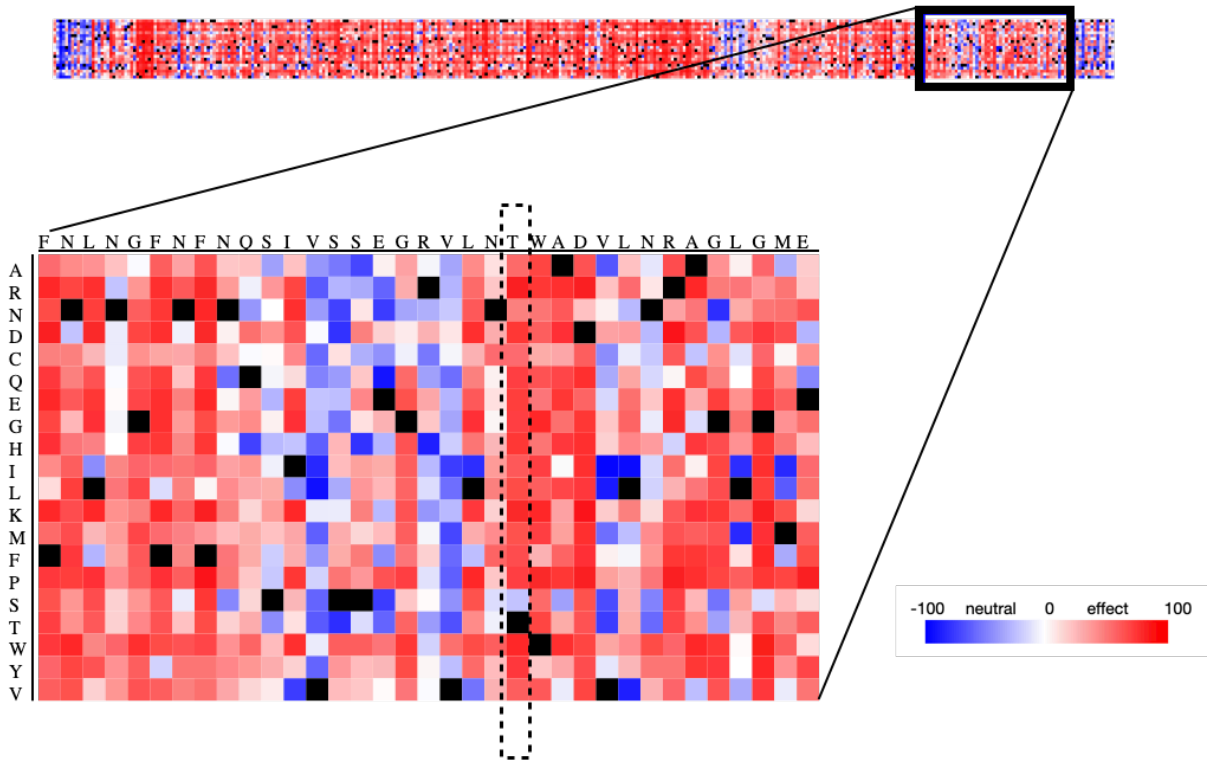


Figure 4.11 Heatmap showing the potential effect of mutations in each amino acid position along the length of the wild type cyanophage S-PM2d PsbA protein. The enlarged section is a close-up of the region containing the threonine residue that is altered in the mutant, highlighted in the dotted box.

After interrogating this further in Phyre2 (Kelley *et al.* 2015) it was evident that the SNP identified would likely cause a change in the structure of the PsbA protein in the mutant (Figure 4.12). The predicted structure of the mutated protein appears to introduce a small beta strand just before where the isoleucine residue is found, at amino acid residue 316. This is not an amino acid that has previously been identified as important in the structure of the PsbA protein in cyanobacterial host PsbA proteins, or in its interaction with other proteins and ligands (Umena *et al.* 2011). However, the prediction results suggest that changing this amino acid could indeed have an effect on the structure of the protein.

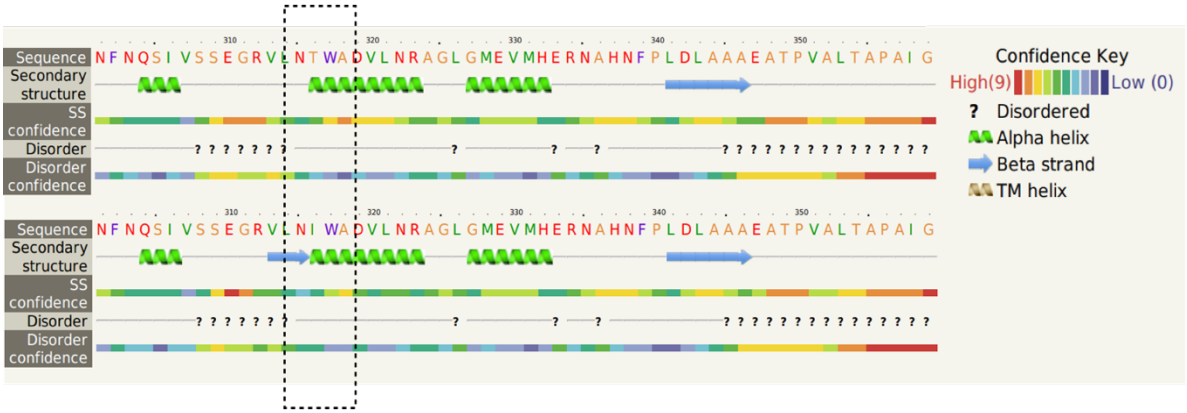


Figure 4.12 Predicted secondary structure of the region of the cyanophage S-PM2dd PsbA protein that contains the identified SNP, as predicted using Phyre2. The region of interest, containing the amino acid change is highlighted in the dotted box.

4.4 Discussion

4.4.1 Mutations in bacteriophage that change capsid size/morphology.

Due to their use as model systems in molecular biology and genetics there are a plethora of experiments on the assembly, and general structure, of the T4 capsid. As such, a number of genes that influence the process and affect the morphology of the capsid have been identified. In the T4 genome a number of genes have a mutant phenotype that leads to deformity in the head structure; all of these are involved in the assembly/maturation of the phage capsid (Miller *et al.* 2003). In the model T4 phage there are a total of 24 proteins assigned to functions involved with head morphogenesis. Of these, 16 are involved in prohead formation/maturation, five are involved in DNA packaging, and three stabilize and complete the assembly of the phage head (Miller *et al.* 2003). Mutations in a number of these genes can cause distinctly abnormal phage head phenotypes. For example, mutations in gene 20, the portal vertex protein of the head, has been found to lead to the formation of 'polyheads' (Laemmli *et al.* 1970), where the phage capsid takes on an extended and tubular form without any hemispherical cap on either end (Figure 4.13, panel A). Mutations that were identified near gene 23, the precursor to the major head subunit, were found to cause 'giant heads' that were from 1.5 – 10 times the normal length of the T4 capsid (Figure 4.13, panel B) (Doermann *et al.* 1973). Mutations in what was originally identified as gene 66 can lead to a 'petite' phenotype, where capsids are significantly smaller, and exhibit an isometric shape rather than the standard prolate head shape of T4 (Doermann *et al.* 1973). This gene was described as being positioned between genes 23 and 24, a gene that upon sequencing of the genome was identified as *segD,* a homing endonuclease (Miller *et al.* 2003; Sokolov *et al.* 2018). This 'petite' phenotype

appears to have similar characteristics to the effect of the mutation in the S-PM2d mutant d85. The appearance of a similar phenotype could support the idea that this mutated gene is a novel structural gene in cyanophage that is involved in the assembly/maturation of the capsid.

The SNP in mutant d85 is located in a gene that is in a region next to a number of hypothetical proteins, and a gene annotated as a structural protein. It is not uncommon for structural genes in phages to be arranged near each other (Hatfull and Hendrix 2011), so that they can be expressed together as a unit. The positioning of this gene near another structural gene could hint at a common arrangement, due to co-expression and a similar function. This suggests that this gene could be involved in the morphology of the phage head/prohead.

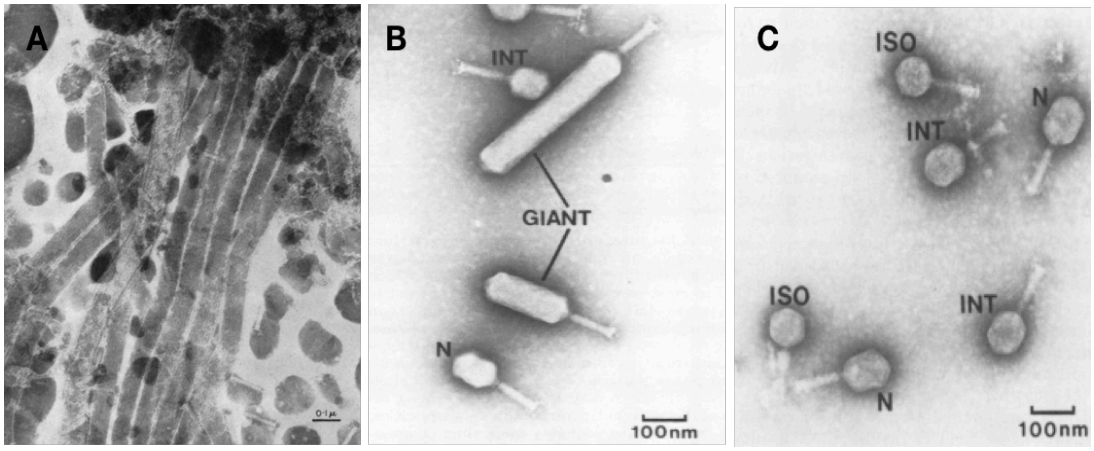


Figure 4.13 Variations in T4 capsid morphology that have previously been identified. Panel A shows the formation of a polyhead. Panel B shows a giant head morphology, whereas panel C shows a petite capsid morphotype (Laemmll *et al.* 1970; Doermann *et al.* 1973)

4.4.2 Changes in plaque size as an indication of phage fitness.

Plaque size of phages is largely determined by two things: the burst size (number of virions produced by each infected cell) and the time it takes the phage to lyse

an infected cell. In the case of the former, the more virion particles that are produced, the wider the potential area that they can diffuse into will be. In the case of latent period, the longer the phage progeny spends inside the cell, the less time it has to diffuse to a new host leading to a smaller plaque size (Gallet *et al.* 2011). Therefore, plaque size measurements have previously been used as a proxy for general phage fitness; where an increase in plaque size has been conferred by an advantageous mutation, and a smaller plaque size by a deleterious one (Gallet *et al.* 2011).

This has been shown in the phenotypes presented by various mutant phages that have been associated with changes in fitness. For example, mutations in the major capsid protein F of bacteriophage lambda have been identified as altering the plaque size of the mutant progeny. These mutants produced larger plaques that also contained more phages per unit area, therefore conferring an increase in phage fitness (Roychoudhury *et al.* 2013). Inversely, several mutants of bacteriophage lambda were isolated whose plaque-forming ability was more sensitive than that of the wild type to high temperature, extreme pH, or the presence of suppressor genes in the host. All of these mutants exhibited a small plaque phenotype that was attributable to a low burst size, and therefore, a general decrease in fitness. In addition to this, some of these small-plaque mutants were found to lead to the lysogeny of some of the cells and lysis of others, but with the production of such a small amount of progeny phage that a self-sustaining lysis process could not be achieved (Campbell 1961). Whilst host density has been found to influence plaque size, this should not be a factor in these experiments, as all phage were plated for plaques at the same time using the same culture. So,

differences in host cell density will be minimal. This strongly suggests that the small plaque phenotype of S-PM2dd mutant d80 is indicative of a decrease in phage fitness. However, this is not reflected in the results from the one-step growth analysis (see Figure 4.7). Mutant d80 is the only mutant that shows a similar latent period and an increased burst size, representative of an increase in phage virulence. This mutant appears to produce a higher number of progeny in a shorter amount of time when compared to the wild type, but these progeny exhibit a significantly reduced plaque size. In this case it appears that the smaller plaque size is actually indicative of an increase in phage fitness; it may be that the phage sacrifices individual virulence (i.e. plaque size) in favour of a larger number of progeny represented by the greater burst size, increasing the odds of further infection. It is therefore interesting that this SNP, at position 15951, does not appear in nature (as discussed in Chapter 3) as this would be a small change with a potentially greatly beneficial result. The fact that this SNP is not identified elsewhere suggests that this mutation must have a fitness cost in an environment different from that tested here.

It is interesting that whilst mutants d80 and d85 share the same SNP at position 100793 in the S-PM2d148 protein, only d85 shows the mutated head phenotype. The S-PM2dd mutant d80 does not, and shows a phenotype of reduced plaque size instead. There are a few reasons as to why this could be, one of which is that the additional SNP in d80 acts as a suppressor mutation. In this way we see the wild type capsid shape phenotype as the second mutation alleviates the phenotype of the other. It has been identified in bacteriophage ϕX174 that bacteriophages that obtain a fixed deleterious mutation are more likely to acquire a compensatory

mutation instead of a back mutation to silence the mutagenic phenotype. These mutations were just as likely to be intergenic as they were intragenic, and the likelihood of obtaining a compensatory mutation increased linearly with the severity of the prior mutation (Poon and Chao 2005).

Another reason could be due to the rate that the phage head is assembled and the balance of morphogenic components present; this was something that was identified by Floor (1970) as a potential reason for the lack of a previously mutagenic phenotype due to the amber mutation in gene 23 (major head protein) in phage T4, when accompanied by a second suppressor mutation in the genome. When the phage head protein is present at low abundance due to a deleterious mutation, the normal assembly process it follows may take place too slowly and so may allow for alternative pathways of head component interaction to occur; resulting in incorrectly assembled heads being formed. The second amber mutation in their work was in a gene involved in baseplate assembly. Tails (containing the baseplate second mutation) join to heads in a 1:1 ratio during assembly. As such, attaching to misassembled heads could be seen to 'trap' them in this form. If the number of tails was reduced due to the second amber mutation Floor postulated that this would give a greater amount of time and possibility for misassembled heads to break down and for correctly assembled ones to accumulate before tail attachment (Floor 1970). It is possible that the mutation at 15951 acts in a similar way, causing an imbalance in the components required for correct assembly or by changing the rate at which the assembly process happens so that the malformed heads caused by the SNP at 100793, could be broken down and correctly assembled heads could accumulate. However, if this were correct

you would expect the mutant d80 to have a longer latent period than mutant d85, something that is not reflected in the one-step results, where mutant d85 has a latent period of 18 hours compared to 13 hours in d80.

4.4.3 A SNP in a virion structural protein that is important for successful phage infection.

Mutant d39 has a single SNP in S-PM2d protein 173 at position 115714, annotated as a virion structural protein. Whilst this mutant appears to show no morphological differences from the results of TEM and plaque size analysis, it shows a drastically different infection profile. For this mutant there is a substantially extended latent period where over the course of this 24 hour experiment a significant burst was not observed. To initiate infection, bacteriophage virions undergo vast amounts of structural remodelling to enable efficient phage adsorption onto the cell surface and further infection (Hu *et al.* 2015). If the structural protein mutated here is involved in this remodelling, the SNP introduced may prevent this from happening, or weaken the interactions following this step. Although the results from one step analysis do not appear to show any differences in adsorption compared to WT, that would reflect changes affecting phage attachment. Previous SNPs in structural genes such as gp5, a part of the virion baseplate in T4, have been identified to enable lytic activity for the structural gene (Kao and McClain 1980). Previous experiments have shown that strains of T4 lacking a lysozyme, but that contain a SNP in gp5 are capable of lysis. Experiments showed that the lytic action of mutant gp5 caused cell lysis. It was therefore concluded that gp5 is involved in the early stages of tail assembly, as well as having a role in cell wall digestion to enable penetration of the tail tube

through the cell envelope. It is therefore not unthinkable that the SNP identified in this work may have a similar but opposite role, causing the suppression of lytic activity of this virion structural gene. This would explain the phenotype displayed in the one-step analysis where a significant burst and end of the latent period was not observed.

4.4.4 Assessing the effect of a SNP in *psbA* in the S-PM2dd mutant.

The structure of photosystem II has been reported at a resolution of 1.9 A (Umena *et al.* 2011). This has provided important information on the arrangement of protein subunits and cofactors, as well as the catalytic core of the water splitting centre. From crystal structures such as these and a number of targeted mutagenesis experiments, the functions of numerous amino acid residues in PsbA (a key component of the PSII complex) have been determined. PsbA is initially synthesized as a precursor form, with an extension on the C-terminal end of around 16 amino acids in cyanobacteria (Nixon *et al.* 1992). After insertion of the preprotein into the thylakoid membrane the extension is cleaved to form a functional PSII complex. PsbA contains 5 transmembrane helices (A-E), including a DE stromal surface helix and a CD luminal surface helix (Pagliano *et al.* 2013). The PsbA protein contains, or is responsible for the binding of, the majority of the ligands involved in the electron transport chain. Some of the most important residues involved in this are: Yz, which is Tyr161 and the special pair chlorophyll A P680 that is ligated by His198. A molecule of pheophytin A is bound by hydrogen bonds to Tyr126 and Gln130. Plastoquinone is ligated by His215 & Ser264. Most of the ligands of the Mn4CaO5 cluster are amino acids of PsbA: Asp170, Glu189,

His332, Glu333, Asp342 and Ala344 (Umena *et al.* 2011; Pagliano *et al.* 2013). These and other amino acid residues that have been implicated were identified as important through mutagenesis or structural modelling and are highlighted in green in Figure 4.14. The SNP in the psbASNP mutant at position 126728 causes a change from amino acid T-I at position 316, highlighted in blue in Figure 4.13. This is not a residue that has been subject to mutagenesis before and so it's importance in the structural formation and functionality of the protein has not been determined.

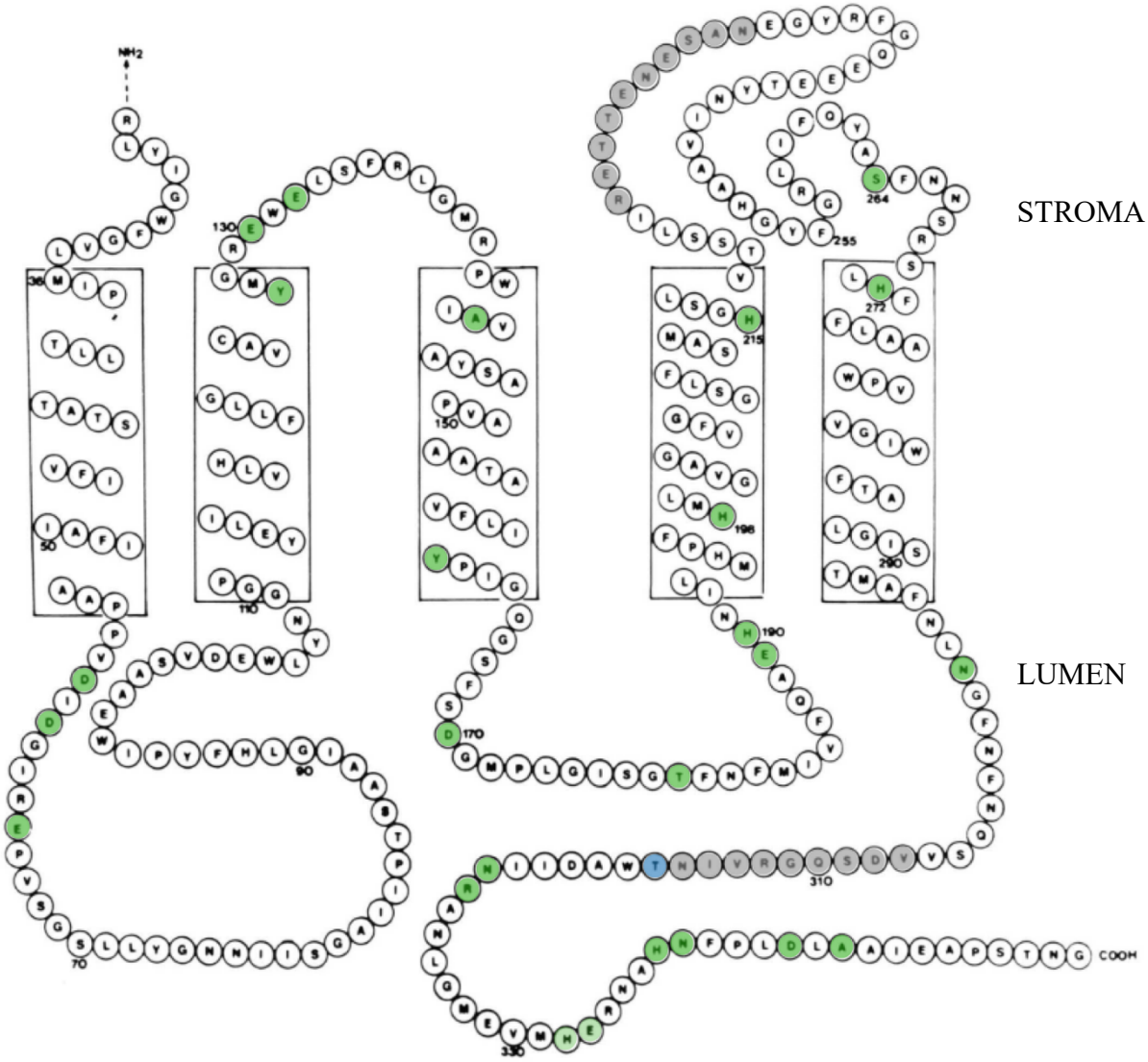


Figure 4.14 Amino acid structure of the *Synechocystis* sp. PCC6803 PsbA protein. Green residues are those that have been mutated in previous studies; the grey regions are those that are different between cyanobacteria and the viral PsbA. The blue residue is the tyrosine mutated in this work(Trebst 1986).

Whilst the viral and cyanobacterial PsbAs have high levels of similarity, there are a few regions in the viral version that noticeably differ from that of the host. The first of these is the PEST domain that occurs between transmembrane helices D and E (the first grey region in Figure 4.14) (Sharon, Tzahor, Williamson, Shmoish, Man-Aharonovich, Rusch, Yooseph, Zeidner, Golden, MacKey, *et al.* 2007). This is a region that has been implicated as the initial site of cleavage for turnover of the D1 protein, although this is debated (Greenberg *et al.* 1987; Puxty *et al.* 2015). The

second region that differs from the cyanobacterial PsbA is a variable region in between transmembrane helix E and the C-terminal end of the mature protein (second grey region in Figure 4.14), a site that is thought to aid in binding Mn^{2+} ions of the oxygen-evolving complex (Nixon *et al.* 1992; Sharon *et al.* 2007). The change in amino acid achieved in this work (threonine to isoleucine at position 316) occurs at the very end of this second variable region. PHYRE2 and SNAP2 analyses on this change in residue both suggested that this alteration would likely cause a change in the structure of the protein. If this amino acid is also involved in the binding of the Mn^{2+} ion to the oxygen evolving complex, as the region immediately next to it is, then a change in structure could be deleterious to PsbA function. If this gene is expressed by the phage to maintain host photosynthesis during infection, this change could cause a change in infection dynamics. The results of the one-step analysis indeed show that the infection kinetics of this mutant are delayed in comparison with that of the wild type, ultimately resulting in a reduced burst size (8 in the psbASNP mutant compared to 33 in the wild type). Together, these results strongly suggest that the SNP in the *psbA* gene causes a change in a residue that is important for phage infection, potentially through the mechanism previously suggested. As the phage *psbA* has been shown to be beneficial under high light conditions it would interesting to repeat the one-step experiment following a shift of host culture growth to high light conditions, and see if the phenotype is exacerbated, signifying its importance in 'viral photosynthesis'.

The psbASNP mutant contains three further SNPs in other regions of the genome, all of which are non-synonymous changes. It is unclear if these SNPs would have

any phenotypic change as one is intergenic, and the other two are in hypothetical proteins. The SNP at 49638 is in a protein surrounded by structural proteins which could give a clue as to its possible function but this is not confirmed. However, it is important to remember that any change in phenotype in this mutant may be the product of any of these mutations, or a combination of them and cannot be linked to a single SNP.

It is evident from this work that there are a large number of genes in S-PM2dd whose function remains unclear but that physiological characterisation of mutants created by random chemical mutagenesis is a potential method for illuminating these functions. This approach allows for the creation of SNPs that would not be selected for in the natural environment likely due to the resulting detrimental effect that they have. However, this allows us to directly infer a role that the gene may play a part in. In just the small pool of mutants characterised here a range of phenotypes were evident, all of which shed light on possible gene functions. These included mutants potentially identifying a new cyanophage structural gene/protein (d85) as well as those involved in cyanophage fitness (psbASNP, d39, d80).

Chapter 5: Investigation into the role of the cyanophage *psbA* gene and associated regulatory elements

5.1. Introduction

All cyanobacterial strains sequenced to date have been found to encode a small family of *psbA* genes (Mulo 2009). This gene encodes the D1 protein that forms part of the core photosystem II complex – that is vital for oxygenic photosynthesis. The cyanobacterial strain *Synechocystis* sp. PCC6803 was used in this work as a heterologous system due to the difficulties posed by constructing mutants in marine *Synechococcus*. *Synechocystis* is naturally transformable and a much more genetically amenable strain that is capable of growing photoheterotrophically (Barten and Lill 1995; Kufryk *et al.* 2002). In this model *Synechocystis* strain there are three copies of the *psbA* gene that are differentially expressed according to the growth conditions (Mohamed *et al.* 1993). *psbAII* is the most active copy of the gene, with *psbAIII* being upregulated to maintain total transcript levels if *psbAII* is damaged or under higher light conditions. *psbAI* was previously thought to be silent, but has been found to be expressed in microaerobic or anaerobic conditions (Mohamed *et al.* 1993; Sicora *et al.* 2009b).

The *psbA* gene is also found in cyanophages where it is considered an 'almost core' auxiliary metabolic gene (Millard *et al.* 2009; Sullivan *et al.* 2010b). It is widespread in nature and has been found in all but a minority of sequenced cyanophages (Crummett *et al.* 2016). The cyanophage *psbA* gene is known to be transcribed during infection and is capable of being translated into protein (Lindell *et al.* 2005). Therefore, it is thought that the viral copy of the gene is induced during phage infection of the host and used to maintain host photosynthesis, thereby enabling a more successful infection. Recent work has shown that cyanophage S-PM2 actively up-regulates its copy of *psbA* under high

light conditions (Puxty *et al.* 2018), which supports this view. However, the function of this gene has yet to be proven and an investigation into this is the core aim of this chapter.

By transforming variations of the S-PM2d cyanophage *psbA* region described previously (see section 1.3.2) into *Synechocystis* sp. PCC6803 I aimed to investigate whether the viral copy of the gene was capable of maintaining photosynthesis in the host, providing evidence it could act as a 'viral life support' and being greatly advantageous during infection. The inclusion of associated regulatory components in these constructs would also allow me to shed light on the functions of these regions.

The aims of this chapter were:

a) To construct mutants of *Synechocystis* sp. PCC6803 that test the function of the cyanophage *psbA* gene, and its associated regulatory elements. This includes the intron, homing endonuclease, and antisense RNA.
b) To determine if the cyanophage *psbA* gene is capable of maintaining photoautotrophic growth in the host.
c) To assess whether the associated regulatory elements have any effect on expression of the *psbA* gene, and its photosynthetic capacity.

5.2. Methods

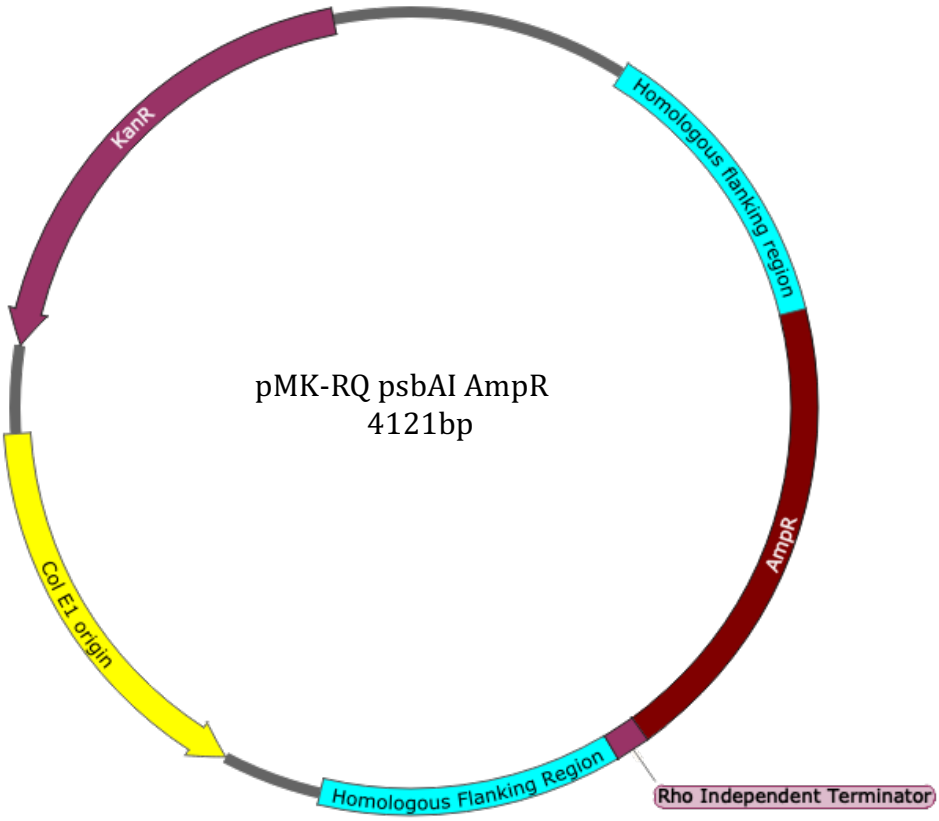
5.2.1 Construct design in glucose tolerant and wild type *Synechocystis* sp. PCC6803 strains.

In initial experiments constructs for use in both the wild type and glucose tolerant strains of *Synechocystis* sp. PCC6803 (Chapter 2, Table 2.1) were designed to determine the function of the S-PM2d phage version of the *psbA* gene, as well as the intron within the gene, the antisense RNA and the homing endonuclease adjacent to *psbA*. These strains initially required the replacement of additional *psbA* genes (I & III) with antibiotic resistance cassettes, so that any photosynthetic activity detected was due to the only remaining copy of the *psbA* gene, *psbAII.* Therefore, constructs were designed to replace *psbAI* & *III* (Figure 5.1A and C). An intron-less S-PM2 *psbA* construct with gentamicin resistance was also designed (Figure 5.1B). In this way I hoped to sequentially replace the *psbA* genes, to be able to identify if the intron-less phage copy of *psbAII* was capable of solely maintaining photosynthesis in *Synechocystis*. Constructs were synthesised by GeneArt (ThermoFisher Scientific) (Sequences can be found in appendix 9) and primers were designed to check the insertion of constructs and are listed below (Table 5.1). PCR conditions were as follows: 95°C 1 min, 95°C 30secs, 55°C 1 min, 72°C 3 min, repeated 32 times from stage two, then 72°C for 5 min before being held indefinitely at 4°C, and the product Sanger sequenced (Appendix 10).

