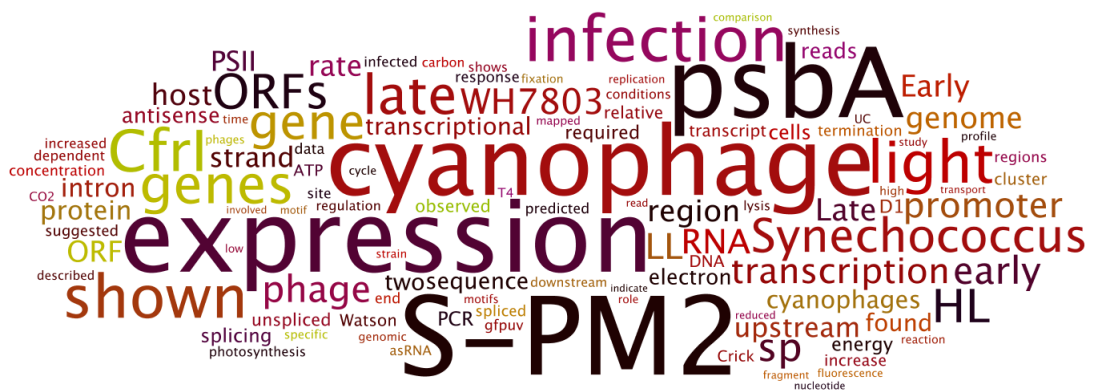# THE ROLE OF LIGHT IN PHOTOSYNTHETIC CYANOPHAGES: FROM PHYSIOLOGY TO GENE EXPRESSION

Richard John Puxty

A thesis submitted for the degree of Doctor of Philosophy

School of Life Sciences, University of Warwick
May 2014

Contents

Acknowledgements

I would like to thank my supervisors Prof. David Scanlan and Prof. David Evans for all of their kind advice and encouragement throughout this PhD. To my PhD committee Prof. David Hodgson, Dr. Kevin Purdy and Dr. Hendrik Schäfer whose discussions have kept me pointing in the right direction. I am indebted to Dr. Andrew Millard for discussion of ideas, technical assistance and his 'tireless optimism'.

I would also like to thank all past and current members of C126, especially Joseph, Fran, Carolina, Blanca, Branko and Monica for all their support and encouragement both in and out of the lab.

Thanks go to the Wolfman, Badger, Stag and Batty, whose friendships have made this PhD an enjoyable experience.

This thesis is dedicated to my family and friends who have always encouraged me and without their support, this would not be possible. Finally I would like to thank Lauren, whose ability to withstand endless one-way conversations about cyanophages is truly remarkable.

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

- Calculation of RNA minimum folding energies presented in Chapter 4. Produced by Prof. David Evans, University of Warwick.
- Sequencing of the PCR fragment amplified from the deletion region of S-PM2 presented in Chapter 4. PCR product was generated and sent for sequencing by Branko Rihtman. University of Warwick.

Parts of this thesis have been published by the author:

Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Shedding new light on viral photosynthesis. *Photosynthesis Research,* submitted March 2014.

Summary

It is estimated that there are approximately 10^{30} ocean virioplankton (Suttle 2007; Parsons et al. 2012). A large component of the oceanic viriosphere are the cyanophages, viruses that specifically infect cyanobacteria. Recent advances in genomics has revealed such viruses encode a multitude of genes, often acquired horizontally, that act to redirect metabolism for their own gains (Mann et al. 2003; Lindell et al. 2004a; Millard et al. 2009; Sullivan et al. 2010; Hurwitz et al. 2013; Enav et al. 2014). These genes have been named auxiliary metabolic genes (AMGs). They include multiple subunits of complexes involved with photosynthetic electron transport (PET) and CO_2 fixation (Mann et al. 2003; Lindell et al. 2004; Millard et al. 2009; Sullivan et al. 2010; Thompson et al. 2011; Puxty et al. *submitted*), leading to the hypothesis that cyanophages directly participate in photosynthesis to provide carbon and energy for their own replication.

Cyanophages face a dynamically changing light environment during their rather lengthy infection cycles ~12hrs. Therefore, it was hypothesised that changes in light intensity may affect the physiology of phage infection in terms of photosynthesis, CO_2 fixation and infection dynamics. During infection of the marine cyanobacterium *Synechococcus* sp. WH7803 with the well characterised cyanophage S-PM2 I show that decoupling of the photochemical and CO_2 fixation reactions of photosynthesis occurs (Chapter 3), which presumably redirects metabolism towards energy generation and away from growth. Moreover, S-PM2 acts to modify the PET which results in improved functioning of PSII at HL. The result is that the lytic cycle is significantly shortened during infection of the *Synechococcus* host under HL compared with low light (LL) conditions. To understand whether this early lysis is a regulated process, whole transcriptome sequencing of S-PM2 was performed in HL and LL (Chapter 5). This revealed a general increase in expression of all genes in HL but only the cyanophage *psbA* gene was significantly up-regulated above this background. This AMG encodes a core complex of photosystem II (PSII) of the PET and therefore plays a vital role in supplying energy through photophosphorylation. It is concluded that light poses a metabolic constraint on cyanophage development that requires large amounts of energy for synthesis and assembly of the structural components of the virion. Cyanophages have therefore acquired and evolved coordinated expression of PSII genes to maintain this supply of energy.

I further hypothesise that gene expression may pose a significant barrier in the acquisition of AMGs from their host due to incompatible gene regulation. To test this, the phage transcriptome was analysed (Chapter 4) to validate the model of temporal transcriptional regulation in cyanophage S-PM2 as previously proposed by comparison to enterobacteriophage T4. It is shown that the experimental data is largely congruent with the proposed model. This also revealed unpredicted characteristics of the transcriptome, including genome wide transcriptional read-through and antisense expression. It is suggested that this is facilitated by either inefficient transcriptional termination or pervasive transcription initiation and may be a biologically relevant process that allows for moderate expression of recently acquired genes. In addition, genome-wide antisense transcription may act to regulate the inventory or temporal expression of specific mRNAs in these regulatory limited phages. Attempts were therefore made to characterise a previously detected non-coding RNA (ncRNA) antisense to the light regulated S-PM2 *psbA* gene (Chapter 6). A model is proposed suggesting that the asRNA may act to tweak *psbA* expression under LL conditions to prevent accumulation of unnecessary PSII proteins. This mechanism has an interesting effect on the rate of splicing of a group I intron encoded by the *psbA* gene.

This study provides an important leap forward in our understanding of the factors that regulate the infection dynamics and therefore ecology of cyanophages. In so doing it also reveals transcriptional constraints and adaptations that go some way to explaining the evolution of cyanophage genomes.   .

List of abbreviations

ADP	Adenosine DiPhosphate	MWCO	Molecular Weight Cut Off
AMG	Auxiliary Metabolic Gene	NADP	Nicotinamide Adenine Dinucleotide Phosphate
asRNA	AntiSense RNA	NCBI	National Centre for Biotechnology Information
ASW	Artificial Sea Water	ncRNA	Non Coding RNA
ATP	Adenosine TriPhosphate	OEC	Oxygen Evolving Complex
BAM	Binary Alignment/Map format	PAR	Photosynthetically Available Radiation
BLAST	Basic Local Alignment Search Tool	PCR	Polymerase Chain Reaction
CM	Cytoplasmic Membrane	PCR	Polymerase Chain Reaction
CPM	Counts Per Minute	PET	Photosynthetic Electron Transport
C_T	Crossing Threshold	PM	Photomultiplier
Cyt b_6f	Cytochrome b6f complex	PMF	Proton Motive Force
DCMU	3-(3,4-dichlorophenyl)-1,1-dimethylurea	POC	Particulate Organic Carbon
DIC	Dissolved Inorganic Carbon	PPP	Pentose Phosphate Pathway
DNA	DeoxyRibonucleic Acid	PQ	PlastoQuinone
F_0	basal fluorescence	PSI	Photosystem I
FC	Fold Change	PSII	Photosystem II
FD	Ferredoxin	PTOX	Plastoquinol Terminal Oxidase
FIMO	Find Individual Motif Occurrences	PWSM	Position Weighted Score Matrix
FISH	Fluorescent In Situ Hybridisation	QA	Quinone binding site A
F_M	maximal fluorescence	QB	Quinone binding site B
FNR	Ferredoxin NADP Reductase	RACE	Rapid Amplification of CDNA Ends
F_V	variable fluorescence	RBS	Ribosome Binding Site
GCN	Genome Copy Number	REST	Relative Expression Software Tool
GFP	Green Fluorescent Protein	RNA	RiboNucleic Acid
GO	Gene Ontology	RNAP	RNA Polymerase
GOS	Global Ocean Survey	RNA-SEQ	RNA Sequencing
GSP	Gene Specific Primer	RPKM	Reads Per Kilobase of exon per Million reads
GTP	Guanosine TriPhosphate	RSS	Read Start Site
HGT	Horizontal Gene Transfer	RT-GSP	Reverse Transcription GSP
HL	High Light	SAM	Sequence Alignment/Map format
HLIP	High Light Inducible Polypeptide	TA	Toxin-Antitoxin
HMW	High Molecular Weight	TAP	Tobacco Acid Pyrophosphatase
LCA	Last Common Ancestor	TEM	Transmission Electron Microscopy
LL	Low Light	TMD	TransMembrane Domain
LOD	Limit Of Detection	TSS	Transcription Start Site
MEME	Multiple Em for Motif Elicitation	UC	UnClustered
MFE	Mean Folding Energy	UTR	UnTranslated Region
MFED	Mean Folding Energy Differential	VBR	Virus Bacteria Ratio
MIDAS	Metal Ion Dependent Adhesion Site	VWF	van Wilibrand Factor
MPN	Most Probable Number	WT	Wild Type

List of figures

List of tables

List of appendices

Chapter One: Introduction

1.1. Marine Cyanobacteria and the global Ocean

The Ocean is responsible for the approximately half of the fixed carbon in the biosphere (Field 1998). Recent estimates suggest that marine Cyanobacteria may be responsible for between 8.5-25 % of the net Oceanic CO_2 fixation (Flombaum et al. 2013) and up to 65 % in some Oceanic provinces (Jardillier et al. 2010). The marine Cyanobacteria are primarily composed of the genera *Prochlorococcus* and *Synechococcus*. Together, these genera are the most numerous photoautotrophs on our planet ($\sim7\pm0.3$ x 10^{26} and 2.9 ± 0.1 x 10^{27} cells respectively Flombaum et al. (2013)). *Prochlorococcus* has a limited geographic range, constrained between 40˚N and 40˚S (Johnson et al. 2006) and being outcompeted in high nutrient waters (Partensky et al. 1999). Thus this genus dominates the largest ecosystem on Earth, the Oceanic gyres (Zwirglmaier et al. 2008). In comparison *Synechococcus* has a much wider geographic range (Scanlan et al. 2009) extending to polar waters (Vincent et al. 2000) and high nutrient coastal waters (Martin et al. 2005). Molecular ecological studies have revealed distinct niche partitioning in both genera (Scanlan and West 2002; Johnson et al. 2006; Bouman et al. 2006; Six et al. 2007; Zwirglmaier et al. 2007; Zwirglmaier et al. 2008). The factors that control both the abundance, distribution and composition of marine picocyanobacterial communities include both abiotic (light (Six et al. 2007), temperature (Zinser et al. 2007), nutrient availability (Johnson et al. 2006)) and biotic factors (grazing (Liu et al. 1994; Baudoux et al. 2008; Zwirglmaier et al. 2009), viral lysis (Proctor and Fuhrman 1990; Suttle 2005)). The estimation of the contribution of biotic factors and in particular viral lysis is poorly constrained (Suttle and Chan 1993; Waterbury and Valois 1993; Wommack and Colwell 2000).

1.2. The role of viruses in marine ecosystems, biogeochemistry and evolution

1.2.1. Viral abundance

To begin to understand their importance, many studies have attempted to quantify the abundance of virus-like particles (VLPs) in time and space in marine environments. In general, VLPs are most abundant in the euphotic zone and decrease exponentially with depth below the surface mixed layer (Boehme et al. 1993; Suttle 2007; Parsons et al. 2012). In addition, surface abundances vary horizontally, with concentrations greater in coastal waters and decreasing towards open ocean environments (Cochlan et al. 1993; Marchant et al. 2000; Culley and Welschmeyer 2002). Moreover, VLPs vary over temporal scales (Parsons et al. 2012; Clasen et al. 2013) and are largely correlated with prokaryotic biomass. The mechanisms that drive the spatio-temporal variability in VLPs are largely unknown and likely differ between specific viral families. For instance, the transient blooms of the coccolith bearing *Emiliania huxleyi* in shelf seas are thought to be terminated by abundant lytic coccolithoviruses of the family *Phycodnaviridae* (Jacquet et al. 2002; Frada et al. 2008; Bidle and Vardi 2011). As such their abundance is largely scaled to that of their bloom forming host with viral-host dynamics largely following a predator-prey model. In comparison, marine *Synechococcus* and *Prochlorococcus* that dominate the photoautotrophic component of open ocean communities (Flombaum et al. 2013), are frequently found to co-occur with their respective phages in high numbers (Waterbury and Valois 1993; Suttle 2005; Millard et al. 2006; Avrani et al. 2011; Parsons et al. 2012; Clasen et al. 2013). Thus, the abundance and, by association, the function of marine viruses may be specifically related to the ecosystem in question. The mechanisms and consequences of this are described below.

1.2.2 The impact of viruses on the marine carbon cycle

Life on this planet is facilitated and sustained by a series of redox reactions (transfer of electrons and protons) between comparatively few elements (Falkowski et al. 2008, Fig.

1.1). These reactions, catalysed mainly by microbes and, in addition to the geochemical acid/base reactions, drive the biogeochemical cycles of Earth. Since electrons are never produced or consumed from the system, redox reactions are paired such that the forward reaction from one pathway is complemented by the reverse of another (Falkowski and Godfrey 2008). This is most ably demonstrated in the balance of photosynthesis and respiration that together mediate the biospheric fluxes of CO_2 and O_2. On a planetary scale, in the order of millions of years, the O_2 concentration is largely unchanged (Bender and Sowers 1994; Falkowski et al. 2005; Falkowski and Godfrey 2008). Yet O_2 has accumulated in the biosphere and, since it's production by oxygenic photosynthesis, has undergone many fluctuations (Falkowski et al. 2005; Holland 2006). This is related to the flux of organic carbon export to deep water (biological pump), the subsequent burial of the reductant (Hedges and Keil 1995; Aller 1998) and the action of the Wilson Cycle (Wilson 1965).
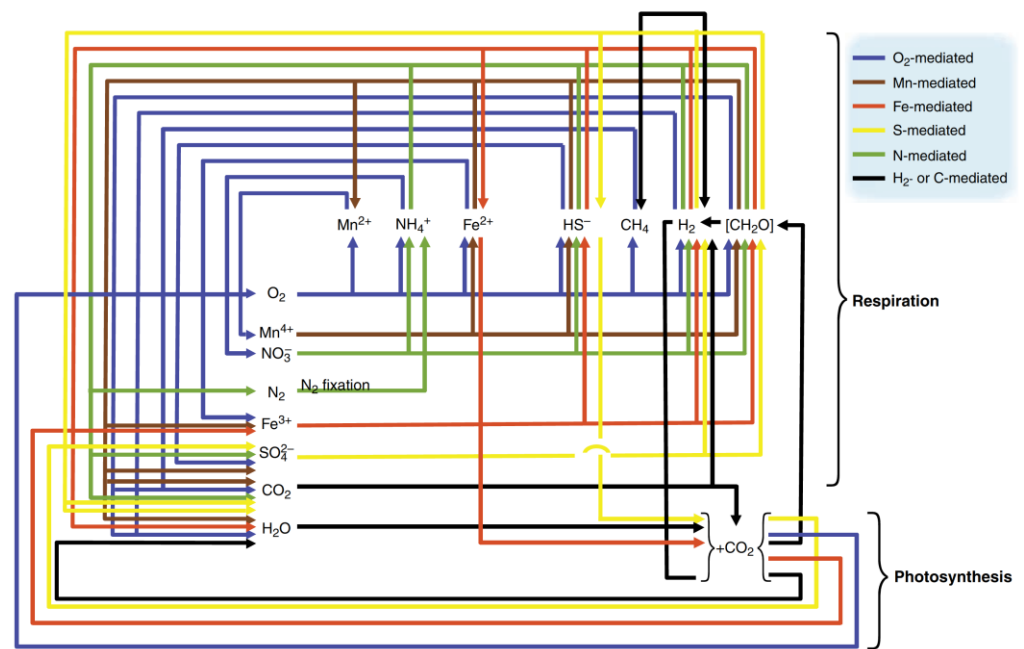
Figure 1.1: Electron transfer network of global biologically mediated cycles of hydrogen, carbon, nitrogen, oxygen, sulfur, and iron. Taken from Falkowski et al. (2008).

The majority of carbon export occurs in shelf seas and tropical upwellings (Siegel et al. 2014) compared with the open ocean, and in particular mid-ocean gyres. Yet mid-ocean gyres (defined as not exceeding 0.07mg chl *a* m^{-3}), are the largest contiguous ecosystems on

the planet and occupy approximately 40% of the surface of Earth (Polovina et al. 2008). These regions have been shown to be expanding, most likely as a result of global climate change (Behrenfeld et al. 2006a; Polovina et al. 2008). In the sunlit, deep mixed surface layers of the ocean much speculation has arisen as to the balance of photosynthesis and respiration (del Giorgio et al. 1997; Williams 1998; Westberry et al. 2012; Duarte et al. 2013; Ducklow and Doney 2013; Williams et al. 2013). Viruses act to mediate the balance of photosynthesis and respiration through cell mortality. Mortality results in the release of organic carbon, the quantity and quality of which supports community respiration through uptake by heterotrophic consortia (Wilhelm and Suttle 1999; Suttle 2005; Suttle 2007; Breitbart 2012). This so-called 'viral shunt' decreases the availability of organic carbon to higher trophic levels and the biological pump. It has been estimated that the lysis of plankton by viruses causes the release of 150 Gt C yr^{-1}(Suttle 2005). To put this into context, the burning of fossil fuels generates 5 Gt C yr^{-1} (Suttle 2005). Since these early estimates global carbon emissions have increased to ~13 Gt yr^{-1} as of 2012 (EDGAR, http://www.sec.gov/edgar.shtml#.U2CWy_ldWSo). Thus, the assessment of the magnitude and sensitivity of viral infection in time and space is essential for understanding the biosphere's response to climate change (for review see Danovaro et al. 2011).

1.2.3 The role of viruses in structuring marine communities and evolution

Through the lysis of host cells, marine viruses shape the composition of the community and therefore diversity (Weinbauer and Rassoulzadegan 2003; Frada et al. 2008; Bidle and Vardi 2011; Martiny et al. 2014). Lytic-viral infection is a density-dependent process such that abundant genotypes are more likely to encounter a virus. This has led to the 'killing the winner' model (Martiny et al. 2014). Within this model genotype succession occurs through stepwise destruction of abundant genotypes and selection of rarer types (Thingstad 2000; Winter et al. 2010). Evidence for the functioning of this model in the environment is sparse. In the case of *E. huxleyi,* it is clear that blooms are frequently terminated by viral lysis (Jacquet et al. 2002; Frada et al. 2008). However, evidence to support this theory in the most

abundant prokaryotic picophytoplankton is lacking. Here, members of the genera *Prochlorococcus and Synechococcus* and their cyanophages are often found to coexist (Waterbury and Valois 1993) and their numbers are often positively correlated (Suttle et al. 1994; Parsons et al. 2012). The explanation most invoked to explain this phenomena is through selected resistance to cyanophages and subsequent overcoming of the resistance by cyanobacteria, subsequently driving antagonistic evolution and sustaining the virus-host 'arms race' (Waterbury and Valois 1993; Lennon et al. 2007; Stoddard et al. 2007; Marston et al. 2012; Martiny et al. 2014). Various studies have attempted to detect rapid fluctuations in marine cyanobacterial and cyanophage populations in terms of abundance and diversity. Molecular studies have proved most powerful and recently succession of genotypes has been observed in marine myoviral communities. (Marston and Sallee 2003; Mühling et al. 2005; Chow and Fuhrman 2012; Clasen et al. 2013). The mechanistic basis for these fluctuations is not fully understood but has been suggested to be related to availability of host genotypes. Recently, Marston et al. (2012) has demonstrated antagonistic coevolution in a single cyanophage-host system in chemostatic growth. Here, between 4 and 13 new viral and between 4 and 11 host phenotypes were evolved differing in host range and resistance range respectively.

1.3 The advent and functioning of photosynthesis

1.3.1 The oxygen cycle and origins of oxygenic photosynthesis

Today, oxygenic photosynthesis is the predominant endergonic process in the biosphere, providing virtually all organic matter to food webs (Field 1998; Falkowski and Raven 2007) as well as the electron donor (oxygen) for aerobic respiration (Falkowski et al. 2005). Yet, prior to ~2.5 Ga ago there was virtually no oxygen on our planet (Falkowski and Raven 2007). Primordial organisms relied on scarce electron donors like H_2S, NH_3, organic acids and Fe^{2+} (Barber 2008) to provide energy. The high positive reduction potential of oxygen means metabolism (energy generation) is approximately 20 times more efficient (Barber

2008). The transition between anaerobic and aerobic metabolism has been complemented by gradual oxidation of the biosphere and has allowed for evolution of modern eukaryotes and multi-cellularity (Falkowski et al. 2005). The complex responsible for the production of ~all free oxygen is photosystem II (PSII) of photoautotrophic organisms. The PSII complex is remarkably conserved structurally and at the sequence level between all photosynthetic organisms (Falkowski and Raven 2007) and strongly suggests that this is the only solution to its biological function: the oxidation of H_2O in the presence of light to produce H^+ and O_2. In oxygenic phototrophs, PSII acts in concert with photosystem I (PSI) and intersystem electron transporters (Fig. 1.2) to facilitate proton translocation across the thylakoid membrane. This generates the energy and reductant required for the fixation of inorganic carbon into chemical bond energy (Falkowski and Raven 2007). The complete photosynthetic electron transport (PET) chain comprises several macromolecular complexes (Fig. 1.2). How this machinery evolved is unclear but the best model suggests that large scale lateral gene transfer (LGT) occurred between a quinone-based reaction centre (PSII) from a purple non-sulphur bacterium and an iron-sulfur reaction centre (PSI) of a green sulphur bacterium (Blankenship 2001; Falkowski and Godfrey 2008; Blankenship 2010). Cyanophages may mediate such LGT events given that many contain genes involved in the PET (Mann et al. 2003; Lindell et al. 2004b; Zeidner et al. 2005; Sullivan et al. 2006; Sharon et al. 2009).
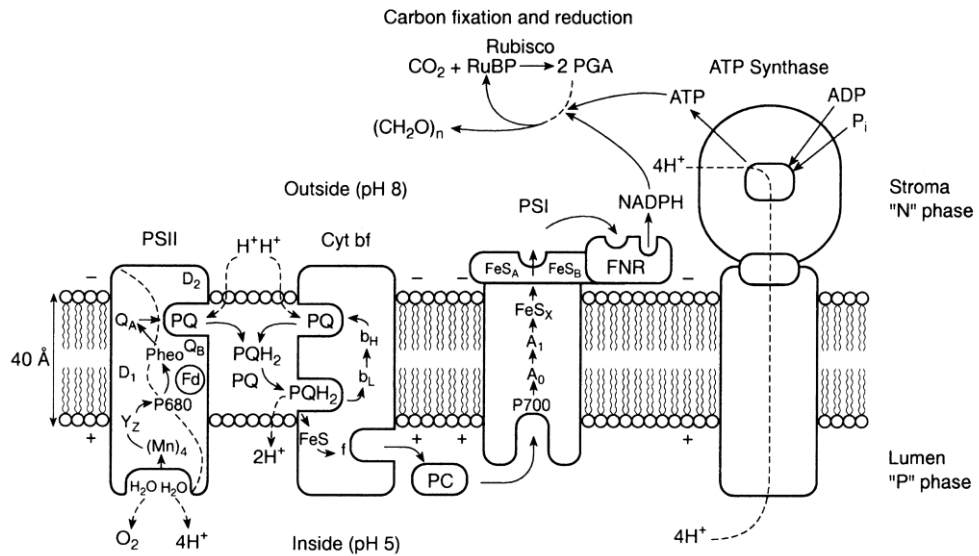
Figure 1.2: Major complexes involved in photosynthetic electron transport in the thylakoid membrane. Electron and proton transport routes are shown by solid and dashed arrows respectively. See section 1.2.2 for description of the process. Taken from Falkowski and Raven (2007).

1.3.2 The light and dark reactions of oxygenic photosynthesis

Photosynthesis is the light dependent conversion of CO_2 into bond energy (Falkowski and Raven 2007). It comprises two independent but related processes: the so-called light dependent reactions (photophosphorylation) and the dark reactions (Calvin cycle). Photophosphorylation provides, through light dependent electron transport, the ATP and NADPH which is subsequently used by the Calvin cycle for the conversion of CO_2 into sugars (Falkowski and Raven 2007). Photophosphorylation occurs in the thylakoid membrane and cyanobacteria are unique amongst prokaryotes containing both a thylakoid and cytoplasmic membrane (Vermaas 2001). The macromolecular complexes required for photophosphorylation are described in Fig. 1.2 whilst the detailed electron transfer kinetics between complexes are shown in Fig. 1.3. The energy for the process ultimately is derived from photons, which when harvested by antennae complexes of PSII and PSI form a quantum electronic state known as an exciton (Fassioli et al. 2014). The energy is transferred and used to reduce the special chlorophyll pair P680 of PSII (Fig. 1.3). P680 rapidly donates its electron to the primary electron acceptor, pheophytin and thence to the quinone binding residue Q_A. The electron hole that forms in P680$^+$ must be filled and therefore H_2O is used as

8

an electron donor. PSIII is the only enzyme that can catalyse this reaction which is thermodynamically extremely unfavourable (Falkowski and Raven 2007). The oxidation of H_2O not only provides electrons for PET but also extrudes H^+ into the thylakoid lumen. Acidification of the lumen is required for generation and maintenance of the proton motive force (PMF). The exact biophysical mechanism by which PSII oxidises water is yet to be elucidated but PSII binds 4 Mn atoms and at least one Ca atom thought to be essential for catalysis. P680 actually gets its electron from a tyrosine residue on the D1 protein (Debus et al. 1988), which in turn obtains its electrons from the 4 Mn atoms of the oxygen evolving complex (OEC). The generation of a synthetic PSII that catalyses this reaction promises to provide an unlimited source of clean fuel for future generations (Tachibana et al. 2012). Q_A rapidly passes the electron to Q_B, where it is used to reduce plastoquinone (PQ) (Fig. 1.3). The reduction of PQ to plastoquionol (PQH_2) requires two electrons from PSII and 2 H^+ sourced from the cytoplasm and in so doing translocates them to the lumen and hence contributes to the PMF. The slow reduction kinetics of PQ makes this the rate limiting step (Fig. 1.3). This property is exploited by measurements of fluorescence to infer PSII functioning as will be described below.
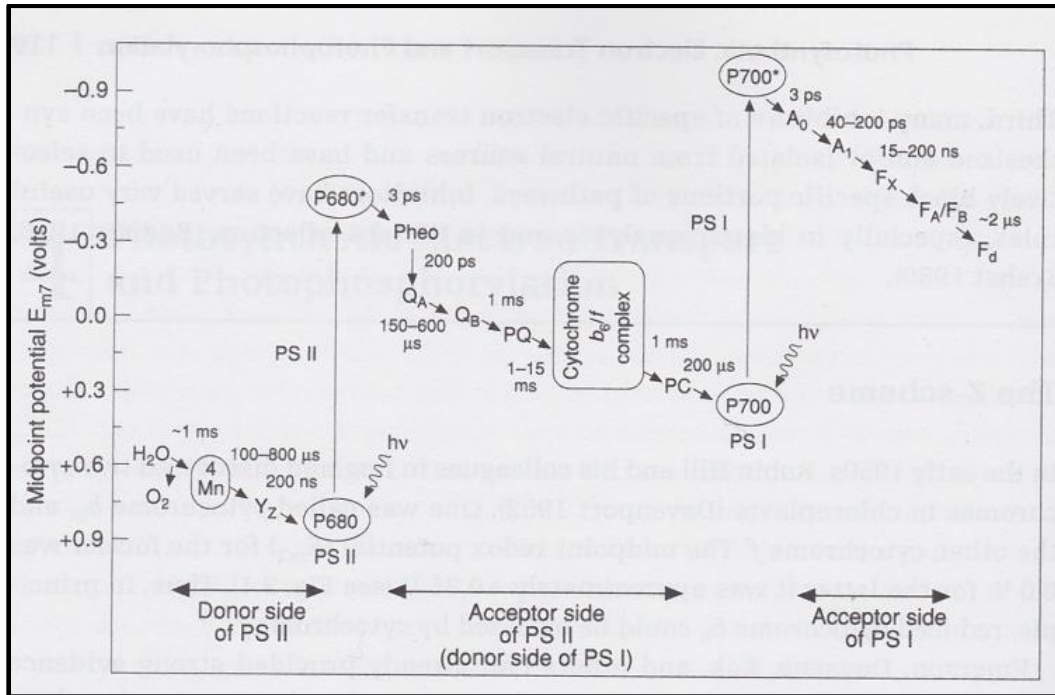
Figure 1.3: Z-scheme of photosynthetic electron transport. Description of electron transfer reactions are given in the text. Taken from Falkowski and Raven (2007).

The PSII reaction centre can be viewed as a H_2O-PQ oxidoreductase. When the PQ pool is completely reduced, both Q_B and Q_A cannot accept another electron from PSII. Therefore the reaction centre is said to be 'closed' (Fig. 1.4a). During dark periods the PQ pool is allowed to re-oxidise due to the activity of PSI hence Q_A becomes completely oxidised and the reaction centre is said to be 'open' (Fig. 1.4a). In the case where Q_A is completely reduced, $P680^+$ has to lose an electron to return to ground state. Since Q_A (photochemistry) is blocked, another pathway is available which allows the electron to be re-released as a photon in the process of fluorescence (Fig. 1.4b). The re-emitted photon has a longer wavelength due to Stokes-shift and therefore can be readily detected using a photomultiplier (PM). When PSII is open, the majority of electrons from PSII reduce Q_A and hence fluorescence yields are low (or quenched). Conversely when PSII is closed, Q_A cannot be reduced and so the majority of the electrons are released as fluorescence. Therefore by placing dark adapted cells in saturating irradiance a transient increase in fluorescence can be detected related to the sequential closing of reaction centres (Fig. 1.4c). This is known as the 'Kautsky' effect and the fluorescence yield (F_v) is an important measure of PSII functioning. For instance, as

relevant for this thesis, high light intensity induces a phenomenon known as photoinhibition that can be probed using variable fluorescence measurements. PSII is a multi-subunit complex where the core comprises the heterodimeric D1 and D2 proteins (Zouni et al. 2001; Umena et al. 2011). These are responsible for binding all the pigments and co-factors required for the primary charge separation and thus are susceptible to damage. D1 is far more susceptible and as such cyanobacteria have evolved an elaborate repair mechanism to continually replace damaged D1 polypeptides with *de novo* synthesised copies (Nixon et al. 2010). Thus, D1 is one of the fastest turned over proteins in the cell. D1 damage is directly proportional to the light intensity (Tyystjärvi and Aro 1996). When the rate of damage is greater than the maximal rate of the repair cycle, photoinhibition occurs. Photoinhibition manifests as a reduction in the quantum yield of PSII photochemistry (Tyystjärvi and Aro 1996; Bailey et al. 2004).
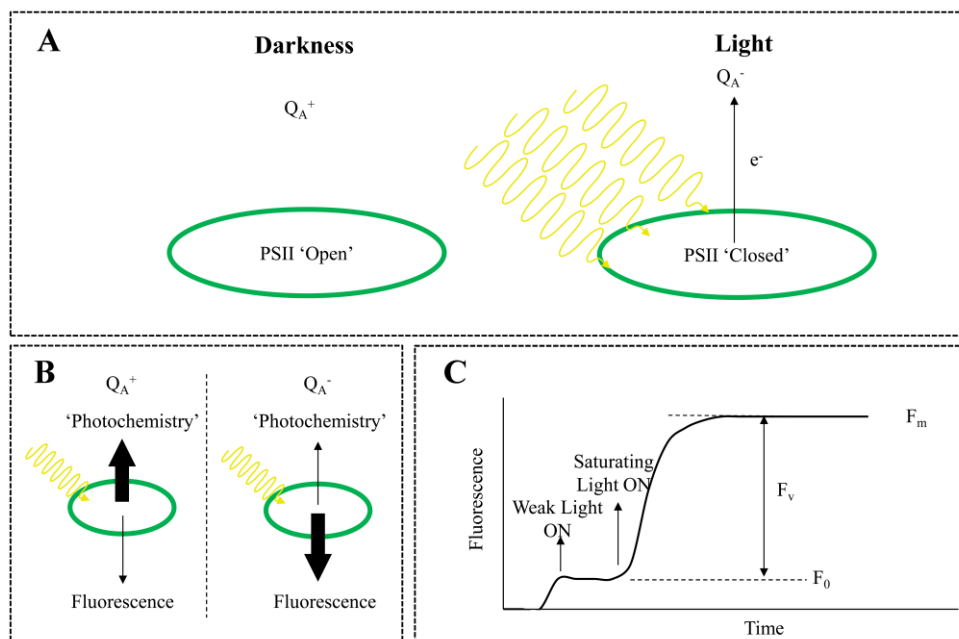
Figure 1.4: The relationship between PSII redox state and chlorophyll fluorescence. A. The effect of light on the redox state of PSII. During darkness QA is completely oxidised and so the reaction centre is 'open'. Upon illumination QA becomes reduced causing closure of the reaction centre. B. The relative balance of photochemical and fluorescence quenching pathways with QA in an oxidised state (QA+) and reduced state (QA-). C. Kautsky induction curve of chlorophyll a fluorescence. A weak light is switched on which causes a basal level of fluorescence (F0). Upon illumination with a saturating light causes sequential closing of reaction centres leading to the balance of excitation energy being shifted toward fluorescence. The maximal fluorescence (Fm) is achieved when QA is fully reduced.

PQH$_2$ donates electrons to the intersystem electron transporter complex cytochrome b_6f, which in turns donates electrons to a luminal mobile electron transporter. In cyanobacteria two proteins facilitate this role: the Cu containing plastocyanin and the iron containing cytochrome c$_6$. Plastocyanin is an evolutionary more recent form and its advent was concomitant with oxidation of iron as a result of oxygenic photosynthesis (Navarro et al. 1997). The mobile electron carrier donates electrons to PSI. PSI can be considered as a plastocyanin (or cytochrome c$_6$):ferredoxin oxidoreductase. PSI contains the reaction centre chlorophyll P700 which when reduced by light, fills its electron hole with electrons from plastocyanin or cytochrome c$_6$. P700$^+$ donates electrons extremely rapidly to ferredoxin (Fig. 1.3). Ferredoxin-NADP$^+$ reductase (FNR) then catalyses the reduction of NADP$^+$ to NADPH, consuming one H$^+$ from the cytoplasm.

Concomitantly, the generation of luminal protons by the PSII complex, the translocation of protons by PQ reduction and the consumption of cytoplasmic protons by the activity of FNR maintain a PMF that is sustained by both an electrochemical proton gradient ($\Delta\Psi$) and proton concentration (ΔpH) (Bailleul et al. 2010). Equilibration of the transmembrane potential is facilitated by the H$^+$ channel of the CF$_0$ subunit of the CF$_0$CF$_1$ complex of ATP synthase (Fig. 1.2). The CF$_1$ subunit acts as the ATP synthase phosphorylating ADP to ATP. Thus, photophosphorylation is an elegant example of Mitchell's chemiosmotic theory (Mitchell 1961). The light dependent reactions of photosynthesis provide both NADPH and ATP at a ratio of 2:3 (Falkowski and Raven 2007).

The Calvin cycle is the process of inorganic carbon fixation to carbohydrate and, elegantly, requires an NADPH:ATP quotient of 2:3 provided by photophosphorylation. The net reaction is:

$$CO_2 + 2NADPH + 2H^+ + 3ATP \rightarrow (CH_2O) + H_2O + 2NADP^+ + 3ADP + 3Pi \qquad \text{Eq. 1.1}$$
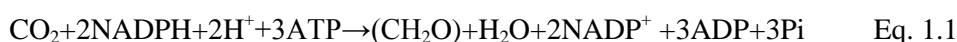
The transformations involved in this cycle are depicted in Fig. 1.6 and described in the legend. However, the cycle is composed of three stages: 1. The fixation step when

carboxylation of Ribulose-1,5-biphosphate (RuBP) with CO_2 occurs to form 3-phosphoglycerate (3-PGA). This reaction is catalysed by the enzyme ribulose-1,5-bisphophate carboxylase oxygenase (RuBisCO, EC 4.1.1.39). 2. The reduction step then catalyses the reduction to G3P. 3. The final stage is the regeneration of RuBP. Cyanobacteria are interesting as they share many components of the Calvin cycle with the pentose phosphate pathway (PPP) that generates reducing equivalents for reductive biosynthesis, as well as R5P as a precursor for nucleotide biosynthesis (Fig. 1.6; Thompson et al. 2011). Cyanophages may hi-jack this connectivity between the two pathways to support PPP over the Calvin cycle as discussed below (Thompson et al. 2011).

1.4 Marine cyanophages and auxiliary metabolic genes (AMGs)

Cyanophages are viruses specifically infecting cyanobacteria. Marine cyanophages were first isolated in the early 1990's (Suttle and Chan 1993; Waterbury and Valois 1993; Wilson et al. 1993a) and since then a genomic approach has dominated the study of cyanophage-host interactions (Sullivan et al. 2005; Mann et al. 2005; Millard et al. 2009; Sullivan et al. 2009; Sullivan et al. 2010; Huang et al. 2012; Labrie et al. 2013). This has culminated in the 71 sequenced cyanophage isolates that are publically available today (Puxty et al. *submitted*). This has led to the realisation that cyanophages frequently encode homologues of bacterial genes involved in metabolism and as such have been designated auxiliary metabolic genes (AMGs *sensu* Breitbart et al. 2007).

Marine cyanophages can be split into three morphological families: *Myoviridae*, *Podoviridae* and *Siphoviridae* (Fig. 1.5a, b, c). The three groups have clear genetic signatures with Myoviruses generally having large genomes and low G+C% (Fig. 1 5d). In comparison, podo- and siphoviruses have smaller genomes (Fig. 1.5d). Interestingly, podo- and siphoviruses have a variable G+C content that is related to the translational efficiency and codon usage of their cyanobacterial host (Bailly-Bechet et al. 2007; Enav et al. 2012). In comparison, Myoviruses act to alter the codon translational efficiency background of their hosts by encoding tRNA genes (Bailly-Bechet et al. 2007; Enav et al. 2012). This has further

led to the suggestion that Myoviruses have much broader host ranges than either podo- and siphoviruses (Sullivan et al. 2006; Enav et al. 2012).
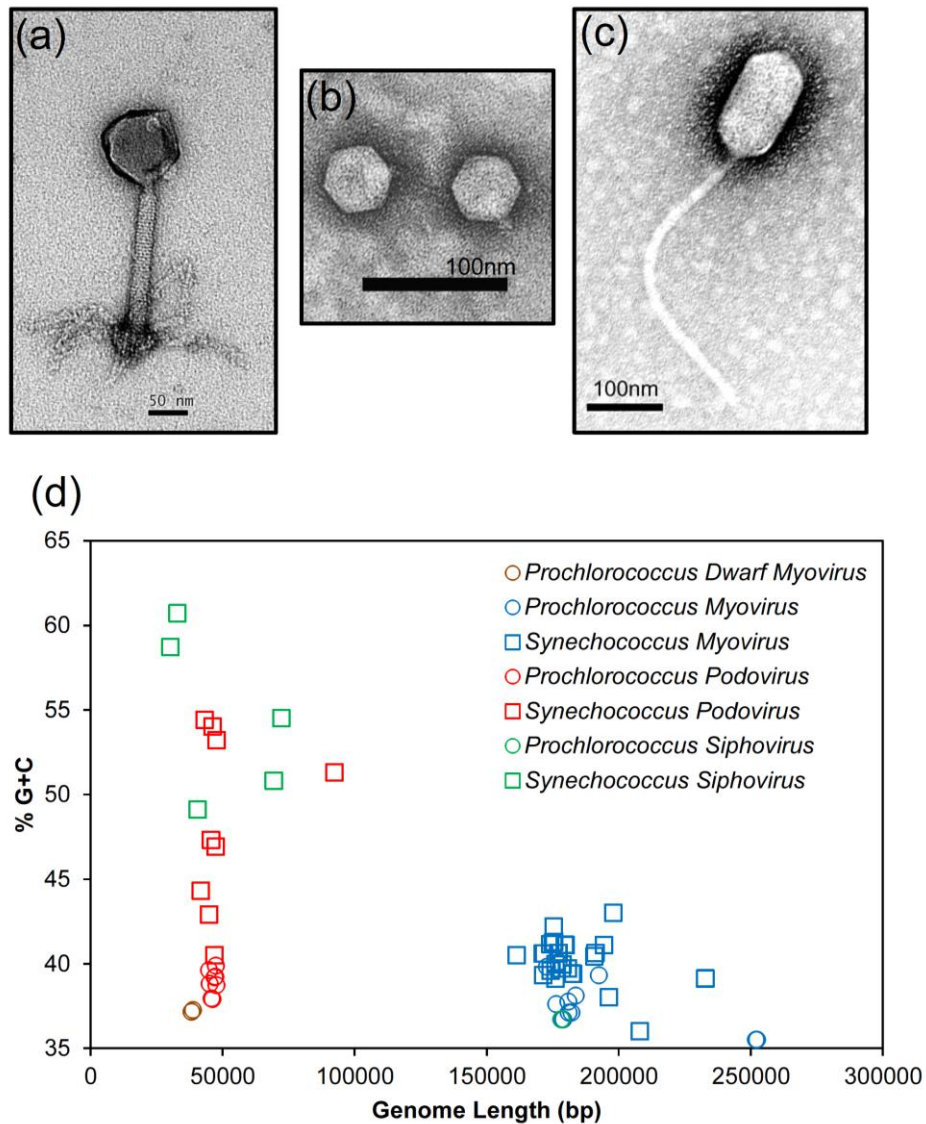
Figure 1.5: Morphological and genomic features of cyanophage families. (a) TEM image of T4-like Cyanomyovirus S-PM2 (b) TEM image of T7-like P-SSP7. (c) TEM image of Cyanosiphovirus P-SS2. TEM images (b) and (c) were taken by Ben Ni and used with permission from Prof. S.W. Chisholm (d) Cyanophage genome length versus nucleotide G+C%, showing separation of cyanophage families. Adapted from Puxty et al., submitted.

1.4.1 AMGs

The majority of the AMGs in the so far sequenced cyanophages are components of, or can be related to, the photosynthetic machinery. This includes light harvesting (Dammeyer et al.

2008; Busch et al. 2011), reaction centres and electron transfer (Mann et al. 2003; Lindell et al. 2004b) and carbon metabolism (Millard et al. 2009; Sullivan et al. 2010; Thompson et al. 2011). These are discussed further below. In addition, genes involved with phosphate acquisition are routinely found in cyanophages and are expressed (Sullivan et al. 2008; Zeng and Chisholm 2012). Genes involved with intracellular signalling (*mazG*) (Bryan et al. 2008), cobalamin biosynthesis (*cobS*) (Sullivan et al. 2010) and also small heat shock proteins (Sullivan et al. 2010; Ignacio-Espinoza and Sullivan 2012) are also routinely found. The functional relevance of these last three are yet to be established. Recently, Sharon et al. (2011) have revealed that the diversity of AMGs in sequenced cyanophages represents only a snapshot of the potential viral pool from metagenomic datasets. Here, 34 separate gene families involved in a variety of cellular processes were reported. In particular, many genes involved with anti-oxidation such as the peroxiredoxin family were found. In addition, many genes involved in translation or co/post translational modifications such as the ribosomal protein S21, translation initiation factor IF-1, phosphorylases and peptide deformylases (PDFs) are present (Sharon et al. 2011).

1.4.2 Cyanophage 'photosynthesis'

The majority of cyanophage encoded AMGs can be related to photosynthesis or carbon metabolism. It has been suggested that many of the AMGs function so as to increase energy (ATP) and reductant (NADPH) for nucleotide biosynthesis and therefore phage genome replication (Sharon et al. 2011; Philosof et al. 2011; Hurwitz et al. 2013; Puxty et al. *submitted*; Enav et al. 2014). A schematic of these genes and the hypothetical role they play in cyanophage infection is shown in Fig. 1.6. Of particular interest are genes involved in the formation of the reaction centres PSII and PSI. The *psbA* gene encoding the D1 protein of PSII was the first to be identified (Mann et al. 2003; Lindell et al. 2004; Millard et al. 2004). Both *psbA* and *psbD* (the latter encoding the D2 polypeptide of PSII) are widespread among cyanophages (Puxty et al. *submitted*). Only 14/71 sequenced cyanophages do not contain *psbA* and of these, 7 belong to the *Podoviridae* and 5 to the *Siphoviridae* (Puxty et al.

submitted). Therefore, all cyanomyoviruses contain the *psbA* gene with the exception of MED4-117 and MED4-184, which belong to the 'dwarf' myoviral family (Ackermann et al. 2011; Comeau et al. 2012).

The functions of the phage encoded *psbA* and *psbD* genes are yet to be explicitly proven, but due to sequence similarity, it is likely that they serve the same function as the host homologue as described above (Sullivan et al. 2006). Phage encoded *psbA* genes have been shown to be expressed during infection (Lindell et al. 2005; Clokie et al. 2006; Millard et al. 2010) and the accumulation of phage encoded *psbA* transcripts is inversely proportional to those of the degraded host *psbA* mRNAs. Moreover, phage encoded *psbA* mRNA is efficiently translated and yields proteins during infection (Lindell et al. 2005).

Incubation in the dark and inhibitors of electron transport downstream of PSII have been shown to reduce the rate of phage genome synthesis in the *psbA*-containing phage P-SSP7, infecting *Prochlorococcus* sp. MED4 (Lindell et al. 2005; Thompson et al. 2011), strongly suggesting that cyanophage development relies on continued photosynthesis (Lindell et al. 2005; Sullivan et al. 2006).

The most consensual view of the function of phage encoded *psbA* genes surrounds the idea of photoinhibition (Bailey et al. 2004). Photoinhibition occurs when the rate of photo-induced damage to the D1 polypeptide exceeds the rate at which *de novo* synthesis can replace it to maintain steady state photosynthesis (Nixon et al. 2010). Typically bacteriophages inhibit translation of host genes rapidly after infection (Miller et al. 2003). The D1 damage rate is significantly faster than the length of the lytic cycle for many cyanophages (Blot et al. 2011; Mella-Flores et al. 2012), with a half-life of less than one hour. In such circumstances it is suggested that an auxiliary source of D1 would be required to maintain PSII functioning. This may, in part, explain the distribution of *psbA* homologues in cyanophage genomes. In particular, podoviruses tend to have a much reduced latent period (defined as the time between entry of viral DNA into the host and the release of the

first progeny viruses detectable in a single step growth curve) (Lindell et al. 2005; Lindell et al. 2007). In particular, the latent period of the podovirus P60 is only 1 hr and this phage does not contain *psbA* (Sullivan et al. 2006).

Bragg and Chisholm (2008) and Hellweger (2009) have sought to model the consequences of maintaining a phage encoded copy of *psbA* with an explicit component for photoinhibition based on experimental data. In both papers, it was concluded that 'fitness benefits' were only found for infections occurring during high light compared with low light conditions. Perhaps, therefore, it is no coincidence that *Synechococcus* phages only encode the high light variant of D1 (Sullivan et al. 2006).

Interestingly, there are cyanophage specific residues within the D1 (Sharon et al., 2007). These residues occur at two locations. Firstly, the stromal loop between transmembrane helices D and E. D1 turnover requires the inactivation of D1 by a specific protease. In *Arabidopsis thaliana,* the DegP2 protease has been shown to be essential for this (Haussühl et al. 2001) and its primary cleavage site is the location of cyanophage specific residues. In cyanobacteria, an entirely different protease is required for the inactivation of D1. The FtsH protein was shown to be essential for maintaining normal rates of D1 degradation *in vivo* with no accumulation of D1 breakdown products observed in a *ftsH* mutant (Silva et al. 2003). FtsH processively degrades D1 form the N-terminus (Komenda et al. 2007) and thus it is unclear what role these stromal loop residues play.

Cyanophage specific residues also occur at the loop at the C-terminal end of transmembrane helix E. This region likely interacts with cytochrome b_{559} (*psbE/F*) and cytochrome c550 (Umena et al. 2011). Whilst the function of the cytochrome c550 complex is largely unknown (Suga et al. 2013), cytochrome b_{559} may act to scavenge electrons from reduced P680 to prevent photo-oxidation of the last antennae chlorophyll (Thompson and Brudvig 1988). As such, cyclic electron flow through cytochrome b_{559} may protect PSII from photoinhibition (Prasil et al. 1996).

To date, genes involved in PSI have not been found in cultured cyanophages. However, Sharon et al. (2009) have found a cluster of genes that are capable of forming a monomeric PSI from metagenomic datasets. These genes are likely of cyanophage origin due to the presence of cyanophage-like genes on the scaffolds. This cluster contains the genes *psaJF-C-A-B-K-E-D*. The genes required to produce a trimeric form of PSI: *psaI, psaL and psaM* (Xu et al. 1995; Naithani et al. 2000; Ben-Shem et al. 2003) are absent. Interestingly the cluster appears to encode a fusion protein between *psaJ* and *psaF*. This fusion results in a truncated C-terminus of PsaF. This region corresponds to the docking site of the electron donor, plastocyanin or cytochrome c_6 (Jordan et al. 2001; Sharon et al. 2009). It was therefore suggested that the phage encoded photosystem may accept electrons from an alternative respiratory donor. Hence, cyanophages may redirect metabolism from the respiratory chain through PSI to generate energy for replication. Recently, Mazor et al. (2014) have shown that the N-terminal region of PsaF is important for the topology of the luminal side of PSI (Mazor et al. 2014). In a mutant of *Synechocystis* sp. PCC6803 where the N-terminal region of PsaF was fused to PsaJ, the kinetics of P700 reduction by respiratory donors were faster compared with the wild type.

The 'rewiring' of the electron transport chain during infection is suggested to occur in other examples of cyanophage AMGs. For instance cyanophages encode a divergent form of plastocyanin (Puxty et al., *submitted*). These cyanophage homologues contain extensive modifications of the N-terminal signal peptide that targets plastocyanin to the thylakoid lumen and have a significantly reduced isoelectric points (Puxty et al., *submitted*). In addition, cyanophages contain conserved residues at sites that are important for catalysis and interaction with the electron donor cytochrome *f* which are absent from any cyanobacterial host (Puxty et al., *submitted*). Taken together it has been suggested that the phage plastocyanin donates electrons to the alternate terminal electron acceptor cytochrome oxidase (COX) instead of PSI (Puxty et al., *submitted*). When reduced, COX translocates 4 H^+ from the cytoplasm to the lumen, contributing to a PMF whilst freeing PSI to accept

electrons from cyclic photophosphorylation (Puxty et al., *submitted*). Cyclic photophosphorylation is the reduction of plastoquinol by reduced ferredoxin instead of $NADP^+$ (Battchikova et al. 2011). Therefore, the electron can cycle through PSI numerous times. The result is continued ATP synthesis at the expense of reductant generation. Cyanophages have been expected to bolster cyclic photophosphorylation around PSI during infection (Sharon et al. 2009, 2011; Philosof et al. 2011; Béjà et al. 2012; Mazor et al. 2012, 2014; Puxty et al., *submitted*). Sharon et al. (2009) showed that cyanophages can encode subunits of the type I NDH-1 complex, which functions in cyclic electron flow in cyanobacteria (Battchikova et al. 2011).

In addition, cyanophages frequently encode the PTOX gene (Millard et al. 2009; Sullivan et al. 2010; Puxty et al., *submitted*). PTOX is suggested to be a plastoquinol terminal oxidase that accepts electrons from PQH_2 and reduces O_2 to H_2O (Bailey et al. 2008). This pseudo water-water cycle acts as an electron safety valve preventing backflow of electrons to PSII under conditions where over-reduction of PQ occurs, e.g. high light or iron limitation (Bailey et al. 2008). The net effect is generation of a trans-thylakoid ΔpH, whilst inhibiting linear electron flow to ferredoxin-NADP(+) reductase (FNR) and concomitant reduction of NADP to NADPH. Thus, cyanophages may still gain energy for development from expressing the PTOX gene. Related is the fact that cyanophages also possess genes of the high-light inducible polypeptide (HLIP) family. While the function of these genes remain controversial (Xu et al. 2002; Havaux 2003; Xu et al. 2004a; Vavilin et al. 2007; Wang et al. 2008; Hernandez-Prieto et al. 2011), mutants lacking all four of the *hli* genes have been shown to be sensitive to high irradiances (He et al. 2001; Havaux 2003).

Recently, Thompson et al. (2011) has suggested that the occurrence of carbon metabolism genes acts to redirect metabolism away from the Calvin cycle and towards the PPP. In particular, an inhibitor of the Calvin cycle, CP12 is frequently found in cyanophages as well as the transaldolase TalC (Thompson et al. 2011). Of the shared PPP and Calvin cycle genes only TalC functions in the direction of the PPP and against the Calvin cycle. Cyanophages

also contain the *zwf* and *gnd* genes encoding glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase, respectively. These two enzymes are key components of the PPP and responsible for reduction of NADP to NADPH. The net result of expression of these genes has been suggested to be the production of $NADPH^+$ and Ru5P at the expense of CO_2 fixation. Both are suggested to enhance the kinetics of nucleotide biosynthesis during infection (Thompson et al. 2011).

The range of hypothetical modifications of the PET and CO_2 fixation machinery during infection are depicted in Fig. 1.6, yet to date many of these hypotheses remain to be tested. A major goal of this thesis is to provide evidence to support the ideas that cyanophages modify the PET to provide increased electron capacity in high light as well as the redirection of metabolism away from CO_2 fixation.
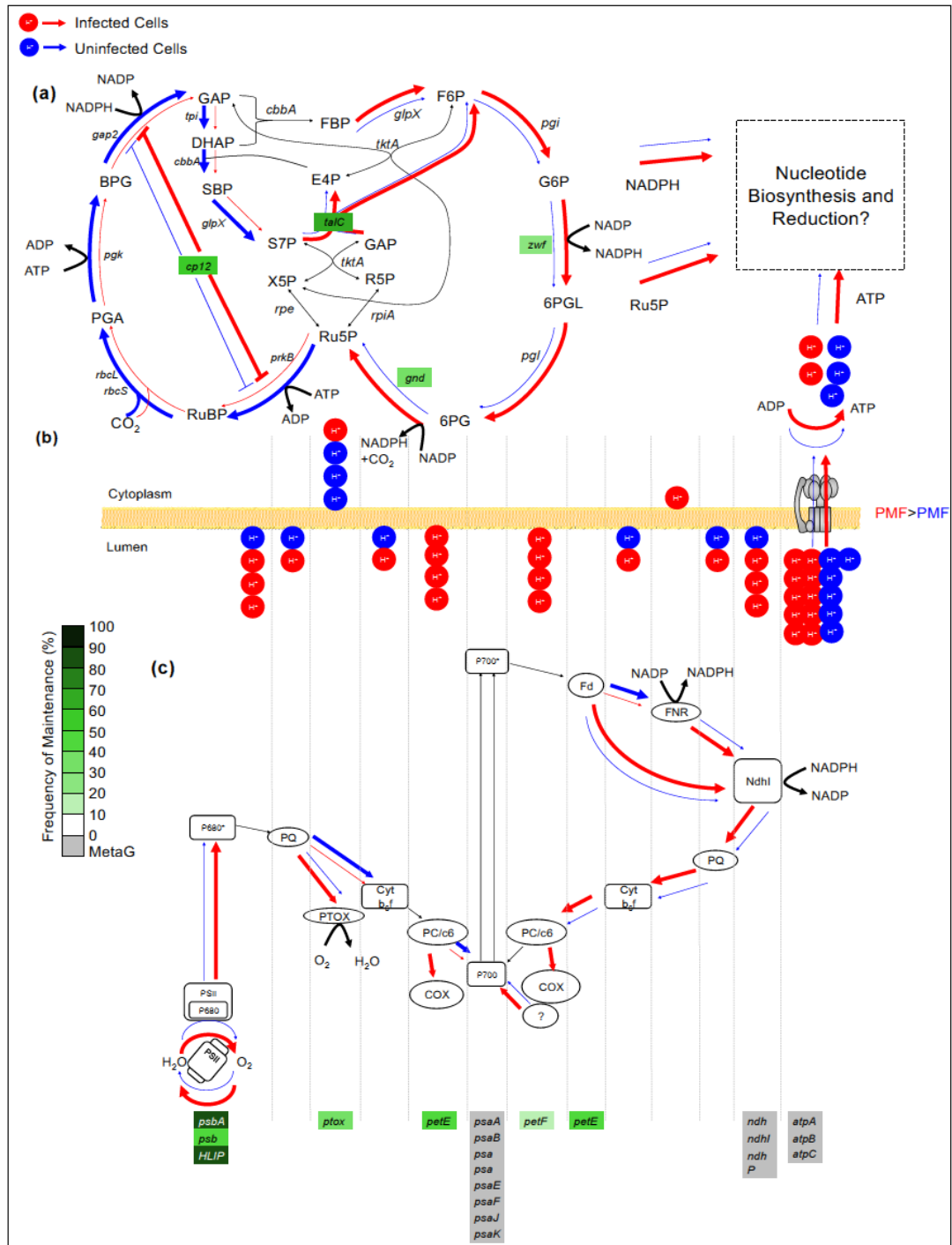
Figure 1.6: Schematic of the hypothetical cyanophage dependent modification of the (PET) chain (c), the Calvin cycle (a), and the resulting effect on the distribution of ions across the thylakoid membrane (b). Genes that have been acquired by cyanophages are highlighted in a shade of green dependent on the frequency of maintenance. The hypothesised magnitude of the pathway under uninfected (blue arrow) and infected (red arrow) cells is represented by the width of the arrow. ? denotes where the electron donor is unknown or hypothetical. PSII (photosystem II), P680 (P680), PQ (plastoquinol), Cytb6f (Cytochrome b6f complex), PC/c6 (plastocyanin/ cytochrome b6), P700 (P700) Fd (ferredoxin), FNR (Ferredoxin-NADP+ reductase), PMF (proton motive force), CP12 (*cp12*), *cbbA* (fructose-1,6-bis-P/sedoheptulose-1,7-bis-P aldolase), *glpX* (fructose-1,6/sedoheptulose-1,7-bisphosphatase), *pgi* (P-glucose isomerase), *pgk* (P-glycerate kinase), *pgl* (6-P-gluconolactonase),

21

prkB (P-ribulokinase), *rbcL/rbcS* (ribulose-1,5-bis-P carboxylase/oxygenase), *rpe* (ribulose-5-P epimerase), *rpiA* (ribulose-5-P isomerase), *tktA* (transketolase), *tpi* (triose-P isomerase), BPG (2,3-bis-P-glycerate), DHAP (DHAP), E4P (erythrose 4-P), FBP (fructose 1,6-bis- P), F6P (fructose 6-P), GAP (glyceraldehyde 3-P), G6P (glucose 6-P), PGA (3-P-glyceric acid), R5P (ribose 5-P), RuBP (ribulose 1,5-bis-P), Ru5P (ribulose 5-P), SBP (sedoheptulose 1,7-bis-P), 6PG (6-P-gluconate), 6PGL (6-P-gluconolactone), S7P (sedoheptulose 7-P), X5P (xylulose 5-P). Taken form Puxty et al., (submitted).