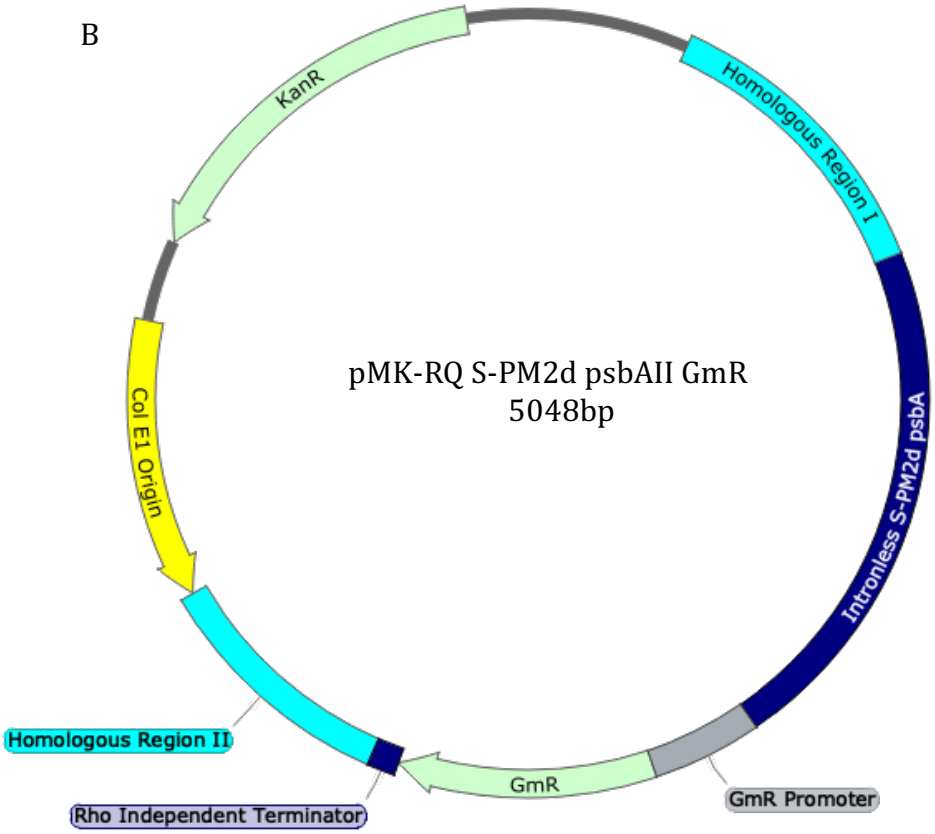
Table 5.1 Sequence of primers used to check the sequences of the *psbAI, II,* and *III* constructs.

Primer Name	Sequence 5'-3'
psbA1_flank_F	**CAGGCTATGTCCCGCTTAAAC**
psbA1_flank_R	**GCCATTGTCATGTTTAGGTTG**
psbA3_flank_F	**GAAAGGACATAGGCCAAGG**
psbA3_flank_R	**GGACGGCAAGTAACAGATC**
psbA2_flanks_F	**GCTTCGTGTATATTAACTTCCT**
psbA2_flanks_R	**GATGGCAATCAAGATCAGC**

A


KanR
Homologous flanking region
pMK-RQ psbAI AmpR
4121bp
AmpR
Col E1 origin
Homologous Flanking Region
Rho Independent Terminator

B


KanR
Homologous Region 1
pMK-RQ S-PM2d psbAII GmR
5048bp
Intronless S-PM2d psbA
Col E1 Origin
Homologous Region II
GmR
GmR Promoter
Rho Independent Terminator

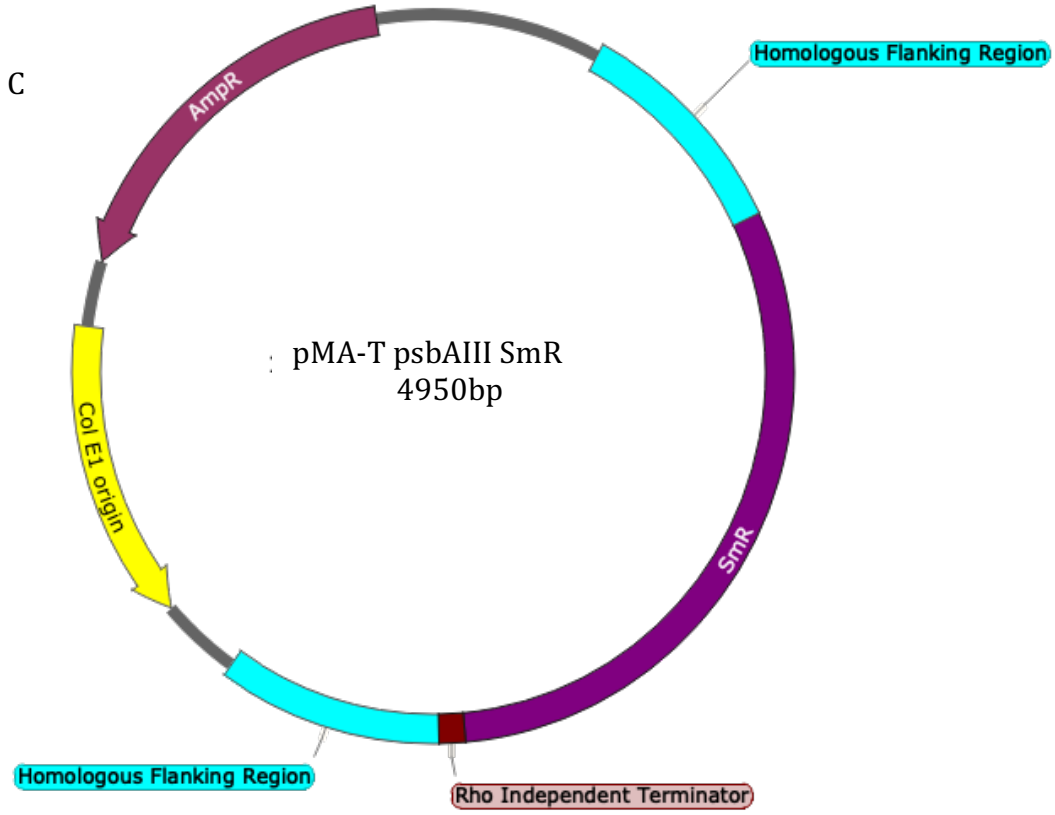


Figure 5.1 Plasmids used to knock out *psbAI* and *III* in *Synechocystis*, and to replace *psbAII* with an intronless S-PM2d phage *psbA* version. Panel A is the plasmid designed with homologous flanking regions to *psbAI* and with apramycin resistance. Panel B is designed with flanking regions homologous to *psbAII* and contains an intron-less version of the S-PM2d phage *psbA* gene along with gentamicin resistance. Panel C is the plasmid designed to replace the final *psbA* gene, and has flanking arms homologous to *psbAIII* and contains spectinomycin resistance.

5.2.2 Construct design in the DA1:DA2Em:DA3-H6 strain of *Synechocystis* sp. PCC6803.

After the *Synechocystis* knock out strains of Rob Burnap became available (Chapter 2, Table 2.1) it became necessary to redesign the constructs due to the resistance cassettes used in the creation of these knock out strains. The gentamicin resistance in my constructs was no longer a viable option, so constructs were designed to include apramycin resistance as well as a variation on the S-PM2d

psbA region, to help characterise the function of the associated regulatory elements in addition to the *psbA* gene itself (Figure 1.5)

The designs of the constructs can be seen in Figures 5.2-5.5, and the sequences are shown in the Appendix (see appendix 9). The first of these constructs (Figure 5.2) contains the homologous flanking regions either side of an apramycin resistance cassette containing a promoter and a resistance gene. The second homologous region is ended by a rho independent terminator. The construct was created to check that there were no phenotypic effects caused by the insertion of the apramycin resistance gene. The second (Figure 5.3) contains the same homologous flanking regions as in the apramycin control construct, as these flank the *psbAII* gene. It also contains the same apramycin resistance cassette that is preceded by the S-PM2d *psbA* gene sequence with the sequence of the intron removed. This should be the same as the sequence of the gene after the intron has been spliced out. The third construct (Figure 5.4) is exactly the same sequence as the second construct, except that the sequence for the intron has not been removed. Finally, construct four (Figure 5.5) contains the entire S-PM2d *psbA* region before the same apramycin resistance cassette. This region is the sequence from the start of the *psbA* gene through until the end of the homing endonuclease (F-cphI), this should allow it to splice the intron of *psbA* to allow the expression of the *psbA* gene. The psbA2_flanks primer pair (Table 5.2) were used to check the sequence of constructs, with the same PCR conditions as described in section 5.2.4, sequences can be found in appendix 11. The size of each insert sequence can be found in the description of each plasmid map, and where these primers target is shown in Figure 5.12.

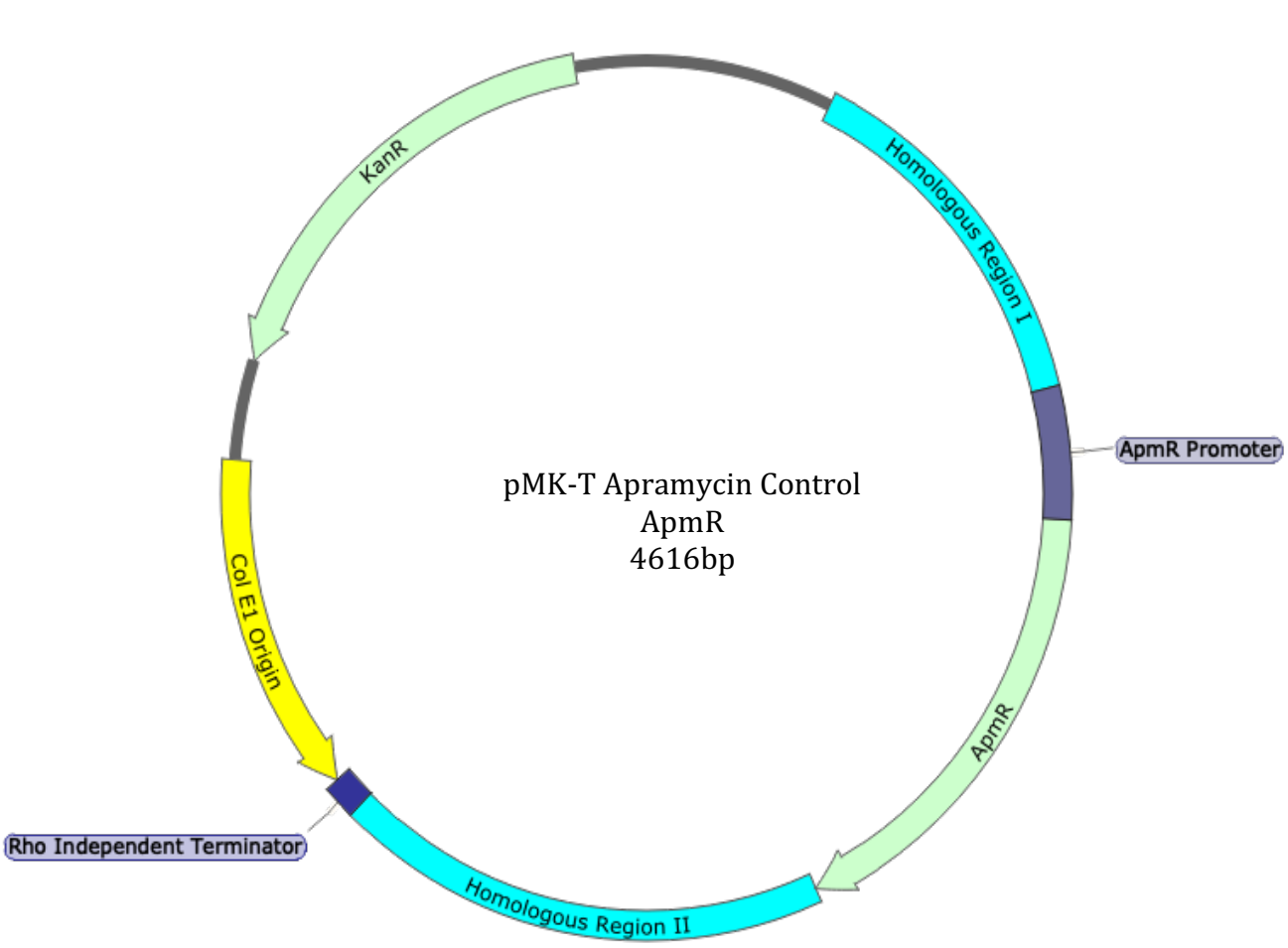


Figure 5.2 Apramycin control construct containing only a region that gives resistance to apramycin. The apramycin resistance cassette is situated downstream of a promoter region. On either end of the insert are regions of homology to either side of the *psbAII* gene in *Synechocystis* sp. PCC6803. The size of the insert region is 2223 bp.

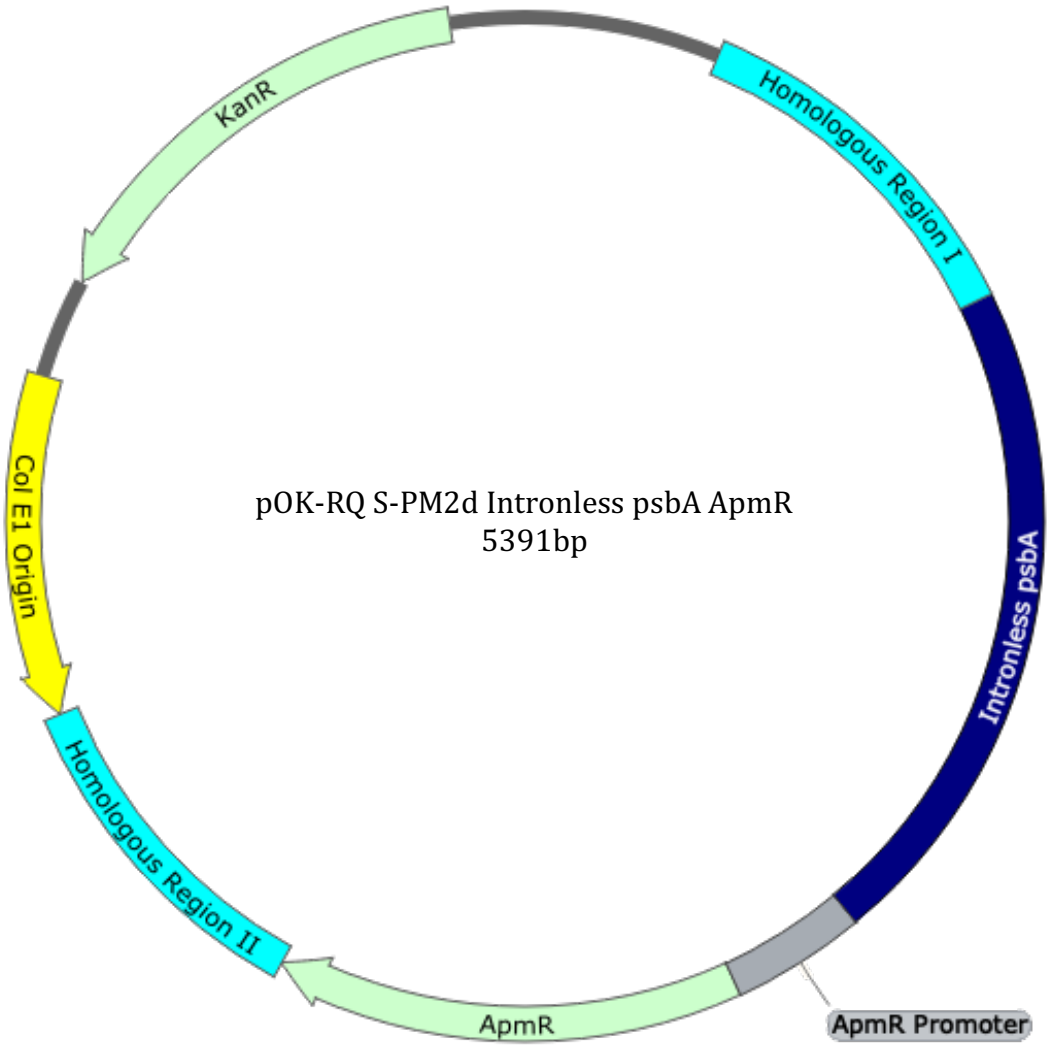


Figure 5.3 Construct containing apramycin resistance and the intron-less version of the cyanophage S-PM2 *psbA* gene. Upstream of the apramycin resistance region that is the same as in Figure 5.2, there is the coding sequence for the *psbA* gene of cyanophage S-PM2, the sequence corresponding to the intron has been removed to form what would be the 'spliced' version of the gene. No promoter has been added so that the insert will be expressed at the same level as the native *psbA* gene would be. The same homologous regions corresponding to either side of the *psbA* gene in *Synechocystis* sp. PCC6803 are used. The size of the insert region is 3302 bp.

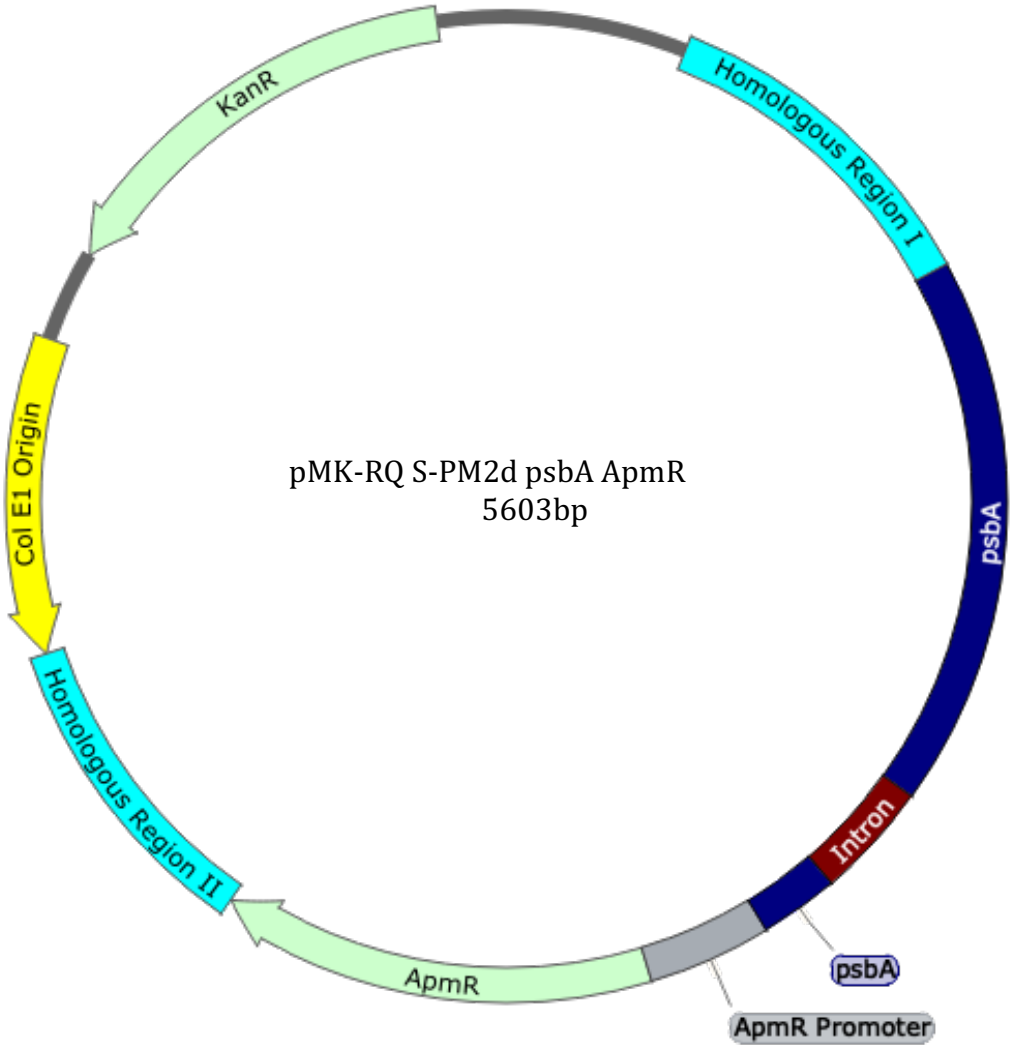


Figure 5.4 Construct containing apramycin resistance and the sequence of the cyanophage S-PM2 *psbA* gene, including the intron that is found before it has been spliced. Upstream of the apramycin resistance cassette is the sequence of the unspliced version of the cyanophage S-PM2 *psbA* gene, therefore it contains the sequence of the intron. The same homologous flanking regions as used in the previous two constructs are used here. The size of this insert with the unspliced intron is 3515 bp.

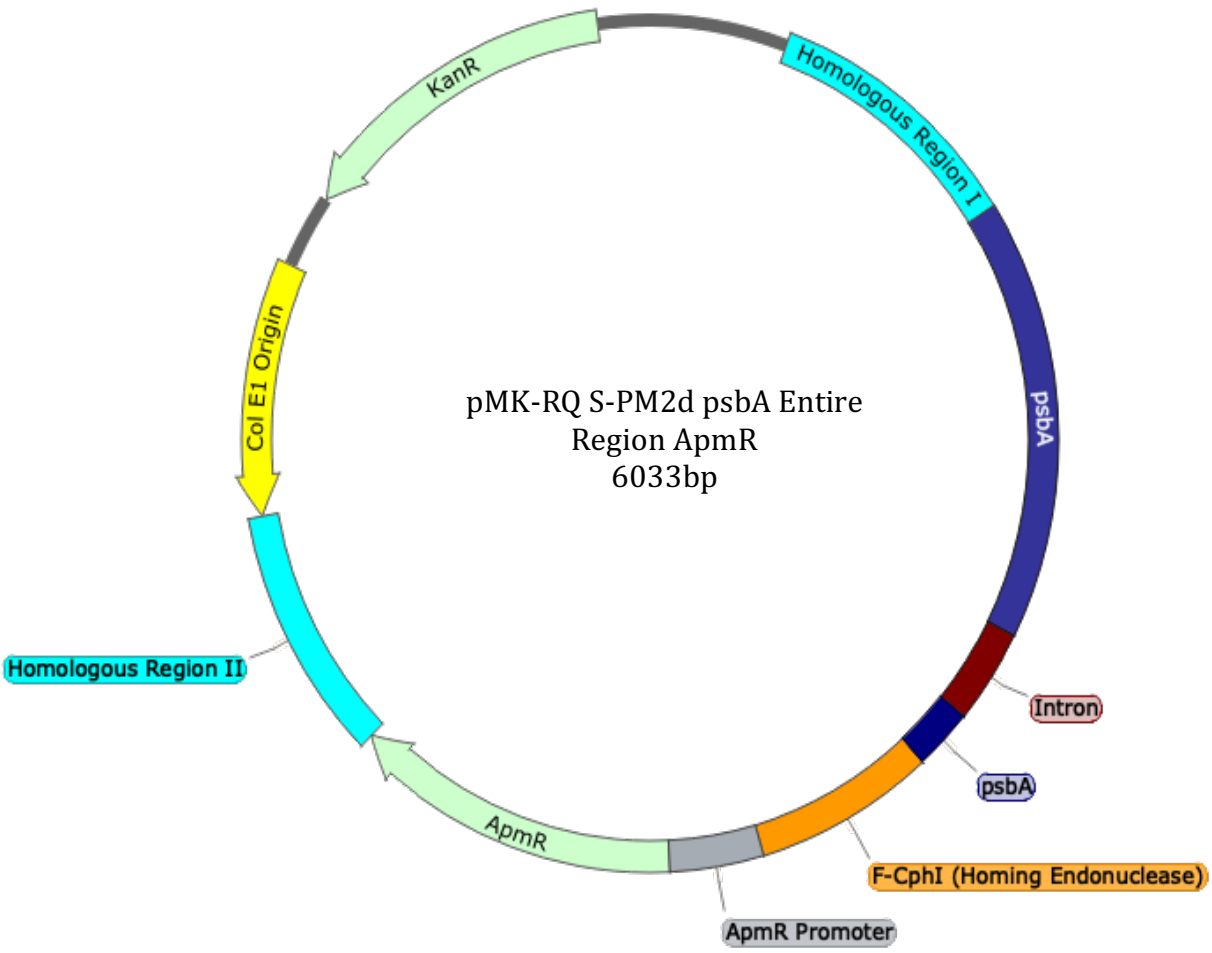


Figure 5.5 Construct containing the entire cyanophage *psbA* region from the start of the *psbA* gene to the end of the homing endonuclease, in addition to the apramycin resistance cassette. This construct will allow me to test whether the homing endonuclease F-CphI is capable of splicing the intron in the S-PM2 *psbA* gene to form a functional gene. The same homologous flanking regions as used previously in the constructs above, are used here, and the size of the inserted region with the unspliced intron is 4010 bp.

5.2.3 Plasmid transformation and purification.

The synthesised plasmids containing the *psbA* constructs were first transformed into *E. coli* DH5α using heat shock (see section 2.8). Colonies were picked using a sterile tip and incubated in a sterile universal containing 5 ml of warmed LB broth and the appropriate antibiotic overnight at 37°C with shaking. The plasmid was then purified from this culture using the QIAprep spin miniprep kit (Qiagen) according to the manufacturer's instructions.

5.2.4 Transformation of *Synechocystis* strains.

Constructs were transformed into all *Synechocystis* strains using an adaptation of the method described by Williams (1988). 50 ml of culture was pelleted by centrifugation with the Avanti J-25/JLA 10.5 rotor at 6000 g for 20 min at 30°C. The supernatant was removed and the pellet washed in 1 ml fresh BG11 medium. The pellet was then centrifuged again for 5 min at 30°C before the cells were resuspended in 100 µl BG11 medium in a 1.5 ml Eppendorf tube. ~1ng of purified plasmid was then added to the *Synechocystis* cells that were then incubated in the dark overnight at 30°C, with shaking. The cell mix was divided and spread onto BG11 solid medium and incubated under constant illumination at 25 µmol m^{-2} s^{-1} for 24 hours. Selective antibiotics were injected under the agar plate and incubated for up to 30 days under constant illumination at 25 µmol m^{-2} s^{-1}, until transformant colonies appeared. Colonies were picked using a sterile loop and streaked onto fresh BG11 solid medium containing the selective antibiotic. This process was repeated multiple times to ensure segregation of the transformants.

The correct insertion and segregation was confirmed using PCR: 95°C 1 min, 95°C 30secs, 55°C 1 min, 72°C 3 min, repeated 32 times from stage two, then 72°C for 5 min before being held indefinitely at 4°C, and the product Sanger sequenced. Primer pair 1 targets the upstream flanking region and start of the construct, primer pair 2 targets the downstream end of the construct and the downstream flanking region for the transformation into WT *Synechocystis* sp. PCC6803 (Table 5.1). Figure 5.8 shows where these primers bind. The psbA2_flanks pair targeted the homologous regions in the constructs and was used for the apramycin control, the intron-less *psbA,* and the entire *psbA* region transformants, figure 5.12 show where these primers bind. These were used when checking for insertion into the DA1:DA2Em:DA3-H6 strain of *Synechocystis*.

Table 5.2 Primer pairs used to check for insertion of the old and new constructs into the WT and the DA1:DA2Em:DA3-H6 strains of *Synechocystis*.

Primer Name	Sequence 5'-3'
psbA_6803M_P1_F	**GATAGGAGCCATCTTGCCC**
psbA_6803M_P1_R	**CTGTTGAGCGATGGATGCA**
psbA_6803M_P2_F	**GATATCGACCCAAGTACCGC**
psbA_6803M_P2_R	**GGATAAACCGCCTGATAGGTG**
psbA2_flanks_F	**GCTTCGTGTATATTAACTTCCT**
psbA2_flanks_R	**GATGGCAATCAAGATCAGC**

5.2.5 Antibiotic sensitivity and transformation efficiency of the WT & GT *Synechocystis* sp. PCC6803.

To enable efficient selection using gentamicin resistance, the sensitivity of the WT and GT *Synechocystis* strains to the antibiotic gentamicin was determined. To do this, increasing concentrations of gentamicin were added to the same volume of

log-phase *Synechocystis* cells in a 24 well plate. Cells were visually inspected for signs of cell death for up to 5 days afterwards.

Whilst *Synechocystis* sp. PCC6803 GT is a naturally competent strain, a rolling method of transformation was used to identify the optimal period of transformation during the growth cycle. Cells were transformed as described above, every day for 14 days using the intron-less phage *psbA* construct (Figure 5.1, panel B). The number of gentamicin resistant transformants that appeared after 30 days were counted and the transformation efficiency calculated using the following formula:

*Transformation efficiency = (number of colonies on plate)/(ng of vector) * (10^3ng /μg) * (final dilution)*

5.2.6 Growth curves

Growth of wild type and transformant strains was measured by taking spectrophotometric measurements at $OD_{750\ nm}$ every 24 hours, for at least a week. For each strain to be measured an initial starter culture was grown to log phase. 50 ml of this culture was spun down to pellet cells at 3220 g for 10 min at 23°C (Eppendorf Centrifuge 5810 R). The supernatant was removed and the pellet washed in fresh media and the spin repeated. The pellet was then resuspended in ~5 ml of fresh media and the $OD_{750\ nm}$ measured again. Cells were added to a new 100 ml Erlenmeyer flask such that the starting OD was at the pre-determined level (usually around 0.05) and this was recorded as time point 0. For photoautotrophic growth curves this fresh media was BG-11 without any glucose, DCMU, or antibiotics added; for photoheterotrophic growth curves these cultures contained

glucose and antibiotics as appropriate. Cultures were maintained under 25µmol/m^2/sec light conditions at 30°C with shaking and 1 ml aliquots were taken every day for at least 7 days to measure the $OD_{750\ nm}$ of the cultures.

5.2.7 RT-PCR to check gene expression

To check for expression of the transformed genes the GoScript Reverse Transcriptase kit (Promega) was used to synthesise cDNA from RNA for RT-PCR, from the wild type and transformed strains. The kit was used according to manufacturer's instructions with random hexamer primers and ~1 ug of purified RNA. Both a no template and no transcriptase control were used to detect any contamination. *psbA* gene expression was then investigated using the psbA2_flanks_F and 126728_R primers below (Table 5.2) and the following PCR conditions: 95°C 1 min, 95°C 30secs, 55°C 1 min, 72°C 3 min, repeated 32 times from stage two, then 72°C for 5 min before being held indefinitely at 4°C. The same PCR conditions were also used to check the expression of a set of housekeeping genes. These are the *rnpB* and *petB* genes, the primers were taken from Pinto *et al* and were used to confirm the reverse transcription process worked (Pinto *et al.* 2012). These primer sequences are shown below:

Table 5.3 Primers used to target *psbA* and housekeeping genes in *Synechocystis* sp. PCC6803.

Primer Name	Sequence 5'-3'
psbA2_flanks_F	**GCTTCGTGTATATTAACTTCCT**
126728_R	**CGTAGTGGAGGTATGGGTCTG**
SpetB1F	**CCTTCGCCTCTGTCCAATAC**
SpetB1R	**TAGCATTACACCCACAACCC**

5.2.8 SDS-PAGE and Western Blotting

To check for the formation of a functional protein SDS-PAGE and western blotting were used. 5 ml of log-phase cells were spun down at 4000 g for 5 min at room temperature (Eppendorf Centrifuge 5810 R). The pelleted cells were resuspended in 1 ml of fresh medium and spun down again at 13 000 g for 5 min (Eppendorf Centrifuge 5415 R). The pellet was weighed and Laemmli buffer added such that the protein was at the same concentration. 100 µl of the supernatant (mixture) was boiled at 85°C for 5 min before spinning at 13 000 g for 5 min. 40 µl of the supernatant was loaded onto a Novox WedgeWell 10-20% Tris-Glycine gel (ThermoFisher Scientific) along with the colour pre-stained protein standard, broad range 11-245 kDa (New England Biolabs). The gel was run for ~1 hour at 120 Volts in a Mini Gel Tank (ThermoFisher Scientific) and 1X SDS buffer (30 g Tris. 144 g Glycine, 10 g SDS for 1 L 10X solution), until the bands had migrated to the bottom of the gel. The gel was then transferred to a small volume of SimplyBlue SafeStain (Invitrogen) and left to stain for 1 hour. The gel was removed from the stain and washed several times in DI water before being left to soak for 1 hour for the protein bands to develop.