1.5 Transcription in bacteria and bacteriophages

1.5.1 Recent advances in bacterial transcription

Recent advances in the study of bacterial transcriptomes have revealed a large degree of complexity (Georg et al. 2009; Mendoza-Vargas et al. 2009; Güell et al. 2009; Dornenburg et al. 2010; Sharma et al. 2010a; Filiatrault et al. 2010; Lasa et al. 2011; Vijayan et al. 2011; Mitschke et al. 2011; Wilms et al. 2012; Nicolas et al. 2012). Specifically, in addition to transcription of known protein coding genes, ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs), there exists a wealth of potential non-coding RNAs (ncRNAs). Such ncRNAs can be further characterised according to their location relative to mRNAs of open reading frames (ORFs) (Sharma et al. 2010a; Mitschke et al. 2011). ncRNAs found in intergenic regions are referred to as *trans*-acting, whilst those found on the opposite strand and somewhat overlapping an ORF mRNA are termed *cis*-acting or antisense RNAs (asRNAs). There is a great deal of flexibility in the literature with regards to terminology with small RNAs (sRNAs), *cis*-acting RNAs, *trans*-acting RNAs, asRNAs and ncRNAs all being used, often interchangeably. In this thesis, two classes are referred to: *trans*-encoded ncRNAs are any non-coding RNAs that are not found antisense to an ORF and asRNAs are RNAs that are found antisense in some region to an ORF.

There is great diversity in the mechanisms by which ncRNAs may function including complimentary base pairing to a target mRNA (Waters and Storz 2009; Storz et al. 2011), causing both post-transcriptional enhancement (Sakurai et al. 2012) and repression (Dühring et al. 2006) of gene expression. Many *trans*-encoded ncRNAs act by short imperfect base pairing to their target mRNAs (Waters and Storz 2009; Storz et al. 2011), usually in a site

specific manner which causes ribosome occlusion (Brantl 2002; Storz et al. 2011) thereby inhibiting translation. Many of these RNA-RNA interactions are mediated by the Sm-like protein Hfq (Møller et al. 2002). *Trans*-encoded ncRNAs are often conserved between many bacterial families and can act on a multitude of targets (Wright et al. 2013) and therefore act as global regulators of gene expression in response to environmental stimuli.

In comparison, asRNAs usually interact with their cognate sense mRNAs with perfect base-pairing (Georg and Hess 2011; Lasa et al. 2012). The result of this interaction can either be positive or negative with regards to its effect on gene expression. A summary of the well-studied asRNAs is given in Fig. 1.7 and described in Table 1.1.
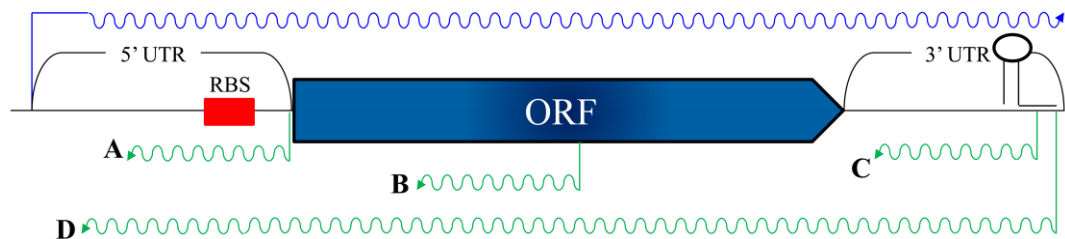
Figure 1.7: Positions of asRNAs relative to a hypothetical ORF. The ORF maintains a 5' untranslated region (UTR) with a ribosome binding site (RBS) and a 3' UTR encoding a transcriptional terminator (stem loop). The ORF mRNA is shown by the blue line. Locations of asRNAs as described so far are shown with green lines. The lettering corresponds to that used in Table 1.1.

asRNA type	Positive Regulation	Negative Regulation
(A) 5'UTR overlap	*Synechocystis* PCC6803 *psbA2/3* and asRNA *psbA2/3R*. *psbAR* protects 5' UTR of *psbA* in concert with ribosome binding to the RBS. This increase stability of the *psbA* mRNA and therefore cellular half-life. (Steglich *et al.,* 2012)	*symR* asRNA of *E. coli* overlaps the 5' UTR of *symE*. *symR* RNA binds to the RBS of the *symE* mRNA preventing initiation of translation. SymE is an SOS-induced antitoxin causing growth inhibition. (Kawano et al. 2007)
(B) Internal antisense	-	IsrR asRNA is found antisense internal to the iron-stress-induced gene *isiA*. Expression of *isrR* and *isiA* leads to co-degradation. The inducible expression of *isiA* has to titrate out *isrR* before a response is observed (Threshold-linear response) (Duhring *et al.,* 2006)

(C) 3' UTR overlap	*E. coli gadY* asRNA overlaps the 3'UTR of the *gadX* gene. *gadY* expression is induced in stationary phase by alternative σ factor RpoS and increases the stability of the *gadX* transcript. GadX encodes a regulator of the acid stress response (Tramonti et al. 2008)	OOP asRNA of λ phage. Overlaps the 3' end of the cII-repressor mRNA. Over-expression of OOP asRNA causes RNase III-dependent cleavage of cII mRNA (Krinke and Wulff 1990).
(D) Whole ORF overlap	Upon phage infection there is an up-regulation of a long (7kb) asRNA to a ribosomal protein operon in *Prochlorococcus* MED4. The asRNA masks RNase E cleavage sites and protects the sense mRNA from degradation (Stazic et al. 2011)	*α-furA* antisense to the *furA* gene in *Anabaena* sp. PCC7120 represses its expression. *furA* encodes a protein involved in the iron-stress response in cyanobacteria (Hernández et al. 2006).

Table 1.1: Description of selected asRNAs identified so far. The classification of the asRNA is given in Fig. 1.7.

The presence of a transcriptional unit antisense to a gene has further implications for gene regulation in addition to a base pairing mechanism. Such effects have been termed transcriptional interference (Fig. 1.8; Georg and Hess 2011; Courtney and Chatterjee 2014). Convergent transcriptional units create a significant amount of 'traffic' at specific loci. This traffic takes the form of accumulating RNAP subunits during the stages of transcription (initiation, elongation, termination). They have been classified into four classes (Fig. 1.8).
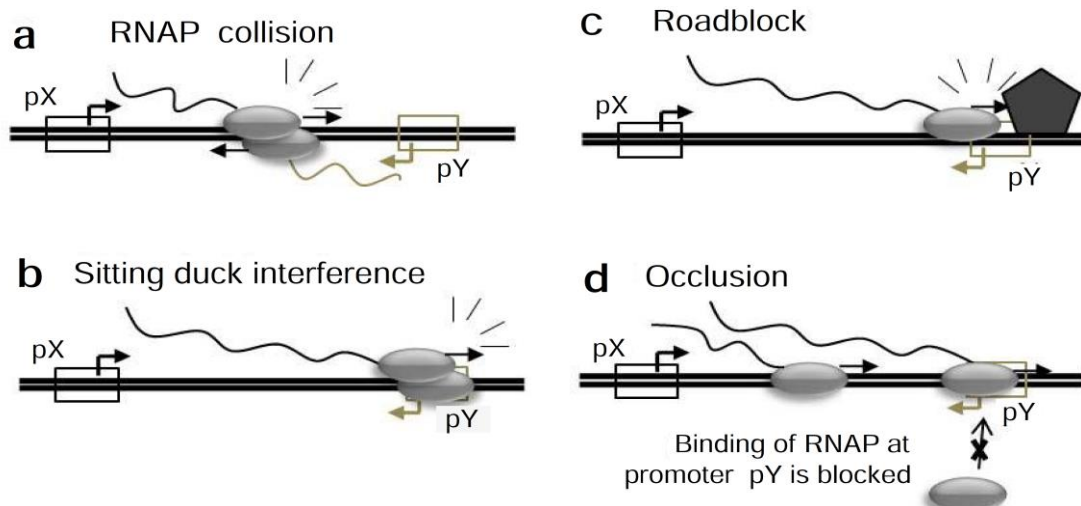
Figure 1.8: Mechanisms of transcriptional interference. Schematic of a general system of convergent promoters pX and pY is shown (a) RNAP collision (b) Sitting duck collision (c) Roadblock (d) Promoter occlusion (Taken from Courtney and Chatterjee (2014).

These mechanisms can have quite unexpected effects on gene expression. Evidence suggests that significant transcriptional interference can result from moderate differences in strengths

of convergent transcriptional units. In coliphage 186, the lytic-lysogenic promoters *pR* and *pL,* respectively, are arranged face-to-face with a 62bp overlap. The presence of the rather stronger *pR* promoter reduces expression of *pL* 5.6 fold through "sitting-duck" interference in a manner which is not reciprocated (Callen et al. 2004). Similar results are observed in the *pR* and *pRE* promoter pair of bacteriophage λ (Palmer et al. 2009).

Mechanistically, the result of a RNAP-RNAP collision can produce interesting consequences, including drop off of both complexes, drop off of one over the other or retraction of the RNAP complex followed by resuming of elongation manifesting as a 'stall' (Crampton et al. 2006; Courtney and Chatterjee 2014). The relative frequencies of each of these events is dependent on numerous factors (Courtney and Chatterjee 2014) and may be difficult to predict without experimentation. It also poses a potential problem for whole transcriptome analyses of gene expression which rely on short probe hybridisation or short read sequencing of targets which may arise from premature termination of transcription.

Courtney and Chatterjee (2014) highlight a hitherto overlooked mechanism that couples transcriptional interference and asRNAs. Transcriptional interference may in-fact create a host of different asRNA isoforms due to variability in the collision site of converging RNAPs. Therefore the abundance of asRNAs relative to their sense mRNAs can be reduced due to multiple sites of interaction. This is exemplified in the *prgQ/prgX* operon of *Enterococcus faecalis,* where both transcriptional interference and the asRNA to *prgQ* coordinate a bistable response to regulate conjugal transfer of plasmid pCF10 (Chatterjee et al. 2011).

In addition to the mechanisms of asRNA-mRNA interactions and transcriptional interference, ncRNAs have shown to bind and function directly on proteins. An elegant example which has relevance to bacteriophage biology comes from the ToxIN abortive infection system of *Pectobacterium atrosepticum.* This is a type III toxin-antitoxin (TA) whereby prior to bacteriophage infection the toxin (ToxN) is inhibited by specific binding of

a ncRNA (*toxI*) (Fineran et al. 2009). Remarkably, *toxI* forms an RNA pseudoknot that binds ToxN monomers into a trimeric 'triangular' structure (Blower et al. 2011; Short et al. 2013). Upon bacteriophage infection, the complex is destabilised either by inhibition of *toxIN* transcription or by interaction with a specific phage product (Blower et al. 2011). The released ToxN monomers function as general endoRNases (Blower et al. 2011) and as such cause cell death preventing the propagation of the bacteriophage throughout the population. Interestingly, *toxI* possesses similarity to the RNA aptamer binding vitamin B_{12} and a queuosine riboswitch, sparking speculation that the RNA itself may sense the stimulus (Blower et al. 2011).

The ability for an RNA to recognise a particular stimulus completes the role that RNAs play in environmental sensing and response. That is, sensing the stimulus, eliciting and regulating a transcriptional response and indeed functioning as the product. RNA sensing of a stimulus is further typified in the bacterial riboswitch (Winkler and Breaker 2005), where catabolite binding by a 5' UTR RNA directly contributes to transcription of a downstream gene.

It is clear from these examples that RNA has a capacity to act as the key regulatory molecule in many bacterial regulatory networks. Yet the same function for RNAs is less understood in bacteriophages.

1.5.2 Transcriptional regulation in bacteriophages

Early studies of bacteriophage transcription focused mainly on the model enterobacteriophages T4, T7 and λ (see description below). These belong to the *Myoviridae*, *Podoviridae* and *Siphoviridae* families, respectively. In all cases the regulation of temporal expression is required such that components of the replication machinery and structural proteins are produced at the correct time. Yet these early studies reveal a remarkable diversity in the transcriptional strategies employed. In particular, in T7, transcription of early genes is governed by host σ^{70}-like promoter sequences upstream of ~20% of genes on the left-hand side of the genome (Chen and Schneider 2005). This includes expression of a

phage specific RNAP which are frequently found among members of the T7-like family (Chen and Schneider 2005). This RNAP coordinates the expression of the remainder of the late genes which are arranged from left to right in genome context. Recently, Zhu et al. (2013) have purified the RNA polymerase from the T7-like cyanophage Syn5. The polymerase specifically initiates transcription downstream from the promoter sequence 5'-ATTGGGCACCCGTAA-3' which only occurs at two loci in the genome (Zhu et al. 2013). A similar prediction was made for the T7-like cyanophage P-SSP7 (Chen and Schneider 2005). This pattern differs from the rest of the T7-like family where multiple occurrences (usually <15) of the phage encoded RNAP promoter sequences are observed (Chen and Schneider 2005). Thus, the transcriptome of the T7-family appears to be rather simplistic. Indeed, Lindell et al. (2007) have demonstrated three extremely distinct temporal clusters of gene expression in the T7-like phage P-SSP7.

Many λ-like bacteriophages can switch between a lytic or lysogenic lifestyle and as such much of the literature has focussed on the transcriptional regulation of both pathways (Rajagopala et al. 2011). The regulatory machinery controlling the lytic/lysogenic switch is extensively reviewed in Oppenheim et al. (2005). The lytic pathway progresses through expression of the early genes *N* and *cro* by the *pL* and *pR* promoters, respectively. Cro inhibits transcription of the repressor required for the lysogenic maintenance, whilst N is a protein required for the phenomenon of anti-termination. Termination of transcription downstream of *pR* and *pL* is facilitated by the *tL1* and *tR1* terminators and prevents expression of downstream genes in the lytic pathway. As N accumulates it forms a complex with host RNAP and host Nus factors (Barik et al. 1987; Oppenheim et al. 2005). Together, this complex allows for read-through of the *tL1* and *tR1* terminators, resulting in the expression of the delayed early operons. These operons encode important components for DNA replication and also the Q protein, a modifier of the RNA polymerase that allows for read-through of terminators downstream of the *pR'* promoter and allows for expression of the late genes (Oppenheim et al. 2005). The regulated read-through of these terminators is

called anti-termination and is the sole mechanism for temporal control of gene expression in λ. Thus, despite the complexity of lytic/lysogenic lifestyle genetics, the lytic pathway is in fact very simple, producing only 3 operons; the early left, the early right and the late operon.

T4-like myoviral genomes appear to be more complex with larger genomes (Fig. 1.5d) and in the case of cyanomyoviruses, acquisition of multiple accessory genes (Millard et al. 2009; Sullivan et al. 2010). Thus coordinated regulation of these genes is expected to be more complex. The transcriptional regulation of T4 genes has been studied extensively (for review see Miller et al. (2003)). Genes expressed early during infection contain host σ^{70} like promoter sequences upstream and thus are recognised by host RNAP (Miller et al. 2003). T4 then undergoes at least two highly regulated modifications of the host RNAP that sequentially favour binding to the middle and late promoters. The middle period involves the activity of two proteins MotA and AsiA, whilst the late period involves modification of the RNAP by the late σ-factor Gp55. Gp55 binds to the late promoter consensus 'TATAAATA' (Geiduschek and Kassavetis 2010), which in the case of cyanophages is reduced to 'ATAAATA' (Mann et al. 2005) or even 'CTAAATA' (Dreher et al. 2011). Gp55 is an extremely divergent σ-factor (Lonetto et al. 1992) which recognises an extremely minimal promoter sequence. A goal of this thesis is to understand how the late period of infection initiates transcription from late promoters given that cyanophages, and particularly the model cyanophage S-PM2 are extremely AT rich.

Millard et al. (2010) have recently discovered an asRNA antisense to the *psbA* gene in the bacteriophage S-PM2. This is the only asRNA to be identified in a bacteriophage with the exception of the λ OOP asRNA (Krinke and Wulff 1990).

1.6 Cyanophage S-PM2 and the genetic organisation of the *psbA* region

Cyanophage S-PM2 was isolated from station L4 in the western English Channel in 1992 (Wilson et al. 1993). Since its original isolation a number of studies have revealed interesting facets of cyanophage-host interactions that may be important for understanding

cyanophage ecology. Of note, Jia et al. (2010) have determined that adsorption of S-PM2 to its host requires light, as is the case for many cyanophages. In addition, Shan et al. (2008) have shown that infection with S-PM2 induces expression of the genes encoding components of the light harvesting apparatus. Moreover, S-PM2 was the first described cyanophage to encode the *psbA* gene (Mann et al. 2003). These facts suggest that S-PM2 may possess an intimate relationship with light and that cyanophage infection may occur specifically during the day to maximise energy from photosynthesis. S-PM2 is an obligately lytic T4-like myovirus containing a ~196Kb genome encoding 244 ORFs (Mann et al., 2005). In comparative genomics with other members of the T4-like cyanophages, it is frequently considered an outlier (Sullivan et al. 2010; Ignacio-Espinoza and Sullivan 2012). This is supported by comparatively few AMGs in this cyanophage (Puxty et al., *submitted*).

The genetic organisation of the S-PM2 *psbA* region is also peculiar compared with other cyanophages (Millard et al. 2004). The *psbA* gene itself contains a 212bp group I intron (Millard et al. 2004). Introns are common in bacteriophage genomes (Belfort 1990), where they are considered genetic parasites (Belfort 1989). Group I introns are self-catalytic, that is, their splicing requires no protein cofactors (Cech 1990). The catalysis is facilitated by the conserved tertiary structure of the folded intron and exogenous guanosine nucleotides that cause the two sequential transesterification reactions leading to excision of the intron and ligation of the two exons (Cech 1990). Introns frequently encode small ORFs that are often endonucleases (Bonocora and Shub 2009). The invasion by the endonuclease facilitates mobility of the intron to intronless alleles in a process known as homing (Belfort et al. 2002; Bonocora and Shub 2009; Hausner et al. 2014). In the case of the S-PM2 *psbA* intron, a homing endonuclease, *F-CphI*, is not encoded within the intron but rather downstream (Fig. 1.9; Millard et al. 2004). *F-CphI* has been purified and shown to cleave intronless alleles of phage *psbAs* (Zeng et al. 2009). Therefore the presence of the intron protects S-PM2 *psbA* from cleavage by *F-CphI*. It is suspected that the introns and homing endonucleases target conserved nucleotide sequences that serve essential biological functions and are well

distributed in the gene pool (Raghavan and Minnick 2009). Therefore, the convergence of the homing endonuclease and the intron on *psbA* were independent and have suggested to be the evolutionary stage before integration of the homing endonuclease into the intron (Zeng et al. 2009). This is interesting because the intron contains a motif extremely similar to the F-CphI cleavage sequence (Zeng et al. 2009). Therefore, why has integration not happened? Millard et al. (2010) report the existence of an asRNA that connects both the 3'end of *psbA* and the 5' end of *F-CphI*. It has been termed Cyanophage functional RNA I (*CfrI*). It is suggested that *CfrI* plays an important regulatory role and that integration of the homing endonuclease into the *psbA* intron would disrupt its function, thus *CfrI* acts to prevent integration and maintain the *status quo* (Millard et al. 2010). Millard et al. (2010) also used bioinformatics to predict several other ncRNAs, suggesting that regulatory RNAs may play an important role during infection. A goal of this thesis is to elucidate whether *CfrI* plays a role in regulation of the cyanophage *psbA* gene as well as to validate the numerous predictions of other ncRNAs.
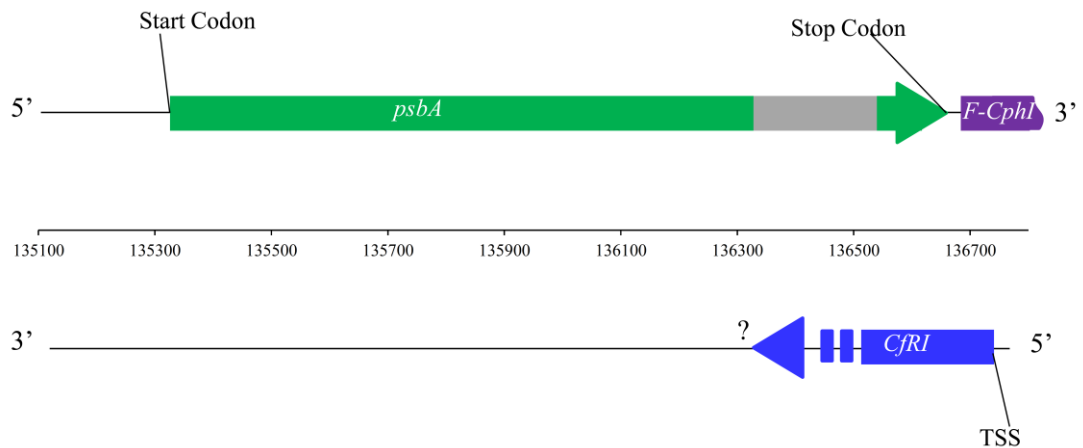
Figure 1.9: Genetic organisation of the cyanophage S-PM2 psbA region. The intron is shown in grey. CfrI has been shown to exist by northern blotting and its TSS mapped to two distinct loci (Millard et al., (2010). The upstream one is shown here. Currently the 3' end of CfrI is unknown.

1.7 Aims and Objectives

The aims of the present study are several-fold:

1) To examine the effect of light intensity in regulating the dynamics of the infection process in cyanophage S-PM2 (Chapter 3). It is hypothesised that an increase in light intensity will increase the rate of photophosphorylation and hence provide more energy for genome replication. Therefore, qPCR was used to target S-PM2 genome copy numbers in both intracellular and extracellular fractions so as to determine the rate of genome replication.

2) To determine whether infection with S-PM2 alters the electron flow in PSII in response to increased irradiance (Chapter 3). It was hypothesised that the range of photosynthetic gene acquisitions by S-PM2 encode specific adaptations that provide improved functioning of the PET at high light irradiances. Chlorophyll fluorescence techniques were therefore used to probe PSII redox chemistry in response to different light irradiances during infection. (Chapter 3).

3) To provide experimental proof that cyanophage infection inhibits CO_2 fixation while maintaining PET. In addition comparisons of the light dependent CO_2 fixation rates between cyanophages that possess genes that are thought to modify central carbon metabolism (S-RSM4, $cp12^+$, $talc^+$, zwf^+, gnd^+) and those that do not (S-PM2, $cp12^-$, $talC^-$, zwf^-, gnd^-).

4) Through the use of RNA-Seq, validate the range of predictions of ncRNAs (Millard et al., 2010) as well as to validate the transcriptional model of cyanophage S-PM2 (Mann et al. 2005) (Chapter 4).

5) To determine the effect of light intensity on gene expression of both cyanophage S-PM2 and *Synechococcus* sp. WH7803 with RNA-Seq. (Chapter 5).

6) Finally, to determine the role of the asRNA, *CfrI* in regulating cyanophage encoded *psbA* expression in response to changes in light intensity (Chapter 6).

Chapter Two: General methods

2.1. General Methods

The methods presented in this chapter are referred to routinely elsewhere in this thesis. More specific methods are referred to within the relevant chapter to maintain coherence and clarity.

2.1.1. Strains and plasmids

The strains and plasmids used in this study are shown in Table 2.1.

Strain/Cyanophage/Plasmid	Genotype	Comments	Reference
Strain			
Synechococcus sp.WH7803	WT	Host strain for S-PM2 and S-RSM4	-
E. coli DH5α (K-12 deriv.)	F⁻ *endA1 gln*V44 *thi*-1 *recA1 relA1 gyr*A96 *deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsd*R17(r$_K^-$ m$_K^+$), λ–	General Cloning Strain	Mendelson & Yuan (1968)
E. coli TOP10 (K-12 deriv.)	F- *mcrA* Δ(*mrr-hsd*RMS-*mcrBC*) φ80*lacZ*ΔM15 Δ*lac*X74 *nupG recA1 ara*D139 Δ(*ara-leu*)7697 *gal*E15 *gal*K16 *rpsL*(Str^R) *endA1* λ⁻	Cloning strain used for one step TOPO TA cloning	Invitrogen
E.coli MC1061 (K-12)	F⁻ Δ(*ara-leu*)7697 [*ara*D139]$_{B/r}$ Δ(*codB-lacI*)3 *gal*K16 *gal*E15 λ⁻ e14⁻ *mcr*A0 *rel*A1 *rps*L150(strR) *spoT1 mcrB*1 *hsd*R2(r⁻m⁺)	Host strain for pRK24 plasmid for conjugal transfer	Casadaban & Cohen (1980)
E.coli BL21 Star^{TM} (DE3) (B deriv.)	F⁻ *ompT hsdS*$_B$(r$_B^-$ m$_B^-$) *gal dcm rne131(DE3)*	Carries the DE3 lysogen encoding the T7 RNA polymerase under the control of the *lac*UV5 promoter	Invitrogen
Cyanophage			
S-PM2	WT		Wilson *et al.*, 1993
S-PM2$^{Δp017:050}$	p017:p050⁻	Contains spontaneous deletion of ORFs p017 to p050	This study
S-RSM4	WT		Millard & Mann (2006)
Plasmid			
pCR2.1	pUC ori Plac *lacZ*α f1 ori Kan^R Amp^R	TOPO TA cloning vector	Invitrogen
pCRpsbASpl	"	Spliced copy of *psbA* from S-PM2 inserted into pCR2.1	This study

pCRpsbAUns	"	Unspliced copy of *psbA* from S-PM2 inserted into pCR2.2	This study
pRK24	Tc^R Apr^R	conjugal transfer plasmid	Meyer (1977)
pRL153	oriV mobC mobB mobA repB repA repC $Kan^{R.}$	Derivative of broad host range plasmid RSF1010. Replicates in *Synechococcus*	Brahamsha (1996)
pcpeTpCfrI	"	pRL153 with the 372bp upstream region of the S-PM2 *cpeT* gene fused to the *cfrI* asRNA inserted into the *Nhe*I/*Hin*dIII sites	This study
pslpCfrI	"	pRL153 with the 400bp upstream region of the S-PM2 S-Layer *gene (p170)* fused to the *cfrI* asRNA inserted into the *Nhe*I/*Hin*dIII sites	This study
pPro33	propionate inducible promoter (*prpR*-P$_{prpB}$) p15A *ori* Cm^R	Used for propionate inducible expression of the anti-intron region of *cfrI*	Lee & Keasling (2005)
pPro24*gfpuv*	propionate inducible promoter (*prpR*-P$_{prpB}$) pBR322 *ori* Amp^R *gfpuv*	Used for amplification of the *gfpuv* gene and propionate inducible control strain	Lee & Keasling (2005)
pET-151D	pBR322 *ori*, Amp^R	Used for cloning and expression of *gfp* variants	Invitrogen
p*gfpuv*	"	Contains the *gfpuv* gene from pPro24*gfpuv* inserted into the one step cloning site of pET-151D	This study
p*gfpuvintL1*	"	Contains the *gfpuv* gene with the S-PM2 *psbA* intron inserted into site 1	This study
p*gfpuvintL2*	"	Contains the *gfpuv* gene with the S-PM2 *psbA* intron inserted into site 2	This study

Table 2.1: Strains, cyanophages and plasmids used in this study

2.1.2. Growth media and conditions

Synechococcus sp. WH7803 was grown in ASW medium (Table 2.2) at 23˚C under constant illumination from white/warm white fluorescent tubes at 5-30 μmol photons m^{-2} s^{-1}. Cultures were routinely maintained in 100 ml volumes in 250 ml Erlenmeyer flasks and larger volumes (1L, 5L and 10L) were grown with the addition of 10 mM $NaHCO_3$.and bubbled with air. Growth of *Synechococcus* on semi-solid media followed Brahamsha (1996) using either agarose or cleaned agar. Agar (Difco Bacto^{TM}, BD Biosciences, San Jose, U.S.A.) was cleaned according to Millard (2009).

ASW		Trace Metal Stock	
Compound	g l-1	Compound	g l-1
NaCl	25	H_3BO_3	2.86
$MgCl_2.6H_2O$	2	$MnCl_2.4H_2O$	1.81
KCl	0.5	$ZnSO_4.7H_2O$	0.222
$NaNO_3$	0.75	$Na_2MoO_4.2H_2O$	0.39
$MgSO_4.7H_2O$	3.5	$CuSO_4.5H_2O$	0.008
$CaCl_2.H_2O$	0.5	$Co(NO_3)_2.6H_2O$	0.005
Tris Base	1.1	$FeCl_3.6H_2O$	3
$K_2HPO_4.3H_2O$	0.03	EDTA	0.5
1 ml trace metal stock			

Table 2.2: Composition of ASW medium

Escherichia coli strains were routinely grown in Luria broth (LB) containing 10 g L^{-1} bacto tryptone, 5 g L^{-1} yeast extract and 10 g L^{-1} NaCl. Growth on solid LB media contained 1.5 % (w/v) Bacto agar. Expression of transformed *E. coli* occurred in SOC medium containing 20 g L^{-1} Bacto tryptone, 5 g L^{-1} yeast extract, 0.5 g L^{-1} NaCl, 0.186 g L^{-1} KCl, 0.952 g L^{-1} $MgCl_2.7H_2O$, 3.603 g L^{-1} glucose.

2.1.3. Concentration and purification of cyanophages

Cyanophages were routinely purified from liquid and plate lysates. Upon confluent plate lysis (see section 2.1.4), the top layer of sloppy agar was removed using a Pasteur pipette and placed in a 2 ml microcentrifuge tube. 1 ml of ASW was added to the agar and vortexed.

The tube was spun at 13,000 rpm in a benchtop centrifuge (Biofuge Pico, Heraeus) and the supernatant collected as a phage stock. To obtain higher cyanophage titres fresh lysates of *Synechococcus* sp. WH7803 (1-10 L) were concentrated by precipitation with polyethylene glycol (PEG). The lysate was spun down (Avanti J-25/JLA 10.5 rotor) at 8,000 r.p.m for 20 min to pellet cell debris. PEG 6000 was added to the supernatant to a final concentration 10 % (w/v) and gently dissolved overnight at 4˚C in the dark. The mixture was precipitated by centrifugation at 10,000 r.p.m for 30 min at 4˚C (Avanti J-25/JLA 10.5 rotor). The pellet was gently resuspended in ASW by pipetting. Resuspended pellets were pooled and treated with an equal volume of chloroform to remove PEG and to further separate remaining debris. The top layer was removed and stored at 4˚C.

Concentrated lysates were purified to high concentration by CsCl ultracentrifugation. Stock concentrations of CsCl were made in ASW to densities $p=1.45, 1.5, 1.7$. Step gradients of CsCl were made in ultracentrifuge tubes using 2ml of $p=1.7$, 3ml of $p=1.5$ and 3ml of $p=1.45$. 6ml of PEG precipitated phage were layered on top of the step gradient and the tubes were subject to ultracentrifugation at 35,000 r.p.m. in a SW40Ti rotor (Beckman Coulter, Pasadena, U.S.A) for 2 hr at 4˚C. The resulting band was removed from the gradient by aspiration. To remove CsCl, phages were dialysed overnight at 4˚C in 2 L of ASW at 4˚C. Dialysis membranes had a 12-14 kDa molecular weight cut-off (MWCO) and were prepared following the manufacturer's instructions (Visking-Medicell, London, UK). ASW was replaced twice during dialysis.

2.1.4. Enumeration of cyanophages

Cyanophages were enumerated using several methods. As an initial estimation of titre, most probable number (MPN) assays were used. *Synechococcus* sp. WH7803 was grown to an OD_{750} between 0.35-0.45, concentrated ten-fold by centrifugation and resuspended in ASW. 90μl was pipetted into the wells of a 96 well plate. 10-fold serial dilutions of a cyanophage stock were made in ASW. 10μl of each dilution were added to the *Synechococcus* wells.

Lysis, as indicated by clearing of the well relative to a no phage control, usually occurred within 5 days. This gave an estimation of the cyanophage titre to the nearest order of magnitude. Subsequently plaque assays were used to give a more precise estimation. *Synechococcus* sp. WH7803 was grown to an OD_{750} between 0.35-0.45. Cells were concentrated 50 fold by centrifugation and re-suspension in fresh ASW. The selected serial dilutions from the MPN assay (+1 dilution greater and less) were added to 0.5ml of the concentrated *Synechococcus* sp. WH7803 and left to adsorb at room temperature in the light for 1hr. For each phage/cell suspension, 2.5 ml of molten 0.4 % (w/v) ASW agar (~40°C) was added to the suspension, mixed, and rapidly poured on top of 1 % (w/v) ASW agar plates. Clean agar was used as described in Millard (2006). Plates were incubated at 23°C under constant illumination of 10-15 μmol photons m^{-2} s^{-1} until plaques had formed. Plaque assays were completed in triplicate and the final titre was calculated as the mean of the three.

2.1.5. Genomic DNA isolation

Genomic DNA was routinely isolated from *Synechococcus* sp. WH7803 for use as templates for PCR. 15ml of cells were centrifuged at 4000 r.p.m. for 10 min at 4°C in a Rotanta 46R centrifuge (Hettich, Tuttlingen, Germany). Pellets were resuspended in 500μl DNA re-suspension buffer (1mM Tris, 1mM EDTA, 1% SDS, pH 8.0). The mixture was left at 65° C for 1 hr. An equal volume of Tris-saturated phenol (pH 8.0) was mixed with the cells and left for 10 min at room temperature. Tubes were spun at 13,000g for 5 min and the top layer was mixed with an equal volume of phenol:chloroform:isoamylalcohol (25:24:1). Tubes were spun at 13,000g for 5 min and the top layer was precipitated according to section 2.1.10.

2.1.6. Phage DNA isolation

CsCl purified cyanophages were used for DNA isolation. An equal volume of Tris-saturated phenol (pH 8.0) was added to the phage suspension, vortexed and left at room temperature for 10 min. The tube was spun down at 4000 r.p.m. for 5 min at room temperature in a

Rotanta 46R centrifuge (Hettich, Tuttlingen, Germany). The top layer was removed and the process repeated again. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added to the tube and left at room temperature for 10 min and centrifuged again. $1/10^{th}$ volume 3M sodium acetate (pH 8.0) was added together with 3 volumes of ice cold 100% (v/v) ethanol and left overnight at -20˚C. The tubes were centrifuged for 30 min at 10,000 rpm at 4˚C (Avanti J-25/JA 25.5 rotor). The supernatant was removed and the pellet washed with 70% ethanol. The pellet was centrifuged as before but for 15mins. The supernatant was removed and the pellet resuspended in H_2O.

2.1.7. RNA isolation

Total RNA was isolated following Logemann et al. (1987). Briefly, 50ml of *Synechococcus* sp. WH7803 was centrifuged at 4000 r.p.m. at 4°C in a Rotanta 46R centrifuge (Hettich, Tuttlingen, Germany) and flash frozen in liquid N_2. Pellets were thawed in 1.5 ml of Z buffer (8M guanidinium hydrochloride; 50mM β-mercaptoethanol; 20mM EDTA) at room temperature for 30min. 2 volumes of acidified phenol (pH 4.5) was added and heated to 65˚C for 30min, followed by the addition of chloroform:isoamyl alcohol for 15min. The aqueous phase was transferred to separate microcentrifuge tubes. RNA was precipitated by addition of 1 vol. isopropanol followed by incubation overnight at 4˚C. The tubes were spun at 13,000 r.p.m. in a benchtop centrifuge (Biofuge Pico, Heraeus) at 4˚C for 30mins. The pellet was then washed with 70% (v/v) ethanol and further centrifuged for 15min. Pellets were resuspended in 50µl H_2O. DNA was removed using the TURBO DNA-*free*™ kit (Ambion®/Life Technologies, Carlsbad, U.S.A.) following the manufacturer's instructions. gDNA contamination was tested by PCR using primers phoH_F/phoH_R (see Appendix 1). Any samples that yielded a PCR product were further treated using the TURBO DNA-*free*™ kit.

2.1.8. Preparation of chemically competent E. coli

10 ml of LB medium containing the relevant antibiotic, if necessary, was inoculated with a single colony of the relevant *E. coli* strain and grown overnight at 37°C. 2 ml of this culture was used to inoculate 50 ml of LB without antibiotics and grown until the OD_{600}=0.35-0.40. The cells were chilled on ice for 10mins. Cells were then centrifuged for 5 min at 4°C at 4000 r.p.m in a Rotanta 46R centrifuge (Hettich, Tuttlingen, Germany). Cells were resuspended in 10 ml ice-cold 0.1M $CaCl_2$ and incubated on ice for 20 min. The cells were then centrifuged as before and then re-suspended in 5 ml 0.1M $CaCl_2$/15% (v/v) glycerol. 100 μl of cells were aliquoted into microcentrifuge tubes, flash frozen in liquid N_2 and stored at -80°C until use.

2.1.9. Chemical transformation of E. coli

Competent cells were thawed on ice. Approximately 1-50ng of DNA were gently mixed with the competent cells and left on ice for 10 min. The mixture was then heat-shocked at 42°C for 1min followed by incubation on ice for 5min. 250μl of SOC medium was added to the tube and incubated at 37°C for 1 hr and shaken at 225 r.p.m. 100μl of the culture was then spread on a LB agar (1.5% (w/v) plate with the relevant antibiotic. Plates were left at 37°C overnight until colonies had formed.

2.1.10. Routine precipitation of nucleic acids

Nucleic acids were routinely precipitated by addition of 1/10 vol. of 3M sodium acetate (pH 8) followed by vortexing. 3 vol. of ice-cold 100% (v/v) ethanol was then added and the mixture left at 4°C overnight. The tubes were spun at 13,000 r.p.m. in a benchtop centrifuge (Biofuge Pico, Heraeus) at 4°C for 30mins. The pellet was then washed with 70% (v/v) ethanol and further centrifuged for 15min.

2.1.11. Enumeration of Synechococcus sp. using flow cytometry

Synechococcus sp. WH7803 were enumerated using a FACScan (Becton Dickson, Franklin, U.S.A.) or a CyFlow Space (Partec, Münster, Germany) flow cytometer. Cells were routinely preserved prior to analysis by treatment with paraformaldehyde (1% w/v, final concentration) and incubation at room temperature in the dark for 1hr. Acquisition was triggered using red fluorescence and cells were counted using red and orange fluorescence filters. Adjustments in gain and gate definitions were made in an *ad hoc* fashion depending on changes in fluorescence properties of the cells during growth.

Chapter Three: The effect of light intensity on the photophysiology, photosynthesis and infection dynamics of cyanophage S-PM2

3.1. Introduction

The kinetics of viral growth are under intense selection. Patterns of gene gain and loss by cyanophages act to modify these kinetics and as such through natural selection drives the evolution of both the virus and host. A mainstay of biological oceanography for over 100 years has been the idea of limitation of primary productivity (*for review see* Moore et al. 2013) . Both abiotic (light, nutrients etc.) and biotic (grazing, viruses, cell death) factors set both the rate of population growth and the amount of biomass that can accumulate (i.e. Blackman vs. Liebig limitation). These factors operate on whole populations as well as individual taxa and therefore play a role in regulating the carbon cycle and biodiversity. Here, this paradigm is extended to study the development of cyanophages in their hosts by asking the question: what factor limits the replication kinetics of cyanophages? Previous studies have implicated phosphate limitation (Wilson et al. 1994; Zeng and Chisholm 2012), CO_2 concentration and related pH changes (Traving et al. 2014) as well as host growth rate (Lindell et al. 2007) as being important for the kinetics of cyanophage development in their hosts. In this study the effect of light was considered. Marine cyanobacteria rely on light to provide energy through photophosphorylation. Thus it was predicted that light intensity would affect the amount of energy available for cyanophage replication. Moreover cyanophages have acquired and modified a range of genes involved in the photosynthetic electron transport chain (Mann et al. 2003; Lindell et al. 2004a; Millard et al. 2009; Sullivan et al. 2010) as well as genes involved in redirection of central carbon metabolism away from anabolism and towards the pentose phosphate pathway (Thompson et al. 2011). Thus one may expect the patterns in photophysiology and rates of CO_2 fixation to be similarly affected by cyanophage infection.

Therefore the aims of this chapter are three-fold:

1) To understand the effect that different light intensities have on the infection cycle of cyanophage S-PM2.

2) To establish differences in photophysiology between cyanophage infected cells and their uninfected hosts that might provide insight into adaptation to their light environment and selection for maintenance of genes involved in photosynthesis.

3) To test the hypothesis that cyanophages decouple the light and dark reactions of photosynthesis to maintain energy generation at the expense of wasteful CO_2 fixation.

4) To compare the carbon fixation rate in cells infected with S-RSM4 (containing genes for redirection of carbon metabolism) and S-PM2 (lacking carbon metabolism genes).

3.2. Methods

3.2.1. Infection dynamics by qPCR

Traditional one-step growth experiments with enumeration by plaque assay prove difficult in phages of marine *Synechococcus*. This is largely associated with slow growth rates of the organism and volumes of cultures required for plaque assays (see section 2.1.4). As such, an alternative strategy was employed that monitors infection dynamics using qPCR to enumerate phage genome copy number (GCN). This provides certain experimental limitations. Namely, in a traditional one-step growth experiment, after adsorption, the infected cell suspension is diluted between 4000 and 40,000 (Hyman and Abedon 2009). This inhibits reinfection of new hosts by the released progeny phage and thus allows correct calculation of burst size. However, dilution would significantly alter the flux of quanta to the cells making the data incomparable with other experiments. Moreover, such a dilution would reduce the concentration of phages to below the limit of detection (LOD) by qPCR, due to small reaction volumes. As such, no dilution was employed and instead a high virus bacteria ratio (VBR) was used. An advantage of the method is that it allows for determination of phage genome replication rate inside the host, which was expected to be affected by light.

Synechococcus sp. WH7803 was grown in ASW medium (see section 2.1.2) to an OD_{750} of 0.372. This was empirically estimated to be ~1-2 x 10^8 cells ml^{-1}. 100 µl of the cell suspension was preserved by addition of paraformaldehyde (PFA, 1% final concentration), to be counted by flow cytometry (see section 2.1.11), so as to determine the actual VBR. 4 ml aliquots were taken into polycarbonate test tubes. 400 µl of cyanophage S-PM2 (stock concentration $2.5x10^{10}$ ml^{-1}, as determined by plaque assay) was then added to the cells. After counting by flow cytometry, the actual VBR was calculated as 11.90. The cells were left to adsorb for one hour at room temperature at a light intensity of 20 µmol photons m^{-2} s^{-1}. Three biological replicates were shifted to a high light (HL) regime of 210 µmol photons m^{-2} s^{-1} (see section 3.3.1) and three were maintained at low light (LL) (20-30 µmol photons m^{-2} s^{-1}) and left at 23˚C for the duration of the experiment. No attempt was made to remove unadsorbed phages.

The method for sampling followed Lindell et al. (2005), Thompson et al. (2011) and Zeng and Chisholm (2012). Samples were taken immediately after addition of cyanophage S-PM2 and then at 1, 2, 3, 5, 7, 9, 12, 15 and 18 hrs. For enumeration of extracellular phage GCN, 100 µl of cell suspension was diluted to 500 µl by addition of ASW medium. This was syringe filtered through a 0.2 µm pore size disposable filter (Minisart, Sartorius, Goettingen, Germany). The filtrate was collected and snap frozen in liquid nitrogen, then stored at -80˚C before quantification. For enumeration of intracellular phage GCN, 100 µl of cell suspension was vacuum filtered onto a 0.2 µm pore size polycarbonate filter (Isopore, Millipore, Billerica, USA) using a glass filter tower on a fritted glass support. All six samples for each time point were filtered simultaneously. The filters were washed three times with 1 ml preservation solution (10 mM Tris-HCl, 100 mM EDTA, 500 mM NaCl, pH 8.0). The filters were removed from the tower and added into a ribolyser tube (Lysing Matrix E, MP Bioproducts, CA, U.S.A.) and snap frozen in liquid nitrogen for DNA extraction.

3.2.1.1. DNA extraction

DNA extraction was extensively optimised and the following was found to yield optimal results. 650 µl of Tris-HCl, pH 8.0 was added to a ribolyser tube (Lysing Matrix E, MP Bioproducts, CA, U.S.A.). The tube was then lysed three times for 30 s in a Reusch Tissue Lyser at 30 Hz. In between disruption the tubes were incubated on ice for 1 min. The supernatant was spun down for 30 s at 10,000g and 100 µl of the supernatant was taken and frozen at -80 ˚C. Samples were diluted 1/10 and 1 µl used as template for qPCR reactions.

3.2.1.2. qPCR

qPCR to enumerate cyanophage S-PM2 GCN was achieved using a PrimeTime 5' nuclease assay (IDT, Coralville, USA). Primer/probe sequences (psbA_intF, psba_intR, psbA_intProbe see Appendix 1) were directed towards the *psbA* intron of S-PM2. 20 µl reaction volumes were used containing 1 x Brilliant III Ultra-Fast qPCR Master Mix (Agilent, CA, USA), 500 nM primers and 250 nM probe, 1 µl template and nuclease free water. Each qPCR plate contained a triplicate dilution series from the initial phage stock from 2.5×10^{10} -2.5×10^{2} ml^{-1}. Cycling conditions were 3 min at 95˚C, followed by 40 cycles of denaturing for 15 s at 95 ˚C and annealing for 15 s at 60 ˚C. GCN was calculated by comparison of experimental C_T values to the linear regression of the standards to infer absolute quantification.

To compare the infection dynamics between the two light regimes a range of parameters were compared. Table 3.1 give definitions for the parameters calculated.

Parameter	Symbol	Units	Description
Initial Concentration	Φ_{t0}	pfu ml^{-1}	Initial concentration of phage. Calculated by plaque assay.

Unadsorbed Phage	Φ_{UA}	GCN ml^{-1}	Concentration of unabsorbed phages. Calculated as the number of extracellular phage GCN in the empirically derived period of genome replication
Fraction of unadsorbed phage	Φ_{FUA}	-	The number of unadsorbed phage divided by the initial concentration.
Intracellular phage concentration at time point 1	ΦINT_{t1}	GCN ml^{-1}	Used in calculation of burst size. Empirically derived point after infection when all phage are estimated to be adsorbed
Maximum intracellular phage concentration	MAX_{INT}	GCN ml^{-1}	Maximum concentration of phage in the intracellular phase during the experiment
Maximum extracellular phage concentration	MAX_{EXT}	GCN ml^{-1}	Maximum concentration of phage in the extracellular phase during the experiment
Rate of genome replication	$\dfrac{\delta G}{\delta t}$	V hr^{-1}	Calculated by regression through the linear part for the genome replication phase
Latent Period 1	LP_1	hrs	Time between adsorption and appearance of extracellular phage genomes
Latent Period 2	LP_2	hrs	The time between adsorption and reduction in intracellular genome copy number
Burst Size 1	BS_1	GCN cell^{-1}	Defined has the number of progeny phage produced per cell. The maximum intracellular phage concentration divided by the intracellular phage concentration at time point 1
Burst Size 2	BS_2	GCN cell^{-1}	Defined has the number of progeny phage produced per cell. The maximum extracellular phage concentration minus the concentration of unadsorbed phages concentration at time point 1

Table 3.1: Definitions of parameters used to compare infection dynamics

3.2.2. Photophysiological measurements

A broader explanation of the electron kinetics that occur in the photosynthetic light reactions is given in section 1.2.2. In these experiments two parameters were quantified. The maximum quantum yield of PSII photochemistry (F_v/F_m) and the effective quantum yield of PSII photochemistry (Φ_{PSII}). F_v/F_m was calculated using the method of Garczarek et al. (2008). All measurements were made using a pulse amplitude-modulated fluorometer

(PhytoPAM, Walz, Effeltrich, Germany). A *Synechococcuss* culture (2 ml) was incubated for 5 min in the dark to completely oxidise the primary electron acceptor Q_A. 500 µl of this culture was added to a PhytoPAM cuvette (Waltz, Effeltrich, Germany) and immediately diluted to 2ml with ASW medium (see section 2.1.2). A weak modulating light was then applied at 520 nm with a magnitude of roughly 1 µmol photons m^{-2} s^{-1} and the basal fluorescence measured (F_0). 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) was added to a final concentration of 100 µM, and an actinic light was supplied at ~1300µE m^{-2} s^{-1}. Chlorophyll fluorescence rose rapidly followed by a slower period of increase until saturation. A saturating pulse of ~2600 µE m^{-2} s^{-1} was then delivered for 200 ms to completely reduce Q_A. The fluorescence was simultaneously measured (F_m) and F_v/F_m calculated as $(F_m-F_0)/F_m$.

'Fast' photosynthesis-irradiance curves were estimated by calculating Φ_{PSII} after a period of illumination at increasing actinic irradiances. An initial maximum quantum yield was determined at a low frequency modulating light as described above. The sample was then subject to 30s periods of illumination at 1, 9, 71, 149, 228, 413, 590, 771, 936, 1279 µE m^{-2} s^{-1}. After each period of actinic illumination the instantaneous fluorescence was measured (F_t). Subsequently a saturating flash of ~2600 µE m^{-2} s^{-1} was delivered and the maximum fluorescence measured (F_m'). Φ_{PSII} was calculated as $(F_m'-F_t/F_m')$ according to (Genty et al. 1989). The actinic irradiances were selected to best match irradiances used in measurements of photosynthesis-irradiance (P-I) curves by the radioisotope method (see section 3.2.3). The relative electron transport rate (rETR) was calculated as Φ_{PSII} x I x 0.5 x 0.84, where I is the incident irradiance and 0.5 and 0.84 are arbitrary constants that account for the amount of photons that reach PSII and the relative absorptivity, respectively.

3.2.3. CO_2 fixation rates

CO_2 fixation rate measurements were made by assessing the amount of uptake of $NaH^{14}CO_3$ (Steeman-Nielsen 1952). Various modifications exist for the basic protocol and are largely

related to the 'type' of primary production one wishes to measure (for review see Falkowski and Raven 2007). In this study no attempt was made to measure the absolute amount of carbon uptake, which, in laboratory cultures, requires sensitive measurements of the concentration of dissolved inorganic carbon (DIC)(Parsons et al. 1984). Further, most protocols will make some attempt to separate particulate organic carbon (POC), so as to separate the radiolabelled carbon that eventually forms biomass from that which is fixed but subsequently released from cells. In this study no attempt was made to separate POC. As such the relative instantaneous rate of radiocarbon uptake is measured between infected and uninfected cells at different irradiances. This rate is governed by two factors: 1. The rate of uptake by inorganic carbon into the cell and 2. The rate of CO_2 fixation by the enzyme RuBisCO (Falkowski and Raven 2007).

Two types of experiments were conducted: 1. Assessment of the rate of carbon uptake over the infection cycle in infected cells and uninfected controls at a constant irradiance. 2. Assessment of the integrated carbon uptake throughout a truncated infection period over a range of irradiances, essentially simulating a photosynthesis-irradiance (P-I) curve.

Synechococcus sp. WH7803 was grown in ASW medium (see section 2.1.2) to an $OD_{750\ nm}$ between 0.35-0.40. Cell concentrations were measured using flow cytometry (see section 2.1.11) and diluted to a concentration of $1x10^8$ cells ml^{-1}. Samples were taken and either cyanophage S-PM2 or S-RSM4 was added to a VBR of 10 in triplicate. The samples were incubated at room temperature under low light (~20 μmol photons $m^{-2}\ s^{-1}$) for 1 hr. The culture was then diluted to a concentration of $1x10^7\ ml^{-1}$ and 0.1 MBq of $NaH^{14}CO_3$ (specific activity: 1.48-2.22 GBq $mmol^{-1}$, Perkin Elmer, Waltham, USA) was added to each culture. In the first experiment the cultures were split into 3 ml volumes in clear polycarbonate tubes and incubated at 25-30 $\mu E\ m^{-2}\ s^{-1}$. The cultures were maintained at 23 ^{o}C using a temperature controlled water bath. An entire 3 ml culture was harvested at 1, 2, 4, 5, 7 and 9 hrs after infection. In the second experiment, samples were again split into 3ml volumes in clear polycarbonate tubes but were instead placed into a laboratory built 'photosynthetron'.

Samples were incubated for 4 hrs before being harvested. 4 hrs was selected as previous experiments had shown that light significantly reduced the latent period (see section 3.3.2). Therefore, shortening the incubation period limited the effect of any early lysis in comparisons between phage and host and between phages. The light delivered to each polycarbonate tube was measured using a Spherical Micro Quantum US-SQS/B light meter (Walz, Effeltrich, Germany), connected to a PhytoPAM (Walz, Effeltrich, Germany). The measurements were made once with cells at $1x10^7$ ml^{-1} and applied to all experiments. This was due to logistical difficulties when using radioactivity.

Once samples were taken PFA was added to a final concentration of 1% and left at room temperature for 1 hr. Fixed samples were added to 20 ml scintillation vials and then acidified by addition of 0.5 vol 6N HCl in a fume hood for 1 hour. This process drives off any inorganic carbon that has not been incorporated. The samples were then neutralised by addition of 0.5 vol 6N NaOH. 5ml scintillation fluid (Ultima Gold, Perkin Elmer, Waltham, USA) was added to scintillation vials and left overnight before scintillation counting using a Packard Tri-Carb 2900TR and collected with QuantaSmart^{TM} software.

3.3. Results

3.3.1. Validation of high light stress

To compare the infection dynamics at a high light (HL) regime compared with low light (LL), one must first establish what a HL regime is. Here, HL is defined as an irradiance where photoinhibition occurs in the host but which is not lethal over the period of illumination. Photoinhibition occurs when the rate of photoinduced damage to photosystem II (PSII) exceeds the rate of repair (Tyystjärvi and Aro 1996) and often sets the limit of the photosynthetic rate at high irradiance. Therefore the rate of damage to PSII was estimated at various light irradiances using a lincomycin amended photophysiological approach (Tyystjärvi and Aro 1996; Garczarek et al. 2008). Lincomycin is an inhibitor of prokaryotic and chloroplast protein biosynthesis (Chang et al. 1966). Therefore, after treatment with

lincomycin (final concentration 500 μg ml⁻¹), F_v/F_m decreases at a rate which is proportional to the PSII damage rate, given that PSII cannot incorporate a *de novo* synthesised D1 polypeptide. By fitting exponential curves to the observed decline one can calculate the coefficient of PSII damage and further deduce the PSII repair rate. The PSII repair rate is calculated as the rate of PSII turnover (-lincomycin) minus the PSII damage rate. The data suggested that a vessel-averaged PAR of 210 μE m^{-2} s^{-1} was sufficient to induce photoinhibition (Fig. 3.1). Further, cells recovered well after 12 hrs under this condition (data not shown). The vessel averaged PAR is calculated given that the incident PAR to the vessel was measured to be 300 μE m^{-2} s^{-1} and the transmitted PAR was measured to be 121 μE m^{-2} s^{-1}. Thus the PAR attenuation coefficient ($\alpha_{400-750nm}$) of the vessel was therefore calculated to be 0.9 m^{-1}, given the vessel's 10 cm diameter. This is valid given the culture vessel was shaken at 120 rpm. In all future experiments this light level was used to define a high light regime.
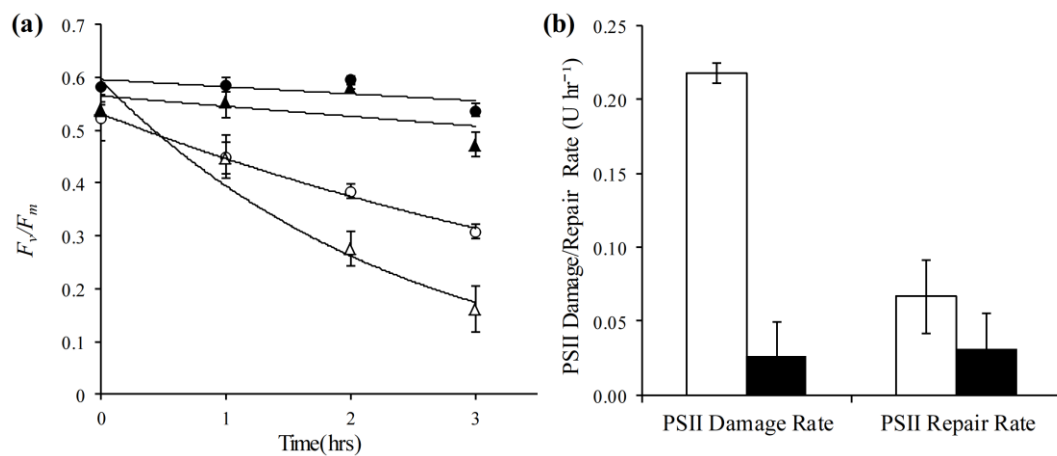
Figure 3.1: Demonstration of photoinhibition of *Synechococcus* sp. WH7803 under HL treatment. (a) The effects of light and lincomycin on Fv/Fm over the period of illumination. Filled circles indicate LL (20 μmol photons m^{-2} s^{-1} - lincomycin. Filled triangles indicate LL + lincomycin. Unfilled circles indicate HL (210 μmol photons m^{-2} s^{-1}) - lincomycin. Unfilled triangles indicate HL + lincomycin. Data are fitted with exponential curves. (b) Relative rates of PSII damage and repair on HL (unfilled bars) and LL (filled bars).