The same process was repeated for western blotting up until the staining step. Instead of this, the gel was loaded into a Mini Blot module (ThermoFisher Scientific), containing a PVDF membrane and filter paper (Novex, life technologies). This membrane had been activated by soaking in methanol for approximately 1 minute and then washed in transfer buffer (3 g Tris, 14 g Glycine, 10% Methanol in 1 L). The gel was transferred onto the membrane for 1.5 hours

at 120 Volts at 4 °C. Once the protein had been transferred the membrane was blocked by incubating in a 5% milk TBS-Tween solution for 1 hour with shaking. After this the membrane was washed in DI water before being incubated in 0.1 % powdered milk (Marvel, Premier Foods group) TBS-Tween with Rabbit anti-*psbA* antibody (raised specifically against cyanobacteria) diluted 1:10,000 according to manufacturer's instructions (Agrisera Antibodies, Sweden) for 30 mins with shaking. The membrane was then washed three times using 50 ml TBS-Tween each time. A 1:2000 dilution of Anti-rabbit IgG was added to 50 ml of 0.1% milk TBS-Tween before being left for 30 mins with shaking. The membrane was then washed with 50 ml TBS-Tween for 5 minutes and the process repeated 3 times. The final wash was conducted in 50 ml TBS for 5 mins with shaking. The membrane was then visualised using the Amersham ECL western blotting analysis system (GE Healthcare), which uses chemiluminescent detection, in conjunction with an ImageQuant LAS 4000 (GE Healthcare) imaging system.

5.2.9 Photosynthetic efficiency measurements.

Photosynthetic efficiency of the *Synechocystis* mutants was determined using 'fast' photosynthesis-irradiance curves. Fq/Fm was calculated after periods of exposure to increasing actinic light irradiances. After this the samples were subjected to 30 second intervals of illumination at 64, 128, 256, 448, 640, 832, 1984 μE m-2 s-2. After each period of exposure to actinic light the instantaneous fluorescence (F') was measured. A saturating flash of light was then delivered to cells in order to measure the maximum fluorescence (Fm'). Fq/Fm was calculated as (Fm'-F'/Fm') according to Genty *et al.* (1989). The relative electron transport

rate (rETR) was then calculated using the following equation, where I is the incident irradiance, 0.5 is a constant for the amount of photons that reach PSII, and 0.84 is a constant of relative absorption.

$$rETR = (Fq/Fm) * I * 0.5 * 0.84$$

5.2.10 Oxygen evolution measurements.

To measure oxygen evolution of the cells the chlorophyll *a* concentration was first measured. 1 ml *Synechocystis* culture was spun down for 10 min at 13,000 g and room temperature (Eppendorf Centrifuge 5415 R). 250 µl cells were then diluted in 1 ml BG11 and extracted in 100% (v/v) methanol. These extracts were left for no longer than 30 minutes in the dark before measuring absorbance at 663 nm and 645 nm. The chlorophyll *a* concentration was then calculated using the following formula:

$$\text{Chlorophyll a conc. (mg mL}^{-1}) = 12.7 * \text{Abs at 663nm} - 2.69 * \text{Abs at 645 nm}$$

Oxygen evolution was measured using an O_2 microsensor (Unisense, Denmark) and phytoPAM (Walz, Effeltrich, Germany). The O_2 electrode was polarised and calibrated using 100% saturated ddH_2O at room temperature and 0% saturated ddH_2O. To make 100% saturated ddH_2O oxygen was bubbled through the ddH_2O using a fish pump (Tetra) for 10 mins at room temperature. For 0% saturated ddH_2O a solution of 0.1M NaOH and 0.1 M sodium ascorbate was used. To make readings samples were diluted to a final chlorophyll *a* concentration of 5 µg in 3 mL of fresh media in the cuvette. $NaHCO_3$ was added to a final concentration of 10 mM and the magnetic stirrer added. O_2 production was measured across a light

curve and recorded using Phytowin software in conjunction with a PhytoUS measuring head.

5.3 Results

5.3.1 Establishing the antibiotic sensitivity and transformation efficiency of the wild type and glucose tolerant *Synechocystis* sp. PCC6803 strains

The antibiotic sensitivity of both the standard and glucose tolerant strains of *Synechocystis* sp. PCC6803 were determined and found to be highly sensitive to gentamicin. The wild type and glucose tolerant strains were both sensitive at the smallest concentration of gentamicin tested which was 2.5µg/ml, as can be seen by the pattern of cell death in Figure 5.6. Consequently, appropriate gentamicin concentrations (10 µg/ml) were applied during the transformation period following. Significantly higher (100 µg/ml) concentrations were also used when transferring into liquid to encourage segregation of transformant strains.

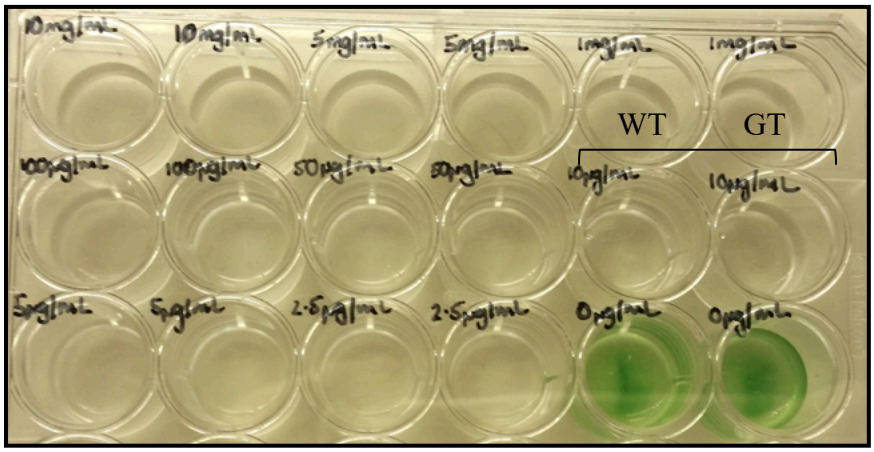


Figure 5.6 Determination of antibiotic sensitivity of the wild type and glucose tolerant *Synechocystis* sp. PCC6803. Each left well in a pair is the wild type, 'WT', each right well is the glucose tolerant strain 'GT'. An example of this has been shown in the image.

The daily transformation of the host *psbA* with an intron-less S-PM2d version was successful and transformants with gentamicin resistance appeared at a maximum of 304 transformants/µg DNA. During the exponential period of *Synechocystis* growth, transformation efficiency appears to follow a cyclical pattern (Figure 5.7).

The peak in transformation correlates with the height of the exponential period and drops as the growth rate plateaus. Therefore, further transformations were carried out during the exponential phase of growth, to increase chances of success.

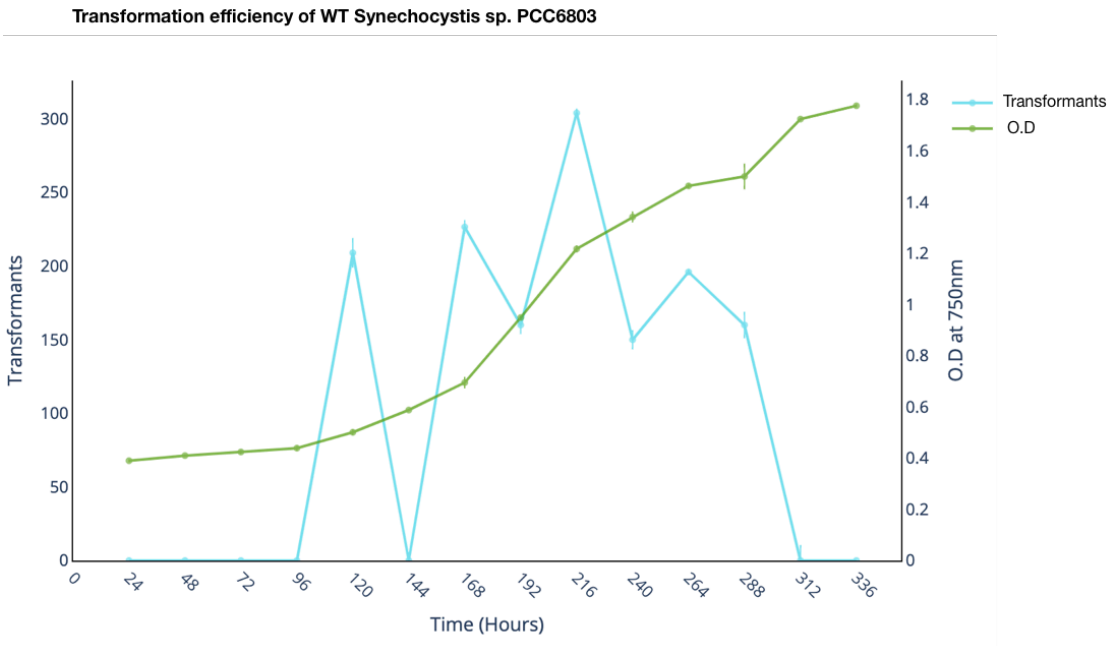


Figure 5.7 Transformation efficiency of WT *Synechocystis* sp. PCC6803 shown with the corresponding optical density of the culture.

Confirmation of insertion of the intron-less S-PM2d *psbA* in to the host was done via the use of two primer pairs targeting outside the flanking region of homology, and within the construct, the regions where these primers target is shown below (Figure 5.8). These products are shown in Figure 5.9 where lanes 2-10 are the products of 9 transformant colonies amplified using primer pair 1 in Table 5.2. The lanes 11-20 are the same colonies amplified using primer pair 2 in Table 5.2, using the PCR condition described in 5.2.4. The products were then confirmed with Sanger sequencing as described in Figure 5.9 and can be found in appendix 12.

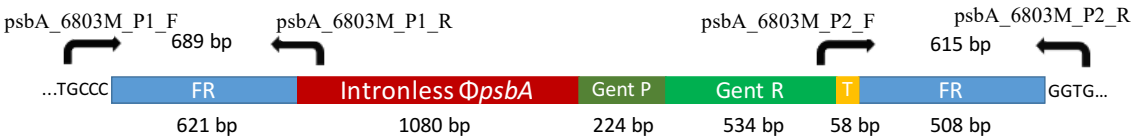


Figure 5.8 Design of primer pairs to confirm insertion of S-PM2 phage *psbA* into *Synechocystis*. Pair one flanks the outside of the homologous region and the start of the *psbA* gene. Pair two covers the end of the gentamicin resistance cassette and just outside the homologous flanking region.

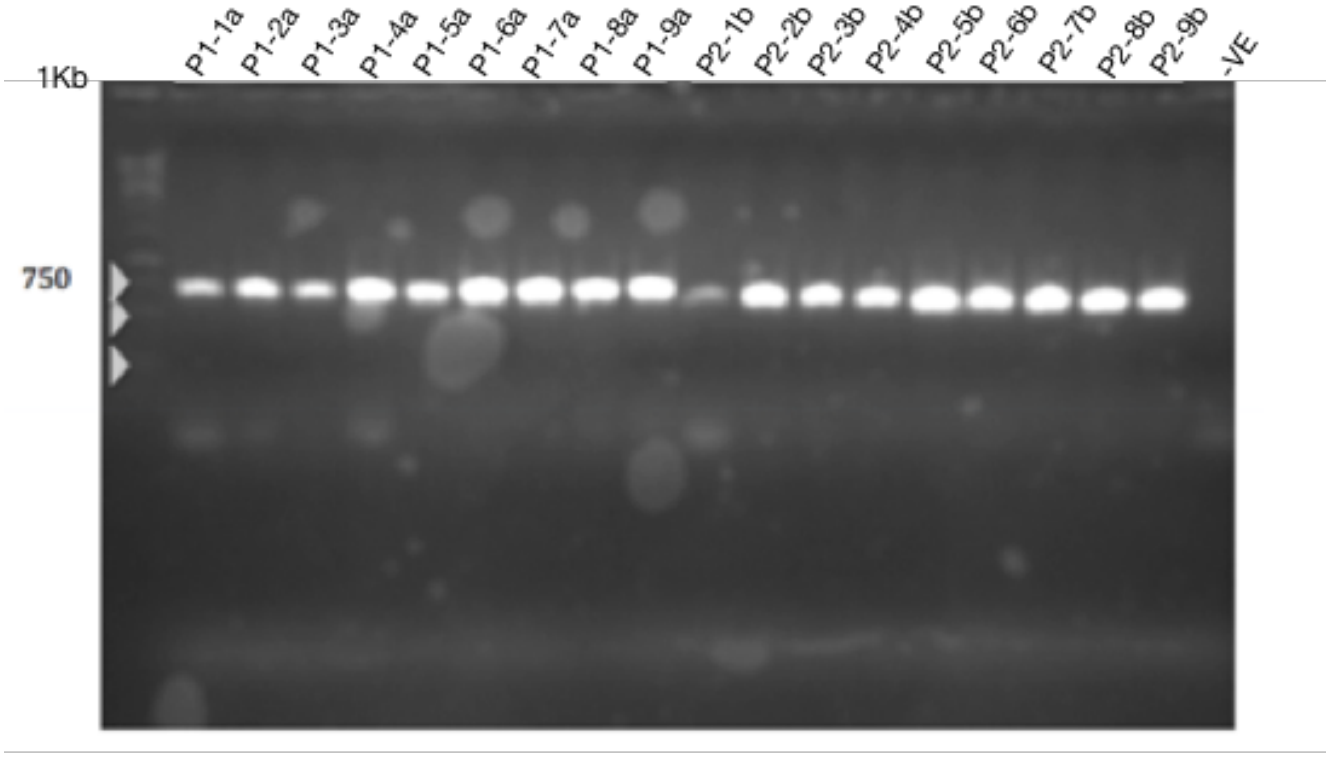


Figure 5.9 Confirmation of insertion of the phage *psbA* construct into *Synechocystis.* Lane 1 is the Fermentas 1 Kb DNA ladder. Lanes 2-10 are nine transformant colonies (referred to as 1a-9a) picked and used in the PCR with primer pair one (referred to as P1) as described in Figure 5.8. Lanes 11-20 are the same colonies used in lanes 2-9 (now referred to as 1b-9b), with primer pair two (referred to as P2) as described in Figure 5.8. Lane 21 is the negative PCR control containing ddH_20 instead of genetic material.

5.3.2 Expression of a phage *psbAII* gene in *Synechocystis*

No bands were present when the extracted RNA was tested for DNA contamination using PCR (gel not shown), so the RNA was used to make cDNA to check for expression of the *psbAII* gene in the mutant. The glucose tolerant *Synechocystis* sp. PCC6803 mutant containing the intron-less S-PM2 phage *psbA* copy shows expression of the gene using RT PCR (Figure 5.10). The no template and no transcriptase controls are blank suggesting no contamination. As was the negative control in lane 5 to check for contamination in PCR reagents.

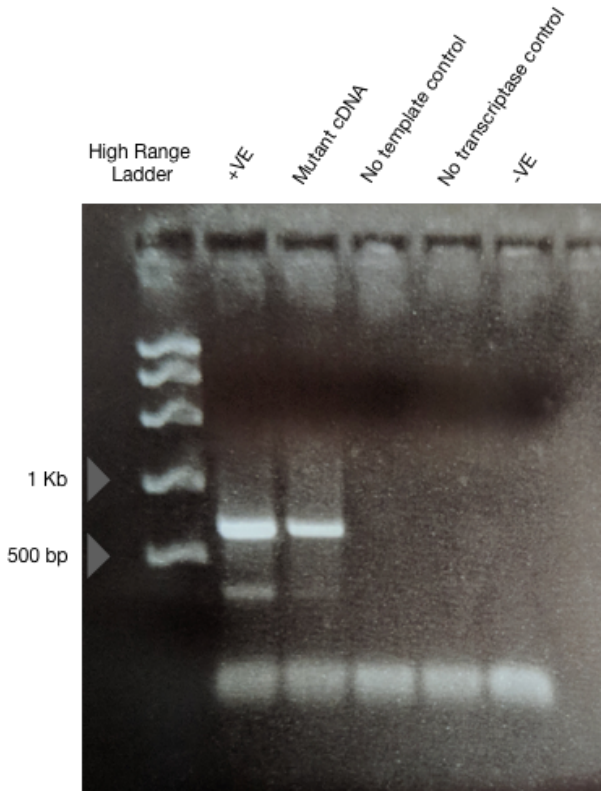


Figure 5.10 Confirmation of expression of intronless phage *psbA* in *Synechocystis* with PCR on the cDNA. Lane one is the Fermentas high range DNA ladder. Lane two is a positive control of construct miniprep to check the PCR worked. Lane three uses the cDNA created from the mutant *Synechocystis* RNA. Lanes four, five and six are the no template, transcriptase and negative control respectively.

5.3.3 Growth of the *Synechocystis* sp. PCC6803 GT mutant containing a phage *psbA*

If the insertion of the phage *psbA* was causing any kind of significant effect either beneficial or detrimental to the host in its ability to photosynthesise, one obvious side effect of this could be a change in growth rate, compared to that of the background strain, the glucose tolerant *Synechocystis* sp. PCC6803. Therefore, growth curves of the mutant, as well as both the glucose tolerant and the WT *Synechocystis* sp. PCC6803 strains were conducted. It is clear from Figure 5.11 that compared to both the wild type, and glucose-tolerant strain there was no significant effect on growth of the mutant *Synechocystis* sp. PCC6803 containing the intron-less phage *psbA* in the place of *psbAII*. All strains reach a similar final $OD_{750\ nm}$ and have very similar growth rates (Table 5.3) suggesting there are no lasting effects on growth.

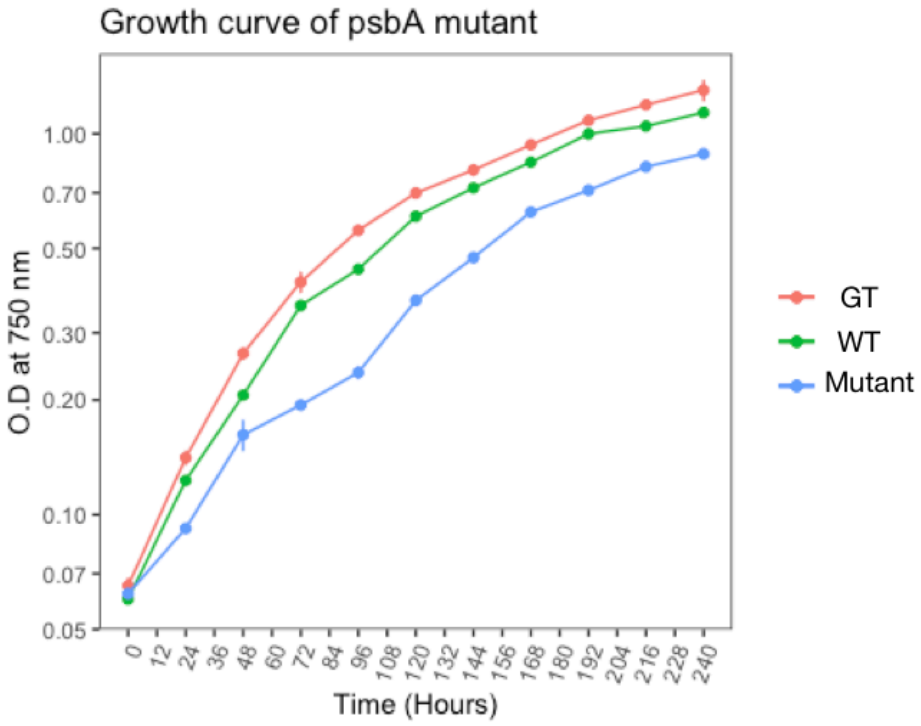


Figure 5.11 Growth curve of wild type *Synechocystis* sp. PCC6803 containing the intron-less phage *psbA* in the place of *psbAII.* Compared to the wild type and glucose tolerant *Synechocystis* sp. PCC6803 strains. 'GT' is the glucose tolerant strain, 'WT' is the wild type, 'Mutant' is the strain containing the phage *psbA* gene replacement. Error bars represent standard deviation of the mean.

Table 5.4 Specific growth rates of all strains of *Synechocystis* sp. PCC6803.

Strain	Growth Rate (h^{-1})
psbA mutant	0.011
WT	0.010
GT	0.011

5.3.4 Construction of *psbA* mutants in the DA1:DA2Em:DA3-H6 strain of *Synechocystis*

Of the *Synechocystis* strains provided by Rob Burnap, the DA1:DA2Em:DA3-H6 strain that contains no functional *psbA* genes was transformed in the same way as

the previous strains, using a rolling method during the exponential phase of growth.

Transformants for the apramycin control, intron-less S-PM2 *psbA*, and entire S-PM2 *psbA* region constructs were observed on plates containing apramycin after ~25 days. These were picked into liquid containing high levels of apramycin (100mg/ml) and checked for correct insertion and segregation using PCR. The primers used target the homologous flanking regions of the construct (shown in figure 5.12) and the products shown in Figure 5.13 were sequenced using Sanger sequencing to confirm the mutants. The entire region construct appears to be the size of the unspliced version of the *psbA* gene, suggesting that it contains the *psbA* gene containing the intron, and the homing endonuclease. This could mean that this construct will not be functional, as the spliced *psbA* is not present. Sanger sequencing results confirmed this (see appendix 13).

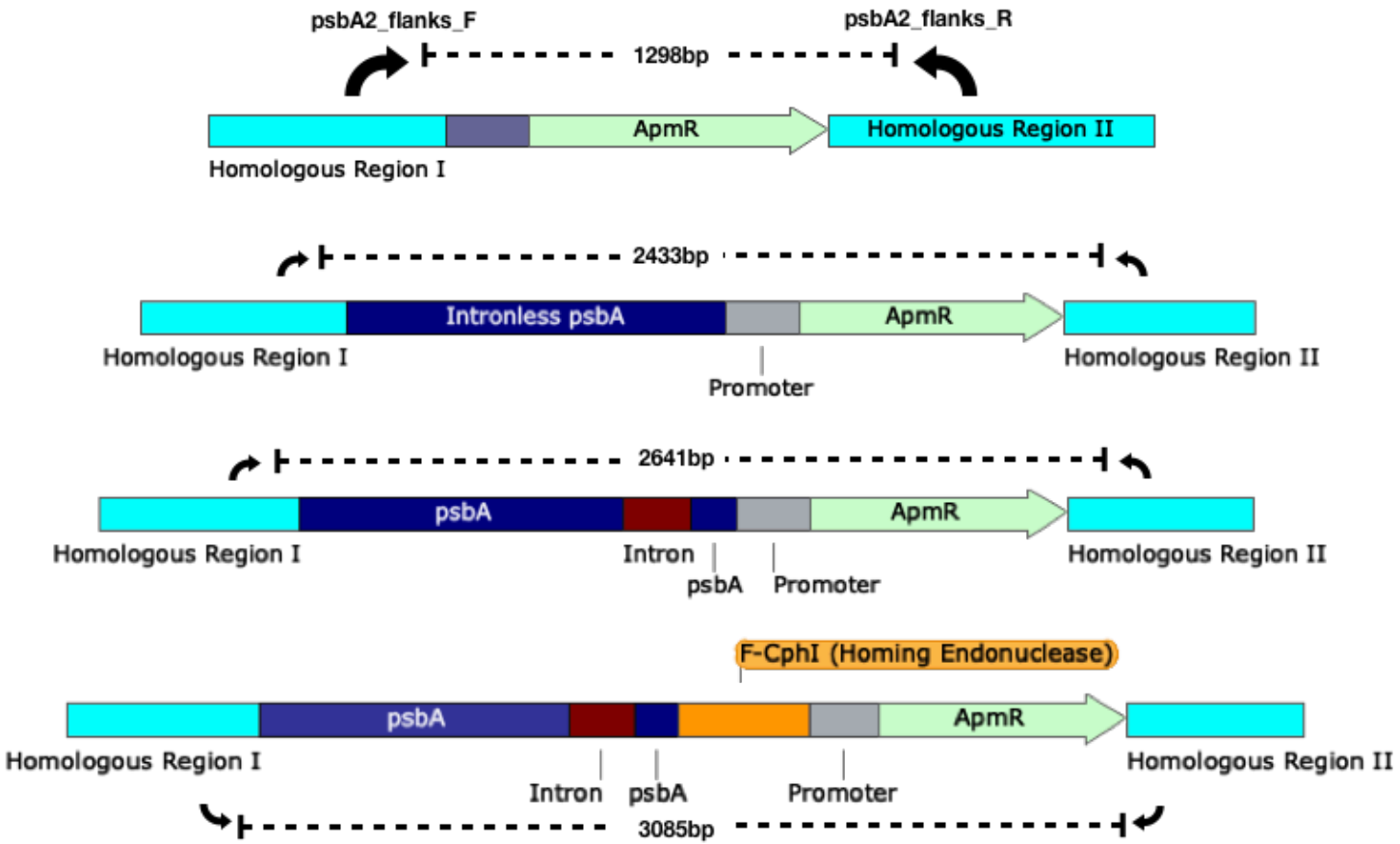


Figure 5.12 Design of the constructs transformed into DA1:DA2Em:DA3-H6 strain of *Synechocystis.* The size of amplicons produced from PCR using the psbA2_flanks primer pair are shown.

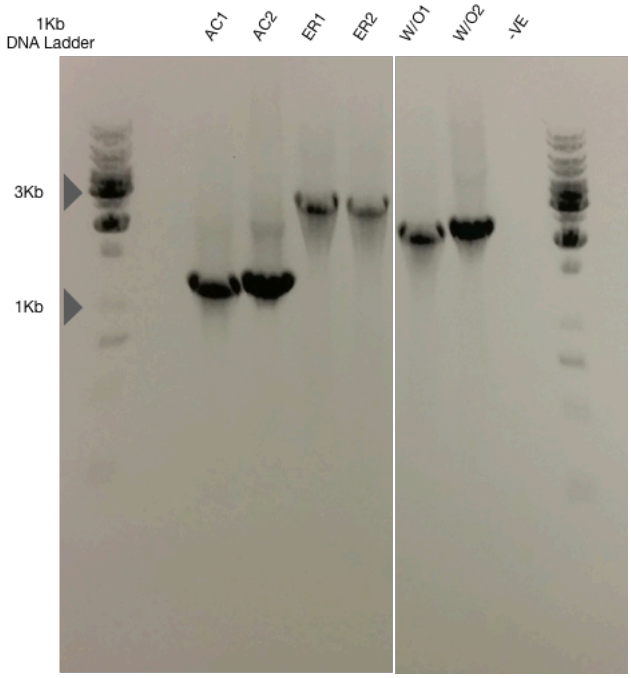


Figure 5.13 Confirmation of construct insertion into the DA1:DA2Em:DA3-H6 strain of *Synechocystis*. On either end is the Fermentas 1 Kb DNA ladder. The first two lanes are the mutant containing the apramycin control construct. Next to this is two lanes of the mutant containing the entire S-PM2 *psbA* region construct. These are followed by two lanes containing the mutant with the intron-less version of the phage *psbA* gene. Finally, is a lane containing the negative PCR control, containing ddH_20 instead of genetic material.

5.3.5 Growth of *Synechocystis* mutants under standard light conditions

Growth curves were conducted at a low concentration of glucose (2.5 mM) as trying to grow the strains purely photoautotrophically did not prove successful. As the apramycin control was essentially a remake of the DA1:DA2Em:DA3-H6 background strain and would not provide any new information about the function of *psbA* or associated regulatory elements, it was not used in any further experiments. It did however, show that the apramycin cassette was not lethal the *Synechocystis* strains and did not prevent growth or transformation.

Whilst there appears to be a short period of initial growth in the mutant strains after day two this plateaus and the rate of growth stays stationary. This is therefore likely due to growth on glucose even though multiple washes were used to prevent carry-over (Figure 5.13). The wild type continues to grow exponentially until day 7 and reaches a significantly higher final yield than either of the mutants. Between the two mutant strains tested there is no significant difference between the final yields and general pattern of growth.

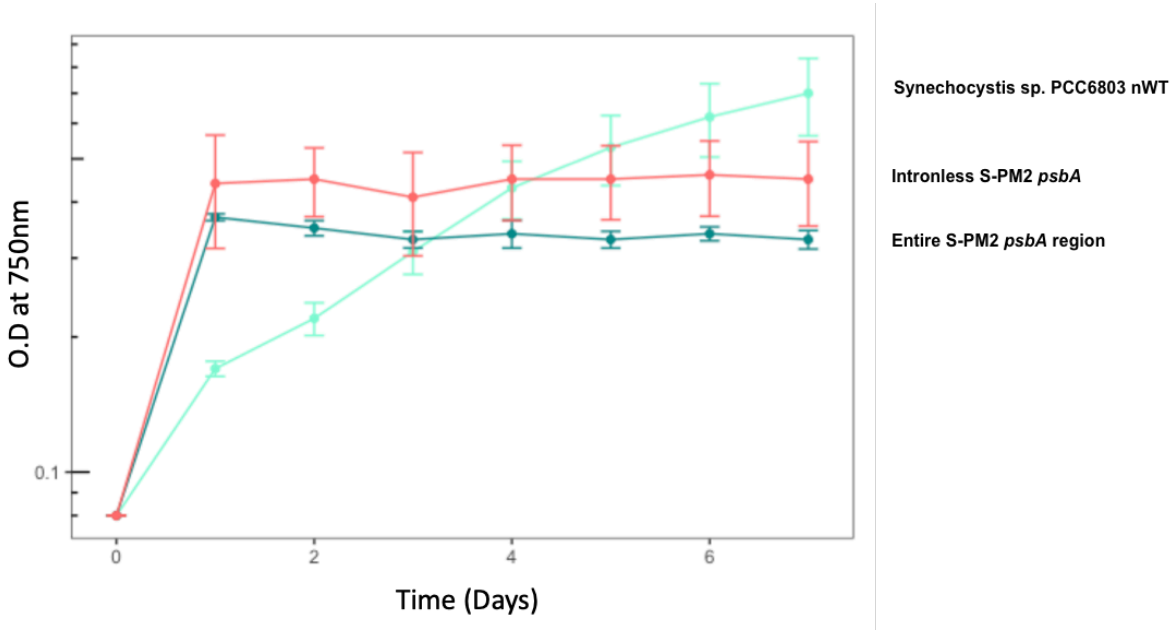


Figure 5.14 Growth curves of the *Synechocystis* mutants containing the intronless S-PM2d *psbA* and the entire S-PM2 *psbA* region, compared to the wild type strain 'nWT'.

5.3.6 Photosynthetic efficiency of mutants.

Oxygen is a product of photosynthesis and so by measuring the amount of oxygen evolution of the mutant compared to wild type, we can determine if there is an effect on the cells ability to photosynthesise. It is clear from Figure 5.15 that both

the intron-less and entire region mutants are producing negligible amounts of oxygen, and so are likely not photosynthesising.