3.3.2. Infection dynamics

The effect of HL treatment on the infection dynamics of cyanophage S-PM2 was tested using a qPCR method to target intracellular as well as extracellular copies of the phage chromosome. Figure 3.2 shows that in both light treatments there is an initial lag of 2 hrs before a linear increase in intracellular GCN. This linear increase occurs between 2-5 hrs. At 7 hrs the maximum concentration of intracellular S-PM2 DNA is reached in both HL and LL. In LL this concentration is maintained until after 12hrs where it decreases rapidly between 12-15 hrs, indicative of a burst. In comparison, in HL, such a rapid decrease is observed some 5 hrs earlier, immediately after the maximum concentration of intracellular S-PM2 DNA is reached. A similar result is observed in the concentration of extracellular S-PM2 DNA. After initial infection this concentration remains extremely low. In HL there is a rapid increase just after 7 hrs, concomitant with a reduction in the concentration of intracellular S-PM2 DNA. In comparison, in LL the low concentration of extracellular S-PM2 DNA remains low until 12 hrs when there appears to be a rapid increase also concomitant with the decrease in intracellular S-PM2 DNA. Fitting of the data shown in Fig. 3.2 to the parameter descriptions in Table 3.1 allows one to approximate the kinetics of viral growth. These are shown in Table 3.2.
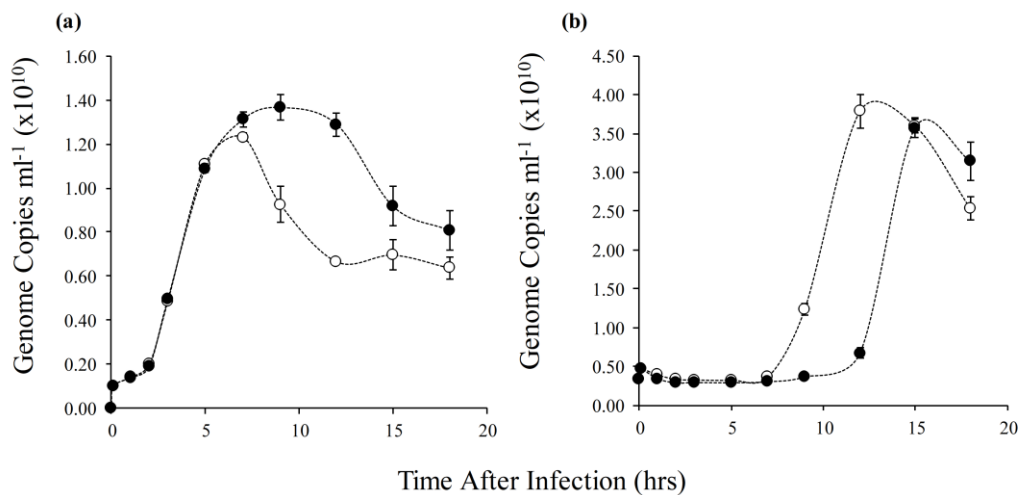
Figure 3.2: Infection kinetics of cyanophage S-PM2 on *Synechococcus* WH7803 in HL and LL. (a) Intracellular genome copies (b) Extracellular genome copies. In both plots unfilled circles are samples subjected to HL treatment and filled circles are LL. Error bars indicate standard error of three biological replicates.

There were very few significant differences in phage growth parameters between HL and LL apart from the burst size according to burst size definition 1. In this case the burst size was 16% higher at LL compared with HL. This was driven by an increase in the intracellular concentration of phages between 7-12 hrs. However, in the alternative measurement of burst size based on the appearance and quantification of extracellular phage GCN there was no significant change in burst size between the two light conditions. Thus it is unclear whether the burst size does in fact change and if so the difference is probably minimal.

Parameter	Symbol	Mean LL	Mean HL	t	p
Initial Concentration	Φ_{t0}	2.50×10^9	2.50×10^9	-	-
Unadsorbed Phage	Φ_{UA}	$3.28 \times 10^8 \pm 0.17$	$2.90 \times 10^8 \pm 0.16$	2.86	0.05
Fraction of unadsorbed phage	Φ_{FUA}	0.13	0.12	2.86	0.05
Intracellular phage concentration at time point 1	ΦINT_{t1}	$1.40 \times 10^9 \pm 0.02$	$1.42 \times 10^9 \pm 0.06$	0.50	0.67
Maximum intracellular phage concentration	MAX_{INT}	$1.4 \times 10^{10} \pm 0.04$	$1.2 \times 10^{10} \pm 0.03$	-7.56	<0.01
Maximum extracellular phage concentration	MAX_{EXT}	$3.6 \times 10^{10} \pm 0.25$	$3.8 \times 10^{10} \pm 0.35$	1.06	0.35
Rate of genome replication	$\dfrac{\delta G}{\delta t}$	1.61 ± 0.13	1.54 ± 0.09	-0.779	0.48
Latent Period 1	LP_1	12	7	-	-
Latent Period 2	LP_2	12	7	-	-
Burst Size 1	BS_1	10.51 ± 0.25	8.83 ± 0.16	-9.850	<0.01

Burst Size 2	BS_2	12.32±0.27	12.16±0.91	-1.581	0.21

Table 3.2: Comparison of cyanophage growth parameters. Definitions and units are given in Table 3.1. t represents the t-statistic from a two-tailed student's t-test. *p* values are also given.

Estimates of the latent period from Fig. 3.2 suggest this is the largest source of difference between the two light conditions. The latent period was approximately 58% longer during infection at LL. These data point to the triggering of an early burst in conditions of HL or rather a delayed burst in conditions of LL.

3.3.3. Photophysiology and photosynthesis

The change in the parameter F_v/F_m with time is proportional to the rate of PSII turnover, which itself is a balance between the PSII damage rate and the repair rate (Tyystjärvi and Aro 1996). As such, the change in F_v/F_m was measured during infection under HL and LL conditions (Fig. 3.3).

Under the growth irradiance of the host strain (LL) there is no change in F_v/F_m during the experimental period. Likewise, there is no change in cells that are infected with cyanophage S-PM2. In comparison, when the host strain is shifted to HL, there is a rapid decrease in F_v/F_m from 0-4hrs after the light shift, which reduces F_v/F_m to approximately 70% of T0. Interestingly, however, when cells infected with cyanophage S-PM2 are shifted to HL, there is an initial rapid decline in F_v/F_m from 0-1 hr after the light shift that is equal to uninfected cells. However, from 1-2 hrs after the light shift there appeared to be an increase in F_v/F_m. After 2hrs, the F_v/F_m of infected cells at HL remained at ~90% of those at LL, whereas uninfected cells dropped to ~70%.
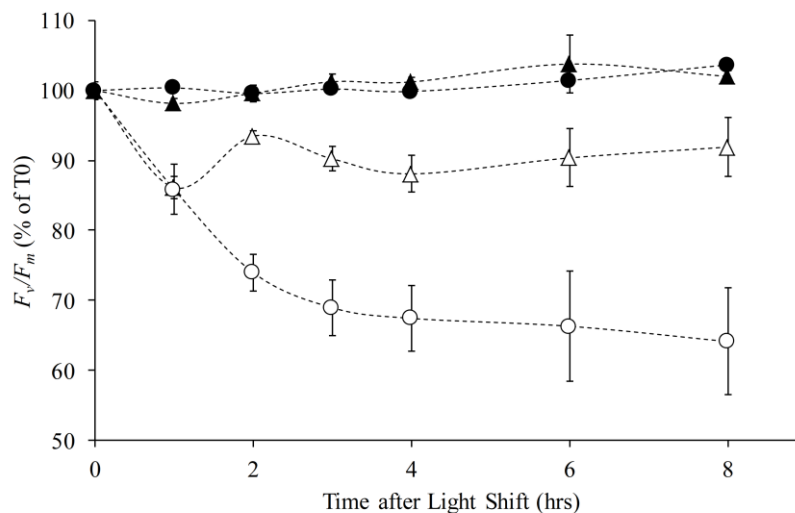
Figure 3.3: Changes in Fv/Fm during infection of *Synechococcus* sp. WH7803 with cyanophage S-PM2 under HL and LL conditions. Filled circles are uninfected cells at LL, filled triangles are infected cells at LL, unfilled triangles are infected cells at HL, unfilled circles are uninfected cells at HL.

In addition to assessing the change in F_v/F_m during infection, so-called 'fast' P-I curves were measured. This method addresses the effective quantum yield of PSII (Φ_{PSII}) after a short period of actinic illumination at a given irradiance (in this case 30s). In this way the relative electron transport rate (rETR) through PSII can be calculated (*see* Suggett et al. 2010). Figure 3.4 compares the $rETR_{PSII}$ in *Synechococcus* sp. WH7803 uninfected and infected with cyanophage S-PM2 throughout infection.

These data show that the shape of the fast P-I curve becomes increasingly linear with progression through infection with S-PM2 whereas the shape remains constant in the uninfected control. The model of Platt et al. (1980) was used to fit curves to these data and therefore derive estimates of $rETR_{PSIIMAX}$, βI, α and I_k (Table 3.3).
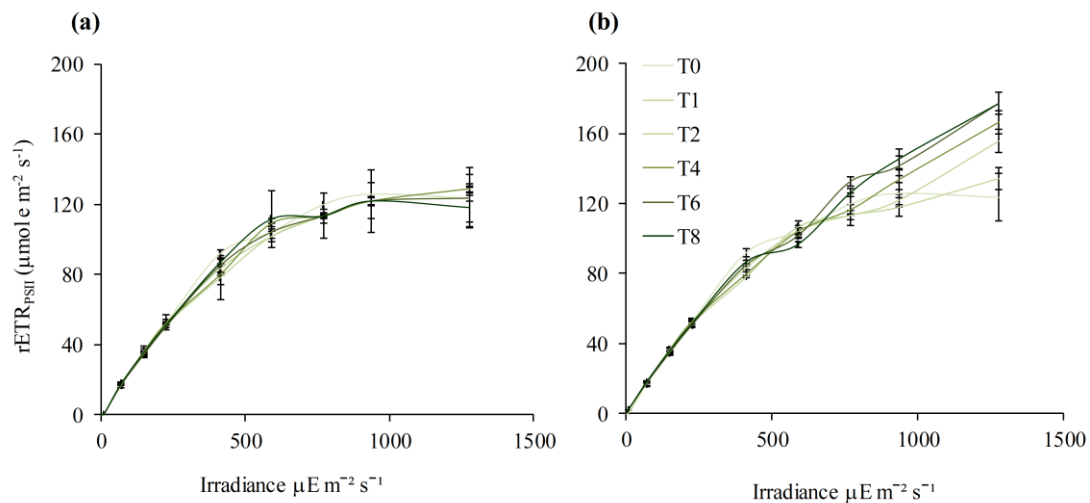
Figure 3.4: Relative electron transport rate through PSII. Measurements were made over a light gradient after 30s acclimation period to the actinic irradiance. (a) *Synechococcus* WH7803 (b) WH7803 infected with cyanophage S-PM2. E represents 1 μmol photons m⁻² s⁻¹.

	WH7803				WH7803+S-PM2			
	$rETR_{PSIIMAX}$ (μmol e $m^{-2} s^{-1}$)	β (μmol photons $m^{-2} s^{-1}$)	α (mol e mol Q^{-1})	I_k (μmol photons $m^{-2} s^{-1}$)	$rETR_{PSIIMAX}$ (μmol e $m^{-2} s^{-1}$)	β (μmol photons $m^{-2} s^{-1}$)	α (mol e mol Q^{-1})	I_k (μmol photons $m^{-2} s^{-1}$)
T0	136.99	0.00	0.33	79.80	136.99	0.00	0.33	79.80
T1	135.14	0.00	0.31	86.18	143.03	0.00	0.29	96.50
T2	142.52	0.00	0.28	100.08	179.65	0.00	0.25	146.64
T4	140.66	0.00	0.30	91.98	210.72	0.00	0.24	180.03
T6	135.95	0.00	0.31	86.83	244.43	0.00	0.24	208.86
T8	130.98	0.00	0.33	74.70	250.01	0.00	0.23	218.64

Table 3.3: Photophysiological parameters derived from fast P-I curves of *Synechococcus* sp. WH7803 uninfected and infected with cyanophage S-PM2. Q is equal to 1 μmol photons $m^{-2} s^{-1}$.

$rETR_{PSIIMAX}$ represents the relative maximum rate of electrons through PSII. This parameter increases throughout infection to a maximum of 250 μmol e $m^{-2} s^{-1}$. In comparison, in uninfected cells, this value remains stable at an average of 137.04 μmol e $m^{-2} s^{-1}$. Therefore, cells infected with S-PM2 can transport electrons through PSII at roughly twice the rate of uninfected cells, at maximal levels. The parameter I_k represents the light intensity where half of $rETR_{PSIIMAX}$ is observed (Platt et al. 1980). This parameter also increases linearly during infection and is ~2.9 times higher in infected cells compared with uninfected cells by 8hrs past infection.

To complement these photophysiological measurements and indeed to test the hypothesis that cyanophages act to redirect energy away from the Calvin cycle, the relative amount of inorganic carbon fixation was measured using a radioisotope technique (Fig. 3.5).
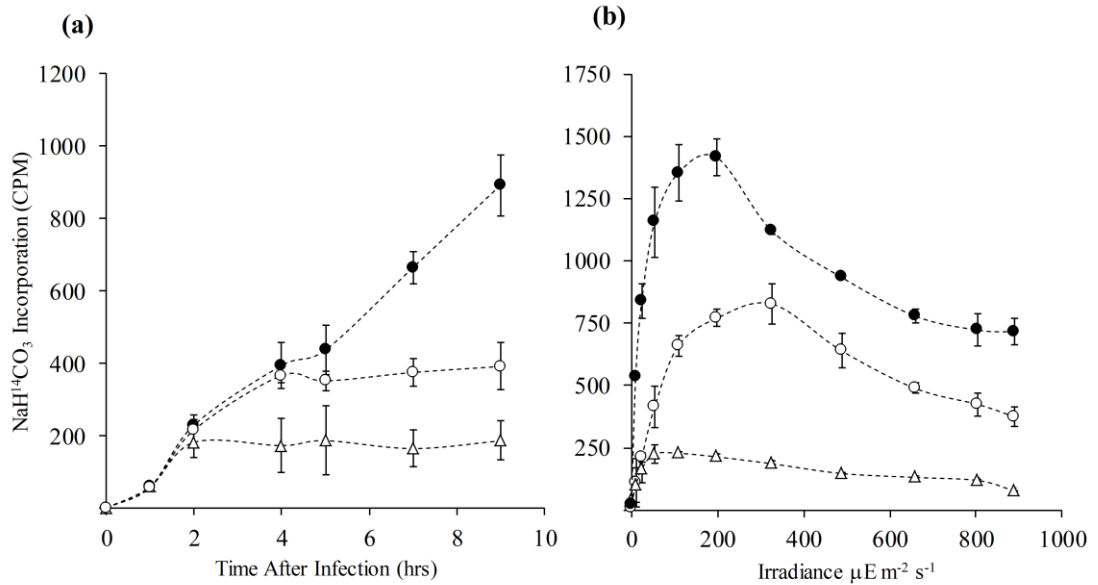
Figure 3.5: Incorporation of ^{14}C labelled sodium bicarbonate by *Synechococcus* sp. WH7803 (filled circles), *Synechococcus* sp. WH7803 infected with cyanophage S-PM2 (unfilled circles) and *Synechococcus* sp. WH7803 infected with cyanophage S-RSM4 (unfilled triangles). **(a)** Time dependent uptake of inorganic carbon throughout infection at a fixed irradiance (20-30 μmol photons m^{-2} s^{-1}). **(b)** 5hr integrated uptake over a light gradient. Error bars indicate standard deviation of three biological replicates. E is equal to 1 μmol photons m^{-2} s^{-1}.

Synechococcus sp WH7803 fixed CO_2 linearly throughout the experimental period at the fixed irradiance of 20-30 μmol photons m^{-2} s^{-1}. *Synechococcus* sp. WH7803 infected with cyanophage S-PM2 showed a linear increase in the amount of CO_2 fixed up until 4hrs. The rate of this uptake was equal in magnitude to that of the uninfected host control. After 4 hours, no more inorganic carbon was fixed. This cessation of CO_2 fixation was more pronounced in *Synechococcus* sp. WH7803 infected with cyanophage S-RSM4, occurring at roughly 2 hrs after infection. At the end of the experimental period S-PM2 infected cells fixed 2.28 times less carbon than the uninfected host, whilst S-RSM4 infected cells fixed 4.79 times less carbon.

In addition, the rate of inorganic carbon fixation was measured over a light gradient over the first 5hrs of the infection cycle (Fig. 3.5b). The data shows clear differences between *Synechococcus* sp. WH7803 and *Synechococcus* sp. WH7803 infected with either cyanophage S-PM2 or S-RSM4. The results of curve fitting of these data to the Platt equation (Platt et al., 1980) are shown in Table 3.4. In uninfected cells the rate of CO_2

fixation is linear up until ~50-100 µ mol photons m^{-2} s^{-1} at a rate (α) of 10.31 CPM hr^{-1} Q^{-1}. In comparison α was much depressed during infection with either cyanophage S-PM2 or S-RSM4, 2.01, 2.36 CPM hr^{-1} Q^{-1}, respectively. In addition, large changes in P_{max} were observed with the host reaching a maximum rate of 314.53 CPM hr^{-1}, whereas infection with cyanophage S-PM2 or S-RSM4 reduced this by ~47 and 87% respectively. An interesting observation is that I_k (the light where half of P_{max} is observed) is increased upon infection with cyanophage S-PM2 by 2.3 fold compared with uninfected cells. This may be a reflection of the observed changes to photophysiology shown above. Meanwhile infection with cyanophage S-RSM4 reduced I_k by approximately half (Table 3.4).

Strain	Rel. P_{max} (CPM hr^{-1})	β (µmol photons m^{-2} s^{-1})	α (CPM hr^{-1} Q^{-1})	I_k (µmol photons m^{-2} s^{-1})
Host	314.53	15.28	10.31	21.81
S-PM2	166.00	7.14	2.01	49.39
S-RSM4	40.91	1.00	2.36	9.57

Table 3.4: Photosynthetic parameters in uninfected cells and cells infected by cyanophage S-PM2 and S-RSM4. Photosynthesis parameters were solved by curve fitting of data to the Platt equation. Data is derived from uptake of NaH^{14}CO$_3$. Q represents 1 µmol photons m^{-2} s^{-1}.

3.4. Discussion

3.4.1. LL lysis delay: Mechanisms and consequences

Previous studies have examined the effect of nutrients on phage replication kinetics. The burst size of cyanophage S-PM2 was significantly reduced and the latent period significantly increased when infecting *Synechococcus* sp. WH7803 grown in phosphate deplete growth medium (Wilson et al., 1994). In addition Zeng and Chisholm (2012) have shown that the rate of extracellular phage GCN accumulation is significantly reduced during growth in phosphate deplete compared with replete media in myovirus S-SM1 infecting *Synechococcus* sp. WH8102. Moreover in a mutant lacking the *phoR* gene of the same strain, extracellular phage GCN following lysis is even further supressed (Zeng and Chisholm 2012). PhoR acts as the sensor in the PhoR/PhoB two-component system

regulating the transcriptional response to phosphate stress (Aiba et al. 1993; Aiba and Mizuno 1994; Hirani et al. 2001). These data point to phosphorus availability being a major limitation to cyanophage development. Unfortunately, there is no data to suggest that this limitation is manifested in phage DNA synthesis as suggested by the authors, as no data is included on the rate of accumulation of intracellular GCN. This is interesting because in the supplementary information of the same paper the infection kinetics of the myovirus P-SSM2 infecting *Prochlorococcus* sp. NATL2A was considered. In this case there was no difference in the rate of phage DNA synthesis between phosphate replete and deplete conditions. However, there was a modest increase in the extracellular GCN in phosphate replete compared with depleted conditions. Further, data from the same experiment and as presented in Lindell et al. (2007) and Thompson et al. (2011), suggests that significant degradation of the host chromosome is observed during early infection. As such, this is thought to be the main source of free nucleotides for phage DNA synthesis. Indeed, this appears to be the case for other marine phages (Wikner 1993) as well as the *E. coli* T4 phage (Kozloff et al. 1951; Miller et al. 2003). As such, it is unclear why phosphate may offer a limitation to phage DNA replication when such a readily available source of nucleotides is apparent.

More recently, acidification of the growth medium by addition of CO_2 was shown to affect cyanophage growth parameters (Traving et al. 2014). Increased acidification decreased the latent period and the burst size significantly. Further, Lindell et al. (2007) have shown a remarkable positive correlation between host specific growth rate and phage genome replication rate in the siphovirus P-SSP7 infecting *Prochlorococcus* sp. MED4. However, this conclusion comes with two caveats. Firstly, it is unclear how the host growth rate was modified. Secondly, the rate of genome replication was calculated as the intracellular GCN at 8 hrs after infection divided by that at 1hr. Clearly, such an analysis precludes any effect of growth rate on the length of the latent period.

In this study the effect of light on the kinetics of infection was examined. Light directly affects cellular energy levels in photosynthetic prokaryotes through coupling of the activity

of the photosynthetic transport chain to ATP synthesis (i.e. photophosphorylation) (Falkowski and Raven 2007). This is true providing light can be effectively utilised by the activity of PSII. Indeed upon actinic illumination, ATP levels in cyanobacteria are rapidly increased (Bornefeld and Simonis 1974; Kallas and Castenholz 1982; Lubberding and Schroten 1984), are correlated with light intensity (Bornefeld and Simonis 1974) and are reduced in the presence of inhibitors of photosynthetic electron transport (Bornefeld and Simonis 1974; Lubberding and Schroten 1984). In addition, in the environment, light availability is unique in that it has a predictable oscillatory pattern. Lindell et al. (2005) have shown that darkness prevents phage genome replication in the podovirus P-SSP7 infecting *Prochlorococcus* sp. MED4, as does treatment with an inhibitor of electron transport downstream of PSII. Thus, it was concluded that cyanophages rely on concurrent photophosphorylation for development. Key to this conclusion is the fact that *post-infecto* phosphorylation of ADP is required, thus these changes are not due to the energy status of the cell prior to infection. Therefore, it was predicted that increasing light intensity and therefore ATP synthesis would fuel an increased rate of cyanophage S-PM2 genome replication. However, the data presented here show the rate of phage genome replication was identical between HL and LL conditions (Fig. 3.2). However, a key difference in the infection dynamics between HL and LL is an earlier burst by approximately 5hrs. It is unclear what the mechanism is for the delay in burst under LL. Perhaps the reduction in cellular ATP levels delays the rate of protein synthesis required to produce the structural components of the virion. Clokie et al. (2008) has studied these major structural components through proteomics. These are listed in Table 3.5. The mol equivalent requirement of ATP for polymerisation of pre-formed amino acids into these structural components is calculated below. This is based on 4 ATP equivalents being used for polymerisation of one amino acid into a growing polypeptide chain by ribosome activity as described in Stouthamer (1973).

Protein	S-PM2 ORF	Predicted copy number	Amino acid length	ATP molecules required for polymerisation	mol ATP required per virion $(\times 10^{-20})$
Gp6	83	12	602	28896	4.80
hyp	85	1	175	700	0.12
Gp8	86	12	634	30432	5.05
hyp	89	1	1251	5004	0.83
hyp	90	1	168	672	0.11
hyp	92	1	306	1224	0.20
hyp	93	1	327	1308	0.22
hyp	94	1	379	1516	0.25
Gp14	97	10	292	11680	1.94
Gp15	98	6	266	6384	1.06
Gp18	105	138	743	410136	68.10
Gp19	106	144	204	117504	19.51
Gp20	107	12	564	27072	4.50
Gp22	110	115	392	180320	29.94
Gp23	111	960	468	1797120	298.42
Gp3	113	6	169	4056	0.67
hyp	148	1	316	1264	0.21
hyp	175	1	1095	4380	0.73
hyp	177	1	1177	4708	0.78
Gp48	203	6	332	7968	1.32
hyp	222	1	295	1180	0.20
hyp	224	1	560	2240	0.37
hyp	226	1	1037	4148	0.69

Table 3.5: Summary of major structural components of the S-PM2 virion as reported in Clokie et al. (2008). The copy number is hypothetical based on comparison to T4 or where homology to T4 is absent the copy number is conservatively assumed to be one. Four mol of ATP is required to add one amino acid to a growing polypeptide chain according to Stouthamer (1973).

These data suggest that 4.4×10^{-17} mol of ATP is required for polymerisation of the amino acids of the major structural components of the S-PM2 virion per infection. This is based on an average burst size of 10. This is an extremely conservative estimate given that 1. It is unlikely that all structural components have been identified 2. The copy number of the hypothetical structural proteins are assumed to be 1. 3. This does not include the number of ATP molecules required to synthesise each amino acid *de* novo. The number of ATP molecules consumed/produced for each *de novo* synthesis of amino acid from glucose is shown in Table 3.6

Amino Acid	Symbol	Frequency of amino acids in structural proteins	ATP Requirement for synthesis of 1 mol of amino acid	ATP molecules required for *de novo* synthesis	mol of ATP required per virion (x10^{20})
Alanine	A	767	1	767	0.13
Arginine	R	424	-3	-1272	-0.21
Asparagine	N	856	-2	-1712	-0.28
Aspartic Acid	D	734	0	0	0.00
Cysteine	C	34	-3	-102	-0.02
Glutamic Acid	E	610	1	610	0.10
Glutamine	Q	365	0	0	0.00
Glycine	G	1120	0	0	0.00
Histidine	H	85	-7	-595	-0.10
Isoleucine	I	833	-1	-833	-0.14
Leucine	L	759	3	2277	0.38
Lysine	K	414	0	0	0.00
Methionine	M	137	-4	-548	-0.09
Phenylalanine	F	523	-2	-1046	-0.17
Proline	P	482	0	0	0.00
Serine	S	1086	0	0	0.00
Threonine	T	1107	-2	-2214	-0.37
Tryptophan	W	106	-5	-530	-0.09
Tyrosine	Y	472	-2	-944	-0.16
Valine	V	838	2	1676	0.28

Table 3.6: Amino acid frequency in virion structural proteins of cyanophage S-PM2 and theoretical ATP requirements of synthesis. Theoretical ATP requirements are from glucose and are reported in Stouthamer (1973).

This calculation suggests that 7.4 x 10^{-20} mol ATP is required per infection to synthesise just the amino acids of the major virion components. This is ~roughly 3 orders of magnitude less than that required for polymerisation so is deemed negligible.

Computing the ATP requirement for phage DNA replication is made difficult due to considerable doubt over the sources of dNTPs. If one assumes that every dNTP of the S-PM2 genome is synthesised *de novo*, according to the predicted purine and pyrimidine biosynthetic pathways of *Synechococcus* sp. WH7803 shown in Fig. 6, it can be calculated that the mol ATP requirement for dNTP synthesis is 2.2 x 10^{-17} mol ATP. This is again based on a burst size of 10 and computed according to the exact dNTP frequency of the S-PM2 genome. However, given that a potential source of dNTPs is host chromosomal

DNA/RNA and free DNA/RNA, the ATP requirement is expected to be much less. Cyanophage S-PM2 encodes multiple enzymes involved in nucleotide metabolism. For example an exonuclease (YP_195231.1), a ribonucleotide reductase A (YP_195185.1), two copies of ribonucleotide reductase B (YP_195176.1, YP_195186.1) and a thymidylate synthetase (YP_195228.2). Therefore, estimation of the ATP requirement for genome replication is probably significantly less than described above. In any case it is clear that by far the most energetically costly process is synthesis of the structural components of the virion. Thus, it can be concluded that under nutrient replete conditions, the rate limiting step in cyanophage development is the energy required to synthesise the structural components of the virion and not genome replication.

In opposition to this argument, data from Traving et al. (2014) suggests that in cyanophage S-PM2 the eclipse period is 5.7 hrs at LL (35 μE m^{-2} s^{-1}). The eclipse period marks the time when the first mature virion is formed inside the infected cell and can be calculated by artificially lysing the cells and through detection of viable phage particles by plaque assay. Further, in this experiment, the latent period under LL was calculated to be 12.3 hrs, which is remarkably similar to data presented here. Thus, the so called maturation period represents 50% of the latent period under LL. In comparison, T4 has been shown to have an eclipse period of 18.2 minutes compared with a latent period of 23.6 min (Rabinovitch et al. 1999). Thus, the maturation period only occupies 23% of the latent period. If one assumes the eclipse period remains at 5.7 hrs in cyanophage S-PM2 under HL and the latent period has been shown to be between 7-8 hr (Fig. 3.2) this would result in the maturation period being considerably more similar to that of T4, between 19- 29% of the latent period. This extended maturation period could therefore suggest that lysis inhibition (Young 1992) may be occurring in cyanophage S-PM2 infections at LL.
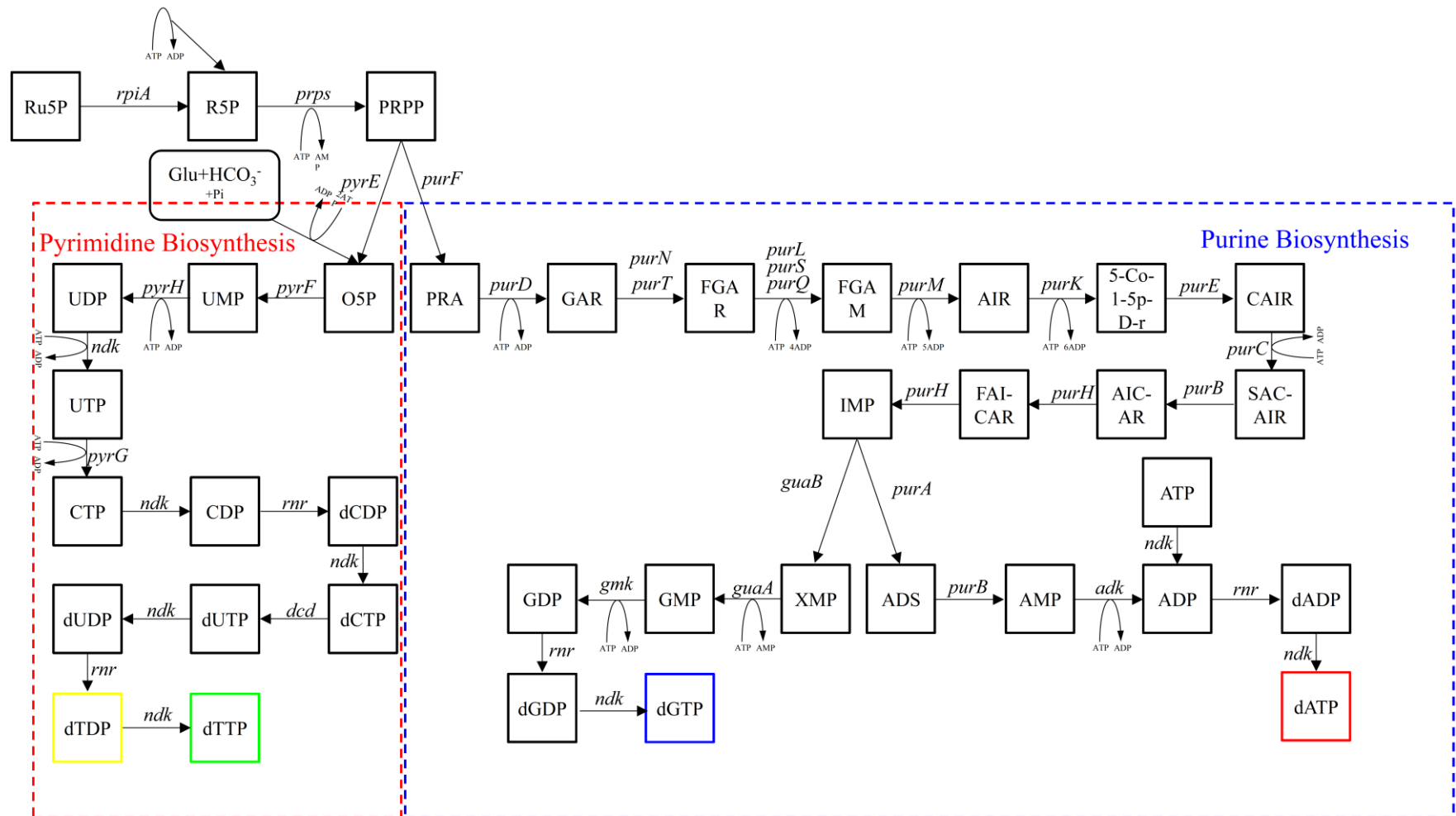
Figure 3.6: Predicted purine and pyrimidine biosynthetic pathways in *Synechococcus* sp. WH7803, used to calculate the theoretical ATP requirement of *de novo* dNTP biosynthesis for phage genome replication. Prediction of the biosynthetic pathways are based on the gene content of *Synechococcus* WH7803 and known biosynthetic pathways as reported in the KEGG database. Substrates: Ru5P= Ribulose-5-phosphate, R5P= Ribose 5-phosphate, PRPP= Phosphoribosyl pyrophosphate, O5P= Orotidine 5'-phosphate, UMP= Uridine monophosphate, UDP= Uridine diphosphate, UTP= Uridine triphosphate, CTP= Cytidine triphosphate, CDP= Cytidine diphosphate, dCDP= Deoxycytidine diphosphate, dCTP= Deoxycytidine triphosphate, dUTP= Deoxyuridine triphosphate, dUDP= Deoxyuridine diphosphate, dTDP= Deoxythymidine diphosphate, dTTP= Deoxythymidine triphosphate, PRA= Phosphoribosylamine, GAR= Glycineamide ribonucleotide, FGAR= Phosphoribosyl-N-formylglycineamide, FGAM= 5'-Phosphoribosylformylglycinamidine, AIR= 5-Aminoimidazole ribotide, 5-Co-1-5p-D-r= 5-Carboxyamino-1-(5-phospho-D-ribosyl)imidazole, CAIR= 5'-Phosphoribosyl-4-carboxy-5-aminoimidazole, SACAIR= Phosphoribosylaminoimidazolesuccinocarboxamide, AICAR= 5-Aminoimidazole-4-carboxamide ribonucleotide, FAICAR= 5-Formamidoimidazole-4-carboxamide ribotide, IMP= Inosine monophosphate, XMP= Xanthine monophosphate, GMP= Guanosine monophosphate, GDP= Guanosine diphosphate, dGDP= Deoxyguanosine diphosphate, dGTP= Deoxyguanosine triphosphate, ADS= Adenylosuccinate, AMP= Adenosine monophosphate, ADP= Adenosine diphosphate, ATP= Adenosine triphosphate, dADP= Deoxyadenosine diphosphate, dATP= Deoxyadenosine triphosphate. Genes: *rpiA*= Ribose-5-phosphate isomerase, *prps*= Phosphoribosyl pyrophosphate synthetase, *pyrE*= Orotate phosphoribosyltransferase, *pyrF*= Orotidine-5'-phosphate decarboxylase, *pyrH*= Uridylate kinase, *ndk*= Nucleoside diphosphate kinase, *pyrG*= CTP synthetase, *rnr*= ribonucleotide reductase, *dcd*= dCTP deaminase, *purF*= Amidophosphoribosyltransferase, *purD*= Phosphoribosylglycinamide synthetase, *purN*= Phosphoribosylglycinamide formyltransferase, *purT*= Glycinamide ribonucleotide transformylase, *purL*= Phosphoribosylformylglycinamide synthase subunit, *purS*= Phosphoribosylformylglycinamidine synthetase subunit, *purQ*= Phosphoribosylformylglycinamidine synthase subunit, *purM*= Phosphoribosyl-aminoimidazole synthase, *purK*= N5-carboxyaminoimidazole ribonucleotide synthase, *purE*= N5-carboxyaminoimidazole ribonucleotide mutase, *purC*= Phosphoribosyl-aminoimidazole-succinocarboxamide synthase, *purB*= Phosphoribosyl-aminoimidazole-succinocarboxamide synthase, *purH*= Phosphoribosylaminoimidazolecarboxamide formyltransferase, *guaB*= Inosine-5'-monophosphate dehydrogenase, *purA*= Adenylosuccinate synthase, *guaA*= GMP synthase, *gmk*= Guanylate kinase, *purB*= Adenylosuccinate lyase, *adk*= Adenylate kinase.

In bacteriophages, lysis is a tightly regulated event involving three processes, each responsible for compromising the components of the cell envelope: the cytoplasmic membrane (CM), the peptidoglycan layer and the outer membrane (for review see Young 2014). These individual processes are facilitated by the holin, endolysin and spanin respectively. The generally accepted model suggests that timing of the lysis event is regulated by the activity of the holin (Catalão et al. 2013; Young 2013; Young 2014). Recently, the precise workings of this mechanism have been elucidated in both canonical (White et al. 2011) and pinholin-SAR type holin-endolysin systems (Xu et al. 2004b; Sun et al. 2009; Kuty et al. 2010). The foundation of temporal control of lysis by holins surrounds the concept of a 'critical concentration'. Holins are the most diverse set of proteins that carry out a single function, with over 50 unrelated families reported (Young 2002). Generally they are small, contain single or multiple transmembrane domains, are localised to the CM and are mobile (Young 2013). The process of lysis proceeds with late transcription of both of the genes encoding the holin and the endolysin. During phage morphogenesis the holin accumulates in the CM, yet appears to have no effect on membrane integrity or proton motive force (PMF). In canonical holin-endolysin systems it is thought that once the critical concentration is met, the holins undergo rapid oligomerisation. This oligomerisation causes a local depolarisation of the membrane. Depolarisation causes a conformational change in the holin structure which promotes formation of micron scale holes in the membrane. The endolysin can exit through this hole and begin to degrade the peptidoglycan layer. In this way, the accumulation of the holin in the CM 'sets the clock' on the timing of lysis.

Cyanophage S-PM2 contains a putative endolysin (YP_195189.2) (Mann et al. 2005) but unsurprisingly, due to their diversity, a holin has not been identified. However, due to the ubiquity of this machinery in dsDNA bacteriophages (Young 2014) it is the likely mechanism of lysis in cyanophage S-PM2. One can envisage several pathways that may lead to inhibition of lysis. The holin critical concentration is unlikely to be changed. However, the rate of accumulation of the holin in the membrane may be slower in LL due to

the aforementioned energy demand of translation. Indeed in bacteriophage λ the critical concentration is thought to be between 1000-3000 monomers (Zagotta and Wilson 1990; Chang and Nam 1995; Young 2014). Alternatively, active repression of holin oligomerisation may occur. In the case of bacteriophage T4, the RI protein is activated upon binding and ectopic injection of the contents of the virion capsid of an exogenous phage particle into the periplasm. RI then binds to the globular periplasmic domain of the single TMD containing T4 holin, T. This prevents CM hole formation and thus lysis is inhibited. Thus RI has been termed an antiholin (Moussa et al. 2012). Clearly, this response is favourable in conditions when a subsequent host for the progeny phage may be limiting. Such a mechanism could be envisaged in cyanophage S-PM2 with the antiholin being light responsive.

The ecological consequences of LL induced delayed lysis are profound. Given that cyanophage development is greatly inhibited by darkness and electron transport inhibitors (Lindell et al. 2005) and that cyanophage S-PM2 itself displays a light-dependent adsorption phenotype (Jia et al. 2010), successful phage infection must only occur during the day. Fig. 3.7 shows a hypothetical model of how LL lysis inhibition may manifest in the infection cycle during a diurnal light regime. Adsorption of a free cyanophage to its host occurs just after dawn when enough quanta affords light dependent adsorption. Successful morphogenesis of virions then occurs during the day when energy derived from photophosphorylation and maintenance of 'cyanophage photosynthesis' can sustain phage production. Mature virions are produced by late evening but reduced quanta characteristic of dusk prevents lysis by LL lysis delay. It is speculated here that a shift from LL to darkness will completely inhibit lysis instead of just a delay. This hypothesis can clearly be tested in future experiments. The cyanophage then maintains in a state of lysis inhibition throughout the night, until when a critical light threshold allows lysis. Thus, progeny phage can then adsorb to hosts through the yet uncharacterised light dependent adsorption mechanism. The antithesis of LL lysis delay would result in lysis during the evening. Enough quanta may still

be available to allow adsorption of progeny phage to their host and successful injection of the virion contents. However, the phage may be inhibited from replicating inside the host because of energy limitation in the absence of photophosphorylation. This hypothesis could be tested in laboratory cultures through the use of a cyclostat (Bruyant et al. 2001; Mella-Flores et al. 2012) and indeed in the environment through the new technology of viral tagging (Deng et al. 2012). Evidence of this process in the environment is conflicting. For instance infections of the freshwater cyanobacteria *Microcystis aeruginosa* by the lytic Ma-LMM01-type cyanophage appear to follow the diurnal pattern described above (Kimura et al. 2012). In comparison, Millard et al. (2006) have shown that in the top 10m of the Indian Ocean, there is a peak in free phages that can infect *Synechococcus* sp. WH7803 that occurs during the night. It has been suggested that these patterns may reflect the light dependent adsorption and development during the day but delayed lysis due to nutrient limitation (Millard et al. 2006). Moreover, reduced viral decay rates during dark periods may play a role (Millard et al. 2006).
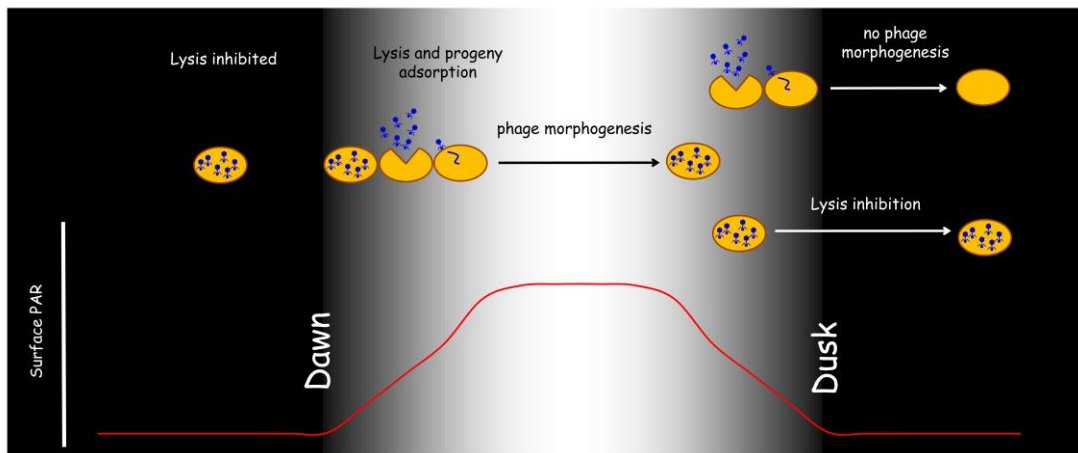
Figure 3.7: Schematic representing the ecological advantage of the proposed LL lysis delay. See text for discussion. Orange ovals represent *Synechococcus* hosts. The model is superimposed on a theoretical equatorial diurnal day. The red line represents the surface photosynthetically active radiation (PAR).

3.4.2. Cyanophage S-PM2 maximises energy generation through modification of the photosynthetic electron transport (PET) chain.

Given the speculated 'lifestyle' of cyanophage S-PM2, infection is predicted to be occurring during the day. The energy required for successful transcription, translation and assembly of

the virion is thought to be derived from photophosphorylation. Given that photophosphorylation is a light dependent process, one might be persuaded that the high light waters of the upper ocean would benefit the development of cyanophages in their hosts. This may be true but only providing that cells can 'make use' of this increased flux of radiation. In practice such high irradiance can severely impair the photosynthetic electron transfer efficiency of marine picophytoplankton as is evidenced in numerous laboratory (Bruyant et al. 2001; Bruyant et al. 2005; Blot et al. 2011; Mella-Flores et al. 2012) and field studies (Behrenfeld 1999; Behrenfeld et al. 2006b). The daily decline in photosynthetic quantum yield in these studies can be explained by the presence of photodamage to PSII, which occurs at a rate that exceeds that of repair (for review see Falkowski and Raven 2007, Nixon et al. 2010). Here, it is shown that shifting of *Synechococcus* sp. WH7803 to ~210 μE m^{-2} s^{-1} from ~20 μE m^{-2} s^{-1} reduces F_v/F_m by approximately 30% (Fig. 3.3) and is comparable to other studies (Garczarek et al. 2008). This decrease is shown to be driven by photoinduced damage to PSII that exceeds the rate of repair (Fig. 3.1) (Garczarek et al. 2008). Interestingly, infection by cyanophage S-PM2 alleviates the photoinduced reduction in F_v/F_m by approximately 20% (Fig. 3.3). This result suggests that infection with cyanophage S-PM2 results in increased tolerance to light damage of PSII.

Cyanophages have been shown to encode their own orthologues of genes involved in PSII functioning (Mann et al. 2003; Lindell et al. 2004b). Two PSII genes are widespread among cyanophages, *psbA* and *psbD* (Puxty et al. *submitted*). Whilst the functions of the phage encoded *psbA* and *psbD* genes are yet to be explicitly proven, it is likely, due to sequence similarity, that they serve the same function as the host homologue (Sullivan et al. 2006). Together, the PsbA (D1) and PsbD (D2) proteins form the heterodimeric core of the PSII reaction centre (Zouni et al. 2001; Umena et al. 2011). Photoinhibition occurs when the rate of photo-induced damage to the D1 polypeptide exceeds the rate at which *de novo* synthesis can replace it to maintain steady state photosynthesis (Nixon et al. 2010). Thus, the degradation and repair rates to D1 regulate the level of photoinhibition. Here, it is speculated

that the either the cyanophage S-PM2 encoded D1 has a decreased rate constant of photoinduced damage, or the rate of D1 repair is greatly increased during infection. Elucidating which one of these may be true is made difficult by the fact that during infection the exact inventory of host D1 vs. S-PM2 D1 is unknown, and is likely to be affected by light intensity (Bragg and Chisholm 2008; Hellweger 2009). Attempts to create a heterologous strain of the freshwater cyanobacterium *Synechocystis* sp. PCC6803 containing only the cyanophage S-PM2 or host D1, were unsuccessful (data not shown). Should this be made possible, the exact rate constant of damage and repair could be measured, albeit within the background of the heterologous strain.

Interestingly, the cyanophage S-PM2 D1 and other cyanophage D1s, including those that can be attributed to cyanophages from metagenomic datasets, differ from cyanobacterial-like D1 proteins in two regions (Sharon et al., 2007): Firstly, the stromal loop between transmembrane helices D and E. This region has been shown to be the site of primary proteolytic cleavage by the protease DegP2 in the model plant *Arabidopsis thaliana* (Haussühl et al. 2001). However, in the model cyanobacterium *Synechocystis* sp. PCC6803, the DegP/HtrA proteases are thought to be unimportant in PSII repair (Nixon et al. 2005) given a mutant lacking all three chromosomally encoded Deg proteases was shown to have an intact PSII repair cycle (Barker et al. 2006). Instead, the FtsH protease has been shown to be essential for maintaining normal rates of D1 degradation *in vivo* with no accumulation of D1 breakdown products observed in a *ftsH* mutant (Silva et al. 2003). This protease has an altogether different mode of action in which damaged D1 is processively degraded from the N-terminus (Komenda et al. 2007). In such a case, the cyanophage specific residues are unlikely to be important in regulating the rate of primary proteolytic cleavage of D1. As such, the function of these cyanophage-specific modifications during infection remains enigmatic. The second site of sequence variability is the loop at the C-terminal end of transmembrane helix E. This variable region occurs at a loop which likely interacts with cytochrome b_{559} (*psbE/F*) and cytochrome c550 (Umena et al. 2011). Whilst the function of

the cytochrome c550 complex is largely unknown (Suga et al. 2013), cytochrome b_{559} may act to scavenge electrons from reduced P680 to prevent photo-oxidation of the last antennae chlorophyll (Thompson and Brudvig 1988). As such, cyclic electron flow through cytochrome b_{559} may protect PSII from photoinhibition. Evidence to support this theory comes from the observed changes in P-I curves during infection with cyanophage S-PM2 (Fig. 3.5). In Fig. 3.3 the mitigation of HL induced damage to the PSII reaction centre by cyanophage S-PM2 infection can be explained by a difference in either D1 damage or D1 repair rate as discussed above. This does not hold true for the differences in the effective quantum yield shown in Fig. 3.4. The acclimation periods in this experiment (30s) are not long enough to induce photodamage to D1. Instead, the observed decline in Φ_{PSII} with increasing periods of actinic illumination, result from serial reduction of plastoquinone (PQ) under that irradiance. Once the PQ pool becomes reduced, P680 fluorescence cannot be quenched by Q_A (Suggett et al. 2010). During infection with cyanophage S-PM2 Φ_{PSII} is greater under higher actinic irradiances compared with uninfected cells suggesting that PQ re-oxidation must be occurring at a rate faster than in the host. Many studies have demonstrated that reduced PQ can donate electrons to cytochrome b_{559} and as such the cyanophage specific residues that are predicted to interact with cytochrome b_{559} may be responsible for accepting electrons from PQ during HL conditions. The re-oxidation of PQ appears to be a general strategy amongst cyanophages as is evidenced by the disproportionate maintenance of the gene encoding plastoquinol terminal oxidase (PTOX) in cyanophage genomes (Millard et al. 2009). PTOX is thought to oxidise PQ and subsequently donate electrons back to O_2 thus forming a pseudo water-water cycle (Bailey et al. 2008). The result is proton translocation across the thylakoid, generating a transmembrane potential for ATP synthesis. The redirection of electrons away from the cytochrome b_6f complex would reduce linear electron flow to NADP and thus suspend CO_2 fixation. Exactly this phenotype is observed in CO_2 fixation experiments (Fig. 3.5). Cyanophage S-PM2 does not encode a homologue of PTOX, and neither does it's *Synechococcus* host strain WH7803, but it is conceivable that an analogous protein may be functioning in the same way.

Regardless of the mechanisms involved, the increased rate of re-oxidation of PQ at HL combined with a PSII that is less susceptible to photodamage suggests that the flow of electrons within the PET chain is increased. Whilst the terminal acceptor(s) remain elusive, the action of charge separation in PSII and PQ reduction-oxidation results in an increased PMF. Thus, it is suggested here that cyanophages act to bolster the light dependent reactions to generate an increased rate of ATP synthesis for phage morphogenesis.

3.4.3. Cyanophage infection decouples photosynthesis: Implications for the assessment of ocean primary production

In recent years it has been shown that cyanophages frequently encode genes involved with redirection of central carbon metabolism (Thompson et al. 2011). This has led to the hypothesis that during infection, energy (ATP) generated from photophosphorylation is used for nucleotide metabolism by cyanophages and is not used for CO_2 fixation in the Calvin cycle. Whilst the activity of at least one of the genes involved in the redirection of carbon metabolism (*talC,* Thompson et al. 2011) has been proven, albeit in *E. coli*, the hypothesis in general remains to be rigorously tested. Here, it is shown that during infection CO_2 fixation is indeed inhibited. In the case of cyanophage S-PM2, this occurs at roughly 4 hrs after initial infection, and therefore *Synechococcus* sp. WH7803 infected with cyanophage S-PM2 fixes at least 2.28 times less carbon than uninfected cells. Cyanophage S-PM2 contains none of the genes required for redirection of carbon metabolism and therefore some other mechanism must be responsible for the cessation of the Calvin cycle. It is speculated that this may result from the cessation of host protein synthesis upon infection, which is common during bacteriophage infection (Nomura et al. 1962). Alternatively, the flow of electrons in the PET may be diverted away from NADP in, for example, cyclic photophosphorylation (Falkowski and Raven 2007) or some other intersystem terminal acceptor. In comparison, cyanophage S-RSM4 does contain the genes *cp12, talC, gnd* and *zwf* (Millard et al. 2009). During infection with this phage, CO_2 fixation is ceases 2hrs after infection and thus there is a reduction of CO_2 fixation of at least 4.79 times during infection with this phage (Fig. 3.5).

Therefore, it is likely that acquisition of these genes by cyanophage S-RSM4 contributes to the rapid shut-down in CO_2 fixation observed. Unfortunately, the kinetics of infection of cyanophage S-RSM4 remain unknown and were not measured during this study. Therefore, it is hard to predict whether acquisition of these genes has significantly benefited the phages 'physiology' in terms of replication kinetics.

The ocean is responsible for approximately 50% of the CO_2 fixation of the biosphere (Field 1998). This estimate is subject to error due to a reliance on sparse discrete *in situ* measurements of ocean production (Longhurst et al. 1995; Behrenfeld et al. 2005; Lawrenz et al. 2013). The current best accepted estimate of primary production comes from *in-situ* measurements of $^{14}CO_2$ uptake (Lawrenz et al. 2013). These measurements themselves are subject to large degrees of technical difficulty which both clouds assessment of their accuracy and limits their use spatially and temporally. A growing trend among researchers is to use fluorescence based measurements of the electron transfer capacity of the PET to derive estimates of ocean productivity (Suggett et al. 2010). Such measurements can be made over large spatial scales and fine temporal resolution (Behrenfeld 1999; Behrenfeld et al. 2006b). Their application relies on an estimate of the rate of linear electron transport, that is the flow of electrons from water to NADPH, with concomitant proton translocation across the thylakoid membrane. Such activities provide the energy (ATP) and reductant (NADPH) to drive the Calvin cycle. Thus, linear electron transport is quantitatively related to the rate of CO_2 fixation. However, many processes act to 'decouple' linear electron transport form CO_2 fixation by consuming electrons, ATP or reductant. These processes are reviewed in Lawrenz et al. (2013) and include alternative electron pathways within the PET, oxygenase activity by RuBisCO and also nutrient assimilation. Here, we extend this growing list of 'photosynthetic decouplers' to include cyanophage infection. The redirection of ATP and reductant to fuel phage morphogenesis would lead to overestimates of fluorescence derived ocean productivity if cyanophage infection occurred at large scales. This problem is exacerbated by the fact that their hosts, *Prochlorococcus* and *Synechococcus* are the most

numerous photoautotrophs on our planet (Flombaum et al. 2013). Estimates of cyanophage infection vary wildly, over 5 orders of magnitude (Suttle and Chan 1993; Waterbury and Valois 1993; Wilhelm and Suttle 1999; Fuhrman 1999; Wommack and Colwell 2000; Suttle 2007). As such more sophisticated methods must be developed to accurately assess the rates of cyanophage infection so as to quantify the magnitude of this alternative electron pathway.

Chapter Four: Transcriptional landscape of the photosynthetic virus S-PM2

4.1. Introduction

The development of bacteriophages in their hosts requires temporal regulation of gene expression. The cascade of expression of bacteriophage genes has been studied in the model myovirus T4 (Luke et al. 2002) and the molecular machinery that regulates the transition from various phases is well understood (see Miller et al. 2003). Interestingly many of the regulatory features of T4 are shared in cyanophages. In particular, all the proteins (Gp55, Gp47 and Gp33) required for modification of the host's RNA polymerase to recognise the T4 late promoter are present (Mann et al. 2005). However, cyanophage S-PM2 has been shown to lack any of the genes required for the so called middle profile of expression (Mann et al. 2005). Thus, it has been suggested that cyanophage S-PM2 possesses a bi-modal regulation of transcription (Mann et al. 2005). Another layer of complexity is added in cyanophages due to the presence of AMGs (Mann et al. 2005; Millard et al. 2009; Sullivan et al. 2010). These AMGs are acquired, often from their hosts (Sullivan et al. 2006), and are suspected to provide some fitness benefit to the phage. However, these genes also require regulation such that expression occurs at a level, and in a temporal fashion, that provides a fitness-benefit. Thus, the genomic context and evolution of AMGs may be constrained by the transcriptional landscape of the phage.