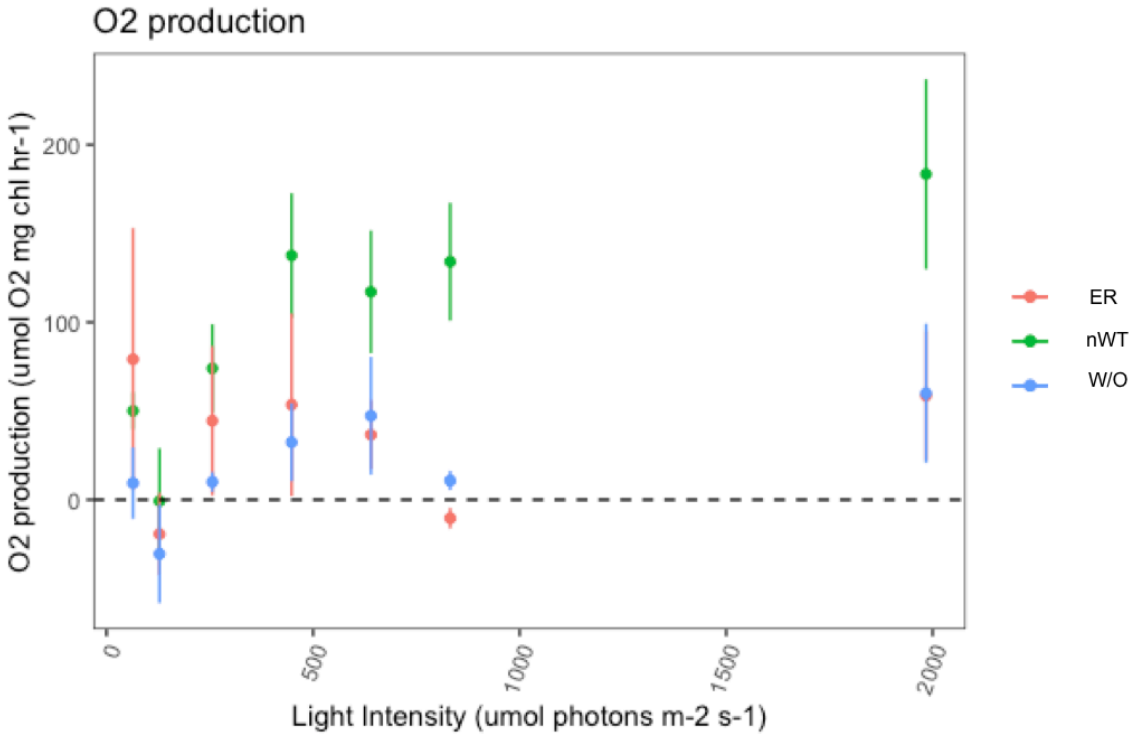


Figure 5.15 Oxygen evolution of the *Synechocystis* mutants containing the intron-less S-PM2d *psbA* (W/O), the entire S-PM2d *psbA* region (ER), compared to the wild type *Synechocystis* strain (nWT).

The relative electron transport rate for the mutants is also zero (Figure 5.16), so the flow of electron through the electron transport chain is non-existent. Limited amounts of oxygen production and no electron transport would suggest that the mutant cells are not capable of photosynthesis, and so are unable to maintain the growth of *Synechocystis* photoautotrophically.

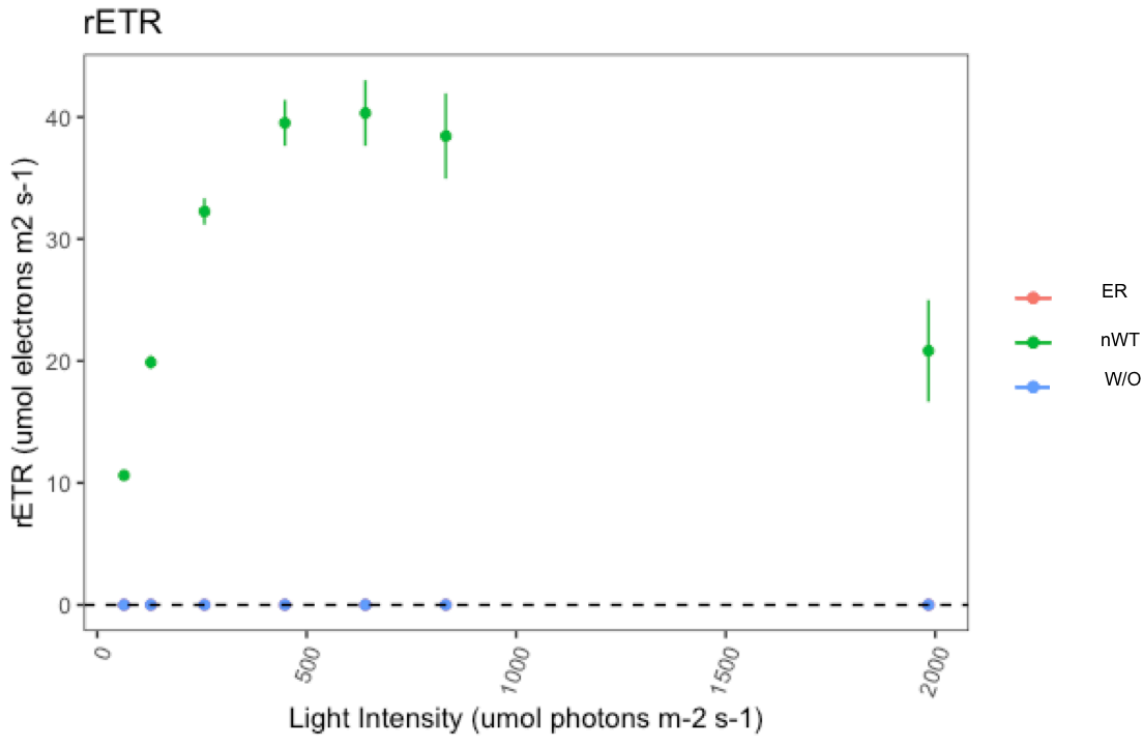


Figure 5.16 Relative electron transport rate measured in the *Synechocystis* mutants containing the intronless S-PM2d *psbA* (W/O), the entire S-PM2d *psbA* region (ER), compared to the wild type *Synechocystis* strain (nWT).

5.3.7 Assessing the expression of the cyanophage S-PM2 *psbA* variants in *Synechocystis*

RT-PCR using primers targeting part of the *psbA* gene listed in Table 5.2 show expression in the strain transformed with the intron-less version of the S-PM2 *psbA* (Figure 5.17) when using the conditions described in section 5.2.7. The no template and no transcriptase controls remain blank, suggesting that there is no DNA contamination. However, no such band can be seen in the other transformed strain, suggesting the gene is not expressed. The expression of the *petB* housekeeping gene (primers in Table 5.2) in both strains would indicate that this is not due to a problem in the synthesis of cDNA, and is indeed a negative result.

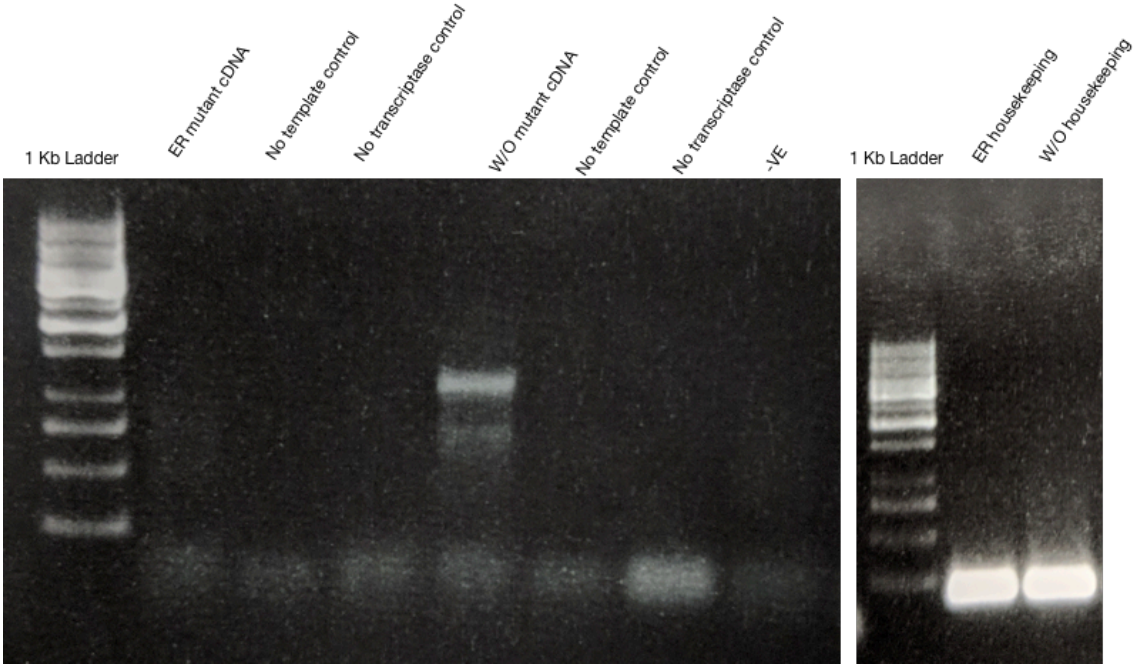


Figure 5.17 PCR of cDNA to confirm expression of *psbA* variants in the transformed *Synechocystis* mutants. The first lane is the Fermentas 1 Kb DNA ladder. On the right of this is the ER mutant cDNA and then it's no template and no transcriptase controls. Next is the intron-less *psbA* cDNA, and then it's no template and transcriptase controls. Next to this is a negative control. In the second gel image the first lane is the same ladder followed by the ER and then W/O mutant cDNA in conjunction with the *petB* primer pair, acting as a housekeeping gene check.

5.3.8 Assessment of cyanophage PsbA protein production in *Synechocystis* by western blotting

Protein analysis was performed on both the nWT strain as a positive control, the DA1:DA2Em:DA3-H6 background strain as a negative control, and the two transformed strains one containing the intron-less S-PM2d *psbA* gene, and the other containing the entire S-PM2d *psbA* region. The PsbA protein is 32 kDa in size, and a band matching this size can be clearly seen in the complemented nWT strain and is lacking in that of the DA1:DA2Em:DA3-H6 (Figure 5.18). The strain transformed with the entire *psbA* region was not included in the image below as there did not appear to be enough material loaded to be able to make an informed

decision about whether a protein was present or not. However, the mutant containing the intronless S-PM2 *psbA* does show evidence of a band at the same position as the wild type, although this could be argued to be inconclusive. A band would suggest that there is a functional copy of *psbA* in the mutant and corroborates the RT PCR results on the two strains.

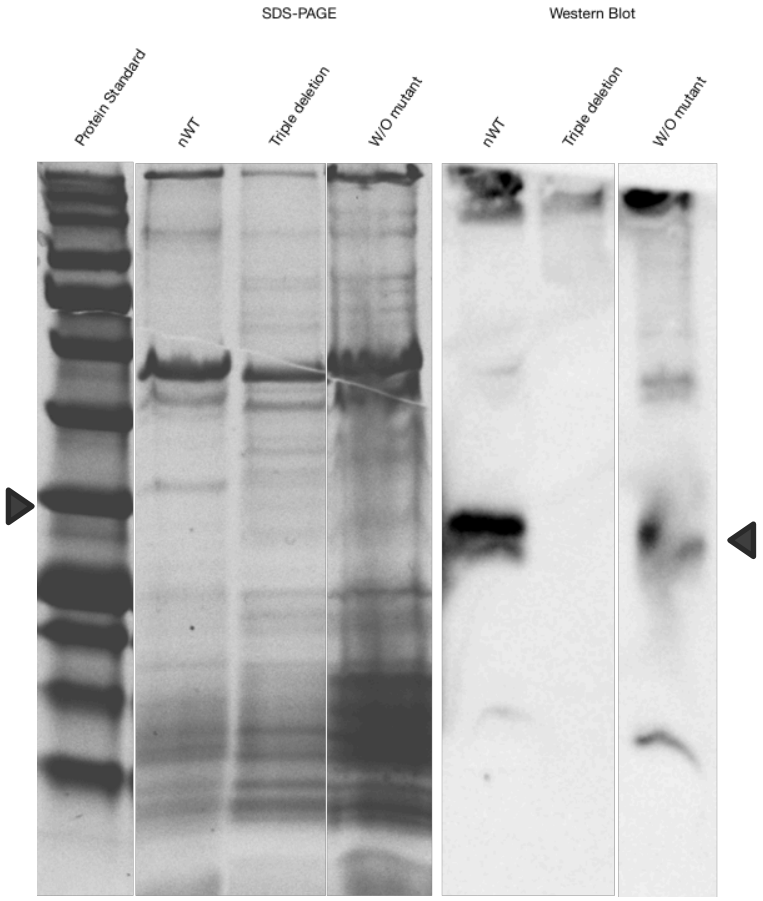


Figure 5.18 SDS-PAGE (left) and Western Blot (right) of the wild type (nWT), the DA1:DA2Em:DA3-H6 mutant that contains no *psbA* genes (referred to here as Triple deletion), and the mutant created in this work that contains the intronless phage *psbA* gene (W/O). The arrows indicate the PsbA protein.

5.4 Discussion

5.4.1 S-PM2d phage variants of the *psbA* gene are not capable of maintaining photosynthesis in *Synechocystis* sp. PCC6803

The cyanophage encoded *psbA* gene is a widespread AMG but whose function has not been proven. Previously, the S-PM2 cyanophage *psbA* gene copy has been shown to be transcribed during infection (Clokie *et al.* 2006; Puxty *et al.* 2016; Millard *et al.* 2010). It has also been shown that viral *psbA* mRNA can be translated (Lindell *et al.* 2005). This led to the hypothesis that the viral copy of *psbA* was acting as a kind of 'life support' to the host. In this way maintaining host photosynthesis whilst the phage redirected cell metabolism towards nucleotide biosynthesis and enabling a more successful infection. Results from Puxty *et al* (2016) support this, where it was shown that when infected with cyanophages S-PM2 and S-RSM4 the CO_2 fixation of the host *Synechococcus* was inhibited, yet PET was unaffected (Puxty *et al.* 2016). This suggests that cyanophages are capable of decoupling photosynthesis, using the ATP and reducing power provided by PET for viral morphogenesis rather than CO_2 fixation. However, the results shown here do not support this hypothesis. Both the relative electron transport rate (Fig. 5.15) and oxygen evolution (Fig. 5.14) are negligible in the *Synechocystis* strains containing the intron-less cyanophage *psbA* or the entire cyanophage *psbA* region showing these transformed *Synechocystis* strains are not photosynthesising. This is also supported by the fact that these strains were unable to be grown without glucose (see Fig 5.13).

It is likely that the mutants made in the *Synechocystis* sp. PCC6803 GT background were capable of growing as normal, and without glucose due to the expression of

other *psbA* copies in the genome. It has previously been shown that when the most active copy of the gene, *psbAII,* is damaged then *psbAIII* is up-regulated to a level that is similar to the undamaged *psbAII* (Mohamed *et al.* 1993). The RT PCR used to show expression of the cyanophage copy of *psbA* was not quantitative, so it could be that the gene was only being expressed at very low levels. However, as the construct was designed to be under the control of the host *psbAII* promoter, it should be expressed at the same levels as the native *psbA*.

The cyanophage *psbA* mutants constructed in the DA1:DA2Em:DA3-H6 strain showed no signs of being able to photosynthesise, even though Western blotting indicates a functional protein is produced in the intron-less phage *psbA* strain, and RT PCR shows that this strain expresses the cyanophage encoded *psbA*. Photosynthesis is driven via the flow of electrons from water splitting through PSII and PSI, and as such the rate of electron transport is used as a measure of the operating efficiency of PSII. In principle in photosynthesis there is a linear relationship between PSII operating efficiency and linear electron flow that allows the rate of noncyclic electron transport through PSII to be calculated using Fq'/Fm' (Baker 2008). In this work this rate of electron transport was zero; without this flow of electrons PSII would not be functional and photosynthesis could not be occurring in the mutants. This is corroborated by the results of the oxygen evolution measurements. Oxygen is a product of photosynthesis, after four successive photoreactions in the same PSII complex a molecule of oxygen is released (van Leeuwen *et al.* 1990). In this way, measuring the production of oxygen is also a method of measuring how well PSII is working. The negligible values found here corroborate the values of relative electron transport rate, and

further confirm the idea that the cyanophage *psbA* containing *Synechocystis* strains are not capable of photosynthesis.

There are a number of reasons why these strains are unable to photosynthesise. One potential reason is that this method uses a heterologous system and as such the host *Synechocystis* sp. PCC6803 contains a freshwater D2 protein that the phage would need to interact with. One way to see if this is an issue would be to express the marine host *Synechococcus* sp. WH7803 *psbA* in the *psbA*-less DA1:DA2Em:DA3-H6 *Synechocystis* and see if this is capable of photoautotrophic growth. Additionally, many of the cyanophage that have been sequenced have also been found to contain other components of the PET. For example, many phage including S-PM2d, also contain their own copies of the *psbD* gene that code for the D2 protein in the photosystem II complex. This complex exists as a dimer, and it may be that both the phage *psbA* and *psbD* are required in combination to make a functional PSII, as the phage *psbA* cannot do so with the host encoded *psbD* gene. Another gene potentially required and found in S-PM2d is *speD*. This gene is known to catalyse the final step in polyamine synthesis in other organisms; polyamines affect the structure and the rate of oxygen evolution of photosystem II in plants and as such, *speD* could potentially be required for the maintenance of PSII during phage infection (Gao *et al.* 2016).

Whilst the cyanophage and cyanobacterial versions of the *psbA* gene have maintained a high-level protein similarity (~87% identity between the S-PM2d PsbA and the *Synechocystis* sp. PCC6803 PsbA), there are regions that differ and it may be these that are responsible for the non-functional phage PsbA. The first region is the stromal loop between transmembrane helices D and E, the so-called

'PEST' domain due to its abundance of proline (P), glutamate (E), serine (S), and Threonine (T). This domain has previously been implicated as the site of proteolytic cleavage to degrade the PSII complex in plants (Haubetauhl *et al.* 2001). However, in other organisms this region of cleavage is different. In *Synechocystis* sp. PCC6803 the complex is degraded by a different protease, FtsH, that processively degrades the D1 protein from the N-terminus (Nixon *et al.* 2004). As such, it is unclear as to whether these differences in the PEST domain would likely alter the degradation of the phage-encoded D1 protein.

The second region that differs is a loop of variable sequence at the C-terminal end of transmembrane helix E. This region likely interacts with cytochrome b559 and cytochrome c550. The function of cytochrome c550 is still largely unknown. However, cytochrome b559 is thought to protect PSII from photoinhibition by scavenging electrons from reduced P680 to prevent photooxidation of the final antennae chlorophyll (Sharon, Tzahor, Williamson, Shmoish, Man-Aharonovich, Rusch, Yooseph, Zeidner, Golden, Mackey, *et al.* 2007b). Whilst the genetic sequences of the cyanophage and cyanobacterial *psbA* genes have remained well conserved, these regions of variability could lead to small conformational changes that may make interactions with PsbD or other proteins for photosynthesis different than usual. There may be other factors in the regulation and expression of the cyanophage copies of these genes that we do not fully understand yet and the process may require more than just the phage copy of the *psbA* gene to work properly.

The results of the Western blot, combined with the RT PCR suggest than in the intron-less S-PM2d *psbA* mutant the phage gene is expressed and a functional

protein produced. One reason why this may not lead to a strain capable of maintaining photosynthesis is due to the localisation of the protein. Normally in cyanobacteria the PSII complex including the PsbA protein is found in the thylakoid membrane. However, differences in structure such as those mentioned earlier could cause problems in the folding that could prevent translocation. Work in *Synechocystis* sp. PCC6803 has shown that the initial steps in the synthesis of PSII takes place in the plasma membrane (Zak *et al.* 2001). The mature, and fully assembled super complexes of PSII were identified here, and are thought to be translocated to the thylakoid membrane through vesicular movement (Zak *et al.* 2001). It may be that the intron-less *psbA* mutant has formed the PsbA protein but that it has not been translocated, and as such is stuck in the plasma membrane. This would explain why the mutant is unable to support photosynthesis. The SDS-PAGE and Western blot performed here were performed on whole cell extracts and so the location of the protein cannot be determined. If fractionation of the cells was completed in future work it would be possible to see if this was the case.

One important thing to note is that all of the experiments involving the cyanophage *psbA* strains were conducted under standard light conditions. Whilst the *psbA* gene has been found to be transcribed in cyanophages it is only thought to be advantageous during high light conditions. Previous studies have modelled the outcome of a hypothetical version of cyanophage P-SS7 containing photosynthesis genes infecting *Prochlorococcus* MED4. 'Fitness benefits' to carrying these cyanophage *psbA* copies during infection under high light conditions were proposed compared to those infections under low light conditions. In addition, the splicing of the intron in the S-PM2 p*sbA* gene has been

found to be responsive to light intensity. Puxty *et al* (2018) determined that during infection of *Synechoccoccus* sp. WH7803, the ratio of spliced:unspliced *psbA* transcripts increased under high light conditions. The lack of activity of the mutant containing the entire *psbA* region may therefore be due to an inability to splice the gene under the light conditions used here. Whilst oxygen evolution and rETR were measured over a light curve that reached increased irradiances, it would be interesting to see if this is something that changed if this strain was acclimated to high light conditions first. However, this would not explain why the mutant containing the spliced version of the cyanophage *psbA* also showed a lack of photosynthetic activity.

5.4.2 Not the first non-functional *psbA*.

This is not the first time a non-functional photosystem II has been described. Wegener *et al* identified a 'sentinel' *psbA,* so called because it is expressed in Cyanothece sp. 51142 exclusively in the dark, leading to the assembly of a non-functional photosystem II. It is thought that when this D1 protein is incorporated into a photosystem II complex, it ensures that water oxidation cannot occur, thereby allowing processes sensitive to oxygen, such as nitrogen fixation, to occur in the cyanobacterial cell (Wegener *et al.* 2015a). However, it is unlikely that this is similar to the action of this viral *psbA.* The viral S-PM2 *psbA* has been shown to be expressed during infection (Clokie *et al.* 2006; Lindell *et al.* 2005; Puxty *et al.* 2016; Millard *et al.* 2010; Sharon *et al.* 2007b), and it has also been shown that S-PM2 requires light to infect (Jia *et al.* 2010). As such there would be no need to prevent water oxidation, this function would be detrimental to the phage.

The results of the work undertaken here poses some new interesting questions to be answered. Whilst it was possible to replace the host copy of *psbA* with a viral version, this appeared non-functional. The reasons for this are unknown and could be due to incorrect localisation of the protein or incorrect light conditions for the splicing/expression of the viral *psbA* versions. Alternatively, the marine nature of these phage PsbA isoforms could be phylogenetically too far removed from the marine *Synechococcus* host PsbA form it would normally interact with to function correctly in the freshwater strain *Synechocystis*. However, this is unlikely as other divergent *psbA* genes (such as from the higher plant Poa annua) have previously been successfully expressed in *Synechocystis* sp. PCC6803 to form a functional hybrid PSII complex(Nixon *et al.* 1991). It was impossible to answer these new questions in the time frame of this work, but future experiments should address these questions, and significantly improve our understanding of how the S-PM2 *psbA* works.

Chapter 6: Conclusions and Future Directions

In this chapter I discuss the results of my thesis in the context of the original aims set out in Chapter 1. Additionally, possible future directions for research are also considered.

6.1 Development of a random chemical mutagenesis system for use in cyanophages.

The first aim of this work was to develop a random chemical mutagenesis system for use in cyanophages (Chapter 3). Having a genetic system such as this would allow the characterisation of potential AMGs, and facilitate the detection of novel gene functions through phenotypic exploration, as was carried out in Chapter 4.

Hydroxylamine was the mutagen selected due to its apparent specificity in creating C-T transition mutations, at a rate of 1 SNP per genome when a 3-log drop in titre is observed (Ernst Freese *et al.* 1961). Whilst the relationship between a 3-log drop in titre and a single SNP per genome was observed when using coliphage Slur29, this pattern was not consistent when using either of the cyanophages S-PM2dd or S-RSM4 where much higher numbers of SNPs were found in the former, and a much lower number in the latter. Additionally, different mutations other than the predicted C-T transitions were identified in the mutant phages produced. The initial characterisation of the mechanism of hydroxylamine mutagenesis was done before the advent of whole genome sequencing, and as such the rest of the genome other than the gene of interest was never fully investigated for additional SNPs. As such, it may be that other SNPs did exist but were not identified using these older methods. In the only other study that explored the use of hydroxylamine mutagenesis with whole genome sequencing

(Robins et al., 2013) similar results to mine were found, where a higher number of SNPs were identified with additional changes other than just the expected C-T transition mutation. Robins *et al.* 2013 dismiss these other mutations as not being caused by hydroxylamine, and suggest they were only identified because of the high sequencing depth obtained. However, the fact that I obtained similar results even though the sequencing depth here was not as high as in Robins et al., with different numbers and types of SNPs being consistently found suggests this explanation is unlikely.

An alternative explanation for the high numbers of mutations in each sequenced S-PM2dd cyanophage is that these mutations were initially caused by hydroxylamine, and this led to a 'transient hypermutability' state (Drake *et al.* 2005). This state leads to the formation of multiple synchronous mutations and appears to be widespread, being found in a range of organisms (Drake *et al* 2005). These mutations can be caused by a number of mechanisms that result in the incorrect production of proteins, such as errors in transcription/translation, or the production of dominant-inactive proteins amongst other things (Drake *et al.* 2005). It is difficult to know if this is the reason behind the increased number of types of SNP present, due to the high percentage proteins in cyanophage genomes being annotated as hypothetical proteins. The functions of these proteins are therefore unknown, so it is unclear if they are involved in the aforementioned mechanisms. However, this hypothesis would explain the appearance of multiple mutations in the mutant phages created and sequenced in this work.

An interesting observation found in Chapter 3 is that the creation of SNPs is not random, as would be expected, and occurs more frequently in identified 'hypervariable' regions of the S-PM2d genome. The SNPs created by hydroxylamine mutagenesis are also not found in nature as determined here by the use of the TARA viral metagenome dataset. One reason for this could be that the majority of the SNPs created in this mutagenesis are detrimental to the phage, and as such are not propagated. This is somewhat corroborated by the phenotypes found in the phages characterised in Chapter 4, where only one mutant appears to show an increase in burst size, indicative of an increase in fitness. The SNPs created here may not be present in nature as they may have a detrimental effect on the phage, and as such propagation of the SNPs would serve no use to the phage.

6.2 Characterisation of novel S-PM2dd mutants.

A major aim of my thesis was to characterise the phenotypes of the cyanophage mutants obtained by random chemical mutagenesis and to assign these phenotypes to a particular SNP. To do this a small pool of S-PM2dd mutants were characterised phenotypically to identify whether SNPs affected phage fitness (Chapter 4). In the four mutants that were characterised, all of them were found to exhibit some kind of phenotypic difference compared to that of the wild type. A hypothetical protein (S-PM2d148) potentially involved as a structural protein in forming the shape of the capsid head was identified through TEM imaging of mutant d85. The protein product of this gene (S-PM2d148) is widespread in cyanophages, but previously nothing was known about its function. Curiously,

mutant d80 that contains an additional SNP to mutant d85 in the S-PM2d075 gene (causing a nonsynonymous alanine to serine amino acid change) displays a completely different phenotype of a reduced plaque size, and increased burst size but no deformed capsid head. This may be due to the additional SNP acting as a suppressor mutation, and so the difference in capsid morphology is not observed. Another SNP found to be detrimental to phage fitness was found in mutant d39; whilst no morphological differences were apparent there were clear differences in infection profile, as no clear burst was observed. It would therefore make sense that this SNP was not found in nature, as it would be detrimental to the phage and as such is unlikely to be propagated.

One of the SNPs identified was found in the S-PM2 *psbA* gene, an AMG that is a major focus of this thesis (see Chapter 5). This mutant displays a clear phenotype of a delayed burst and reduced burst size. It could be that this amino acid change from Threonine to Isoleucine affects PsbA functioning, so that the production of energy via photosynthesis is perturbed. However, this mutant also contains three more SNPs, and so additional work would need to be performed to link this SNP in *psbA* to the observed phenotype. PSII function can be evaluated through measuring rETR and oxygen evolution as done in Chapter 5, following infection of *Synechococcus* sp. WH7803 with this mutant phage. If the phage *psbA* behaves in the way hypothesised, to maintain host photosynthesis during infection, this SNP may cause a clear difference in rETR and oxygen evolution compared to rates of *Synechococcus* infected with wild type S-PM2dd phage. Modelling of the expression of the cyanophage *psbA* gene is thought to bring a fitness benefit to the phage under high light conditions (Bragg and Chisholm 2008; Hellweger 2009).

As such it would be interesting to repeat the photophysiological experiments described above during infection of the *Synechococcus* host with this mutant phage at different light conditions, to see if this SNP has a larger detrimental effect under high-light conditions where the gene is thought to be more beneficial.

An unexpected but interesting result of Chapter 3 is the discovery of a novel deletion in the S-PM2d wild type phage, leading to the detection of a new wild type S-PM2d phage variant termed 'S-PM2dd' for the purpose of this work. The combination of long and short read sequencing technologies allowed false deletions that were an artefact of Illumina sequencing to be overcome, and SNPs to be called. In this process a real, novel deletion was discovered in the S-PM2d wild type phage. This variant had already been used as the wild type in a number of experiments, and as such continued to be used for the rest of this work. Cyanophage S-PM2 originally lost ~10 kb of its genome to become S-PM2d. The result of this loss was an increase in fitness as determined by an increase in plaque size. The contents of this 10 kb were all unknown ORFs, and as such their function could not be determined (Puxty, Perez-Sepulveda, *et al.* 2015). Here, the S-PM2d phage lost an additional ~1 kb to become S-PM2dd. This loss was in the same area as seen previously, and an additional three ORFs of unknown function were lost. The frequency for S-PM2(d) to lose parts of its genome suggests that these are regions that are involved in adaptation to particular environmental conditions, so are not needed when propagated in the stable and 'ideal' conditions of the lab. It would be interesting to compare this new deletion mutant through plaque size assays or one-step infection assays to both S-PM2 and S-PM2d to see if the loss of more ORFs has again led to an increase in fitness. Another interesting possibility

would be to use metagenomic data, such as the TARA viral metagenomes, to see if the ORFs lost in S-PM2(d) are found in all of the natural isolates, or are restricted to different environments. If the genes that are lost are really involved in specific niche adaptation then it is possible that they are only present in isolates from certain environmental regions. The TARA dataset has viral fractions from a number of distinct geographical regions. By mapping reads to this region of the genome it would be possible to identify whether this "ORFanage" region is present in each environment, or whether variations appear in different places.

6.3 Understanding the role of the viral S-PM2 *psbA* and its associated regulatory elements during infection.

The final two aims of my thesis were to further understand the role of the viral *psbA* and associated regulatory elements (intron, antisense RNA, and homing endonuclease) during infection of its *Synechococcus* host (Chapter 5). To do this, variations of the region as described in Chapter 5.1 were heterologously expressed in a *Synechocystis* sp. PCC6803 strain lacking all three endogenous *psbA* genes. Mutants containing the intronless S-PM2d phage *psbA* and the entire *psbA* region from the start of the *psbA* gene to the end of the homing endonuclease, as well as the *psbA* triple mutant containing just the apramycin resistance cassette were successfully made. However, these mutants were only able to be grown photoheterotrophically, and were not capable of surviving without glucose. Photophysiological measurements suggested that there was no photosynthetic electron transport or oxygen evolution in any of the mutants tested, even though the results of RT-PCR and western blotting suggested that the intron-less cyanophage *psbA* gene was expressed. These results leave a range of further

questions to be addressed in future work, to help understand the function of this intron-less phage *psbA*.

Firstly, there is a possibility that the components of the PSII complex in the mutants created here are not being translocated to the thylakoid membrane as they would normally (Zak *et al.* 2001), and so is not functioning correctly. It may be that the cyanophage PsbA protein is still located in the plasma membrane where it is initially made(Zak *et al.* 2001). This question could be answered by preparing different cellular fractions including cytoplasmic and thylakoid membranes as well as soluble extracts and crude cell lysates. These fractions could then be run on a native gel and the location of the PSII complex identified by western blotting. This would directly tell us if it had been translocated to the right region.

Secondly, if the photosystem is correctly located it may be that the D1 monomer requires other cyanophage encoded photosynthetic components for functionality. Many cyanophages also possess a *psbD* gene that encodes the second major protein of photosystem II comprising the D1/D2 dimer (Lindell *et al.* 2004). It would be interesting to see if functionality was restored if the S-PM2d *psbD* gene was additionally supplied. There are also a number of other genes present in cyanophages, and specifically S-PM2d, that are thought to be involved in photosynthesis, and encode different parts of the photosystems or additional supportive components, such as *cpeT* (Mann *et al.* 2005; Gasper *et al.* 2017) and *speD* (Sullivan *et al.* 2005), although the precise function of some of these genes remains to be proven. By removing *psbA* from its own genetic context we may be

preventing interactions that we do not know about yet, but are important for the correct formation and function of the 'cyanophage' photosystem II.