The advent of next generation sequencing methods, and in particular RNA sequencing (RNA-Seq), has revolutionised the study of prokaryotic transcriptomes (Güell et al. 2009; Sharma et al. 2010b; Mitschke et al. 2011). Despite this, few studies have employed this technology to study the transcriptomes of bacteriophages. Currently only two phage transcriptomes have been reported: that of the lytic podovirus LUZ19 infecting *Pseudomonas aeruginosa* PAO1 (Luz et al. 2013) and the temperate mycobacteriophage, Giles (Dedrick et al. 2013). In this chapter the first example of the transcriptome of a T4-like myovirus, the cyanophage S-PM2, is presented. The aims of this chapter are as follows:

1. To examine the temporal expression of phage genes during infection

2. To validate the transcriptional model that has been proposed for cyanophage S-PM2. In particular, the existence of bi-modal global transcriptional regulation.

3. To discover novel *trans* and *cis* acting RNA elements that may fine tune gene expression beyond that of global regulation.

4.2. Methods

4.2.1. Cyanophage infection of Synechococcus sp. WH7803

Synechococcus sp. WH7803 was grown in ASW medium (section 2.12) to 1×10^8 ml^{-1} at 23˚C under continuous illumination of 15 μmol photons m^{-2} s^{-1}. Infection was carried out with cyanophage S-PM2 at a virus: bacteria ratio (VBR) of 10. Samples were incubated under low light (15 μmol photons m^{-2} s^{-1}) conditions and 50 ml samples were taken before infection and at 1, 3, 6, 9 hrs after infection for total RNA extraction. RNA extraction followed the protocol described in section 2.1.7.

4.2.2. Removal of rRNA

Attempts were made to simultaneously remove rRNA and enrich for primary transcripts by using Terminator^{TM} 5' monophosphate dependent nuclease (Epicentre-Illumina, Madison, USA) (Sharma et al. 2010a). 5 μg total RNA, 2 μl 10x Terminator reaction buffer A or B, 0.5μl RiboGuard RNase Inhibitor, 1-2U Terminator Exonuclease and nuclease free water to 20 μl were incubated at 30˚C for 60 min. The reaction was stopped by phenol chloroform extraction and ethanol precipitation (see section 2.1.10). Purified RNA was resuspended in 10 μl nuclease-free water and 1 μl was run on a Bioanalyser Nano-chip (Agilent, Santa Clara, U.S.A.) following the manufacturer's instructions. The amount of rRNA removal was estimated by comparing the area under the 16S and 23S rRNA peaks before and after treatment and are shown in appendix 2. 2U of Terminator^{TM} Exonuclease in buffer A suggested the greatest removal of rRNAs and so was used for all time points.

4.2.3. Library preparation and sequencing

RNA-Seq libraries were prepared for an uninfected control (T0) and for 1, 3, 6 and 9 hr after infection as well as a sample from T9 that contained no treatment with Terminator[TM] Exonuclease as a control. Directional RNA-Seq libraries were prepared using the ScriptSeq[TM] v2 library preparation kit (Fig. 4.1, Epicentre-Illumina, Madison, USA). by the University of Exeter Sequencing Service (http://biosciences.exeter.ac.uk/facilities/sequencing/).
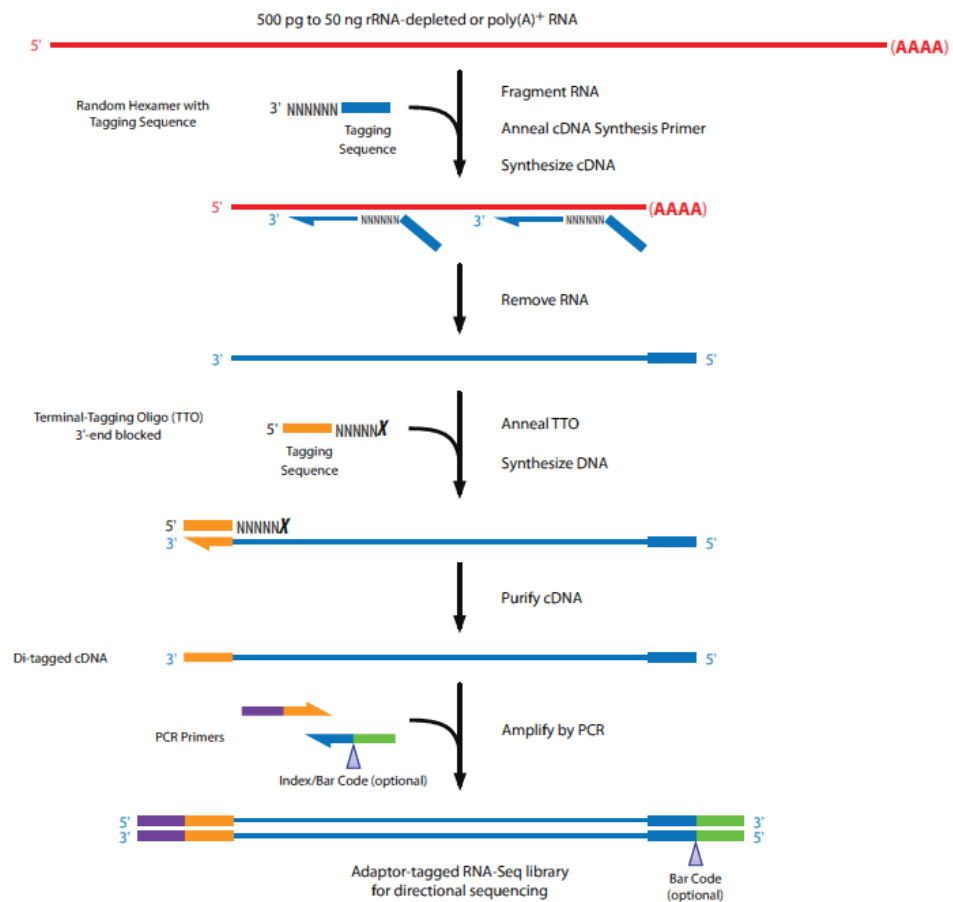
Figure 4.1: Overview of the ScriptSeqTM v2 library preparation protocol. Image courtesy of Epicentre-Illumina and can be found at http://www.epibio.com/docs/default-source/protocols/scriptseq-v2-rna-seq-library-preparation-kit.pdf?sfvrsn=10

The ScriptSeq[TM] v2 library preparation was chosen for several reasons. Firstly, the protocol retains the 'strandedness' of the RNA fragment. This is achieved through a 5' tagging

sequence on the random hexamer used to prime initial reverse transcription (Fig. 4.1). This capability was exploited in an attempt to discover novel *cis* acting RNA elements. Secondly, the protocol should maintain the 5' end of the RNA fragment. This is achieved through the use of the terminal tagging oligonucleotide (TTO). The TTO anneals randomly to the cDNA fragment. However, the presence of the 5' tagging sequence of the TTO prevents stable annealing to the cDNA apart from at the 3' end of the cDNA fragment. The TTO contains a blocking nucleotide on its 3' end such that DNA polymerase activity cannot create a double stranded DNA. Instead, the 5'-3' polymerase activity of DNA polymerase acts to 'tag' the end of the cDNA fragment which is then amplified by PCR (Fig. 4.1). 5' ends of cDNA fragments should occur through three processes: Firstly, the fragmentation of the RNA in the fragmentation stage. Chemical RNA fragmentation is a random process (i.e. is sequence independent). Therefore, one would not expect the 5' end of the read to map with any bias. The second is the result of RNA processing and in particular by RNase cleavage (Lasa et al. 2011). Lastly, the transcription initiation at a TSS is a directed process and occurs at specific nucleotides. Therefore using this protocol, it was hypothesised that TSSs could be mapped. Lastly, the library preparation allows for paired end sequencing which offers significant power when attempting to reconstruct full length transcripts.

Libraries were sequenced on the HiSeq 2000 platform using paired end sequencing generating reads of 100 bp. Sequencing was performed by the University of Exeter Sequencing service (see above).

4.2.4. Sequence mapping

Reads were initially mapped to the cyanophage S-PM2 genome sequence (acc. no. NC_006820) using Bowtie 2 (Langmead and Salzberg 2012). Mapping was done in –end-to-end mode using the following parameters: -D 20 -R 3 -N 0 -L 20 -i S,1,0.50. Mapping was in paired-end mode using the –fr tag assuming no maximum fragment length. SAM files were converted to BAM files and sorted using SAMtools (Li et al. 2009). Initial analysis

revealed a large deletion that had occurred within cyanophage S-PM2 (see section 4.3.1) and thus reads were re-mapped to a new cyanophage S-PM2 genome file that lacked the deleted nucleotides. RNA fragments were reconstituted from paired-end reads using a custom script written in MATLAB (MathWorks, Natick, U.S.A.), and is available upon request. The script first filters reads that were mapped in a 'proper pair' whereby the first of the pair was on the opposite strand to the second and where both reads have a MAPQ score greater than 30. For each paired-end read, the genomic coordinate from the start site of the first mapped read to the end of the second mapped read was written to a fasta file. This fasta file was then re-aligned to the genome using the same parameters as previous. This created read fragments where base and read quality are lost. The same process was applied to map reads to the *Synechococcus* sp. WH7803 genome (acc. no. NC_009481).

4.2.5. Validation of deletion by PCR

To validate the identified deletion in S-PM2 (section 4.3.1.), PCR was used. The strategy for PCR is described in section 4.3.1. PCR primers are listed in appendix 1 (Flank1_F/R, Flank2_F/R, Flank3_F/R, Flank4_F/R, Del1_F/R, Del2_F/R, Del3_F/R). PCRs reactions were in 50 μl and contained ~45 ng S-PM2 DNA/1 μl random primed cDNA (section 6.2.2.), 1x MyTaq master mix (BioLine, London, UK), 0.4 μM each primer and water to 50 μl. PCRs on isolated plaques were accomplished by aspiration of the sloppy agar containing the plaque into 100 μl of ASW in each well of a 96 well plate. 1 μl was subsequently used for PCR template. Annealing was at 55°C and ran for 35 cycles.10 μl of PCR product was run on a 1% (w/v) agarose gel.

4.2.5. Transcript and TSS prediction

It is desirable to assemble RNA-Seq reads into a 'transcriptional map'. This can be viewed as a description of the transcriptional units of an organism (McClure et al. 2013) or indeed a list of genome wide TSSs (Sharma et al. 2010a; Mitschke et al. 2011).

Very few tools exist that can assemble fragmented cDNA reads to predict transcript boundaries for bacteria. Recently, the Rockhopper pipeline has been developed (McClure et al. 2013). This acts to use transcript 'seeds' from known annotated protein coding genes. The seeds are then extended both upstream and downstream from the ORF using a Bayesian principal. The probability of the up/downstream read is based on comparison to the level of transcription antisense to protein coding genes. The Rockhopper pipeline was run on paired end reads mapped to the cyanophage S-PM2 genome with default settings. The results of Rockhopper processing were not suitable for analysis of the cyanophage S-PM2 transcriptome given that *a priori* knowledge of ORF boundaries biased transcript boundary assignment. This was manifested in the fact that transcript boundaries were assigned to -1 to +1 relative to the ORF start and stop codon respectively. Instead, a tool was sought that attempted to map transcriptional boundary *de novo* i.e. with no *a priori* knowledge of protein coding genes. Thus, the Cufflinks pipeline (Trapnell et al. 2012) was implemented using default settings with the –multi-read-correct option.

TSSs were predicted based on a read start site bias score (RSS score) as introduced in Voss et al. (2013). For each base in the genome the number of reads whose 5' end mapped to that nucleotide was counted. The RSS is a ratio of this number to the number of reads that cover the base.

4.2.6. Estimation of expression

Expression estimates are inferred from read counts mapped to specific loci. Several methods for read counting are used in this study. For the estimation of expression of ORFs and genomic regions the program HTSeq-count was used (Anders et al. 2014). This program features 3 counting methods: Union, Intersection-Strict and Intersection-Nonempty (Anders et al. 2014). These methods differ in the treatment of reads that overlap multiple genomic regions. Due to uncertainty in the correct method to use, all three were employed. Unless indicated, the read count represents the mean of all three estimates. For the calculation of per

base read counts the method 'getBaseCoverage' was employed in MATLAB (MathWorks, Natick, U.S.A.). This calculates the numbers of reads that overlap the specific locus. Read counts were normalised to the library size and indeed the size of the genomic region using the reads per kilobase of exon per million mapped reads (RPKM) model (eq. 4.1.) giving an estimate of relative expression (R.E.).

$$R.E. = RPKM = \frac{(R_c/(l/1000))}{(T_R/1x10^6)} \text{ [Eq. 4.1]}$$

Where R_c is the read count of the locus, l is the length of the locus and T_R is the total number of mapped reads.

4.2.7. Clustering of gene expression estimates

Genes were clustered according to their RPKMs from T0 (Uninfected), T1, T3, T6 and T9. Clusters were visualised using hierarchical clustering implemented on relative expression estimates within the MATLAB environment. Dissimilarity matrices were generated using the 'euclidean' distance measure and agglomerative trees were generated with the 'average' linkage method. Data were standardized such that the mean was 0 and the standard deviation was 1. Initial analysis suggested the existence of multiple clusters. To test the reliability of cluster formation *k-means* clustering was performed where *k* was varied from 2 to 5 using the 'sqEuclidean' distance metric implemented in MATLAB. Resulting clusters were tested for stability using silhouetting analysis (Rousseeuw 1987). The value of *k* where a decrease in the average silhouette score was observed was selected to define the number of clusters. Assignment of ORFs to clusters was accomplished by the sum of silhouette scores for the three HTSeq counting methods. If the sum of scores was less than 1.5 the ORF was considered to be unassigned.

4.2.8. Prediction of asRNAs

asRNAs were predicted using the SeqMonk package (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). This program functions by

first determining the overall level of antisense expression to protein coding genes. A binomial distribution is then used to determine how likely the observed number of antisense reads in a region occurs, given the total number of reads mapped to that region the global rate of antisense transcription. After calculating this value for each feature the set of p-values is corrected using a Benjamini and Hochberg multiple testing correction.

Regions of high RNA secondary structure were predicted using the method reported in Millard et al. (2010). Where used the parameters used are indicated. Data is courtesy of Prof David Evans, University of Warwick.

4.2.9. Calculation of termination frequency

Potential rho-independent terminators for the Watson strand of S-PM2 were made using ARNold (http://rna.igmors.u-psud.fr/toolbox/arnold/). This predicted 47 Watson strand terminators of which 25 were intergenic (Appendix 3). For each of the intergenic terminators the frequency of termination per time point was calculated as:

$$TF = \left(\frac{mC_{(t-25:t-1)}}{C_{(t+1)}}\right) x100$$

Where $mC_{(t-25:t-1)}$ is the arithmetic mean read coverage from 25 to 1bp upstream of terminator start coordinate and $C_{(t+1)}$ is the read coverage at 1bp downstream from the terminator stop coordinate.

4.2.10. BLAST analysis

The deleted ORFs from p017-p050 and downstream ORFs p051, p053 and p053 were subject to BLAST analysis against the NCBI–nr database using BLASTp with an e-value of 1e-5. ORFs whose alignment length was greater or equal to 50% of the protein length were considered. Further ORFs, p017-p053 were compared to the Global Ocean Survey (GOS)

database (Yooseph et al. 2007) using tBLASTn with an e-value of 1e-5. Again, only ORFs whose alignment was greater than 50% of the protein length were considered.

4.3. Results

4.3.1. Transcriptome sequencing reveals a large deletion in cyanophage S-PM2

Mapping statistics and coverage estimation for each sequenced library are shown in appendix 4a. Initial alignments of sequencing reads to the cyanophage S-PM2 genome revealed a large region of the genome whereby one pair of the sequenced cDNA fragment mapped ~10kb upstream of the other (Fig. 4.2), suggesting transcripts of approximately this size extending through the region. However, due to fragmentation during library preparation, this scenario was unlikely. Instead, it was possible that a large region of the genome had been deleted during laboratory propagation. Therefore, a PCR strategy was devised to test for presence or absence of this region (Fig. 4.3).
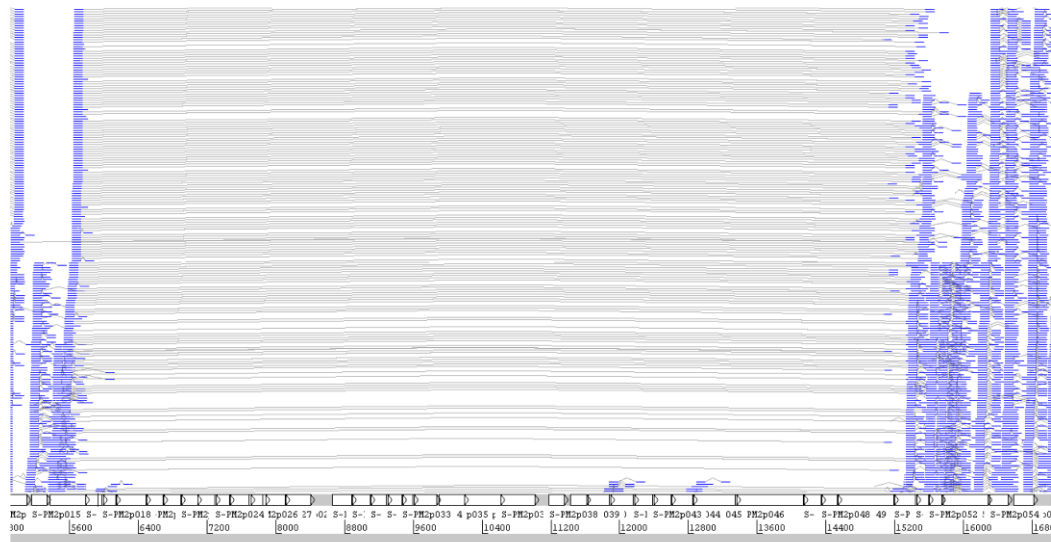
Figure 4.2: Artemis display of the cyanophage S-PM2 deletion region. Blue bars indicate reads mapped to the cyanophage S-PM2 genome. Grey lines indicate connection between paired reads. Cyanophage S-PM2 ORFs are shown below. Numerical markers indicate genome coordinates in bp.
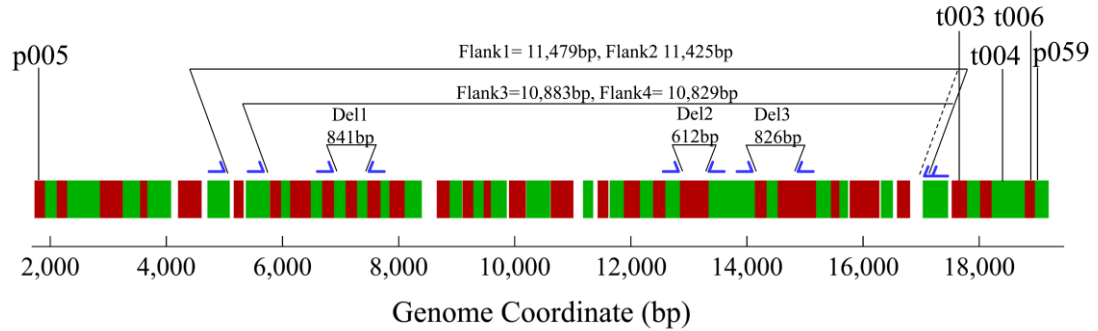
Figure 4.3: PCR strategy for detection of the suspected deletion in cyanophage S-PM2. Alternating coloured bars indicate ORFs starting with S-PM2p005 to S-PM2p059. The blue arrows indicate positions of PCR primers. The five amplicons tested are shown above the blue arrows.

Results of these PCRs show that the apparent deleted ORFs were present in cyanophage S-PM2 lysates (Fig 4.4a, lanes 6, 7, 8). Interestingly, they also show that products were obtained for the targets Flank1-4. This was surprising given that these amplicons should be greater than 10kb in length. These amplicons were estimated to be ~1,250-2,000 bp by gel electrophoresis (Fig. 4.4a, lanes 1-4). Thus it was clear that that a deletion had occurred in a fraction of the population of cyanophage S-PM2 virions from lysates. Sequencing of the Flank1 amplicon revealed that the deletion had occurred between nucleotides 5,798 and 15,340 of the original cyanophage S-PM2 sequence. Thus 9,542 nucleotides had been deleted from the genome encompassing 33 ORFs. This represents 4.9% of the genome and 13.5% of the cyanophage S-PM2 ORFs. Therefore two cyanophage S-PM2 variants existed that here are referred to as S-PM2^{WT} and S-PM2$^{\Delta p017:050}$. To estimate the frequency within the population and indeed to isolate clones of both cyanophage S-PM2^{WT} and S-PM2$^{\Delta p017:050}$, PCR was carried out to target the 'Del1' and 'Flank1' amplicons on a range of isolated plaques. This was accomplished on both recent cyanophage S-PM2 lysates and also a lysate produced in 2006 (Fig 4.4b).
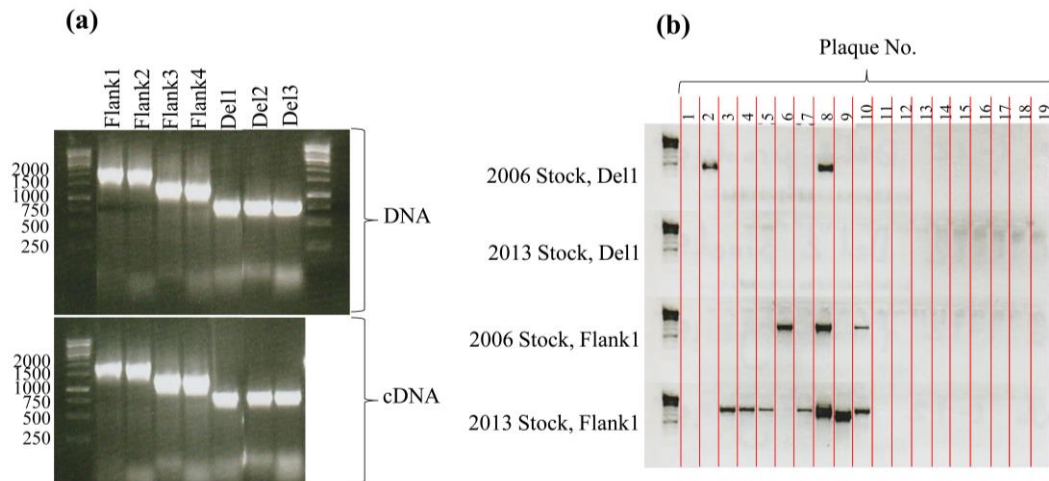
Figure 4.4: Confirmation of the cyanophage S-PM2 deletion. (a) Results of PCRs using the strategy highlighted in Fig. 4.3 on a 2013 lysate and on RNA extracted during infection. (b) PCRs directed at the Del1 and Flank1 amplicons on isolated plaques of cyanophage S-PM2 from 2013 and 2006.

The results of these PCRs demonstrate that deletion had occurred in a fraction of the population from 2006. 2/3 of the PCR positive clones were cyanophage S-PM2$^{\Delta p017:050}$. One plaque (8) contained products from both Flank1 and Del1 PCRs, probably indicative of a mixed plaque. In comparison, no plaques of cyanophage S-PM2^{WT} were obtained from the 2013 stock, whereas 7 were obtained that were cyanophage S-PM2$^{\Delta p017:050}$. Taken together with the absence of reads that mapped to the deleted region indicates that the frequency of S-PM2^{WT} is extremely low in recent lysates of cyanophage S-PM2. Thus, throughout this thesis cyanophage S-PM2 is in fact a mixture of cyanophage S-PM2$^{\Delta p017-p050}$ and S-PM2^{WT} but is referred to as cyanophage S-PM2 for simplicity. All references to genomic coordinates in this chapter are relative to the cyanophage S-PM2$^{\Delta p017-p050}$ genome sequence.

4.3.2. Absent ORFs

33 ORFs were absent in the cyanophage S-PM2$^{\Delta p017:050}$ genome as well as what appeared to be a fusion between the N-terminus of p017 and the C-terminus of p050. The fusion is out of frame of p050 and therefore is predicted to extend to the next in-frame stop codon creating an ORF with 29 residues. BLAST results of this ORF revealed no significant similarities to anything in the NCBI–nr database. For the remaining deleted ORFs both the NCBI–nr and GOS databases were searched for similar sequences. Fig 4.5 shows the numbers of returned

sequences for each ORF and suggests that whilst many of these ORFs are absent from the NCBI–nr database they appear to be present in metagenomic datasets.
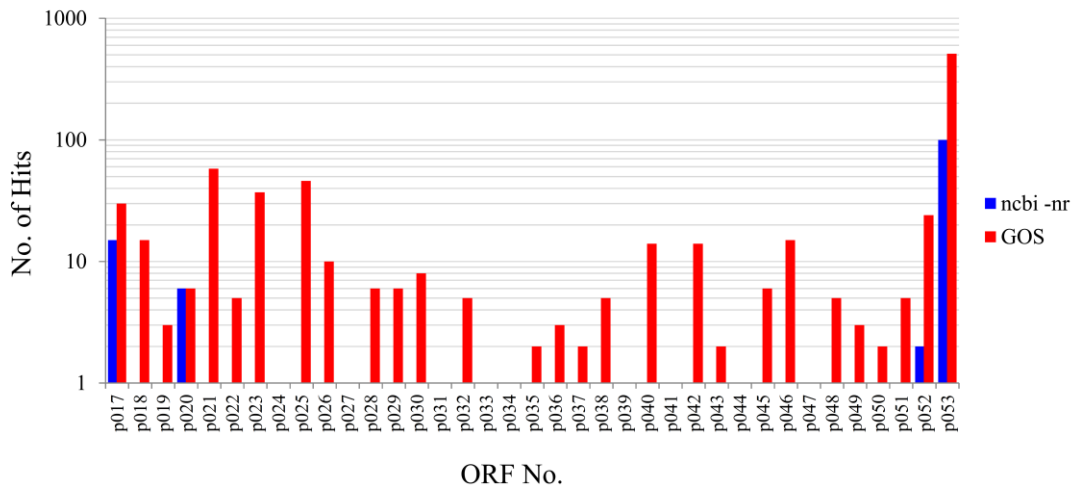
Figure 4.5: Numbers of returned BLAST hits for each of the deleted cyanophage S-PM2 ORFs.

Many of the retrieved hits from the GOS database were found on the same DNA fragment. Thus, it appears that these genes are frequently found in the same region of the genome. Fig. 4.6 shows a synteny plot of the returned GOS hits that were found on the same fragment. The synteny of the ORFs in cyanophage S-PM2 are rarely conserved in metagenomic fragments. A more common organisation is for ORF p017 to be found upstream of p025 or for ORFs p021 or p022 to also be found upstream of p025 (Fig. 6). This strongly suggests intense re-organisation of these ORFs in the environment.
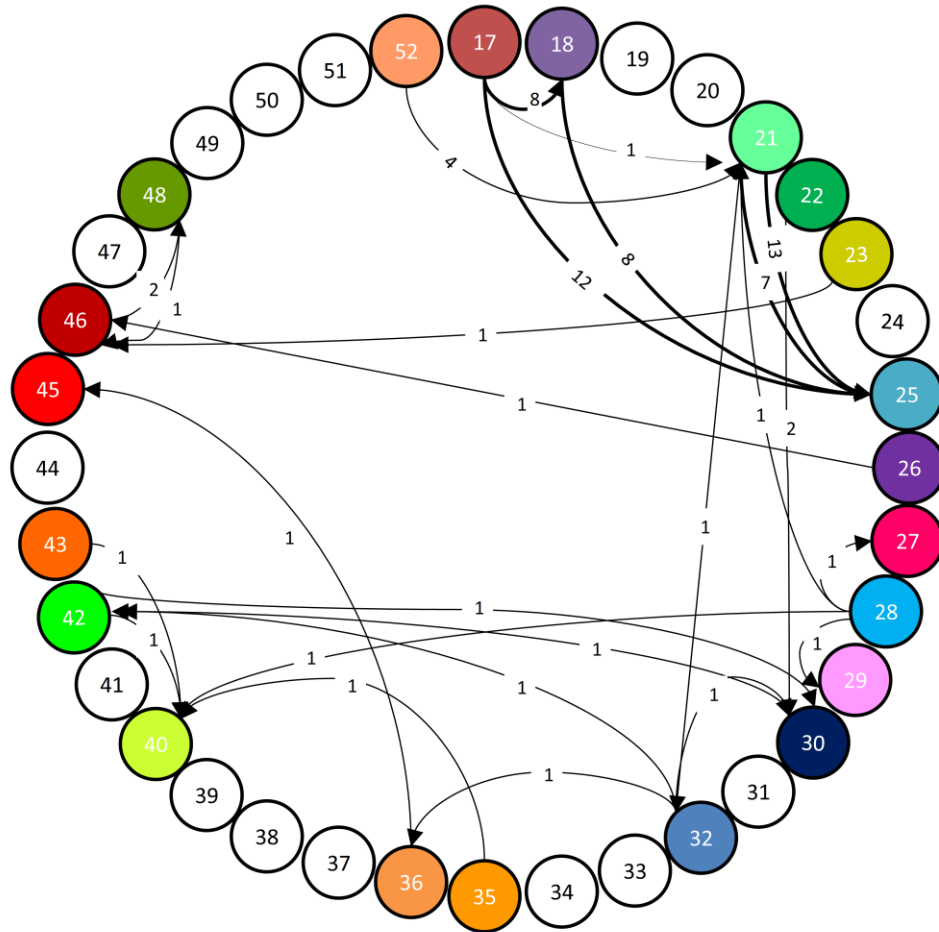
Figure 4.6: Synteny plot of the cyanophage S-PM2 deleted ORFs as found in the GOS dataset. Each circle represents a deleted ORF as indicated by the number in the circle. The synteny in S-PM2 is from p017-p053. Synteny of ORFs on GOS fragments are indicated by arrows that link the circles. The number of instances of this link is shown by the numbers that break the arrow. The width of the arrow is also dependent on the frequency of the observed synteny.