The mutant containing the entire S-PM2d *psbA* region did not appear functional in this work. One potential reason for this is due to incorrect/absent splicing of the group I intron contained within the cyanophage *psbA* gene. This could be due to the light levels used here. Whilst the intron has previously been shown capable of self-splicing in vitro, when the asRNA is not present (Millard *et al.* 2010), *in vivo* data suggests there may be additional levels of regulation. Thus, splicing has been shown to be responsive to light levels, with high light conditions promoting increased ratios of spliced:unspliced transcripts in S-PM2d (Puxty *et al.* 2018). Therefore, higher light may be needed to initiate the splicing of the intron in the 'entire region' mutant. Future work could be done to investigate this, such as growth curves of the mutant compared to the wild type at different light levels containing minimal glucose (so as to maintain growth if the *psbA* gene is unspliced/non-functional). Alternatively, a similar assay to the RT-qPCR developed in Puxty *et al* (2018) could be used to measure the ratio of spliced:unspliced transcripts in the mutant, under different light conditions. In this way we could determine whether incorrect/absent splicing is a product of the light levels used, or if there are alternative reasons for this.

In conclusion, this thesis greatly contributes towards our understanding of cyanophage biology and especially interactions with their host. A potential method for creating single nucleotide polymorphisms in cyanophages has been developed, though highlighting that hydroxylamine mutagenesis can lead to

mutation types other than the C-T transitions traditionally associated with bacteriophage mutagenesis work. Even so, hydroxylamine mutagenesis led to the creation of several novel S-PM2dd mutants. One of these mutants helped identify a novel structural protein previously identified as 'hypothetical', which is widespread in cyanophages. The function of the vast majority of cyanophage genes are still unknown, and as such a greater understanding of even a few genes provides insight into how these cyanophages work.

One gene in particular has been the focus of this work, namely *psbA*. The current hypothesis is that the phage copy of PsbA acts as a 'life support' during infection, to optimise the infection of the host. For this to be correct, a functional copy of the protein would be required. Whilst this work has been able to insert the cyanophage S-PM2d version of *psbA* into *Synechocystis* sp. PCC6803, it does not appear to produce a functional/active protein/photosystem. Further work is required to distinguish whether this is due to incorrect localisation of the protein, a lack of other photosynthesis components required for PSII function, or if the cyanophage protein itself is in fact non-functional in the host, such as the 'sentinel *psbA*' in *Cyanothece* sp. ATCC 51142 (Wegener *et al.* 2015b), and the current hypothesis incorrect. Whichever is the case this work still highlights vital steps that have been taken towards understanding the role of this gene, paving the way for future work to solve this.

Bibliography

Ackermann, H.-W. (2003). Bacteriophage observations and evolution. *Research in Microbiology* **154**:245–251.

Alcamán, M.E., Alcorta, J., Bergman, B., Vásquez, M., Polz, M. and Díez, B. (2017). Physiological and gene expression responses to nitrogen regimes and temperatures in Mastigocladus sp. strain CHP1, a predominant thermotolerant cyanobacterium of hot springs. *Systematic and Applied Microbiology* **40**:102–113.

Allen, J.F. (2003). Cyclic, pseudocyclic and noncyclic photophosphorylation: New links in the chain. *Trends in Plant Science* **8**:15–19.

Alperovitch-Lavy, A., Sharon, I., Rohwer, F., Aro, E.-M., Glaser, F., Milo, R., Nelson, N., *et al.* (2011). Reconstructing a puzzle: existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. *Environmental Microbiology* **13**:24–32.

Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K.A., Toner, B.M. and Dick, G.J. (2014). Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. *Science* **344**:757–760.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., *et al.* (2006). The marine viromes of four oceanic regions. *PLoS biology* **4**:368.

Avrani, S., Wurtzel, O., Sharon, I., Sorek, R. and Lindell, D. (2011). Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474**:604–8.

Axmann, I.M., Kensche, P., Vogel, J., Kohl, S., Herzel, H. and Hess, W.R. (2005). Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biology* **6**:73.

Baker, N.R. (2008). Chlorophyll fluorescence: A probe of photosynthesis *in vivo*. *Annual Review of Plant Biology* **59**:89–113.

Baltz, R.H., Bingham, P.M. and Drake, J.W. (1976). Heat mutagenesis in bacteriophage T4: the transition pathway. *Proceedings of the National Academy of Sciences of the United States of America* **73**:1269–73. A

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., *et al.* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**:1709–12.

Barten, R. and Lill, H. (1995). DNA-uptake in the naturally competent cyanobacterium, *Synechocystis* sp. PCC 6803. *FEMS Microbiology Letters* **129**:83–88.

Bebenek, A., Dressman, H.K., Carver, G.T., Ng, S., Petrov, V., Yang, G., Konigsberg,

W.H., *et al.* (2001). Interacting fidelity defects in the replicative DNA Polymerase of bacteriophage RB69. *Journal of Biological Chemistry* **276**:10387–10397.

Benzer, S. and Freese, E. (1958). Induction of specific mutations with 5-Bromouracil. *Proceedings of the National Academy of Sciences of the United States of America* **44**:112–9.

Berg, J., Tymoczko, J. and Stryer, L. (2002). The Calvin cycle synthesizes hexoses from carbon dioxide and water. In: *Biochemistry*. 5th ed. New York: Freeman, WH.

Bingham, P.M., Baltz, R.H., Ripley, L.S. and Drake, J.W. (1976). Heat mutagenesis in bacteriophage T4: the transversion pathway. *Proceedings of the National Academy of Sciences of the United States of America* **73**:4159–63.

, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., *et al.* (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**:1453–1462.

Blot, N., Mella-Flores, D., Six, C., Le Corguillé, G., Boutte, C., Peyrat, A., Monnier, A., *et al.* (2011). Light history influences the response of the marine cyanobacterium *Synechococcus* sp. WH7803 to oxidative stress. *Plant Physiology* **156**:1934–54.

Bragg, J.G. and Chisholm, S.W. (2008). Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene *PLoS ONE* **3**:e3550.

Breimer, L.H. (1988). Ionizing radiation-induced mutagenesis. *British Journal of Cancer* **57**:6–18.

Breitbart, M., Bonnain, C., Malki, K. and Sawaya, N.A. (2018). Phage puppet masters of the marine microbial realm. *Nature Microbiology* **3**:754–766.

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

Brotcorne-Lannoye, A. and Maenhaut-Michel, G. (1986). Role of RecA protein in untargeted UV mutagenesis of bacteriophage lambda: evidence for the requirement for the *dinB* gene. *Proceedings of the National Academy of Sciences of the United States of America* **83**:3904–8.

Buettner, V.L., Hill, K.A., Scaringe, W.A. and Sommer, S.S. (2000). Evidence that proximal multiple mutations in Big Blue transgenic mice are dependent events. *Mutation research* **452**:219–29.

Buick, R. (1992). The antiquity of oxygenic photosynthesis: evidence from stromatolites in sulphate-deficient Archaean lakes. *Science* **255**:74–7.

Bull, J.J., Badgett, M.R. and Wichman, H.A. (2000). Big-benefit mutations in a bacteriophage inhibited with heat. *Molecular Biology and Evolution* **17**:942–950.

Campbell, A. (1961). Sensitive mutants of bacteriophage λ. *Virology* **14**:22–32.

Castenholz, R.W. (2017). The biogeography of hot spring algae through enrichment cultures. *SIL Communications, 1953-1996* **21**:296–315.

Chen, J.M., Férec, C. and Cooper, D.N. (2009). Closely spaced multiple mutations as potential signatures of transient hypermutability in human genes. *Human Mutation* **30**:1435–1448.

Chisholm, W., Ahlgren, N.A., Rocap, G. and Chisholm, S.W. (2006). Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environmental Microbiology* **8**:441–454.

Christaki, U., Jacquet, S., Dolan, J.R., Vaulot, D. and Rassoulzadegan, F. (1999). Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. *Limnology and Oceanography* **44**:52–61.

Clarke, A.K., Soitamo, A., Gustafsson, P. and Oquist, G. (1993). Rapid interchange between two distinct forms of cyanobacterial photosystem II reaction-center protein D1 in response to photoinhibition. *Proceedings of the National Academy of Sciences of the United States of America* **90**:9973–7.

Clasen, J., Hanson, C., Ibrahim, Y., Weihe, C., Marston, M. and Martiny, J. (2013). Diversity and temporal dynamics of Southern California coastal marine cyanophage isolates. *Aquatic Microbial Ecology* **69**:17–31.

Clokie, M.R.J., Shan, J., 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**:827–835.

Clokie, M.R.J., Thalassinos, K., Boulanger, P., Slade, S.E., Stoilova-McPhie, S., Cane, M., Scrivens, J.H., *et al.* (2008). A proteomic approach to the identification of the major virion structural proteins of the marine cyanomyovirus S-PM2. *Microbiology* **154**:1775–1782.

Colgin, L.M., Hackmann, A.F.M., Emond, M.J. and Monnat, R.J. (2002). The unexpected landscape of in vivo somatic mutation in a human epithelial cell lineage. *Proceedings of the National Academy of Sciences of the United States of America* **99**:1437–42.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**:819–23.

Cottingham, K.L., Ewing, H.A., Greer, M.L., Carey, C.C. and Weathers, K.C. (2015). Cyanobacteria as biological drivers of lake nitrogen and phosphorus cycling. *Ecosphere* **6**:1.

Crummett, L.T., Puxty, R.J., Weihe, C., Marston, M.F. and Martiny, J.B.H. (2016). The

genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **499**:219–229.

Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W. and Frankenberg-Dinkel, N. (2008). Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Current Biology* **18**:442–448.

Defais, M., Fauquet, P., Radman, M. and Errera, M. (1971). Ultraviolet reactivation and ultraviolet mutagenesis of λ in different genetic systems. *Virology* **43**:495–503.

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**:1286–1299.

Deveau, H., Garneau, J.E. and Moineau, S. (2010). CRISPR/Cas System and its role in phage-bacteria interactions. *Annual Review of Microbiology* **64**:475–493.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research* **41**:4336–4343.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., *et al.* (2008). Functional metagenomic profiling of nine biomes. *Nature* **452**:629–632.

Doermann, A.H.H., Eiserling, F.A.A., Boenher, L. and Boehner, L. (1973). Genetic control of capsid length in bacteriophage T4 I. Isolation and preliminary description of four new mutants. *Journal of virology* **12**:374–385.

Drake, J.W., Bebenek, A., Kissling, G.E. and Peddada, S. (2005). Clusters of mutations from transient hypermutability. *Proceedings of the National Academy of Sciences* **102**:12849–12854.

Dreher, T.W., Brown, N., Bozarth, C.S., Schwartz, A.D., Riscoe, E., Thrash, C., Bennett, S.E., *et al.* (2011). A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. *Environmental Microbiology* **13**:1858–1874. A

Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., Paulsen, I.T., *et al.* (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biology* **9**:R90.

Duhring, U., Axmann, I.M., Hess, W.R. and Wilde, A. (2006). An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proceedings of the National Academy of Sciences* **103**:7054–7058.

Ebina, H., Misawa, N., Kanemura, Y. and Koyanagi, Y. (2013). Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Scientific Reports* **3**:2510.

Enav, H., Béjà, O. and Mandel-Gutfreund, Y. (2012). Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts. *The ISME Journal* **6**:619–628.

Enav, H., Kirzner, S., Lindell, D., Mandel-Gutfreund, Y. and Béjà, O. (2018). Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. *Nature Communications* **9**:4698.

Fang, X., Jiang, Y., Li, K. and Zeng, Q. (2018). F-CphI represents a new homing endonuclease family using the Endo VII catalytic motif. *Mobile DNA* **9**:27.

Farrant, G.K., Doré, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M., Pitt, F.D., *et al.* (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences* **113**:3365–3374.

Fassioli, F., Dinshaw, R., Arpin, P.C. and Scholes, G.D. (2014). Photosynthetic light harvesting: excitons and coherence. *Journal of the Royal Society, Interface* **11**:20130901.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., *et al.* (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**:500–507. A

Fike, D.A., Gammon, C.L., Ziebis, W. and Orphan, V.J. (2008). Micron-scale mapping of sulfur cycling across the oxycline of a cyanobacterial mat: a paired nanoSIMS and CARD-FISH approach. *The ISME Journal* **2**:749–759.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincon, J., Zabala, L.L., Jiao, N., Karl, D.M., *et al.* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences* **110**:9824–9829.

Floor, E. (1970). Interaction of morphogenetic genes of bacteriophage T4. *Journal of Molecular Biology* **47**:293–306.

Frangeul, L., Quillardet, P., Castets, A.-M., Humbert, J.-F., Matthijs, H.C., Cortez, D., Tolonen, A., *et al.* (2008). Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genomics* **9**:274.

Frank, J.A., Lorimer, D., Youle, M., Witte, P., Craig, T., Abendroth, J., Rohwer, F., *et al.* (2013). Structure and function of a cyanophage-encoded peptide deformylase. *The ISME Journal* **7**:1150–1160.

Freese, Ernst, Bautz-Freese, E. and Bautz, E. (1961). Hydroxylamine as a mutagenic and inactivating agent. *Journal of Molecular Biology* **3**:133–143.

Freese, E, Bautz, E. and Freese, E.B. (1961). The chemical and mutagenic specificity of hydroxylamine. *Proceedings of the National Academy of Sciences of the United States of America* **47**:845–55.

Fridman, S., Flores-Uribe, J., Larom, S., Alalouf, O., Liran, O., Yacoby, I., Salama, F., *et al.* (2017). A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nature Microbiology* **2**:1350–1357.

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**:541–8.

Gallet, R., Kannoly, S., Wang, I.-N., d'Hérelle, F., d'Hérelle, F., Yin, J., Yin, J., *et al.* (2011). Effects of bacteriophage traits on plaque formation. *BMC Microbiology* **11**:181.

Gao, E.-B., Huang, Y. and Ning, D. (2016). Metabolic genes within cyanophage genomes: Implications for diversity and evolution. *Genes* **7**:80.

Gasper, R., Schwach, J., Hartmann, J., Holtkamp, A., Wiethaus, J., Riedel, N., Hofmann, E., *et al.* (2017). Distinct features of cyanophage-encoded T-type phycobiliprotein lyase ΦCpeT. *The Journal of Biological Chemistry* **292**:3089–3098.

Genty, B., Briantais, J.-M. and Baker, N.R. (1989). The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochimica et Biophysica Acta (BBA) - General Subjects* **990**:87–92.

Georg, J. and Hess, W.R. (2011). cis-antisense RNA, another level of gene regulation in bacteria. *Microbiology and Molecular Biology Reviews* **75**:286–300.

Georg, J., Voß, B., Scholz, I., Mitschke, J., Wilde, A. and Hess, W.R. (2009). Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Molecular Systems Biology* **5**:305.

Gierga, G., Voss, B. and Hess, W.R. (2012). Non-coding RNAs in marine *Synechococcus* and their regulation under environmentally relevant stress conditions. *The ISME Journal* **6**:1544–1557.

Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**:759–769. A

Greenberg, B.M., Gaba, V., Mattoo, A.K. and Edelman, M. (1987). Identification of a primary in vivo degradation product of the rapidly-turning-over 32 kd protein of photosystem II. *The EMBO journal* **6**:2865–9.

Hall, D. and Tessman, I. (1966). T4 mutants unable to induce deoxycytidylate deaminase activity. *Virology* **29**:339–345.

Hatfull, G.F. and Hendrix, R.W. (2011). Bacteriophages and their genomes. *Current Opinion in Virology* **1**:298–303.

Haubetauhl, K., Andersson, B. and Adamska, I. (2001). A chloroplast DegP2 protease performs the primary cleavage of the photodamaged D1 protein in plant photosystem II. *The EMBO Journal* **20**:713–722.

Hecht, M., Bromberg, Y. and Rost, B. (2015). Better prediction of functional effects for sequence variants. *BMC Genomics* **16**:S1.

Hellweger, F.L. (2009). Carrying photosynthesis genes increases ecological fitness of cyanophage *in silico*. *Environmental Microbiology* **11**:1386–1394.

Hershey, A, D. and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology* **36**:39–56.

Hodyra, K. and Dąbrowska, K. (2015). Molecular and chemical engineering of bacteriophages for potential medical applications. *Archivum Immunologiae et Therapiae Experimentalis* **63**:117–127.

Hu, B., Margolin, W., Molineux, I.J. and Liu, J. (2015). Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences of the United States of America* **112**:4919-28.

Hu, J., Li, T., Xu, W., Zhan, J., Chen, H., He, C. and Wang, Q. (2017). Small antisense RNA RblR positively regulates RuBisCo in *Synechocystis* sp. PCC 6803. *Frontiers in Microbiology* **8**:231.

Huang, L., McCluskey, M.P., Ni, H. and LaRossa, R.A. (2002). Global gene expression profiles of the cyanobacterium *Synechocystis* sp. strain PCC 6803 in response to irradiation with UV-B and white light. *Journal of Bacteriology* **184**:6845–58.

Humphreys, G.O., Willshaw, G.A., Smith, H.R. and Anderson, E.S. (1976). Mutagenesis of plasmid DNA with hydroxylamine: Isolation of mutants of multi-copy plasmids. *Molecular & General Genetics* **145**:101–108.

Hurwitz, B.L., Brum, J.R. and Sullivan, M.B. (2015). Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *The ISME Journal* **9**:472–84.

Hutchins, D.A., Walworth, N.G., Webb, E.A., Saito, M.A., Moran, D., McIlvin, M.R., Gale, J., *et al.* (2015). Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. *Nature Communications* **6**:8155.

Jia, Y., Shan, J., Millard, A., Clokie, M.R.J. and Mann, N.H. (2010). Light-dependent adsorption of photosynthetic cyanophages to *Synechococcus* sp. WH7803. *FEMS Microbiology Letters* **310**:120–126.

Jiang, Wenyan, Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**:233–239.

Jiang, Wenzhi, Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D.P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Research* **41**:188

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**:1737–40.

Jou, W.M., Haegeman, G., Ysebaert, M. and Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* **237**:82–88.

Kada, T., Tutikawa, K. and Sadaie, Y. (1972). In vitro and host-mediated "rec-assay" procedures for screening chemical mutagens; and phloxine, a mutagenic red dye detected. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **16**:165–174.

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**:12343–8.

Kao, S.H. and McClain, W.H. (1980). Baseplate protein of bacteriophage T4 with both structural and lytic functions. *Journal of virology* **34**:95–103.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**:845–858.

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. *The ISME Journal* **7**:1827–1841.

Kiro, R., Shitrit, D. and Qimron, U. (2014). Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biology* **11**:42–4.

Kirschvink, J.L., Gaidos, E.J., Bertani, L.E., Beukes, N.J., Gutzmer, J., Maepa, L.N. and Steinberger, R.E. (2000). Paleoproterozoic snowball earth: extreme climatic and geochemical global change and its biological consequences. *Proceedings of the National Academy of Sciences of the United States of America* **97**:1400–5.

Koenig, A.C. (2002). A study of mutation methods for evolutionary algorithms.

Kot, W., Kilstrup, M., Vogensen, F.K. and Hammer, K. (2016). Clear plaque mutants of Lactococcal phage TP901-1 Semsey, S. (ed.). *PLOS ONE* **11**:e0155233.

Kricker, M.C. and Drake, J.W. (1990). Heat mutagenesis in bacteriophage T4: another walk down the transversion pathway. *Journal of Bacteriology* **172**:3037–9.

Kufryk, G.I., Sachet, M., Schmetterer, G. and Vermaas, W.F.J. (2002). Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: Optimization of efficiency. *FEMS Microbiology Letters* **206**:215–219.

Kulkarni, R.D. and Golden, S.S. (1994). Adaptation to high light intensity in *Synechococcus* sp. strain PCC 7942: regulation of three *psbA* genes and two forms of the D1 protein. *Journal of Bacteriology* **176**:959–65.

Lacap-Bugler, D.C., Lee, K.K., Archer, S., Gillman, L.N., Lau, M.C.Y., Leuzinger, S., Lee, C.K., *et al.* (2017). Global diversity of desert hypolithic cyanobacteria. *Frontiers in Microbiology* **8**:867.

Laemmll, U.K., Mölbert, E., Showe, M. and Kellenberger, E. (1970). Form-determining function of the genes required for the assembly of the head of bacteriophage T4. *Journal of Molecular Biology* **49**:99–113.

van Leeuwen, P.J., Vos, M.H. and van Gorkom, H.J. (1990). Photosynthetic oxygen evolution. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1018**:173–176.

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**:86–89.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F. and Chisholm, S.W. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proceedings of the National Academy of Sciences of the United States of America* **101**:11013–8.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000). Molecular Analysis of Photosystems. In: *Molecular Cell Biology*. 4th ed. New York: W. H. Freeman.

Loessner, M.J., Rees, C.E., Stewart, G.S. and Scherer, S. (1996). Construction of luciferase reporter bacteriophage A511::luxAB for rapid and sensitive detection of viable *Listeria* cells. *Applied and Environmental Microbiology* **62**:1133–40.

Longhurst, A.R. (2007). *Ecological Geography of the Sea*. Academic Press.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., *et al.* (2013). RNA-guided human genome engineering via Cas9. *Science* **339**:823–826.

Mann, N.H., Clokie, M.R.J., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A., *et al.* (2005). The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine *Synechococcus* strains. *Journal of Bacteriology* **187**:3188–200.

Mann, N.H., Cook, A., Bailey, S., Clokie, M., Amanullah, A., Azam, N., Balliet, A., *et al.* (2003). Bacterial photosynthesis genes in a virus. *Nature* **424**:741–742.

Marine, R., Polson, S.W., Ravel, J., Hatfull, G., Russell, D., Sullivan, M., Syed, F., *et al.* (2011). Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Applied and Environmental Microbiology* **77**:8071–9.

Marinelli, L.J., Piuri, M., Swigoňová, Z., Balachandran, A., Oldfield, L.M., van Kessel, J.C. and Hatfull, G.F. (2008). BRED: A simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS ONE* **3**:e3957.

Martel, B. and Moineau, S. (2014). CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. *Nucleic acids research* **42**:9504–13.

Martínez-Pérez, C., Mohr, W., Löscher, C.R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., *et al.* (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology* **1**:16163.

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

Mella-Flores, D., Six, C., Ratin, M., Partensky, F., Boutte, C., Le Corguillé, G., Marie, D., *et al.* (2012). *Prochlorococcus* and *Synechococcus* have evolved different adaptive mechanisms to cope with light and UV stress. *Frontiers in Microbiology* **3**:285.

Meselson, M., and Yuan, R., (1968) DNA restriction enzyme from *E. coli Nature* **23:** 1110-4.

Millard, A.D., Gierga, G., Clokie, M.R.J., Evans, D.J., Hess, W.R. and Scanlan, D.J. (2010). An antisense RNA in a lytic cyanophage links *psbA* to a gene encoding a homing endonuclease. *The ISME Journal* **4**:1121–1135.

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**: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**:2370–87.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. and Rüger, W. (2003). Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews* **67**:86–156.

Mizuno, C.M., Ghai, R. and Rodriguez-Valera, F. (2014). Evidence for metaviromic islands in marine phages. *Frontiers in Microbiology* **5**:27.

Mohamed, A., Eriksson, J., Osiewacz, H.D. and Jansson, C. (1993). Differential expression of the *psbA* genes in the cyanobacterium *Synechocystis* 6803. *MGG Molecular & General Genetics* **238**:161–168.

Mulo, P. (2009). Cyanobacterial *psbA* gene family: optimization of oxygenic photosynthesis - Springer. *Cellular and Molecular Life Sciences* **66**.

Nagarajan, A., Winter, R., Eaton-Rye, J., and Burnap, R. (2011) A synthetic DNA and fusion PCR approach to the ectopic expression of high levels of the D1 protein of photosystem II in *Synechocystis* sp. PCC 6803. *Journal of Photochemistry and Photobiology B: Biology* **104:**212-219.

Nixon, P.J., Barker, M., Boehm, M., de Vries, R. and Komenda, J. (2004). FtsH-mediated repair of the photosystem II complex in response to light stress. *Journal of Experimental Botany* **56**:357–363.

Nixon, P.J., Rögner, M. and Diner, B.A. (1991). Expression of a higher plant *psbA* gene in *Synechocystis* 6803 yields a functional hybrid photosystem II reaction center complex. *The Plant Cell* **3**:383–395.

Nixon, P.J., Trost, J.T. and Diner, B.A. (1992). Role of the carboxy-terminus of polypeptide D1 in the assembly of a functional water-oxidizing manganese cluster in photosystem II of the cyanobacterium *Synechocystis* sp. PCC 6803:assembly requires a free carboxyl group at C-terminal position 344. *Biochemistry* **31**:10859–10871.

Nobrega, F.L., Costa, A.R., Santos, J.F., Siliakus, M.F., van Lent, J.W.M., Kengen, S.W.M., Azeredo, J., *et al.* (2016). Genetically manipulated phages with improved pH resistance for oral administration in veterinary medicine. *Scientific Reports* **6**:39235.

Nowaczyk, M.M., Wulfhorst, H., Ryan, C.M., Souda, P., Zhang, H., Cramer, W.A. and Whitelegge, J.P. (2011). NdhP and NdhQ: two novel small subunits of the cyanobacterial NDH-1 complex. *Biochemistry* **50**:1121–4.

Pagliano, C., Saracco, G. and Barber, J. (2013). Structural, functional and auxiliary proteins of photosystem II. *Photosynthesis Research* **116**:167–188.

Pandolfini, T., Molesini, B. and Spena, A. (2013). AUCSIA: an ancestral green plant miniprotein and the emergence of auxin transport. *Plant Signaling & Behavior* **8**:e22928.

Partensky, F., Blanchot, J. and Vaulot, D. (1999). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Marine Cyanobacteria* **19**:419–434.

Partensky, F, Hess, W.R. and Vaulot, D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews : MMBR* **63**:106–27.

Paulsen, M.L., Doré, H., Garczarek, L., Seuthe, L., Müller, O., Sandaa, R.-A., Bratbak, G., *et al.* (2016). *Synechococcus* in the Atlantic gateway to the Arctic Ocean. *Frontiers in Marine Science* **3**:191.

Payne, J.L., McClain, C.R., Boyer, A.G., Brown, J.H., Finnegan, S., Kowalewski, M., Krause, R.A., *et al.* (2011). The evolutionary consequences of oxygenic photosynthesis: a body size perspective. *Photosynthesis Research* **107**:37–57.

Pessi, I.S., Lara, Y., Durieu, B., Maalouf, P. de C., Verleyen, E. and Wilmotte, A. (2018). Community structure and distribution of benthic cyanobacteria in Antarctic lacustrine microbial mats. *FEMS Microbiology Ecology* **94**.

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

Philosof, A. and Béjà, O. (2013). Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environmental Microbiology Reports* **5**:475–482.

Pinto, F., Pacheco, C.C., Ferreira, D., Moradas-Ferreira, P. and Tamagnini, P. (2012). Selection of suitable reference genes for RT-qPCR analyses in cyanobacteria *PLoS ONE* **7**:e34983.

Pires, D.P., Cleto, S., Sillankorva, S., Azeredo, J. and Lu, T.K. (2016). Genetically engineered phages: a review of advances over the last decade. *Microbiology and Molecular Biology Reviews* **80**:523–43.

Polovina, J.J., Howell, E.A. and Abecassis, M. (2008). Ocean's least productive waters are expanding. *Geophysical Research Letters* **35**:03618.

Poon, A. and Chao, L. (2005). The rate of compensatory mutation in the DNA bacteriophage φX174. *Genetics* [**170**:989–999.

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

Puxty, R.J., Evans, D.J., Millard, A.D. and Scanlan, D.J. (2018). Energy limitation of cyanophage development: implications for marine carbon cycling. *The ISME Journal* **12**:1273–1286.

Puxty, R.J., Millard, A.D., Evans, D.J. and Scanlan, D.J. (2015). Shedding new light on viral photosynthesis. *Photosynthesis Research* **126**:71–97.

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**:1585–

1589.

Puxty, R.J., Perez-Sepulveda, B., Rihtman, B., Evans, D.J., Millard, A.D. and Scanlan, D.J. (2015). Spontaneous deletion of an "ORFanage" region facilitates host adaptation in a "Photosynthetic" cyanophage *PLOS ONE* **10**:e0132642. A

Rang, F.J., Kloosterman, W.P. and de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology* **19**:90.

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**:e2055.

Ritchie, D.A. (1964). Mutagenesis with light and proflavine in phage T4. *Genetical Research* **5**:168.

Robins, W.P., Faruque, S.M. and Mekalanos, J.J. (2013). Coupling mutagenesis and parallel deep sequencing to probe essential residues in a genome or gene. *Proceedings of the National Academy of Sciences* **110**:848–857.

Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pasić, L., Thingstad, T.F., Rohwer, F. and Mira, A. (2009). Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7**:828–36.

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

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos, B.T., *et al.* (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**:689–693.

Roychoudhury, P., Shrestha, N., Wiss, V.R. and Krone, S.M. (2013). Fitness benefits of low infectivity in a spatially structured population of bacteriophages. *Proceedings of the Royal Society B: Biological Sciences* **281**:20132563–20132563.

Sakurai, I., Stazic, D., Eisenhut, M., Vuorio, E., Steglich, C., Hess, W.R. and Aro, E.-M. (2012). Positive regulation of *psbA* gene expression by cis-encoded antisense RNAs in *Synechocystis* sp. PCC 6803. *Plant physiology* **160**:1000–10.

Sandaa, R.-A. and Larsen, A. (2006). Seasonal variations in virus-host populations in Norwegian coastal waters: focusing on the cyanophage community infecting marine *Synechococcus* spp. *Applied and Environmental Microbiology* **72**:4610–8.

Sane, P. V, Ivanov, A.G., Sveshnikov, D., Huner, N.P.A. and Oquist, G. (2002). A transient exchange of the photosystem II reaction center protein D1:1 with D1:2 during low temperature stress of *Synechococcus* sp. PCC 7942 in the light lowers

the redox potential of QB. *The Journal of Biological Chemistry* **277**:32739–45.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., *et al.* (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature* **265**:687–695.

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**:181–191.

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

Schindelin, J., Rueden, C.T., Hiner, M.C. and Eliceiri, K.W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development* **82**:518–529.

Schofield, D.A., Molineux, I.J. and Westwater, C. (2009). Diagnostic bioluminescent phage for detection of Yersinia pestis. *Journal of clinical microbiology* **47**:3887–94.

Schopf, J.W. (1993). Microfossils of the early Archean apex chert: new evidence of the antiquity of life. *Science (New York, N.Y.)* **260**:640–6.

Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R.Y., *et al.* (2009). Photosystem I gene cassettes are present in marine virus genomes. *Nature* **461**:258–62.

Sharon, I., Battchikova, N., Aro, E.M., Giglione, C., Meinnel, T., Glaser, F., Pinter, R.Y., *et al.* (2011). Comparative metagenomics of microbial traits within oceanic viral communities. *The ISME Journal* **5**:1178–1190.

Sharon, I., Tzahor, S., Williamson, S., Shmoish, M., Man-Aharonovich, D., Rusch, D.B., Yooseph, S., Zeidner, G., Golden, S.S., Mackey, S.R., *et al.* (2007). Viral photosynthetic reaction center genes and transcripts in the marine environment. *The ISME Journal* **1**:492–501. A

Shen, J.-R. (2015). The Structure of Photosystem II and the Mechanism of Water Oxidation in Photosynthesis. *Annual Review of Plant Biology* **66**:23–48.

Sicora, C.I., Ho, F.M., Salminen, T., Styring, S. and Aro, E.-M. (2009a). Transcription of a "silent" cyanobacterial *psbA* gene is induced by microaerobic conditions. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1787**:105–112.

Sokolov, A.S., Latypov, O.R., Kolosov, P.M., Shlyapnikov, M.G., Bezlepkina, T.A., Kholod, N.S., Kadyrov, F.A., *et al.* (2018). Phage T4 endonuclease SegD that is similar to group I intron endonucleases does not initiate homing of its own gene. *Virology* **515**:215–222.

Song, K., Li, L. and Zhang, G. (2016). Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology. *Scientific Reports* **6**:35736.

Stal, L.J. (1995). Physiological ecology of cyanobacteria in microbial mats and other communities. *New Phytologist* **131**:1–32.

Stanier, R.Y., Kunisawa, R., Mandel, M. and Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriological Reviews* **35**:171–205.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F. and Chisholm, S.W. (2005). Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations *PLoS Biology* **3**:0790–0806.

Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P.R., DeFrancesco, A.S., *et al.* (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology* **12**:3035–3056.

Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P. and Chisholm, S.W. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts *PLoS Biology* [**4**:1344–1357.

Sullivan, M.B., Waterbury, J.B. and Chisholm, S.W. (2003). Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**:1047–1051.

Summerfield, T.C., Toepel, J. and Sherman, L.A. (2008). Low-oxygen induction of normally cryptic *psbA* genes in cyanobacteria †. *Biochemistry* [**47**:12939–12941.

Sundby, B., Gobeil, C., Silverberg, N. and Alfonso, M. (1992). The phosphorus cycle in coastal marine sediments. *Limnology and Oceanography* **37**:1129–1145.

Suttle, C.A. (2007). Marine viruses--major players in the global ecosystem. *Nature reviews. Microbiology* **5**:801–12.

Suttle, Curtis A (2005). Viruses in the sea. *Nature* **437**:356–61.

Suttle, C.A. and Chan, A.M. (1993). Marine cyanophages infecting oceanic and coastal strains of Synechococcus. *Marine Ecology Progress Series* **92**:99–109.

Tessman, I. (1959). Mutagenesis in phages ϕX174 and T4 and properties of the genetic material. *Virology* **9**:375–385.

Tessman, I. (1968). Mutagenic treatment of double- and single-stranded DNA phages T4 and S13 with hydroxylamine. *Virology* **35**:330–333.

Tessman, I., Poddar, R.K. and Kumar, S. (1964). Identification of the altered bases in mutated single-stranded DNA: I. In vitro mutagenesis by hydroxylamine, ethyl

methanesulfonate and nitrous acid. *Journal of Molecular Biology* **9**:352–363.

Thingstad, T. F., and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol*. **13**: 19–27.

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* [**108**:E757–E764.

Tindall, K.R., Stein, J. and Hutchinson, F. (1988). Changes in DNA base sequence induced by gamma-ray mutagenesis. *Department of Molecular Biophysics and Biochemistry* **118**:551–560.

Trebst, A. (1986). The topology of the plastoquinone and herbicide binding peptides of photosystem II in the thylakoid membrane. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* **41**:240–246.

Tyystjarvi, T., Herranen, M. and Aro, E.-M. (2001). Regulation of translation elongation in cyanobacteria: membrane targeting of the ribosome nascent-chain complexes controls the synthesis of D1 protein. *Molecular Microbiology* **40**:476–484.

Umena, Y., Kawakami, K., Shen, J.R. and Kamiya, N. (2011). Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9Å. *Nature* **473**:55–60.

Venza Colon, C.J., Vasquez Leon, A.Y. and Villafañe, R.J. (2004). Initial interaction of the P22 phage with the *Salmonella typhimurium* surface. *Puerto Rico Health Sciences Journal* **23**:95–101.

Vermaas, W.F.J. (2001). Photosynthesis and respiration in cyanobacteria. *Encyclopedia of Life Sciences*.

Villafane, R. (2009). Construction of phage mutants. Humana Press, pp. 223–237. A

Waterbury, J.B., Watson, S.W., Valois, F.W., and Franks, D.G. (1986) Biological and ecological characterisation of the marine unicellular cyanobacterium *Synechococcus. Can J Fish Aquat Sci* **214:** 71-120.

Wegener, K.M., Nagarajan, A. and Pakrasi, H.B. (2015). An atypical *psbA* gene encodes a sentinel D1 protein to form a physiologically relevant inactive photosystem II complex in cyanobacteria. *The Journal of Biological Chemistry* **290**:3764–74.

Weigele, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., King, J., *et al.* (2007). Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environmental Microbiology* **9**:1675–

1695.

Weitz, J.S. and Wilhelm, S.W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports* **4**:17.

Wilhelm, S.W. and Suttle, C.A. (1999). Viruses and nutrient cycles in the sea. *BioScience* **49**:781.

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**: 3736-3743.

Zak, E., Norling, B., Maitra, R., Huang, F., Andersson, B. and Pakrasi, H.B. (2001). The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes. *Proceedings of the National Academy of Sciences* **98**:13443–13448.

Zeidner, G., Bielawski, J.P., Shmoish, M., Scanlan, D.J., Sabehi, G. and Béjà, O. (2005). Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environmental Microbiology* **7**:1505–1513.

Zeng, Q. and Chisholm, S.W. (2012). Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Current Biology : CB* **22**:124–8.

Zhang, Q., Ou, T., Ou, T., Gao, X., Li, S. and Zhang, Q. (2015). Genome analysis and gene nblA identification of Microcystis aeruginosa myovirus (MaMV-DC) reveal the evidence for horizontal gene transfer events between cyanomyovirus and host. *Journal of General Virology* **96**:3681–97.

Zheng, Q., Jiao, N., Zhang, R., Chen, F. and Suttle, C.A. (2013). Prevalence of *psbA*-containing cyanobacterial podoviruses in the ocean. *Scientific Reports* **3**:3207.

Zinser, E.R., Johnson, Z.I., Coe, A., Karaca, E., Veneziano, D. and Chisholm, S.W. (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnology and Oceanography* **52**:2205–2220.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., Not, F., *et al.* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology* **10**:147–161.

Appendices

Appendix 1: Perl script used to identify regions of no coverage in S-PM2dd mutants.

```
use Bio::DB::Sam;
use Statistics::Basic qw(:all);
#tamsin version for quick checking
my $usage = "calc_bam_coverage_tamsin.pl BAMFILE \n OUTFILE WILL BE FOR
OUTPUT , BAM FILE SHOULD BE SORTED BAM FILE  \n " ;

open (OUT,">$ARGV[1]") or die $usage;

#general script to print coverage to file
#reads in a sorted bam file to get the bam file as an object. Call this a
object SAM just to be confusing appraently
 # high level API
my $sam = Bio::DB::Sam->new(-bam  =>"$ARGV[0]",
                            -autoindex =>1,
                            );


#get the sequence ID of all sequences
my @targets    = $sam->seq_ids;
my $contig_num =  @targets;


foreach my $contig (@targets) {
 push my @input, $contig;

my $c;

my @cov = &coverage(@input);

foreach $_ (@cov)  {


$c++;

#print "$c\t$_\n" ;


if ($_ < 2 ) { print OUT "$c\t$_ \n "} }


}




######################
```

```
sub coverage {

my $contig = @_[0];

my ($coverage) = $sam->features(-type=>'coverage',-seq_id=>"$contig");

my @data       = $coverage->coverage;



return @data;

};
```

Appendix 2: List of accession numbers of the 91 viral metagenomes used in this work, originally from the TARA oceans cruise metagenomes.

ERR599345: ERR599355: ERR599340: ERR599363: ERR599338: ERR599337: ERR599367: ERR599354: ERR599343: ERR2752151: ERR2752161: ERR599369: ERR599357: ERR599349: ERR599346: ERR2750829: ERR599359: ERR2750827: ERR2752159: ERR2752158: ERR2752157: ERR2752156: ERR2752150: ERR599374: ERR2752148: ERR2752147: ERR2752155: ERR599339: ERR599362: ERR2752154: ERR599376: ERR599358: ERR599344: ERR599356: ERR2752153: ERR599379: ERR2752152: ER599350: ERR599383: ERR594357: ERR594412: ERR2752144: ERR2752143: ERR594411: ERR594355: ERR594354: ERR594388: ERR594379: ERR594364: ERR2752145: ERR594353: ERR594415: ERR594391: ERR594404: ERR594389: ERR594362: ERR594382: ERR594359: ERR594385: ERR594392: ERR594394: ERR594376: ERR594413: ERR594403: ERR594366: ERR594384: ERR594363: ERR594365: ERR594399: ERR594400: ERR594402: ERR594369: ERR594390: ERR594368: ERR594360: ERR594393: ERR594410: ERR594375: ERR594396: ERR594406: ERR594352: ERR594358: ERR594405: ERR594408

Appendix 3: R script used to define regions with coverage of less than 20% of the mean, to define hypervariable regions.

```
x2 <- T2$Coverage
y2 <- T2$Position

hvrcov2 <- vector("numeric", length(T2))
for (i in 1:nrow(T2)) {
  if(x2[i] <= 239){
   hvrcov2[[i]] <- print(y2[i])}
  else {
    hvrcov2[[i]] <- print(0)}
  }

str(hvrcov2)
str(unlist(hvrcov2))
s2 <- split(hvrcov2, cumsum(c(TRUE, diff(hvrcov2) != 1)))
```

Appendix 4: PCR products showing sequences of S-PM2d mutants verified. Both the wild type and mutant sequences are shown, with changes in bases highlighted.

Mutant d39 SNP sequences.

115714 C-T

WT

```
TGGATCTGGTGGTGATGATGTAGTATACTCATTTGATCCTCCACCCAGAATGCTATTAGAGATTTCTTT
AACGGATTCCAAGTTATAACTCTGGGGAACAGCATTATCTTGCGATCCACCAGCATAGAATCCGTAAAT
AACTTCACCGCCTATAAAATCAACATCAGATCTTCCATCAAAGTTTGAATATTCTGCTAGTGATGTGCC
GCCAACAGGCAACCATTCCGTAGAAATTGTTGGAATTGGATTGAGAACTAGTTCCACAAAAAACTGACC
TTCCGATAAAACGTCCATACTACGGAATACCAACTGCATTCTGTTGATGAGTTCCCTGGATCCAAATGA
TCCAATAATGCCATTATCAACAGATGGAGATATTCTAAGAGCAAGAACAGAACGAGTTCTATTACCACC
AGTCCCAATAGTTCTTCTTAACTTAGTACCAACAGTATAAACATATGCTCGGTCATCATCATAAAGACC
ATCCATAATAACAGACGAACCCCAGTGTGAAATCT
```

MT

```
TGGATCTGGTGGTGATGATGTAGTATACTCATTTGATCCTCCACCCAGAATGCTATTAGAGATTTCTTT
AACGGATTCCAAGTTATAACTCTGGGGAACAGCATTATCTTGCGATCCATCAGCATAGAATCCGTAAAT
AACTTCACCGCCTATAAAATCAACATCAGATCTTCCATCAAAGTTTGAATATTCTGCTAGTGATGTGCC
GCCAACAGGCAACCATTCCGTAGAAATTGTTGGAATTGGATTGAGAACTAGTTCCACAAAAAACTGACC
TTCCGATAAAACGTCCATACTACGGAATACCAACTGCATTCTGTTGATGAGTTCCCTGGATCCAAATGA
TCCAATAATGCCATTATCAACAGATGGAGATATTCTAAGAGCAAGAACAGAACGAGTTCTATTACCACC
AGTCCCAATAGTTCTTCTTAACTTAGTACCAACAGTATAAACATATGCTCGGTCATCATCATAAAGACC
ATCCATAATAACAGACGAACCCCAGTGTGAAATCT
```

Mutant d80 SNP sequences.

15951 G-T

WT

```
GCCGATTTAATGCAGGATGTAGCATTAGATTACTACTCAGGCACACAAAAAGCCCCTGAGATTAGTCTA
GAGACTGATTTCAGTGGTATTGTACAAAAGGAGACAATTTCTAATGGACAGCGAACTGATTCAACGATT
GAATTTCTTGGAGACTAAACTCAAAGAGTTAGAAGAACCAGGACCTTTAATGTATCGTCGTCCAGGATC
AGAAGAACACGAAAATTTAGTAGATTTTCTTAATGACACGTATATACAATTACAAGAAACAAGAGCATG
```

TCTAAACCAATTGCTGATATTACATCAACAGACTCAGGTGGATCGGGCTCATGTAATACACCGCCTCTA
GTGCTAGGTAATGCGTCAGGAACGCCTGTTAAAGTATATGTGGGTAAGCAGATGGTCGTAAAGGACCAG
GACGTGTTTAGCGAGGCGCCTGGAACGACCCCCAAAGGCGATCCATGTGCCAGTCAAAGGATTTTACGT
GCTACGAGTCGTGTGAAGG

MT
GCCGATTTAATGCAGGATGTAGCATTAGATTACTACTCAGGCACACAAAAATCCCCTGAGATTAGTCTA
GAGACTGATTTCAGTGGTATTGTACAAAAGGAGACAATTTCTAATGGACAGCGAACTGATTCAACGATT
GAATTTCTTGGAGACTAAACTCAAAGAGTTAGAAGAACCAGGACCTTTAATGTATCGTCGTCCAGGATC
AGAAGAACACGAAAATTTAGTAGATTTTCTTAATGACACGTATATACAATTACAAGAAACAAGAGCATG
TCTAAACCAATTGCTGATATTACATCAACAGACTCAGGTGGATCGGGCTCATGTAATACACCGCCTCTA
GTGCTAGGTAATGCGTCAGGAACGCCTGTTAAAGTATATGTGGGTAAGCAGATGGTCGTAAAGGACCAG
GACGTGTTTAGCGAGGCGCCTGGAACGACCCCCAAAGGCGATCCATGTGCCAGTCAAAGGATTTTACGT
GCTACGAGTCGTGTGAAGG

100793 T-A
WT
TTCGGTGATGGATCAAAACAGAATATAGATAATTTCTCTGCTAAAAGAATTGCTGCTTGTTATGCACAA
CTTGATGATGAAAAGCAAAATCAGTTTCGTTATATGTTGAATAGAGATGCTGCTACATTCCAGTCAGCT
CTGGACTTCGCAGTTCGCAACGTTTGATACGGAGGATGACATGCCATTCGGGTTTGGAAAAGAAATCGC
GGTCCTAGAAGGAAAGTTCCAGATTTATGAAGATCTCTCTAAAGAGATGCTTGACAAACTTGAGAGGGC
AGTTGAAAAAATAAGTGAGGGTAACCAGAAGGTTGCCCTTATCCTTGCTAAGCATGAAGAAAAGATAGA
ACAAAGTATTCGTGCCGATCAACTTATTGTCAAAATGGTTGATGAAATGAGAGAAGCAAACAGCAAAGA
ACATGCTGCTGTTATTGATAGGATCGGAAGTGTTGAAAGTAAAATCAGCGATCTATCAAAGTTCAGATG
GATCACCGCTGGCGTCGCAGCAGCAATTGTGTTGGTCATCGGATCTGCAGAATTCTTTGGTGGCGTCTT
GACAATGGGCAGTA

MT
TTCGGTGATGGATCAAAACAGAATATAGATAATTTCTCTGCTAAAAGAATTGCTGCTTGTTATGCACAA
CTTGATGATGAAAAGCAAAATCAGTTTCGTTATATGTTGAATAGAGATGCTGCTACATTCCAGTCAGCT
CTGGACTTCGCAGTTCGCAACGTTTGATACGGAGGATGACATGCCATTCGGGTTTGGAAAAGAAATCGC
GGTCCTAGAAGGAAAGTTCCAGATTTATGAAGATCTCTCTAAAGAGATGCTTGACAAACTTGAGAGGGC
AGTTGAAAAAATAAGTGAGGGTAACCAGAAGGTTGCCCTTATCCTTGCTAAGCATGAAGAAAAGATAGA
ACAAAGTATTCGTGCCGATCAACATATTGTCAAAATGGTTGATGAAATGAGAGAAGCAAACAGCAAAGA
ACATGCTGCTGTTATTGATAGGATCGGAAGTGTTGAAAGTAAAATCAGCGATCTATCAAAGTTCAGATG
GATCACCGCTGGCGTCGCAGCAGCAATTGTGTTGGTCATCGGATCTGCAGAATTCTTTGGTGGCGTCTT
GACAATGGGCAGTA

Mutant d85 SNP sequences.

100793 T-A
WT
TTCGGTGATGGATCAAAACAGAATATAGATAATTTCTCTGCTAAAAGAATTGCTGCTTGTTATGCACAA
CTTGATGATGAAAAGCAAAATCAGTTTCGTTATATGTTGAATAGAGATGCTGCTACATTCCAGTCAGCT
CTGGACTTCGCAGTTCGCAACGTTTGATACGGAGGATGACATGCCATTCGGGTTTGGAAAAGAAATCGC
GGTCCTAGAAGGAAAGTTCCAGATTTATGAAGATCTCTCTAAAGAGATGCTTGACAAACTTGAGAGGGC
AGTTGAAAAAATAAGTGAGGGTAACCAGAAGGTTGCCCTTATCCTTGCTAAGCATGAAGAAAAGATAGA
ACAAAGTATTCGTGCCGATCAACTTATTGTCAAAATGGTTGATGAAATGAGAGAAGCAAACAGCAAAGA
ACATGCTGCTGTTATTGATAGGATCGGAAGTGTTGAAAGTAAAATCAGCGATCTATCAAAGTTCAGATG
GATCACCGCTGGCGTCGCAGCAGCAATTGTGTTGGTCATCGGATCTGCAGAATTCTTTGGTGGCGTCTT
GACAATGGGCAGTA

MT
TTCGGTGATGGATCAAAACAGAATATAGATAATTTCTCTGCTAAAAGAATTGCTGCTTGTTATGCACAA
CTTGATGATGAAAAGCAAAATCAGTTTCGTTATATGTTGAATAGAGATGCTGCTACATTCCAGTCAGCT
CTGGACTTCGCAGTTCGCAACGTTTGATACGGAGGATGACATGCCATTCGGGTTTGGAAAAGAAATCGC
GGTCCTAGAAGGAAAGTTCCAGATTTATGAAGATCTCTCTAAAGAGATGCTTGACAAACTTGAGAGGGC
AGTTGAAAAAATAAGTGAGGGTAACCAGAAGGTTGCCCTTATCCTTGCTAAGCATGAAGAAAAGATAGA

ACAAAGTATTCGTGCCGATCAACATATTGTCAAAATGGTTGATGAAATGAGAGAAGCAAACAGCAAAGA
ACATGCTGCTGTTATTGATAGGATCGGAAGTGTTGAAAGTAAAATCAGCGATCTATCAAAGTTCAGATG
GATCACCGCTGGCGTCGCAGCAGCAATTGTGTTGGTCATCGGATCTGCAGAATTCTTTGGTGGCGTCTT
GACAATGGGCAGTA

psbASNP Mutant.

SNPs at 126728 and 100717 were confirmed with PCR and sequencing results, SNPs at 12031 and 49638 were confirmed with sequencing results.

100717 G-A

WT

TTCGGTGATGGATCAAAACAGAATATAGATAATTTCTCTGCTAAAAGAATTGCTGCTTGTTATGCACAA
CTTGATGATGAAAAGCAAAATCAGTTTCGTTATATGTTGAATAGAGATGCTGCTACATTCCAGTCAGCT
CTGGACTTCGCAGTTCGCAACGTTTGATACGGAGGATGACATGCCATTCGGGTTTGGAAAAGAAATCGC
GGTCCTAGAAGGAAAGTTCCAGATTTATGAAGATCTCTCTAAAGAGATGCTTGACAAACTTGAGAGGGC
AGTTGAAAAAATAAGTGAGGGTAACCAGAAGGTTGCCCTTATCCTTGCTAAGCATGAAGAAAAGATAGA
ACAAAGTATTCGTGCCGATCAACTTATTGTCAAAATGGTTGATGAAATGAGAGAAGCAAACAGCAAAGA
ACATGCTGCTGTTATTGATAGGATCGGAAGTGTTGAAAGTAAAATCAGCGATCTATCAAAGTTCAGATG
GATCACCGCTGGCGTCGCAGCAGCAATTGTGTTGGTCATCGGATCTGCAGAATTCTTTGGTGGCGTCTT
GACAATGGGCAGTA

MT

TTCGGTGATGGATCAAAACAGAATATAGATAATTTCTCTGCTAAAAGAATTGCTGCTTGTTATGCACAA
CTTGATGATGAAAAGCAAAATCAGTTTCGTTATATGTTGAATAGAGATGCTGCTACATTCCAGTCAGCT
CTGGACTTCGCAGTTCGCAACGTTTGATACGGAGGATGACATGCCATTCGGGTTTGGAAAAGAAATCGC
GGTCCTAGAAGGAAAGTTCCAGATTTATGAAGATCTCTCTAAAGAGATGCTTGACAAACTTGAGAGGGC
AGTTGAAAAAATAAGTAAGGGTAACCAGAAGGTTGCCCTTATCCTTGCTAAGCATGAAGAAAAGATAGA
ACAAAGTATTCGTGCCGATCAACTTATTGTCAAAATGGTTGATGAAATGAGAGAAGCAAACAGCAAAGA
ACATGCTGCTGTTATTGATAGGATCGGAAGTGTTGAAAGTAAAATCAGCGATCTATCAAAGTTCAGATG
GATCACCGCTGGCGTCGCAGCAGCAATTGTGTTGGTCATCGGATCTGCAGAATTCTTTGGTGGCGTCTT
GACAATGGGCAGTA

126728 C-T

WT

TACATGCTTGTTTTCCAGGCAGAACACAACATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGG
GTCTTCGGTGGTTCTCTGTTCTCTGCTATGCATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACG
ACTGAGGTAGAATCCCAGAACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCT
GCTCATGGTTACTTCGGTCGTCTGATCTTCCAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTT
TTCCTTGCTGCTTGGCCTGTTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAAC
CTCAATGGATTCAACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGAT
GTGCTGAATCGTGCTGGTCTGGGTATGGAGGTGATGCACGAGCGTCAAGTTGTGCTTTGCGCTCTTTAA
ATCGGATGAATTGCTGGAAACCCCAAGTGGGCAATCAGCAGCCAAGTCTCAGATACATCTGAGAAAGGT
TCAGAGACTACCTGAGGGATATAGTTCCCTTAATAACAGGAATAAGCGTCCGACACCAGAAATGGTGAT
GATATAGTCCAATCCTGGTAGTAATACCAGATAGTTAAGGGAAGTTTAAGAATGCACACAACTTCCCTC
TTGACCTTGCAGCAGCAGAAGCAACACCAGTTGCCTTGACTGCACCCGCAATCGGTTGATATAATGTTA
AGGAACTCTTCGGAGTTCCTTTTTTTATAAATATTTAAGCACGAAAGAAAACACGAATGACTAAACTAT
ACTCTGATTTGTATAGAACTTGTATGACTTGTGGGGAGGAA

MT

TACATGCTTGTTTTCCAGGCAGAACACAACATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGG
GTCTTCGGTGGTTCTCTGTTCTCTGCTATGCATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACG
ACTGAGGTAGAATCCCAGAACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCT
GCTCATGGTTACTTCGGTCGTCTGATCTTCCAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTT
TTCCTTGCTGCTTGGCCTGTTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAAC

CTCAATGGATTCAACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACATCTGGGCAGAT
GTGCTGAATCGTGCTGGTCTGGGTATGGAGGTGATGCACGAGCGTCAAGTTGTGCTTTGCGCTCTTTAA
ATCGGATGAATTGCTGGAAACCCCAAGTGGGCAATCAGCAGCCAAGTCTCAGATACATCTGAGAAAGGT
TCAGAGACTACCTGAGGGATATAGTTCCCTTAATAACAGGAATAAGCGTCCGACACCAGAAATGGTGAT
GATATAGTCCAATCCTGGTAGTAATACCAGATAGTTAAGGGAAGTTTAAGAATGCACACAACTTCCCTC
TTGACCTTGCAGCAGCAGAAGCAACACCAGTTGCCTTGACTGCACCCGCAATCGGTTGATATAATGTTA
AGGAACTCTTCGGAGTTCCTTTTTTTATAAATATTTAAGCACGAAAGAAAACACGAATGACTAAACTAT
ACTCTGATTTGTATAGAACTTGTATGACTTGTGGGGAGGAA

Appendix 5: Table showing the 336 variable positions of the S-PM2d genome found in nature, when mapped at 90% identity.

Position	Base in Reference	Variants Identified
2389	A	T
2391	C	Y
2399	T	W
2402	T	W
2410	G	C
2417	C	S
2426	A	R
2474	C	T
2495	C	T
5453	A	G
31221	T	C
31260	T	C
31269	C	A
61703	G	R
61709	A	R
61715	T	G
61727	A	T
61728	G	C
61733	T	A
61748	T	C
61751	T	Y
61775	T	W
61790	G	R
61817	A	G
61827	A	C
61832	T	C
61838	T	Y
61841	G	R
61844	C	Y
61856	A	R
61865	C	Y

61868	A	W
61871	T	K
61872	T	K
61883	A	M
61915	G	R
61919	G	R
61919	G	S
61946	G	C
67034	T	Y
67052	A	W
67055	A	R
67060	T	Y
68093	C	A
68096	C	T
68549	T	Y
68555	C	M
68567	A	W
68573	A	R
68591	T	W
70327	A	W
70338	G	R
70339	C	Y
70354	T	W
70354	T	K
70372	A	M
70372	A	W
70375	C	Y
70891	A	G
71152	T	C
71188	C	T
71197	A	W
71203	T	C
71212	A	M
71212	A	R
71227	T	Y
71230	T	W
71239	G	R
71248	T	Y
71251	G	S
71251	G	K
71262	A	R
71265	C	M
71275	C	Y

71281	T	Y
71293	T	Y
71296	T	Y
71308	A	W
71311	G	S
73660	T	C
73666	C	T
73700	T	Y
73714	A	G
73741	T	A
73750	T	Y
73753	A	G
73756	T	Y
75157	C	Y
75166	T	A
75172	C	T
75175	C	M
77370	T	C
77422	T	C
79103	T	G
81489	A	C
81495	G	A
81501	C	T
81532	C	G
81549	A	M
82526	C	M
82535	C	Y
82542	G	A
89333	G	R
89333	G	K
89345	T	W
89348	C	Y
89525	C	T
89747	T	G
89756	T	A
91939	T	Y
91951	A	C
91986	G	T
92567	C	A
92828	C	Y
92867	A	G
92885	A	T
92891	T	Y

92906	C	S
92906	C	Y
92909	C	Y
92918	G	R
92921	T	Y
92930	T	Y
92931	C	M
92933	T	K
92940	A	R
92954	A	W
92966	G	R
92976	C	M
92976	C	S
92981	T	Y
93015	A	G
93026	T	Y
98178	T	G
101729	C	A
102613	A	R
102620	A	R
102622	T	Y
102631	A	M
102633	A	R
102640	G	A
102670	G	R
102673	T	K
102674	A	M
102680	A	C
102697	G	T
102710	A	G
102717	G	K
102718	C	Y
102736	T	Y
102744	C	S
102748	C	Y
102766	G	R
102775	A	W
102779	C	M
102781	C	S
102781	C	Y
102787	A	M
102806	C	M
102818	G	A

102826	C	T
102833	C	M
102841	A	W
102847	T	K
102848	A	C
102859	T	Y
102863	A	G
102865	T	W
102865	T	C
102868	A	R
102868	A	W
102889	A	R
102901	C	Y
102922	T	W
102934	A	R
102949	G	R
102949	G	S
102967	C	M
102968	G	S
102973	C	Y
102976	C	M
102982	A	W
103003	T	A
103027	A	M
103105	T	C
103108	T	C
103168	C	T
103178	A	T
103204	A	W
103220	T	C
103262	G	A
103876	T	W
103876	T	G
103879	A	G
103885	C	Y
103889	G	R
103903	G	R
103924	A	T
103933	T	C
103945	G	R
103960	G	R
103969	G	R
104035	C	Y

104037	C	S
104059	A	T
104074	T	C
104083	G	R
104089	C	T
104116	A	R
104117	A	M
104119	C	Y
104126	A	R
104128	T	Y
104135	G	R
104603	C	T
104622	G	R
104639	A	M
104639	A	W
104654	G	T
105784	T	C
109954	G	A
110694	A	M
110694	A	W
110746	T	W
110753	G	A
110760	A	W
110764	C	Y
113471	T	C
113513	G	T
119877	T	C
119902	G	A
119907	C	T
119937	C	T
119972	A	C
120568	G	A
120569	G	T
120591	A	G
120606	G	A
120630	G	A
120645	G	A
126081	T	Y
126087	T	Y
126099	A	G
126360	C	A
126396	C	M
126405	T	W

126474	A	G
126498	C	Y
126516	T	C
126525	T	Y
126531	T	W
126534	T	W
126537	T	Y
126540	T	A
126546	C	Y
126564	A	R
126570	G	A
126570	G	K
126573	G	S
126585	T	Y
126591	G	R
126594	A	R
126747	T	C
126756	T	W
126762	T	W
126771	G	R
127402	A	G
127475	G	A
127484	G	A
127490	C	T
127497	A	G
127511	A	C
127517	T	C
128307	A	C
128323	C	Y
128330	T	Y
128342	A	C
128353	T	Y
128356	C	Y
128362	A	C
128371	T	Y
128381	C	Y
128384	C	Y
128386	A	W
128405	T	W
128405	T	Y
128410	C	M
128411	C	A
128416	T	Y

128434	A	R
128437	G	S
128437	G	K
128446	A	R
128455	C	Y
128479	T	Y
128485	A	R
128488	C	Y
128500	C	Y
128503	A	R
128506	A	R
128512	T	Y
128512	T	G
128515	G	A
128518	T	K
128527	T	Y
128549	C	Y
128572	A	G
128575	A	M
128579	A	T
128591	T	K
128592	T	K
128612	A	R
128613	C	M
128632	G	R
128632	G	K
129025	T	W
129033	A	W
129043	A	R
129067	T	Y
129082	C	Y
129205	T	W
129205	T	Y
129235	T	Y
129268	T	A
129271	T	A
129283	T	W
130896	C	T
130905	T	C
130926	A	G
130950	T	A
130968	T	C
132786	T	W

154050	T	Y
154987	G	A
158831	T	Y
158864	C	M

Appendix 6: S-RSM4 spot assay to validate the 3-log drop in titre. Between hours 1 (the top three rows on the plate on the left) and hour 6 (the bottom three rows on the plate on the left) there is a 3-log drop in the number of plaques of the cyanophage S-RSM4.

Appendix 7: Summary of statistical results of ANOVA on capsid diameter of mutant S-PM2dd phages.

Source	*Sum Sq*	*df*	*Mean Sq*	*F value*	*Sig.*
Between	2236.5	4	559.14	30.726	<2.2e-16
Within	4058.1	223	18.20		
Total	6294.6	227			

Appendix 8: Summary of statistical results of ANOVA on the plaque sizes of S-PM2dd mutant phages.

Source	*Sum Sq*	*df*	*Mean Sq*	*F value*	*Sig.*
Between	61.2	4	15.30	12.73	3.16E-09
Within	234.3	195	1.20		
Total	295.5	199			

Appendix 9: Sequence of the Invitrogen synthesised construct in Figures 5.1-5.5 in Chapter 5.

pMK-RQ psbAI AmpR

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCG
CTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCT
TCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT
TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATT
GAAGGAAGGCCGTCAAGGCCGCATTCATCGTTACTGACAAGGAGATTGAGGACCGGGGATGGGGTCATA
GGACACCAAAGAATCGAGATTGTGCCTGCCAGCCTTAACGATATCAGGCTATGTCCCGCTTAAACTCTG
ATTCTTACCAGGTAACTCCTGGGCTCCATGGGGCCACAACCAGGCAGTATTTTGTTCCTTTGGCCAATG
GGGCGATCGGGGAAAAATGGCTTGATCTGGCATTTACGAGAAAAATTTTTATTTTTTAATGATTTATTT
TTTCCTATTAAAATCTTTTTTTTACCTTTGGAAACCAACTGCAATCTGAGAAACCATCTTGTTTTTTTA
AAGAAATATTATTAATCTGAAATTCAAGGGAAGTTAATCAATGCCAATAATTATCTCGCATTATTAATC
CCCCTTTATCTATCTGGTTGAGTTGGATTTAGCTGATAGTTTATCACCAAAATAACAAGCAAAATCAAA
TCCAAGCTTAAACCCAAAATCTTACTTCGTAATTATTCGCTTGCAATACGAATGGCGAAAAGCCGAGCT
CATCGGTCAGCTTCTCAACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAG
CGTCCGGCCCCTCGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGG
GACGCTCGTCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTAC
ACCGGACCTTGGAGTTGTCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATT
TGCCTTTGCGGCAGCGGGGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTC
GCCTGCAAGCCCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGA
TGCCAACACGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCAC
CATTCTTCAGGATGGCAAGTTGGTACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGC
CTTGGCGGACAGGTGGCTCAAGGAGAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCG
GTTGATCCGCTCCCGCGACATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCT
GCATCCGCCAGAGGCGGGATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATGAAAT
GTCTAACAATGCTTCAAGCCGACGCCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTGCCATTGCCATAA
CTGCTTTCGGTTAGACTTCGTTTCATTTGGTTAATCAAGGGCACTCTCGCAATGGGGTGCCTTTTATGG
TCCAAGGTTAAAGTTAAGCCAGTAAAGTTTAAGTCTATTTCTAGGGTGAAATGTAATGAATCAATTTAG
GGACTGGGGTTTTTCCACTGATTGGTGGCAAGGCAAAAAAGGGGAATATTGGGTACTTGGGCAAACAAT
TCTCTCCGTTGGATTTGTACTGTTACCGGTTTATACTCCTACTAATTTACAGTTACTAGCTGAACAAAA
TCAGTGGATTGCCTGGGCTGGTACTCTGTTTTTTGGCATGATTGCGGCAGTTTTGCTGATTGGGGGGGG
TCTGCATTTAGGGGAAAATCTGACTCCCTTGCCCCATCCTAAAAAGGATAGTCAATTGGTCACCACTGG
TATTTACAGCATGGTCCGTCATCCCCTCTACGGCGGCGTGGTTTTTCTGGCGATCGCCTATGGCTTTTG
GCAACTGGGCCTCATGGGCCTTCCTTTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGC
ATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCTCACTGACTCGCTGC
GCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAA
AAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT
AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG

TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG
TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTT
GATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATT
ATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATA
TGAGTAAACTTGGTCTGACAGTTATTAGAAAAATTCATCCAGCAGACGATAAAACGCAATACGCTGGCT
ATCCGGTGCCGCAATGCCATACAGCACCAGAAAACGATCCGCCCATTCGCCGCCCAGTTCTTCCGCAAT
ATCACGGGTGGCCAGCGCAATATCCTGATAACGATCCGCCACGCCCAGACGGCCGCAATCAATAAAGCC
GCTAAAACGGCCATTTTCCACCATAATGTTCGGCAGGCACGCATCACCATGGGTCACCACCAGATCTTC
GCCATCCGGCATGCTCGCTTTCAGACGCGCAAACAGCTCTGCCGGTGCCAGGCCCTGATGTTCTTCATC
CAGATCATCCTGATCCACCAGGCCCGCTTCCATACGGGTACGCGCACGTTCAATACGATGTTTCGCCTG
ATGATCAAACGGACAGGTCGCCGGGTCCAGGGTATGCAGACGACGCATGGCATCCGCCATAATGCTCAC
TTTTTCTGCCGGCGCCAGATGGCTAGACAGCAGATCCTGACCCGGCACTTCGCCCAGCAGCAGCCAATC
ACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACACGGAACACCGGTGGTGGCCAGCCAGCTCAG
ACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCTCAGATCGGTTTTCACAAACAGCACCGGACG
ACCCTGCGCGCTCAGACGAAACACCGCCGCATCAGAGCAGCCAATGGTCTGCTGCGCCCAATCATAGCC
AAACAGACGTTCCACCCACGCTGCCGGGCTACCCGCATGCAGGCCATCCTGTTCAATCATACTCTTCCT
TTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA
GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

pMK-RQ S-PM2d psbAII GmR

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCG
CTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCT
TCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT
TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGTTTAGAC
TTTGACATTAGTTAATTTTTCCCCATTGCCCCAAAATACATCCCCCTAAAAATATCAGAATCCTTGCCC
AGATGCAGGCCTTCTGGCGATCGCCATGGTGAGCAACGATTGCGGCTTTAGCGTTCCAGTGGATATTTG
CTGGGGGTTAATGAAACATTGTGGCGGAACCCAGGGACAATGTGACCAAAAAATTCAGGGATATCAATA
AGTATTAGGTATATGGATCATAATTGTATGCCCGACTATTGCTTAAACTGACTGACCACTGACCTTAAG
AGTAATGGCGTGCAAGGCCCAGTGATCAATTTCATTATTTTTCATTATTTCATCTCCATTGTCCCTGAA
AATCAGTTGTGTCGCCCCTCTACACAGCCCAGAACTATGGTAAAGGCGCACGAAAAACCGCCAGGTAAA
CTCTTCTCAACCCCCAAAACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACT
TATGTCATCTATAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCT
TTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAATGACTG
CATCCATCGCTCAACAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGAGTGGGTTACCAGCACCGACA
ACCGCCTCTATGTGGGTTGGTTTGGAACGCTGATGATTCCTACTCTTCTCGCAGCTGCTATCTGCTTTA
TTGTTGCTTTCATCGCTGCACCTCCCGTCGATATTGACGGCATCCGCGAACCTGTTGCTGGTTCTCTAA
TGTACGGAAATAACATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGCCATCGGCCTTCACTTCTATC
CCATCTGGGAAGCTGCTTCTCTTGATGAGTGGCTATATAATGGTGGTCCTTATCAACTAGTGGTCTTCC
ACTTCCTTATCGGTGTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTCTTACCGACTGGGTATGCGCC
CTTGGATTTGTGTCGCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGTTTTCCTGGTCTACCCCTTTG
GACAAGGTTCTTTCAGTGATGGTATGCCGCTTGGCATTTCTGGAACTTTTAACTACATGCTTGTTTTCC
AGGCAGAACACAACATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGGGTCTTCGGTGGTTCTC
TGTTCTCTGCTATGCATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACGACTGAGGTAGAATCCC
AGAACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCTGCTCATGGTTACTTCG
GTCGTCTGATCTTCCAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTTTTCCTTGCTGCTTGGC
CTGTTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAACCTCAATGGATTCAACT
TCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGATGTGCTGAATCGTGCTG
GTCTGGGTATGGAGGTGATGCACGAGCGTAATGCACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAG
CAACACCAGTTGCCTTGACTGCACCCGCAATCGGTTGAAACCCAGTTGACATAAGCCTGTTCGGTTCGT
AAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGT
AACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCC
AAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAG
CAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTCG
CACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTTGATCTTTTCGGTCGTGAGTT
CGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCCGATTACCTCGGGAACTTGCTCCGTAGTAA
GACATTCATCGCGCTTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCC
CAGGTTTGAGCAGCCGCGTAGTGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCA

GGGCATTGCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATGTGATCTA
CGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAAGTTGGGCATACGGGAAGAAGT
GATGCACTTTGATATCGACCCAAGTACCGCCACCTAAAATGTCTAACAATGCTTCAAGCCGACGCCGCT
TCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTC
TCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATCGAT
CGCCGGGGAGTCCGGGGCAGTTACCATTAGAGAGTCTAGAGAATTAATCCATCTTCGATAGAGGAATTA
TGGGGGAAGAACCTGTGCCGGCGGATAAAGCATTAGGCAAGAAATTCAAGAAAAAAAATGCCTCCTGGA
GCATTGAAGAAAGCGAAGCTCTGTACCGGGTTGAGGCCTGGGGGGCACCTTATTTTGCCATTAATGCCG
CTGGTAACATAACCGTCTCTCCCAACGGCGATCGGGGCGGTTCGTTAGATTTGTTGGAACTGGTGGAAG
CCCTGCGGCAAAGAAAGCTCGGCTTACCCCTATTAATTCGTTTTTCCGATATTTTGGCCGATCGCCTAG
AGCGATTGAATAGTTGTTTTGCCAAGGCGATCGCCCGTTACAATTACCCCATTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATA
CCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT
GTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGT
GTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAA
CAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG
CTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCG
CAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAA
CTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAA
ATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTAGAAAAATTCATC
CAGCAGACGATAAAACGCAATACGCTGGCTATCCGGTGCCGCAATGCCATACAGCACCAGAAAACGATC
CGCCCATTCGCCGCCCAGTTCTTCCGCAATATCACGGGTGGCCAGCGCAATATCCTGATAACGATCCGC
CACGCCCAGACGGCCGCAATCAATAAAGCCGCTAAAACGGCCATTTTCCACCATAATGTTCGGCAGGCA
CGCATCACCATGGGTCACCACCAGATCTTCGCCATCCGGCATGCTCGCTTTCAGACGCGCAAACAGCTC
TGCCGGTGCCAGGCCCTGATGTTCTTCATCCAGATCATCCTGATCCACCAGGCCCGCTTCCATACGGGT
ACGCGCACGTTCAATACGATGTTTCGCCTGATGATCAAACGGACAGGTCGCCGGGTCCAGGGTATGCAG
ACGACGCATGGCATCCGCCATAATGCTCACTTTTTCTGCCGGCGCCAGATGGCTAGACAGCAGATCCTG
ACCCGGCACTTCGCCCAGCAGCAGCCAATCACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACA
CGGAACACCGGTGGTGGCCAGCCAGCTCAGACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCT
CAGATCGGTTTTCACAAACAGCACCGGACGACCCTGCGCGCTCAGACGAAACACCGCCGCATCAGAGCA
GCCAATGGTCTGCTGCGCCCAATCATAGCCAAACAGACGTTCCACCCACGCTGCCGGGCTACCCGCATG
CAGGCCATCCTGTTCAATCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCG
AAAAGTGCCAC

pMA-T psbAIII SmR

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCG
CTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCT
TCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT
TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATT
GGCGGAAGGCCGTCAAGGCCACGTGTCTTGTCCGCGGTACCCCCGCAGAGATAGTTGCCGTTGTAACTC
CTGTTTTTGCGCCGGCTCTAGGGTTTGAATTTGGATTTGATAGCGACGGGCTAAATCTAAAGCCACATC
TGTAAAGGAACGCTTACCCAGTTCCATTAAAAAATTAAAGGCATTGCCCCCAGCCCCACAACCAAAGCA
ATAATACATTTGCTTGGCTGGACTGACACTAAAACTCGGTGACTTTTCATCGTGGAAAGGACATAGGCC
AAGGTGATCCCGCCCCCTTTTCTTGAGCACGACATAATCGCCGATAATTTCGACAATGTCTATTCTTTG
CTTAATTTCTTGGATGGTGTCTGGATGCAGACGGAGGTTATCCATGATAAATCAACCTACTCTCCTTCT
AAGAGTGACGATGGGGAGAGATACTTGTCAACTTAAGCTCTACTCAAGCATGGTCAACCAGCAAAAAGC
ACCCCCCGGTGAAGGAGAGTGCAATTTGCAGATTATTCAGTTGGCATTACACCAAGGAACACCAAACAC
CCCCCAAAACCAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCAT
CACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTG
GTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAA
GAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTT
CAGGAGCTAAGGAAGCTAAAATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTA
ACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGGGGTACAGTCTATGCCTCGGGCAT
CCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGC
AGGGCAGTCGCCCTAAAACAAAGTTAAACATCATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAAC

TATCAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCGTACATTTGTACGGCT
CCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTTG
ATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGA
TTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGC
GCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAGGTATCTTCGAGCCAGCCACGATCG
ACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTTGCCTTGGTAGGTCCAGCGGCGG
AGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGA
ACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAATGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCG
CAGTAACCGGCAAAATCGCGCCGAAGGAGGTCGCTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGT
ATCAGCCCGTCATACGTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCGCG
CAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAATTTT
TTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTGCTACGCCTGAATAAGTGATAATAAGCGGATGAA
TGGCAGAAATTCGAAAGCAAATTCGACCCGGTCGTCGGTTCAGGGCAGGGTCGTTAAATAGCCGCTTAT
GTCTATTGCTGGTTTACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGA
GGCCAGTTTGCTCAGGCTCTCCCCGTGTGGTGTGGTGTGGTTGTGTGTTGGTTATAATTCCTTATGTAT
TTGTCAATGTTCAAAGGATTTGGCCTCAAGCTCAGAGATTTCTTTGCTTGGGAGCCGGTTATGTATATG
AGTAATATAACAAGATTGTAACAATTTGTAAATAGCTTGCTGAAAAAAATCTGTCTAGCCGGTGGTTGC
TAGGGGTTGCCGATCGCCATTAGAGGGGTGATTGATTCCGTTGGGGCTTCAGTCTGTGGTGGTGGGGAG
ATGGAAAGCCACTAATCCTGCAACACAACTTCCCCTTGGCGCAGAATTTGGAAGATTTCCTTTGCTCCG
TCCCATTGGGCAACGGTGGACGGTAGACCTTTGCCAATTGGCCCCTGTTGCTCCAGGCTGAGGCAATCT
AGACAATGAACGGCAGGGAAAGTTTCGGCAATATCGGCCAACTTTTCCAGGGGGGGCTGTCCCGACAAA
TTAGCACTGGTGGTGGCCAGGGGTCCCGTTTGCAGCAAAATTTCTAGGGAGATCGCCTGGCGGGGAGCT
CGGAGCACAAGACTGGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGC
CAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCTCACTGAC
TCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGA
ACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAA
GCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGG
GCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATG
TAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTA
TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA
CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAG
ATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCA
TGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCT
GTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTAC
CATCTGGCCCCAGTGCTGCAATGATACCGCGAGAACCACGCTCACCGGCTCCAGATTTATCAGCAATAA
ACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGC
GAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAA
GTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCA
TCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGT
AACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAA
CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCC
TTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

pMK-T Apramycin control AmpR

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCG
CTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCT
TCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT
TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATT

GAAGGAAGGCCGTCAAGGCCACGTGTCTTGTCCGCTTAATTAATTTAGACTTTGACATTAGTTAATTTT
TCCCCATTGCCCCAAAATACATCCCCCTAAAAATATCAGAATCCTTGCCCAGATGCAGGCCTTCTGGCG
ATCGCCATGGTGAGCAACGATTGCGGCTTTAGCGTTCCAGTGGATATTTGCTGGGGGTTAATGAAACAT
TGTGGCGGAACCCAGGGACAATGTGACCAAAAAATTCAGGGATATCAATAAGTATTAGGTATATGGATC
ATAATTGTATGCCCGACTATTGCTTAAACTGACTGACCACTGACCTTAAGAGTAATGGCGTGCAAGGCC
CAGTGATCAATTTCATTATTTTTCATTATTTCATCTCCATTGTCCCTGAAAATCAGTTGTGTCGCCCCT
CTACACAGCCCAGAACTATGGTAAAGGCGCACGAAAAACCGCCAGGTAAACTCTTCTCAACCCCCAAAA
CGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTATAAGCTTC
GTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAG
TTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAAACCCAGTTGACATAAGCCTGTTCGG
TTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCG
GTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGG
CATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTCATCA
GCGGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAACCTT
GGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAGATGGG
CCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCTCGTGG
TCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCT
GACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCA
CAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCGCCCGT
GTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGCATCTT
GCCGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCAAGTTG
GTACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAG
GAGAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACATT
GTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGGGATGC
GAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAAGCCGA
CGCCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAAT
TGCACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGC
CATCGATCGCCGGGGAGTCCGGGGCAGTTACCATTAGAGAGTCTAGAGAATTAATCCATCTTCGATAGA
GGAATTATGGGGGAAGAACCTGTGCCGGCGGATAAAGCATTAGGCAAGAAATTCAAGAAAAAAAATGCC
TCCTGGAGCATTGAAGAAAGCGAAGCTCTGTACCGGGTTGAGGCCTGGGGGGCACCTTATTTTGCCATT
AATGCCGCTGGTAACATAACCGTCTCTCCCAACGGCGATCGGGGCGGTTCGTTAGATTTGTTGGAACTG
GTGGAAGCCCTGCGGCAAAGAAAGCTCGGCTTACCCCTATTAATTCGTTTTTCCGATATTTTGGCCGAT
CGCCTAGAGCGATTGAATAGTTGTTTTGCCAAGGCGATCGCCCGTTACAATTACCCCAGGCGCGCCGGA
GCACAAGACTGGCCTCATGGGCCTTCCTTTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGC
TGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCTCACTGACTCGC
TGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCG
TAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACG
CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCT
CGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGC
TGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTAT
ATATGAGTAAACTTGGTCTGACAGTTATTAGAAAAATTCATCCAGCAGACGATAAAACGCAATACGCTG
GCTATCCGGTGCCGCAATGCCATACAGCACCAGAAAACGATCCGCCCATTCGCCGCCCAGTTCTTCCGC
AATATCACGGGTGGCCAGCGCAATATCCTGATAACGATCCGCCACGCCCAGACGGCCGCAATCAATAAA
GCCGCTAAAACGGCCATTTTCCACCATAATGTTCGGCAGGCACGCATCACCATGGGTCACCACCAGATC
TTCGCCATCCGGCATGCTCGCTTTCAGACGCGCAAACAGCTCTGCCGGTGCCAGGCCCTGATGTTCTTC
ATCCAGATCATCCTGATCCACCAGGCCCGCTTCCATACGGGTACGCGCACGTTCAATACGATGTTTCGC
CTGATGATCAAACGGACAGGTCGCCGGGTCCAGGGTATGCAGACGACGCATGGCATCCGCCATAATGCT
CACTTTTTCTGCCGGCGCCAGATGGCTAGACAGCAGATCCTGACCCGGCACTTCGCCCAGCAGCAGCCA
ATCACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACACGGAACACCGGTGGTGGCCAGCCAGCT
CAGACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCTCAGATCGGTTTTCACAAACAGCACCGG
ACGACCCTGCGCGCTCAGACGAAACACCGCCGCATCAGAGCAGCCAATGGTCTGCTGCGCCCAATCATA
GCCAAACAGACGTTCCACCCACGCTGCCGGGCTACCCGCATGCAGGCCATCCTGTTCAATCATACTCTT
CCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT
TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

pOK-RQ S-PM2d Intronless psbA ApmR

GCTAATAACAGGCCTGCTGGTAATCGCAGGCCTTTTTATTTGGTGTATTTAGAAAAATAAACAAATAGG
GGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCG
TTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCA
AAAGAATAGACCGAGATAGGGTTGAGTGGCCGCTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAA
CTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTG
CAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGC
GCGACGTAATACGACTCACTATAGGGCGAATTGTAGAAGGAAGGCCGTCAAGGCCGCATTTTAGACTTT
GACATTAGTTAATTTTTCCCCATTGCCCCAAAATACATCCCCCTAAAAATATCAGAATCCTTGCCCAGA
TGCAGGCCTTCTGGCGATCGCCATGGTGAGCAACGATTGCGGCTTTAGCGTTCCAGTGGATATTTGCTG
GGGGTTAATGAAACATTGTGGCGGAACCCAGGGACAATGTGACCAAAAAATTCAGGGATATCAATAAGT
ATTAGGTATATGGATCATAATTGTATGCCCGACTATTGCTTAAACTGACTGACCACTGACCTTAAGAGT
AATGGCGTGCAAGGCCCAGTGATCAATTTCATTATTTTTCATTATTTCATCTCCATTGTCCCTGAAAAT
CAGTTGTGTCGCCCCTCTACACAGCCCAGAACTATGGTAAAGGCGCACGAAAAACCGCCAGGTAAACTC
TTCTCAACCCCCAAAACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTAT
GTCATCTATAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTA
GACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAATGACTGCAT
CCATCGCTCAACAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGAGTGGGTTACCAGCACCGACAACC
GCCTCTATGTGGGTTGGTTTGGAACGCTGATGATTCCTACTCTTCTCGCAGCTGCTATCTGCTTTATTG
TTGCTTTCATCGCTGCACCTCCCGTCGATATTGACGGCATCCGCGAACCTGTTGCTGGTTCTCTAATGT
ACGGAAATAACATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGCCATCGGCCTTCACTTCTATCCCA
TCTGGGAAGCTGCTTCTCTTGATGAGTGGCTATATAATGGTGGTCCTTATCAACTAGTGGTCTTCCACT
TCCTTATCGGTGTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTCTTACCGACTGGGTATGCGCCCTT
GGATTTGTGTCGCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGTTTTCCTGGTCTACCCCTTTGGAC
AAGGTTCTTTCAGTGATGGTATGCCGCTTGGCATTTCTGGAACTTTTAACTACATGCTTGTTTTCCAGG
CAGAACACAACATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGGGTCTTCGGTGGTTCTCTGT
TCTCTGCTATGCATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACGACTGAGGTAGAATCCCAGA
ACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCTGCTCATGGTTACTTCGGTC
GTCTGATCTTCCAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTTTTCCTTGCTGCTTGGCCTG
TTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAACCTCAATGGATTCAACTTCA
ACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGATGTGCTGAATCGTGCTGGTC
TGGGTATGGAGGTGATGCACGAGCGTATGCACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAGCAAC
ACCAGTTGCCTTGACTGCACCCGCAATCGGTTGAAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAAC
TGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACG
GCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGC
AGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTCATCAGCGGTGGAG
TGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAACCTTGGGGTTACC
CCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAGATGGGCCACTTGGA
CTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCTCGTGGTCAGGTCTG
GACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCTGACACATTC
TGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCACAGGCAGAG
CAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCGCCCGTGTCCATGAA
CTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGCATCTTGCCGAGTTG
ATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTC
GATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAAGAGC
CTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACATTGTGGCGACA
GCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGGGATGCGAAGAATGC
GATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAAGCCGACGCCGCTTC
GGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTC
CTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATCGATCG
CCGGGGAGTCCGGGGCAGTTACCATTAGAGAGTCTAGAGAATTAATCCATCTTCGATAGAGGAATTATG
GGGGAAGAACCTGTGCCGGCGGATAAAGCATTAGGCAAGAAATTCAAGAAAAAAAATGCCTCCTGGAGC
ATTGAAGAAAGCGAAGCTCTGTACCGGGTTGAGGCCTGGGGGGCACCTTATTTTGCCATTAATGCCGCT
GGTAACATAACCGTCTCTCCCAACGGCGATCGGGGCGGTTCGTTAGATTTGTTGGAACTGGTGGAAGCC
CTGCGGCAAAGAAAGCTCGGCTTACCCCTATTAATTCGTTTTTCCGATATTTTGGCCGATCGCCTAGAG
CGATTGAATAGTTGTTTTGCCAAGGCGATCGCCCGTTACAATTACCCCACTGGGCCTCATGGGCCTTCC
TTCTATCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTT
TCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGC

CTGGGGTGCCTAAATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTCGTTTTATTGAGCAAAAGGC
CAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGAC
GAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
TTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTC
GTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAAC
TATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATT
AGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGA
AGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAA
AAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGT
TAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGT
TTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTAGAAAAATTCATCCAGCAGA
CGATAAAACGCAATACGCTGGCTATCCGGTGCCGCAATGCCATACAGCACCAGAAAACGATCCGCCCAT
TCGCCGCCCAGTTCTTCCGCAATATCACGGGTGGCCAGCGCAATATCCTGATAACGATCCGCCACGCCC
AGACGGCCGCAATCAATAAAGCCGCTAAAACGGCCATTTTCCACCATAATGTTCGGCAGGCACGCATCA
CCATGGGTCACCACCAGATCTTCGCCATCCGGCATGCTCGCTTTCAGACGCGCAAACAGCTCTGCCGGT
GCCAGGCCCTGATGTTCTTCATCCAGATCATCCTGATCCACCAGGCCCGCTTCCATACGGGTACGCGCA
CGTTCAATACGATGTTTCGCCTGATGATCAAACGGACAGGTCGCCGGGTCCAGGGTATGCAGACGACGC
ATGGCATCCGCCATAATGCTCACTTTTTCTGCCGGCGCCAGATGGCTAGACAGCAGATCCTGACCCGGC
ACTTCGCCCAGCAGCAGCCAATCACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACACGGAACA
CCGGTGGTGGCCAGCCAGCTCAGACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCTCAGATCG
GTTTTCACAAACAGCACCGGACGACCCTGCGCGCTCAGACGAAACACCGCCGCATCAGAGCAGCCAATG
GTCTGCTGCGCCCAATCATAGCCAAACAGACGTTCCACCCACGCTGCCGGGCTACCCGCATGCAGGCCA
TCCTGTTCAATCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
GGATACATATTTGAA

pMK-RQ S-PM2d psbA ApmR

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCG
CTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCT
TCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT
TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATT
GAAGGAAGGCCGTCAAGGCCGCATTTTAGACTTTGACATTAGTTAATTTTTCCCCATTGCCCCAAAATA
CATCCCCCTAAAAATATCAGAATCCTTGCCCAGATGCAGGCCTTCTGGCGATCGCCATGGTGAGCAACG
ATTGCGGCTTTAGCGTTCCAGTGGATATTTGCTGGGGGTTAATGAAACATTGTGGCGGAACCCAGGGAC
AATGTGACCAAAAAATTCAGGGATATCAATAAGTATTAGGTATATGGATCATAATTGTATGCCCGACTA
TTGCTTAAACTGACTGACCACTGACCTTAAGAGTAATGGCGTGCAAGGCCCAGTGATCAATTTCATTAT
TTTTCATTATTTCATCTCCATTGTCCCTGAAAATCAGTTGTGTCGCCCCTCTACACAGCCCAGAACTAT
GGTAAAGGCGCACGAAAAACCGCCAGGTAAACTCTTCTCAACCCCCAAAACGCCCTCTGTTTACCCATG
GAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTATAAGCTTCGTGTATATTAACTTCCTGT
TACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGAC
AAATACATAAGGAATTATAACCAAATGACTGCATCCATCGCTCAACAGCGTGGTAGCAACACTTGGGAA
CAGTTCTGCGAGTGGGTTACCAGCACCGACAACCGCCTCTATGTGGGTTGGTTTGGAACGCTGATGATT
CCTACTCTTCTCGCAGCTGCTATCTGCTTTATTGTTGCTTTCATCGCTGCACCTCCCGTCGATATTGAC
GGCATCCGCGAACCTGTTGCTGGTTCTCTAATGTACGGAAATAACATCATCTCTGGTGCTGTTATCCCT
TCTAGCAACGCCATCGGCCTTCACTTCTATCCCATCTGGGAAGCTGCTTCTCTTGATGAGTGGCTATAT
AATGGTGGTCCTTATCAACTAGTGGTCTTCCACTTCCTTATCGGTGTCTTCTCTTACATGGGTCGCGAG
TGGGAACTCTCTTACCGACTGGGTATGCGCCCTTGGATTTGTGTCGCTTACTCAGCCCCTGTTGCCGCT
GCTACTGCTGTTTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTCAGTGATGGTATGCCGCTTGGCATT
TCTGGAACTTTTAACTACATGCTTGTTTTCCAGGCAGAACACAACATTCTCATGCATCCTTTCCACATG
TTGGGAGTTGCTGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATGCATGGATCTCTTGTCACCTCTAGT
CTTGTACGTGAGACGACTGAGGTAGAATCCCAGAACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACT
TACAACATTGTTGCTGCTCATGGTTACTTCGGTCGTCTGATCTTCCAATATGCGTCGTTCAACAACTCT
CGTTCGCTACACTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCT
ACGATGGCGTTCAACCTCAATGGATTCAACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTC
AACACCTGGGCAGATGTGCTGAATCGTGCTGGTCTGGGTATGGAGGTGATGCACGAGCGTCAAGTTGTG
CTTTGCGCTCTTTAAATCGGATGAATTGCTGGAAACCCCAAGTGGGCAATCAGCAGCCAAGTCTCAGAT

ACATCTGAGAAAGGTTCAGAGACTACCTGAGGGATATAGTTCCCTTAATAACAGGAATAAGCGTCCGAC
ACCAGAAATGGTGATGATATAGTCCAATCCTGGTAGTAATACCAGATAGTTAAGGGAAGTTTAAGAATG
CACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAGCAACACCAGTTGCCTTGACTGCACCCGCAATCG
GTTGAAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGC
AACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTAT
GACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGT
TTGATGTTATGGAGCAGCAACGATGTCATCAGCGGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAA
AGCCGAGCTCATCGGTCAGCTTCTCAACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTC
CTTCCGTAGCGTCCGGCCCCTCGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGG
TCCGGGAGGGACGCTCGTCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTC
GCCCGTTACACCGGACCTTGGAGTTGTCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGC
CCATCCATTTGCCTTTGCGGCAGCGGGGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCC
ACCTCACTCGCCTGCAAGCCCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGT
GGGACACGATGCCAACACGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAG
ACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTCGATTATCTCGAGAATGACCACTGCTGTGA
GCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGC
CTTTGCTCGGTTGATCCGCTCCCGCGACATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTT
GATCTTCCTGCATCCGCCAGAGGCGGGATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGC
TCATAAAATGTCTAACAATGCTTCAAGCCGACGCCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCT
TGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACT
GAGTAATCTTCTGATTGCTGATCTTGATTGCCATCGATCGCCGGGGAGTCCGGGGCAGTTACCATTAGA
GAGTCTAGAGAATTAATCCATCTTCGATAGAGGAATTATGGGGGAAGAACCTGTGCCGGCGGATAAAGC
ATTAGGCAAGAAATTCAAGAAAAAAAAATGCCTCCTGGAGCATTGAAGAAAGCGAAGCTCTGTACCGGGT
TGAGGCCTGGGGGGCACCTTATTTTGCCATTAATGCCGCTGGTAACATAACCGTCTCTCCCAACGGCGA
TCGGGGCGGTTCGTTAGATTTGTTGGAACTGGTGGAAGCCCTGCGGCAAAGAAAGCTCGGCTTACCCCT
ATTAATTCGTTTTTCCGATATTTTGGCCGATCGCCTAGAGCGATTGAATAGTTGTTTTGCCAAGGCGAT
CGCCCGTTACAATTACCCCACTGGGCCTCATGGGCCTTCCTTTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAA
AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATC
ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCC
CTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCC
CTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT
CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAAT
CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTAGAAAAATTCATCCAGCAGACGATAAA
ACGCAATACGCTGGCTATCCGGTGCCGCAATGCCATACAGCACCAGAAAACGATCCGCCCATTCGCCGC
CCAGTTCTTCCGCAATATCACGGGTGGCCAGCGCAATATCCTGATAACGATCCGCCACGCCCAGACGGC
CGCAATCAATAAAGCCGCTAAAACGGCCATTTTCCACCATAATGTTCGGCAGGCACGCATCACCATGGG
TCACCACCAGATCTTCGCCATCCGGCATGCTCGCTTTCAGACGCGCAAACAGCTCTGCCGGTGCCAGGC
CCTGATGTTCTTCATCCAGATCATCCTGATCCACCAGGCCCGCTTCCATACGGGTACGCGCACGTTCAA
TACGATGTTTCGCCTGATGATCAAACGGACAGGTCGCCGGGTCCAGGGTATGCAGACGACGCATGGCAT
CCGCCATAATGCTCACTTTTTCTGCCGGCGCCAGATGGCTAGACAGCAGATCCTGACCCGGCACTTCGC
CCAGCAGCAGCCAATCACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACACGGAACACCGGTGG
TGGCCAGCCAGCTCAGACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCTCAGATCGGTTTTCA
CAAACAGCACCGGACGACCCTGCGCGCTCAGACGAAACACCGCCGCATCAGAGCAGCCAATGGTCTGCT
GCGCCCAATCATAGCCAAACAGACGTTCCACCCACGCTGCCGGGCTACCCGCATGCAGGCCATCCTGTT
CAATCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

pMK-RQ S-PM2d psbA Entire Region ApmR

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCG
CTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCT

TCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTT
TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATT
GAAGGAAGGCCGTCAAGGCCGCATTTTAGACTTTGACATTAGTTAATTTTTCCCCATTGCCCCAAAATA
CATCCCCCTAAAAATATCAGAATCCTTGCCCAGATGCAGGCCTTCTGGCGATCGCCATGGTGAGCAACG
ATTGCGGCTTTAGCGTTCCAGTGGATATTTGCTGGGGGTTAATGAAACATTGTGGCGGAACCCAGGGAC
AATGTGACCAAAAAATTCAGGGATATCAATAAGTATTAGGTATATGGATCATAATTGTATGCCCGACTA
TTGCTTAAACTGACTGACCACTGACCTTAAGAGTAATGGCGTGCAAGGCCCAGTGATCAATTTCATTAT
TTTTCATTATTTCATCTCCATTGTCCCTGAAAATCAGTTGTGTCGCCCCTCTACACAGCCCAGAACTAT
GGTAAAGGCGCACGAAAAACCGCCAGGTAAACTCTTCTCAACCCCCAAAACGCCCTCTGTTTACCCATG
GAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTATAAGCTTCGTGTATATTAACTTCCTGT
TACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGAC
AAATACATAAGGAATTATAACCAAATGACTGCATCCATCGCTCAACAGCGTGGTAGCAACACTTGGGAA
CAGTTCTGCGAGTGGGTTACCAGCACCGACAACCGCCTCTATGTGGGTTGGTTTGGAACGCTGATGATT
CCTACTCTTCTCGCAGCTGCTATCTGCTTTATTGTTGCTTTCATCGCTGCACCTCCCGTCGATATTGAC
GGCATCCGCGAACCTGTTGCTGGTTCTCTAATGTACGGAAATAACATCATCTCTGGTGCTGTTATCCCT
TCTAGCAACGCCATCGGCCTTCACTTCTATCCCATCTGGGAAGCTGCTTCTCTTGATGAGTGGCTATAT
AATGGTGGTCCTTATCAACTAGTGGTCTTCCACTTCCTTATCGGTGTCTTCTCTTACATGGGTCGCGAG
TGGGAACTCTCTTACCGACTGGGTATGCGCCCTTGGATTTGTGTCGCTTACTCAGCCCCTGTTGCCGCT
GCTACTGCTGTTTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTCAGTGATGGTATGCCGCTTGGCATT
TCTGGAACTTTTAACTACATGCTTGTTTTCCAGGCAGAACACAACATTCTCATGCATCCTTTCCACATG
TTGGGAGTTGCTGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATGCATGGATCTCTTGTCACCTCTAGT
CTTGTACGTGAGACGACTGAGGTAGAATCCCAGAACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACT
TACAACATTGTTGCTGCTCATGGTTACTTCGGTCGTCTGATCTTCCAATATGCGTCGTTCAACAACTCT
CGTTCGCTACACTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCT
ACGATGGCGTTCAACCTCAATGGATTCAACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTC
AACACCTGGGCAGATGTGCTGAATCGTGCTGGTCTGGGTATGGAGGTGATGCACGAGCGTCAAGTTGTG
CTTTGCGCTCTTTAAATCGGATGAATTGCTGGAAACCCCAAGTGGGCAATCAGCAGCCAAGTCTCAGAT
ACATCTGAGAAAGGTTCAGAGACTACCTGAGGGATATAGTTCCCTTAATAACAGGAATAAGCGTCCGAC
ACCAGAAATGGTGATGATATAGTCCAATCCTGGTAGTAATACCAGATAGTTAAGGGAAGTTTAAGAATG
CACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAGCAACACCAGTTGCCTTGACTGCACCCGCAATCG
GTTGATATAATGTTAAGGAACTCTTCGGAGTTCCTTTTTTTATAAATATTTAAGCACGAAAGAAAACAC
GAATGACTAAACTATACTCTGATTTGTATAGAACTTGTATGACTTGTGGGGAGGAAAAACTTTCTACAG
AGTTTTATGTTCGCAATAAAAAAACAGGAGTTCGTCATTCTTCTTGTAAAGAATGTGATAAAGTGCGAG
TGAAATCAAGACACAAAGAAAATCCAGAACGCACAAAAAATAATGACTTAAAAAGACTTTATGGTATTA
CTCTTGACGAACATACGCAAATGTATGAGGAGCAGAATGGAGTTTGTGCTATTTGTAAAGGAGAAGGTG
ATGGTAAATGGAAAAAACTTTGCGTAGATCATGACCATGAAACTGGTAAAGTTCGTCAGTTGCTTTGTA
GGAACTGCAATATGATGTTGGGTCAAGTAAATGATAATGTTAATTTACTTTCTGAAATGATTAAATACT
TAAAGAGGTATCAATGAAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGT
ATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTC
ATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCC
GTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTCATCAGCGGTGGAGTGCAATGTCGTGCAATA
CGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAACCTTGGGGTTACCCCCGGCGGTGTGCTGCT
GGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCG
TGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTCGA
TCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCTGACACATTCTGGCGCCTGCCAAATGT
AAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCACAGGCAGAGCAGATCATCTCTGATCC
ATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACT
TCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTA
TGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTCGATTATCTCGAGAATGA
CCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAAGAGCCTTCAGAAGGAAGGTCC
AGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACATTGTGGCGACAGCCCTGGGTCAACTGGG
CCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGGGATGCGAAGAATGCGATGCCGCTCGCCAGTC
GATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAAGCCGACGCCGCTTCGGCGCGGCTTAACTCAA
GCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTCCTTCAATGGGGGGTGCT
TTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATCGATCGCCGGGGAGTCCGGGGCA
GTTACCATTAGAGAGTCTAGAGAATTAATCCATCTTCGATAGAGGAATTATGGGGGAAGAACCTGTGCC
GGCGGATAAAGCATTAGGCAAGAAATTCAAGAAAAAAAATGCCTCCTGGAGCATTGAAGAAAGCGAAGC
TCTGTACCGGGTTGAGGCCTGGGGGGCACCTTATTTTGCCATTAATGCCGCTGGTAACATAACCGTCTC
TCCCAACGGCGATCGGGGCGGTTCGTTAGATTTGTTGGAACTGGTGGAAGCCCTGCGGCAAAGAAAGCT
CGGCTTACCCCTATTAATTCGTTTTTCCGATATTTTGGCCGATCGCCTAGAGCGATTGAATAGTTGTTT
TGCCAAGGCGATCGCCCGTTACAATTACCCCACTGGGCCTCATGGGCCTTCCTTTCACTGCCCGCTTTC

CAGTCGGGAAACCTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTC
TCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAA
AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC
CTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACC
AGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT
CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT
AGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG
GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACA
GGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACA
CTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACT
CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAAT
GAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTAGAAAAATTCATCCA
GCAGACGATAAAACGCAATACGCTGGCTATCCGGTGCCGCAATGCCATACAGCACCAGAAAACGATCCG
CCCATTCGCCGCCCAGTTCTTCCGCAATATCACGGGTGGCCAGCGCAATATCCTGATAACGATCCGCCA
CGCCCAGACGGCCGCAATCAATAAAGCCGCTAAAACGGCCATTTTCCACCATAATGTTCGGCAGGCACG
CATCACCATGGGTCACCACCAGATCTTCGCCATCCGGCATGCTCGCTTTCAGACGCGCAAACAGCTCTG
CCGGTGCCAGGCCCTGATGTTCTTCATCCAGATCATCCTGATCCACCAGGCCCGCTTCCATACGGGTAC
GCGCACGTTCAATACGATGTTTCGCCTGATGATCAAACGGACAGGTCGCCGGGTCCAGGGTATGCAGAC
GACGCATGGCATCCGCCATAATGCTCACTTTTTCTGCCGGCGCCAGATGGCTAGACAGCAGATCCTGAC
CCGGCACTTCGCCCAGCAGCAGCCAATCACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACACG
GAACACCGGTGGTGGCCAGCCAGCTCAGACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCTCA
GATCGGTTTTCACAAACAGCACCGGACGACCCTGCGCGCTCAGACGAAACACCGCCGCATCAGAGCAGC
CAATGGTCTGCTGCGCCCAATCATAGCCAAACAGACGTTCCACCCACGCTGCCGGGCTACCCGCATGCA
GGCCATCCTGTTCAATCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA
TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCAC

Appendix 10: PCR confirmation products of the insertion of constructs transformed into the GT *Synechocystis* sp. PCC6803.

pMA-T psbAIII SmR insertion confirmation.

GAAAGGACATAGGCCAAGGTGATCCCGCCCCCTTTTCTTGAGCACGACATAATCGCCGATAATTTCGAC
AATGTCTATTCTTTGCTTAATTTCTTGGATGGTGTCTGGATGCAGACGGAGGTTATCCATGATAAATCA
ACCTACTCTCCTTCTAAGAGTGACGATGGGGAGAGATACTTGTCAACTTAAGCTCTACTCAAGCATGGT
CAACCAGCAAAAAGCACCCCCCGGTGAAGGAGAGTGCAATTTGCAGATTATTCAGTTGGCATTACACCA
AGGAACACCAAACACCCCCCAAAACCAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCA
GTAAGTTGGCAGCATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAG
AATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACG
TTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTG
AGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGCGCTCACGCAACTGGTCCAGAACCTTGACCG
AACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGGGGTACAGT
CTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCA
ACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAAACATCATGAGGGAAGCGGTGATCGCCG
AAGTATCGACTCAACTATCAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTGCTGGCCG
TACATTTGTACGGCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGG
TGACCGTAAGGCTTGATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCC
CTGGAGAGAGCGAGATTCTCCGCGCTGTAGAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGGC
GTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAGGTATCTTCG
AGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGA

pMK-RQ S-PM2d psbAII GmR insertion confirmation.

GCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTA
GTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAATGACTGCATCCATCGCTCAA
CAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGAGTGGGTTACCAGCACCGACAACCGCCTCTATGTG

GGTTGGTTTGGAACGCTGATGATTCCTACTCTTCTCGCAGCTGCTATCTGCTTTATTGTTGCTTTCATC
GCTGCACCTCCCGTCGATATTGACGGCATCCGCGAACCTGTTGCTGGTTCTCTAATGTACGGAAATAAC
ATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGCCATCGGCCTTCACTTCTATCCCATCTGGGAAGCT
GCTTCTCTTGATGAGTGGCTATATAATGGTGGTCCTTATCAACTAGTGGTCTTCCACTTCCTTATCGGT
GTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTCTTACCGACTGGGTATGCGCCCTTGGATTTGTGTC
GCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGTTTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTC
AGTGATGGTATGCCGCTTGGCATTTCTGGAACTTTTAACTACATGCTTGTTTTCCAGGCAGAACACAAC
ATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATG
CATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACGACTGAGGTAGAATCCCAGAACTATGGTTAC
AAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCTGCTCATGGTTACTTCGGTCGTCTGATCTTC
CAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATC
TGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAACCTCAATGGATTCAACTTCAACCAGTCCATT
GTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGATGTGCTGAATCGTGCTGGTCTGGGTATGGAG
GTGATGCACGAGCGTAATGCACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAGCAACACCAGTTGCC
TTGACTGCACCCGCAATCGGTTGAAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAA
GTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGC
GGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCG
TTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGCAACGATGTTACGC
AGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTCGCACATGTAGGCTCG
GCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTTGATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCA
CCTACTCCCAACATCAGCCGGACTCCGATTACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCGCGC
TTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAGGTTTGAGCAGC
CGCGTAGTGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGCATTGCCACCG
CGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATGTGATCTACGTGCAAGCAGATT
ACGGTGACGATCCCGCAGTGGCTCTCTATACAAAGTTGGGCATACGGGAAGAAGTGATGCACTTTGATA
TCGACCCAAGTACCGCCACCTAAAATGTCTAACAATGCTTCAAGCCGACGCCGCTTCGGCGCGGCTTAA
CTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTCCTTCAATGGGGG
GTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATC

Appendix 11: PCR confirmation to check the sequences of the new apramycin-containing constructs.

Apramycin Control

GCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTA
GTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAAACCCAGTTGACATAAGCCTG
TTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACG
CAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCC
TCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGT
CATCAGCGGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCA
ACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAG
ATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCT
CGTGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTG
TCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGG
GGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCG
CCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGC
ATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCA
AGTTGGTACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGC
TCAAGGAGAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCG
ACATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGG
GATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAA
GCCGACGCCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTG
CAAATTGCACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTG
ATTGCCATC

Intronless S-PM2 *psbAII*

ACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAG
GAATTATAACCAAATGACTGCATCCATCGCTCAACAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGA
GTGGGTTACCAGCACCGACAACCGCCTCTATGTGGGTTGGTTTGGAACGCTGATGATTCCTACTCTTCT
CGCAGCTGCTATCTGCTTTATTGTTGCTTTCATCGCTGCACCTCCCGTCGATATTGACGGCATCCGCGA

ACCTGTTGCTGGTTCTCTAATGTACGGAAATAACATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGC
CATCGGCCTTCACTTCTATCCCATCTGGGAAGCTGCTTCTCTTGATGAGTGGCTATATAATGGTGGTCC
TTATCAACTAGTGGTCTTCCACTTCCTTATCGGTGTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTC
TTACCGACTGGGTATGCGCCCTTGGATTTGTGTCGCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGT
TTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTCAGTGATGGTATGCCGCTTGGCATTTCTGGAACTTT
TAACTACATGCTTGTTTTCCAGGCAGAACACAACATTCTCATGCATCCTTTCCACATGTTGGGAGTTGC
TGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATGCATGGATCTCTTGTCACCTCTAGTCTTGTACGTGA
GACGACTGAGGTAGAATCCCAGAACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACTTACAACATTGT
TGCTGCTCATGGTTACTTCGGTCGTCTGATCTTCCAATATGCGTCGTTCAACAACTCTCGTTCGCTACA
CTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTT
CAACCTCAATGGATTCAACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACACCTGGGC
AGATGTGCTGAATCGTGCTGGTCTGGGTATGGAGGTGATGCACGAGCGTATGCACACAACTTCCCTCTT
GACCTTGCAGCAGCAGAAGCAACACCAGTTGCCTTGACTGCACCCGCAATCGGTTGAAACCCAGTTGAC
ATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTT
GACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACA
GTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAG
CAACGATGTCATCAGCGGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTC
AGCTTCTCAACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGC
CCCTCGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCG
TCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACC
TTGGAGTTGTCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTG
CGGCAGCGGGGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAA
GCCCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACA
CGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTC
AGGATGGCAAGTTGGTACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGG
ACAGGTGGCTCAAGGAGAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCC
GCTCCCGCGACATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGC
CAGAGGCGGGATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACA
ATGCTTCAAGCCGACGCCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTG
AATAATCTGCAAATTGCACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTG
CTGATCTTGATTGCCATC

Intron-containing S-PM2 *psbAII*

GCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTA
GTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAATGACTGCATCCATCGCTCAA
CAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGAGTGGGTTACCAGCACCGACAACCGCCTCTATGTG
GGTTGGTTTGGAACGCTGATGATTCCTACTCTTCTCGCAGCTGCTATCTGCTTTATTGTTGCTTTCATC
GCTGCACCTCCCGTCGATATTGACGGCATCCGCGAACCTGTTGCTGGTTCTCTAATGTACGGAAATAAC
ATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGCCATCGGCCTTCACTTCTATCCCATCTGGGAAGCT
GCTTCTCTTGATGAGTGGCTATATAATGGTGGTCCTTATCAACTAGTGGTCTTCCACTTCCTTATCGGT
GTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTCTTACCGACTGGGTATGCGCCCTTGGATTTGTGTC
GCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGTTTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTC
AGTGATGGTATGCCGCTTGGCATTTCTGGAACTTTTAACTACATGCTTGTTTTCCAGGCAGAACACAAC
ATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATG
CATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACGACTGAGGTAGAATCCCAGAACTATGGTTAC
AAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCTGCTCATGGTTACTTCGGTCGTCTGATCTTC
CAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATC
TGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAACCTCAATGGATTCAACTTCAACCAGTCCATT
GTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGATGTGCTGAATCGTGCTGGTCTGGGTATGGAG
GTGATGCACGAGCGTCAAGTTGTGCTTTGCGCTCTTTAAATCGGATGAATTGCTGGAAACCCCAAGTGG
GCAATCAGCAGCCAAGTCTCAGATACATCTGAGAAAGGTTCAGAGACTACCTGAGGGATATAGTTCCCT
TAATAACAGGAATAAGCGTCCGACACCAGAAATGGTGATGATATAGTCCAATCCTGGTAGTAATACCAG
ATAGTTAAGGGAAGTTTAAGAATGCACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAGCAACACCAG
TTGCCTTGACTGCACCCGCAATCGGTTGAAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAA
TGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCA
GTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAA
GCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTCATCAGCGGTGGAGTGCAA
TGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAACCTTGGGGTTACCCCCGG
CGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAGATGGGCCACTTGGACTGAT

CGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCTCGTGGTCAGGTCTGGACGA
CGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCTGACACATTCTGGCG
CCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCACAGGCAGAGCAGAT
CATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCGCCCGTGTCCATGAACTCGA
TGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGCATCTTGCCGAGTTGATGGC
AAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTCGATTA
TCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAAGAGCCTTCA
GAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACATTGTGGCGACAGCCCT
GGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGGGATGCGAAGAATGCGATGC
CGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAAGCCGACGCCGCTTCGGCGC
GGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTCCTTCA
ATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATC

Entire S-PM2 *psbAII* region

GCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTA
GTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAATGACTGCATCCATCGCTCAA
CAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGAGTGGGTTACCAGCACCGACAACCGCCTCTATGTG
GGTTGGTTTGGAACGCTGATGATTCCTACTCTTCTCGCAGCTGCTATCTGCTTTATTGTTGCTTTCATC
GCTGCACCTCCCGTCGATATTGACGGCATCCGCGAACCTGTTGCTGGTTCTCTAATGTACGGAAATAAC
ATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGCCATCGGCCTTCACTTCTATCCCATCTGGGAAGCT
GCTTCTCTTGATGAGTGGCTATATAATGGTGGTCCTTATCAACTAGTGGTCTTCCACTTCCTTATCGGT
GTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTCTTACCGACTGGGTATGCGCCCTTGGATTTGTGTC
GCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGTTTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTC
AGTGATGGTATGCCGCTTGGCATTTCTGGAACTTTTAACTACATGCTTGTTTTCCAGGCAGAACACAAC
ATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATG
CATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACGACTGAGGTAGAATCCCAGAACTATGGTTAC
AAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCTGCTCATGGTTACTTCGGTCGTCTGATCTTC
CAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATC
TGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAACCTCAATGGATTCAACTTCAACCAGTCCATT
GTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGATGTGCTGAATCGTGCTGGTCTGGGTATGGAG
GTGATGCACGAGCGTCAAGTTGTGCTTTGCGCTCTTTAAATCGGATGAATTGCTGGAAACCCCAAGTGG
GCAATCAGCAGCCAAGTCTCAGATACATCTGAGAAAGGTTCAGAGACTACCTGAGGGATATAGTTCCCT
TAATAACAGGAATAAGCGTCCGACACCAGAAATGGTGATGATATAGTCCAATCCTGGTAGTAATACCAG
ATAGTTAAGGGAAGTTTAAGAATGCACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAGCAACACCAG
TTGCCTTGACTGCACCCGCAATCGGTTGATATAATGTTAAGGAACTCTTCGGAGTTCCTTTTTTTATAA
ATATTTAAGCACGAAAGAAAACACGAATGACTAAACTATACTCTGATTTGTATAGAACTTGTATGACTT
GTGGGGAGGAAAAACTTTCTACAGAGTTTTATGTTCGCAATAAAAAAACAGGAGTTCGTCATTCTTCTT
GTAAAGAATGTGATAAAGTGCGAGTGAAATCAAGACACAAAGAAAATCCAGAACGCACAAAAAATAATG
ACTTAAAAAGACTTTATGGTATTACTCTTGACGAACATACGCAAATGTATGAGGAGCAGAATGGAGTTT
GTGCTATTTGTAAAGGAGAAGGTGATGGTAAATGGAAAAAACTTTGCGTAGATCATGACCATGAAACTG
GTAAAGTTCGTCAGTTGCTTTGTAGGAACTGCAATATGATGTTGGGTCAAGTAAATGATAATGTTAATT
TACTTTCTGAAATGATTAAATACTTAAAGAGGTATCAATGAAACCCAGTTGACATAAGCCTGTTCGGTT
CGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGT
GGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCA
TCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTCATCAGC
GGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAACCTTGG
GGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAGATGGGCC
ACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCTCGTGGTC
AGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCTGA
CACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCACA
GGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCGCCCGTGT
CCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGCATCTTGC
CGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGT
ACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGA
GAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACATTGT
GGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGGGATGCGA
AGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAAGCCGACG
CCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTG
CACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCA
TC

Appendix 12: Products of PCR confirmation of insertion of the intronless *psbAII* construct into WT *Synechocystis* sp. PCC6803. As all colonies were the same sequence, only one is provided here as an example.

psbA_6803M_P1 primer pair product.

GATAGGAGCCATCTTGCCCCGTGATGCCTGTCAGCAAAACAACTTTAGACTTTGACATTAGTTAATTTT
TCCCCATTGCCCCAAAATACATCCCCCTAAAAATATCAGAATCCTTGCCCAGATGCAGGCCTTCTGGCG
ATCGCCATGGTGAGCAACGATTGCGGCTTTAGCGTTCCAGTGGATATTTGCTGGGGGTTAATGAAACAT
TGTGGCGGAACCCAGGGACAATGTGACCAAAAAATTCAGGGATATCAATAAGTATTAGGTATATGGATC
ATAATTGTATGCCCGACTATTGCTTAAACTGACTGACCACTGACCTTAAGAGTAATGGCGTGCAAGGCC
CAGTGATCAATTTCATTATTTTTCATTATTTCATCTCCATTGTCCCTGAAAATCAGTTGTGTCGCCCCT
CTACACAGCCCAGAACTATGGTAAAGGCGCACGAAAAACCGCCAGGTAAACTCTTCTCAACCCCCAAAA
CGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTATAAGCTTC
GTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAG
TTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAATGACTGCATCCATCGCTCAACAG

psbA_6803M_P2 primer pair product.

GATATCGACCCAAGTACCGCCACCTAAAATGTCTAACAATGCTTCAAGCCGACGCCGCTTCGGCGCGGC
TTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTCCTTCAATG
GGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATCGATCGCCGGGGAG
TCCGGGGCAGTTACCATTAGAGAGTCTAGAGAATTAATCCATCTTCGATAGAGGAATTATGGGGGAAGA
ACCTGTGCCGGCGGATAAAGCATTAGGCAAGAAATTCAAGAAAAAAAAATGCCTCCTGGAGCATTGAAGA
AAGCGAAGCTCTGTACCGGGTTGAGGCCTGGGGGGCACCTTATTTTGCCATTAATGCCGCTGGTAACAT
AACCGTCTCTCCCAACGGCGATCGGGGCGGTTCGTTAGATTTGTTGGAACTGGTGGAAGCCCTGCGGCA
AAGAAAGCTCGGCTTACCCCTATTAATTCGTTTTTCCGATATTTTGGCCGATCGCCTAGAGCGATTGAA
TAGTTGTTTTGCCAAGGCGATCGCCCGTTACAATTACCCCAAACACCTATCAGGCGGTTTATCC

Appendix 13: PCR confirmation of the DA1:DA2Em:DA3-H6 strain construct insertion.

pMK-T Apramycin Control ApmR construct, insertion product.

GCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTA
GTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAAACCCAGTTGACATAAGCCTG
TTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACG
CAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCC
TCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGT
CATCAGCGGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCA
ACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAG
ATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCT
CGTGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTG
TCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGG
GGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCG
CCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGC
ATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCA
AGTTGGTACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGC
TCAAGGAGAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCG
ACATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGG
GATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAA
GCCGACGCCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTG
CAAATTGCACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTG
ATTGCCATC

pOK-RQ S-PM2d Intronless *psbA* ApmR construct, insertion product.

ACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAG
GAATTATAACCAAATGACTGCATCCATCGCTCAACAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGA

GTGGGTTACCAGCACCGACAACCGCCTCTATGTGGGTTGGTTTGGAACGCTGATGATTCCTACTCTTCT
CGCAGCTGCTATCTGCTTTATTGTTGCTTTCATCGCTGCACCTCCCGTCGATATTGACGGCATCCGCGA
ACCTGTTGCTGGTTCTCTAATGTACGGAAATAACATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGC
CATCGGCCTTCACTTCTATCCCATCTGGGAAGCTGCTTCTCTTGATGAGTGGCTATATAATGGTGGTCC
TTATCAACTAGTGGTCTTCCACTTCCTTATCGGTGTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTC
TTACCGACTGGGTATGCGCCCTTGGATTTGTGTCGCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGT
TTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTCAGTGATGGTATGCCGCTTGGCATTTCTGGAACTTT
TAACTACATGCTTGTTTTCCAGGCAGAACACAACATTCTCATGCATCCTTTCCACATGTTGGGAGTTGC
TGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATGCATGGATCTCTTGTCACCTCTAGTCTTGTACGTGA
GACGACTGAGGTAGAATCCCAGAACTATGGTTACAAGTTCGGTCAAGAAGAAGAGACTTACAACATTGT
TGCTGCTCATGGTTACTTCGGTCGTCTGATCTTCCAATATGCGTCGTTCAACAACTCTCGTTCGCTACA
CTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATCTGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTT
CAACCTCAATGGATTCAACTTCAACCAGTCCATTGTATCTTCTGAAGGTCGTGTACTCAACACCTGGGC
AGATGTGCTGAATCGTGCTGGTCTGGGTATGGAGGTGATGCACGAGCGTATGCACACAACTTCCCTCTT
GACCTTGCAGCAGCAGAAGCAACACCAGTTGCCTTGACTGCACCCGCAATCGGTTGAAACCCAGTTGAC
ATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTT
GACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACA
GTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAG
CAACGATGTCATCAGCGGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTC
AGCTTCTCAACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGC
CCCTCGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCG
TCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACC
TTGGAGTTGTCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTG
CGGCAGCGGGGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAA
GCCCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACA
CGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTC
AGGATGGCAAGTTGGTACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGG
ACAGGTGGCTCAAGGAGAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCC
GCTCCCGCGACATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGC
CAGAGGCGGGATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACA
ATGCTTCAAGCCGACGCCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTG
AATAATCTGCAAATTGCACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTG
CTGATCTTGATTGCCATC

pMK-RQ S-PM2d *psbA* Entire Region ApmR construct, insertion product.

GCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTA
GTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAAATGACTGCATCCATCGCTCAA
CAGCGTGGTAGCAACACTTGGGAACAGTTCTGCGAGTGGGTTACCAGCACCGACAACCGCCTCTATGTG
GGTTGGTTTGGAACGCTGATGATTCCTACTCTTCTCGCAGCTGCTATCTGCTTTATTGTTGCTTTCATC
GCTGCACCTCCCGTCGATATTGACGGCATCCGCGAACCTGTTGCTGGTTCTCTAATGTACGGAAATAAC
ATCATCTCTGGTGCTGTTATCCCTTCTAGCAACGCCATCGGCCTTCACTTCTATCCCATCTGGGAAGCT
GCTTCTCTTGATGAGTGGCTATATAATGGTGGTCCTTATCAACTAGTGGTCTTCCACTTCCTTATCGGT
GTCTTCTCTTACATGGGTCGCGAGTGGGAACTCTCTTACCGACTGGGTATGCGCCCTTGGATTTGTGTC
GCTTACTCAGCCCCTGTTGCCGCTGCTACTGCTGTTTTCCTGGTCTACCCCTTTGGACAAGGTTCTTTC
AGTGATGGTATGCCGCTTGGCATTTCTGGAACTTTTAACTACATGCTTGTTTTCCAGGCAGAACACAAC
ATTCTCATGCATCCTTTCCACATGTTGGGAGTTGCTGGGGTCTTCGGTGGTTCTCTGTTCTCTGCTATG
CATGGATCTCTTGTCACCTCTAGTCTTGTACGTGAGACGACTGAGGTAGAATCCCAGAACTATGGTTAC
AAGTTCGGTCAAGAAGAAGAGACTTACAACATTGTTGCTGCTCATGGTTACTTCGGTCGTCTGATCTTC
CAATATGCGTCGTTCAACAACTCTCGTTCGCTACACTTTTTCCTTGCTGCTTGGCCTGTTGTCGGTATC
TGGTTTGCTGCCCTGGGCGTATCTACGATGGCGTTCAACCTCAATGGATTCAACTTCAACCAGTCCATT
GTATCTTCTGAAGGTCGTGTACTCAACACCTGGGCAGATGTGCTGAATCGTGCTGGTCTGGGTATGGAG
GTGATGCACGAGCGTCAAGTTGTGCTTTGCGCTCTTTAAATCGGATGAATTGCTGGAAACCCCAAGTGG
GCAATCAGCAGCCAAGTCTCAGATACATCTGAGAAAGGTTCAGAGACTACCTGAGGGATATAGTTCCCT
TAATAACAGGAATAAGCGTCCGACACCAGAAATGGTGATGATATAGTCCAATCCTGGTAGTAATACCAG
ATAGTTAAGGGAAGTTTAAGAATGCACACAACTTCCCTCTTGACCTTGCAGCAGCAGAAGCAACACCAG
TTGCCTTGACTGCACCCGCAATCGGTTGATATAATGTTAAGGAACTCTTCGGAGTTCCTTTTTTTATAA
ATATTTAAGCACGAAAGAAAACACGAATGACTAAACTATACTCTGATTTGTATAGAACTTGTATGACTT
GTGGGGAGGAAAAACTTTCTACAGAGTTTTATGTTCGCAATAAAAAAACAGGAGTTCGTCATTCTTCTT
GTAAAGAATGTGATAAAGTGCGAGTGAAATCAAGACACAAAGAAAATCCAGAACGCACAAAAAATAATG

ACTTAAAAAGACTTTATGGTATTACTCTTGACGAACATACGCAAATGTATGAGGAGCAGAATGGAGTTT
GTGCTATTTGTAAAGGAGAAGGTGATGGTAAATGGAAAAAACTTTGCGTAGATCATGACCATGAAACTG
GTAAAGTTCGTCAGTTGCTTTGTAGGAACTGCAATATGATGTTGGGTCAAGTAAATGATAATGTTAATT
TACTTTCTGAAATGATTAAATACTTAAAGAGGTATCAATGAAACCCAGTTGACATAAGCCTGTTCGGTT
CGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGT
GGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCA
TCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTCATCAGC
GGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAACCTTGG
GGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAGATGGGCC
ACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCTCGTGGTC
AGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCTGA
CACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCACA
GGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGCCCGGTCGCCCGTGT
CCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGCATCTTGC
CGAGTTGATGGCAAAGGTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGT
ACGCGTCGATTATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGA
GAAGAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACATTGT
GGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGGGATGCGA
AGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTCATAAAATGTCTAACAATGCTTCAAGCCGACG
CCGCTTCGGCGCGGCTTAACTCAAGCGTTAGTTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTG
CACTCTCCTTCAATGGGGGGTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCA
TC