4.3.3. General properties of the cyanophage S-PM2 transcriptional landscape

The distribution of reads mapped to cyanophage S-PM2 from time points 1-9 are shown in Fig. 4.7. At one hour past infection (T1), much of the cyanophage S-PM2 genome does not appear to be transcribed. During progression through infection the amount of the genome that appears to be transcribed increases. Fig.4.8 shows the proportion of the Watson and Crick strands of the cyanophage S-PM2 genome that had mapped reads exceeding 'n' reads and compares these with the
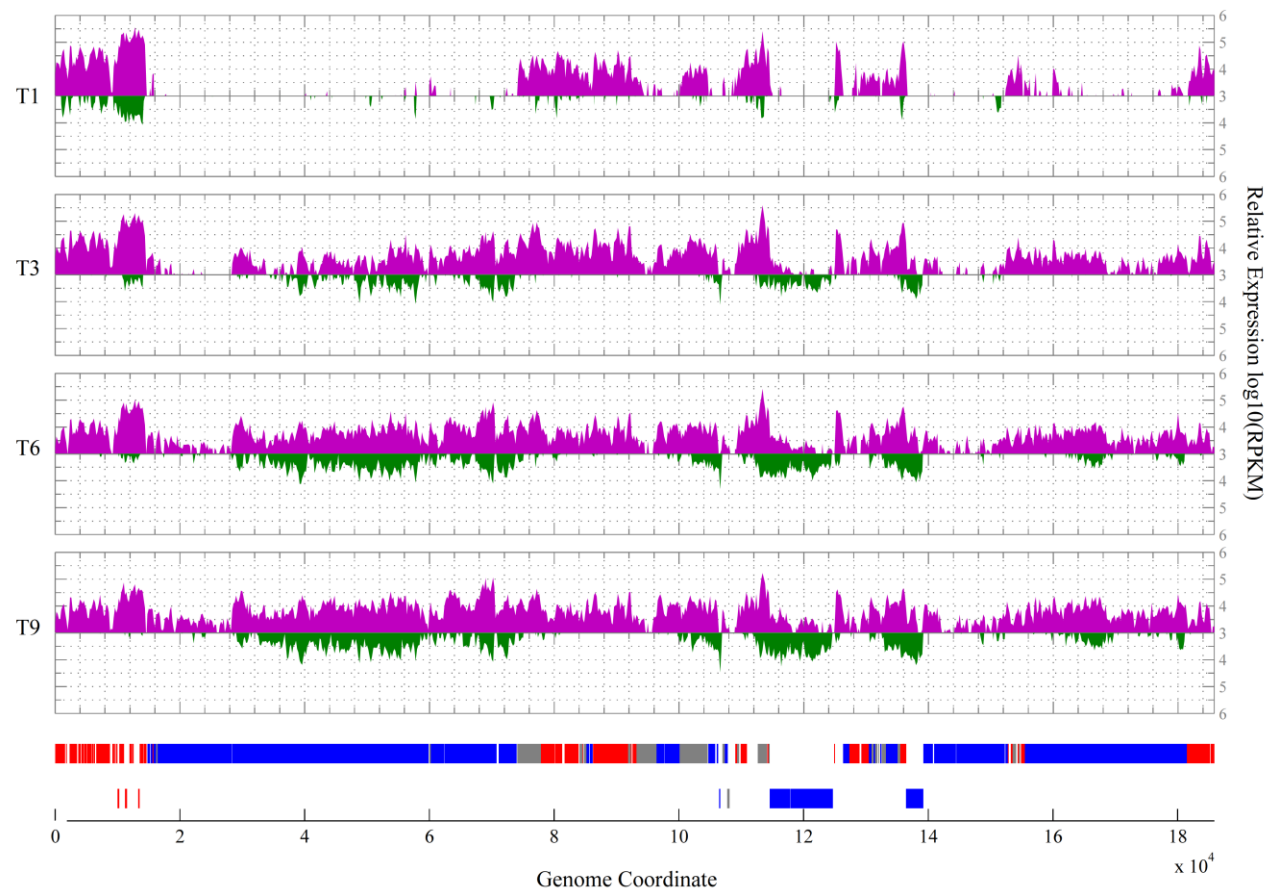
Figure 4.7: Results of RNA-Seq read mapping to the cyanophage S-PM2 genome from time point T1-T9. Watson strand reads are shown in blue and Crick strand reads are shown in red. Data is presented as RPKM. The bottom panel shows a genomic map of cyanophage S-PM2 ORFs. Watson strand ORFs are shown above and Crick ORFs are below. The colour of each ORF is the results of clustering of gene expression shown in section 4.3.4. Red indicates early ORFs, blue indicates late ORFs, grey indicates ORFs that could not be classified.

Synechococcus sp. WH7803 host. The patterns are clearly different. A relatively low proportion of nucleotides on the Watson and Crick strands of *Synechococcus* sp. WH7803 contain at least 1 read. The fraction is below the calculated coding capacity of that strand for both Watson and Crick strands. The proportion of the genome that is covered by n reads decreases dramatically when n is increased to 10 such that only 6.5% of both Watson and Crick strands are covered with greater than 10 reads. In comparison, the percentage of bases that are covered by n reads varies during infection with cyanophage S-PM2. In early infection, the curve largely resembles that of the host with 56.9% of the Watson strand bases being covered by at least 1 read. This value decreases with increasing n. The percentage of bases is always lower than the Watson strand coding capacity during early infection. On the Crick strand, 34.3% of the bases are covered with at least 1 read. This is surprising given that only 7.3% of the Crick strand is coding. During progression through infection the shape of this curve changes dramatically (Fig. 4.8a). By late infection (9 hrs), 92.7% of Watson bases were read with at least 25 reads. This exceeds the coding capacity of the Watson strand by 6.8%. Thus there appears to be a basal level of transcription throughout the Watson strand that is not apparent in the host. Similarly the same trend is observed on the S-PM2 Crick strand. During late infection (6-9 hrs) the percentage of bases that are covered by 25 reads or more is between 43.3 and 43.8% despite the coding capacity of the Crick strand being only 7.3%.
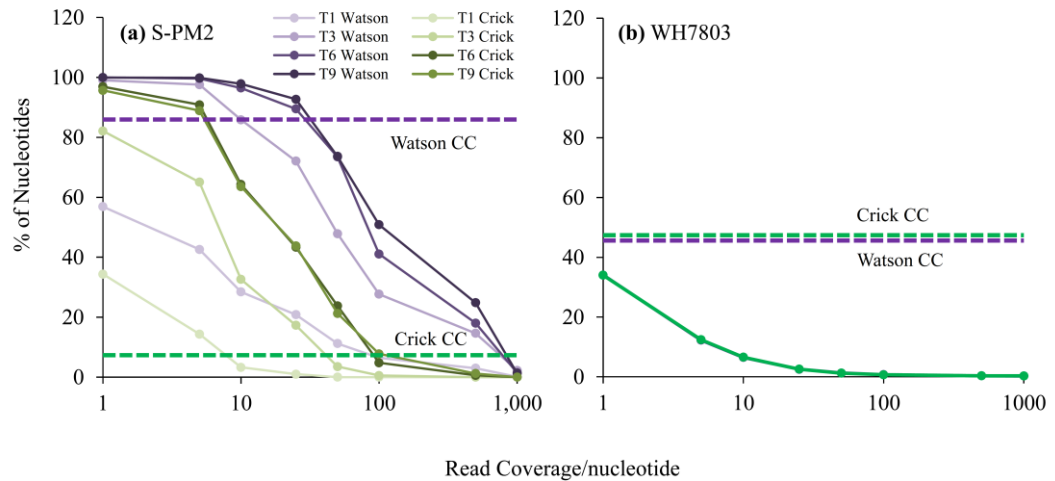
Figure 4.8: Per base coverage of Watson and Crick strands in **(a)** cyanophage S-PM2 and **(b)** *Synechococcus* sp. WH7803. The number of nucleotides that have mapped reads exceeding certain thresholds (x-axis) is shown on the y-axis. In each case dashed purple lines correspond to the coding capacity of the Watson strand and the dashed green line corresponds to the coding capacity of the Crick strand.

The above data suggests a significant proportion of the non-coding cyanophage S-PM2 genome is expressed compared with the host. To test this, the bases that were intergenic or antisense were extracted from the genome and their expression level estimated (Fig. 4.9). The data clearly show that during progression through infection there is a marked increase in the expression level of non-coding bases. A plot of the $\log_{10}$ of the RPKM of non-coding bases suggests a normal distribution for cyanophage S-PM2 non-coding bases during mid to late infection. In comparison, *Synechococcus* sp. WH7803 shows a Poisson distribution with the majority of non-coding bases not being expressed.
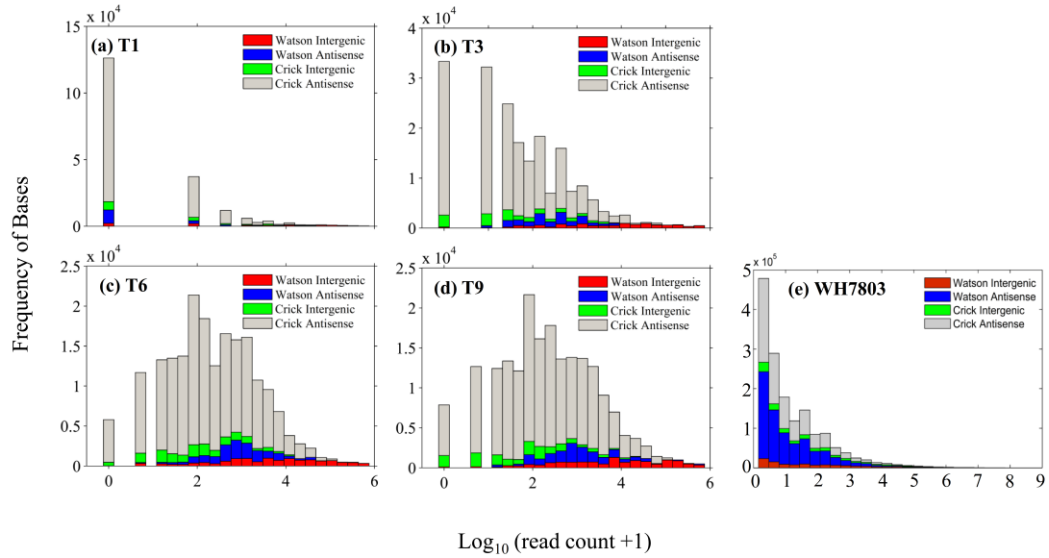
Figure 4.9: Expression of non-coding bases. Non-coding bases are characterised by genomic context. 'Watson Intergenic' are bases on the Watson strand that are intergenic but not antisense to a Crick strand ORF. 'Watson Antisense' are bases on the Watson strand that are antisense to a Crick strand ORF. 'Crick Intergenic' are bases on the Crick strand that are intergenic but not antisense to a Watson strand ORF and 'Crick Antisense' are bases on the Crick strand that are antisense to Watson strand ORFs. **(a-d)** are cyanophage S-PM2 at different point during infection and **(e)** is the *Synechococcus* sp. WH7803 host.

4.3.4. Clustering of cyanophage S-PM2 temporal gene expression

For time points 1-9, expression of each cyanophage S-PM2 ORF was estimated using HTSeq-count (Anders et al. 2014). The results of hierarchical clustering of these expression estimates are shown in Fig. 4.11. These analyses reveal that regardless of the counting method employed, 2 clusters are formed. These correspond to an early profile, or genes that are expressed early during infection but then whose expression is dampened later during infection, and a late profile whose expression is dampened during early infection but then rises to a maximum during the late period between 6-9 hrs (Fig. 4.10). To test the robustness of assignment of each ORF into either cluster, *k*-means clustering was performed with silhouetting analysis to estimate the validity of assignment to either cluster. These analyses suggest that the vast majority of ORFs can be reliably clustered into either the early or late cluster (169/199). The position of each clustered ORF in the genome is shown in Fig. 4.7. These data suggest that ORFs whose expression is temporally clustered are also clustered together in terms of genomic position. There exists a cluster of early genes extending from

genomic coordinate 181,422 starting with ORF p233 and extending to coordinate 14,606 (ORF p075). Another cluster of early genes extend from coordinate 77,867 (p122) to 92,602 (p141). In comparison, two large clusters of late genes are present starting from genomic coordinate 15,783 (p080) to 73,982 (p116) and also from 132,573 (p195 (td)) to 181,434 (p232). The cluster assignment for each ORF is given in Appendix 5. However, of most interest are those ORFs that do not reliably cluster within each group. This method highlighted some 30 genes that could not be assigned to either the late or early cluster. The expression profiles for these unclustered ORFs are shown in Fig. 4.12.
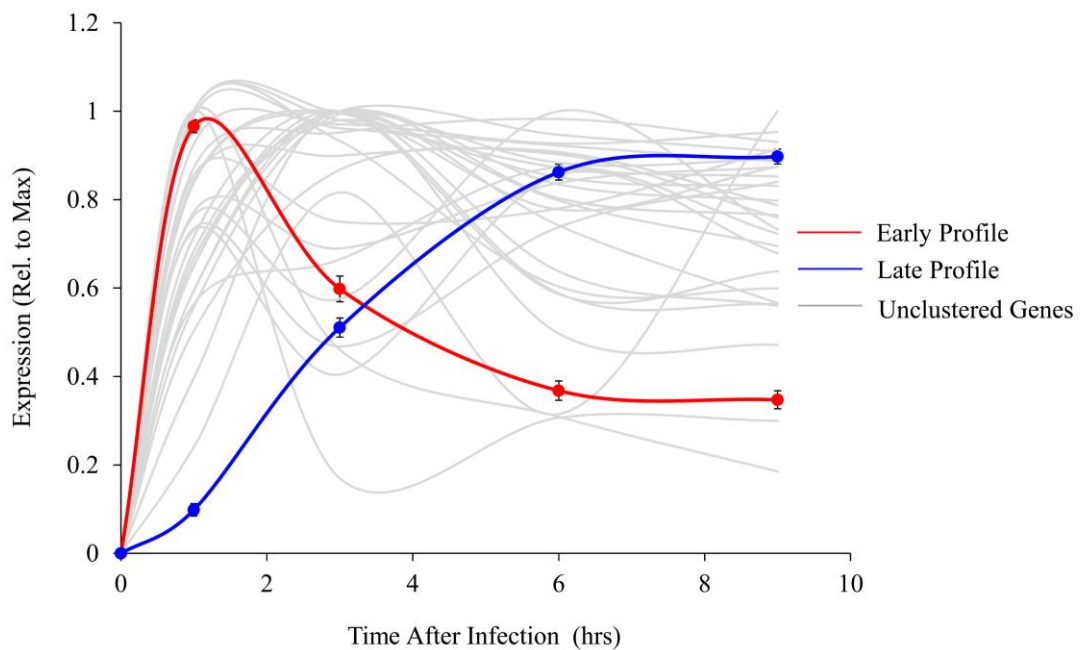
Figure 4.10: Relative expression profiles of clustered genes. The early cluster is shown in red and the late cluster shown in blue. Unclustered ORFs are shown in grey. Expression values are normalised to the maximum expression of that ORF in the time series to allow comparison. Error bars represent the standard error around the mean given the difference in sample sizes between clusters.
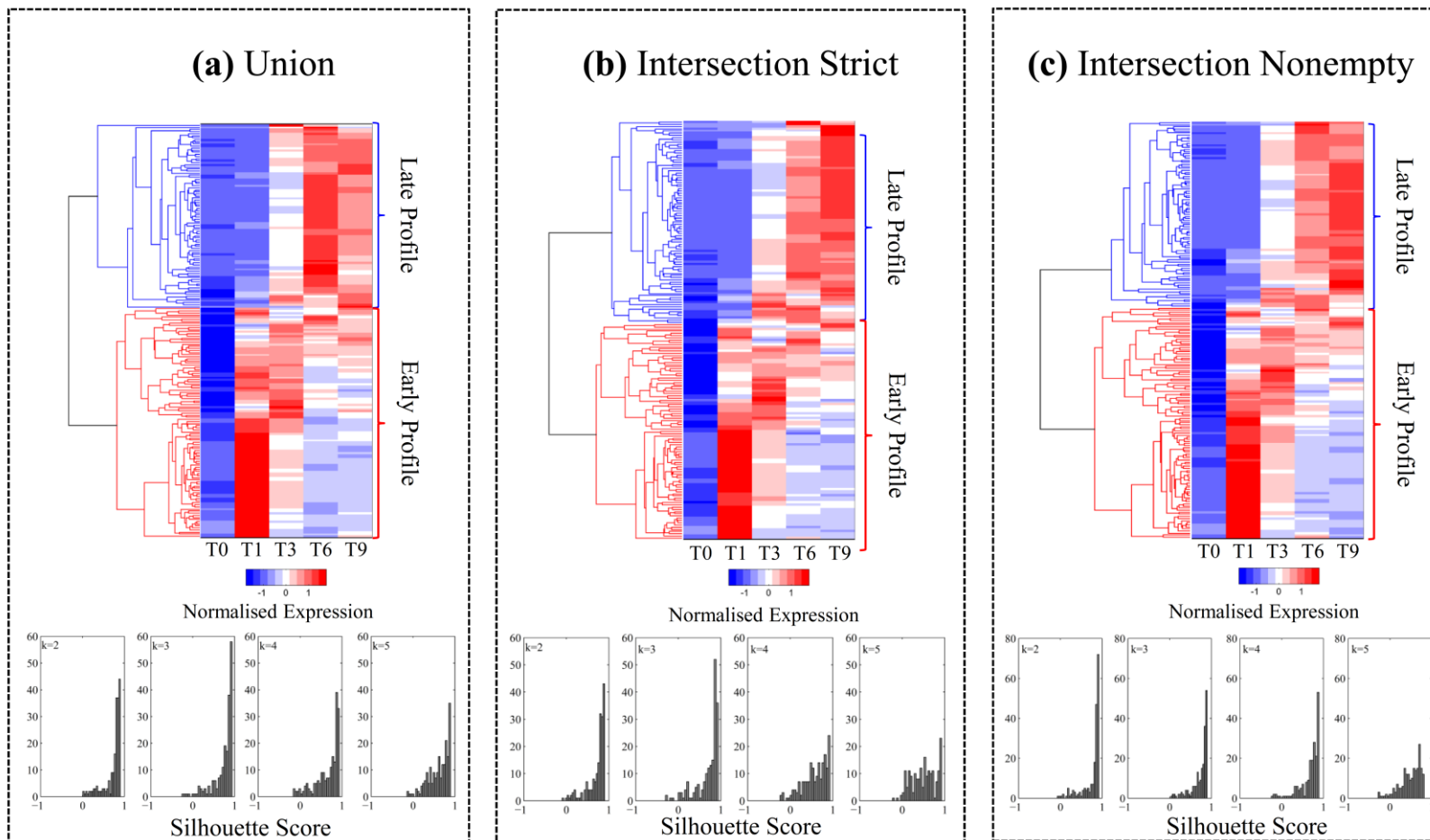
Figure 4.11: Clustergrams demonstrating hierarchical clustering of relative normalised expression estimates from each cyanophage S-PM2 ORF. (a) Results of clustering using the union counting method (b) Results from the intersection-strict method and (c) Results of intersection-nonempty method. In each case the panels below show the distribution of silhouette scores derived from k-means clustering with k values between 2 and 5.

4.3.5. Unclustered ORFs

The expression profiles of each of the unclustered ORFs is shown in Fig. 4.12. The majority of the ORFs that cannot be clustered appear to have an expression profile that is characterised by an early onset of gene expression but with no late period dampening of expression, suggesting these ORFs are actively transcribed throughout infection. Here, these genes are referred to as having a 'continuous' mode. Moreover, subsequent re-clustering of the unclustered ORFs (Fig. 4.13) suggests existence of a 'pseudo-middle' profile. These ORFs are characterised by a peak in gene expression at 3 hr after infection followed by a dampening at 6-9 hr after infection. The 'pseudo-middle' is a well-supported cluster but only when counting using a 'union' or 'intersection strict' method. This is evidenced by increased silhouette scores when k=3 (Fig. 4.11). However, when the 'intersection nonempty' method is employed, the silhouette scores are much reduced when k=3 compared with when k=2. Due to uncertainty in the most appropriate method to employ for counting bacteriophage gene expression by RNA-Seq, a conservative conclusion was drawn that only two clusters exist.
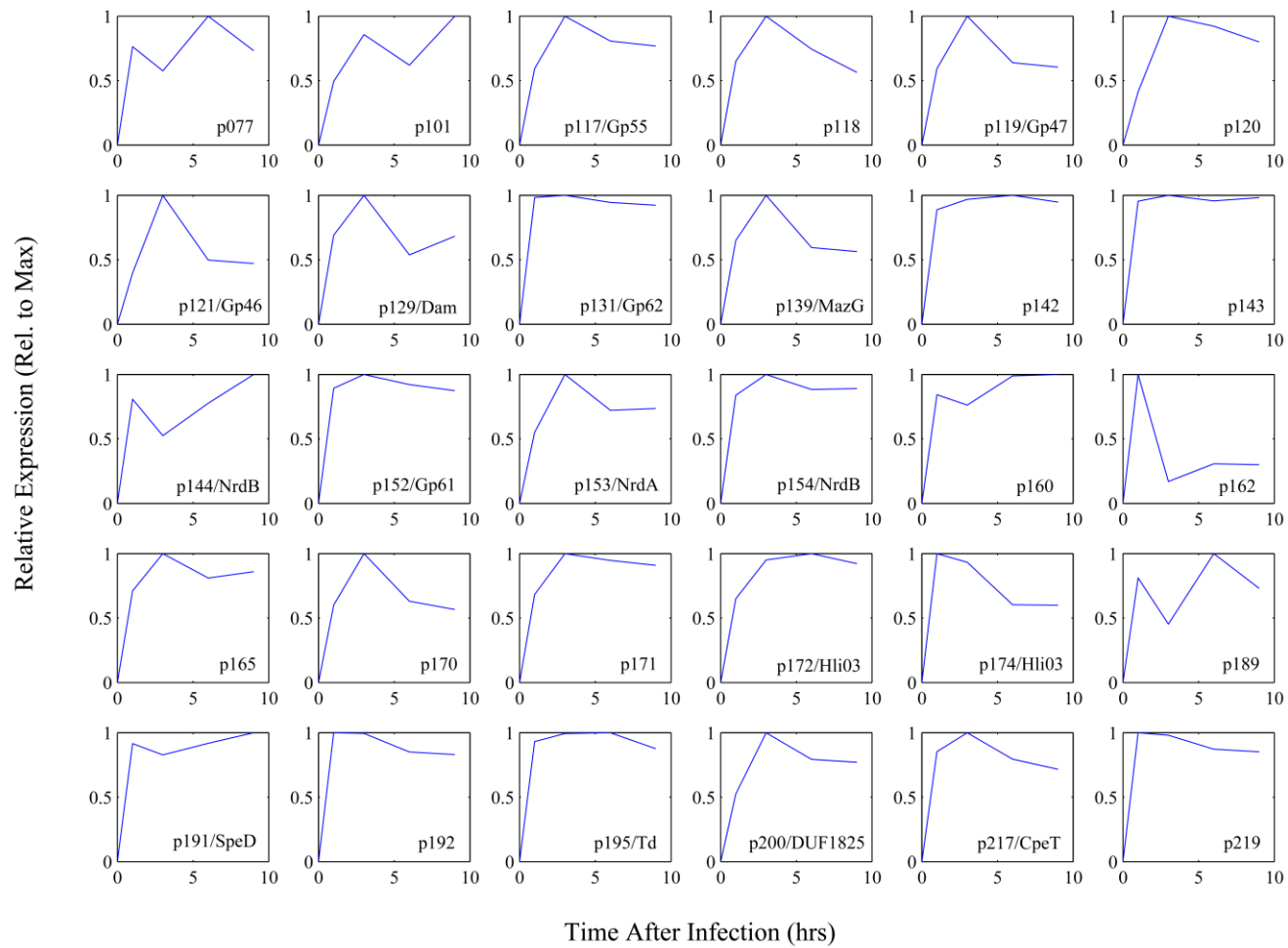
Figure 4.12: Expression profiles of unclustered ORF
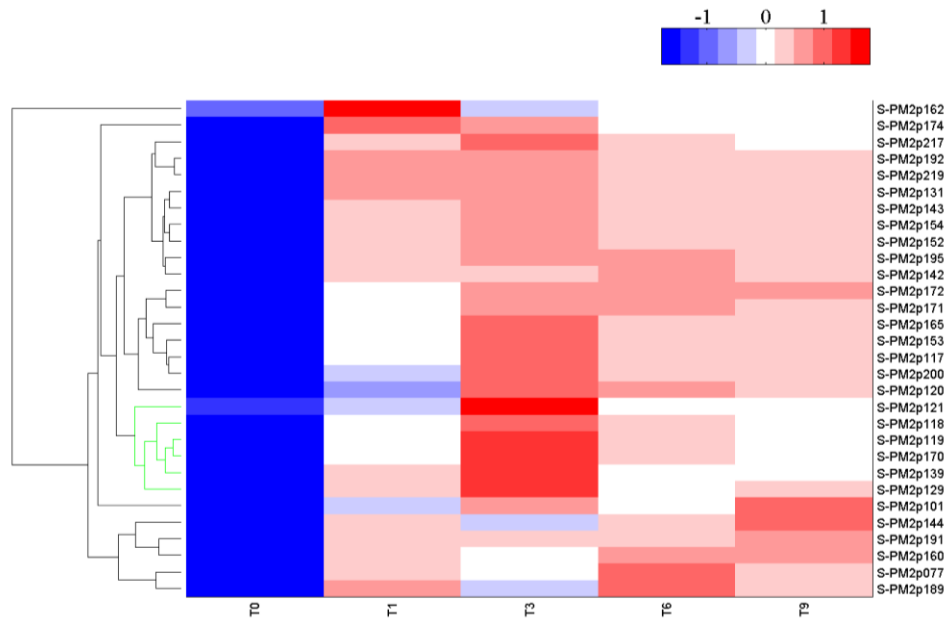
Figure 4.13: Clustergram of hierarchical clustering of normalised relative expression of unclustered ORFs. Green branches symbolise the 'pseudo-middle' profile.

Interestingly, many of the unclustered ORFs are contiguous in their genomic location. In particular, they appear to be found in regions between early and late clusters of ORFs (Fig. 4.7). The genomic context of these ORFs are shown in Fig. 4.14. In the case of the apparent operon shown in Fig. 4.14a, the transcript appears to be driven by an identified early promoter (Mann et al. 2005; Sullivan et al. 2010), yet the ORFs cluster within the 'pseudo-middle' or the 'continuous' group (Fig. 4.12 & Fig. 4.13). Interestingly, the operon is preceded by a validated late gene, p116. There appears to be no termination signal in the intergenic region between p116 and this operon and hence the delayed peak in expression may be the result of transiently active early and late gene expression. Three of the genes of this operon share homology with genes in T4, namely *g*55, *g*47 and *g*46. *g*55 is a small sigma factor that is required for expression of late genes (Miller et al. 2003). *g*47 and *g*46 encode components of the join-copy recombination pathway and are also required for expression of T4 late genes (Miller et al. 2003).

The ORFs contained within the apparent operon shown in Fig. 4.14b have a rather surprising expression profile. A T4-like late promoter is upstream of gene *p*141, yet its expression profile clusters within the early group. Downstream of *p*141, the ORFs *p*142, *p*143 and *nrdB*

were all unclustered despite absence of any promoter or terminator downstream of *p*141. Interestingly, *p*142 and *p*143 contain highly structured regions at their 5' ends (Fig. 4.15a). It could be possible that these structured regions may mask RNase cleavage sites and therefore act to stabilise the transcript. Hence, these ORFs display a 'continuous' expression profile.
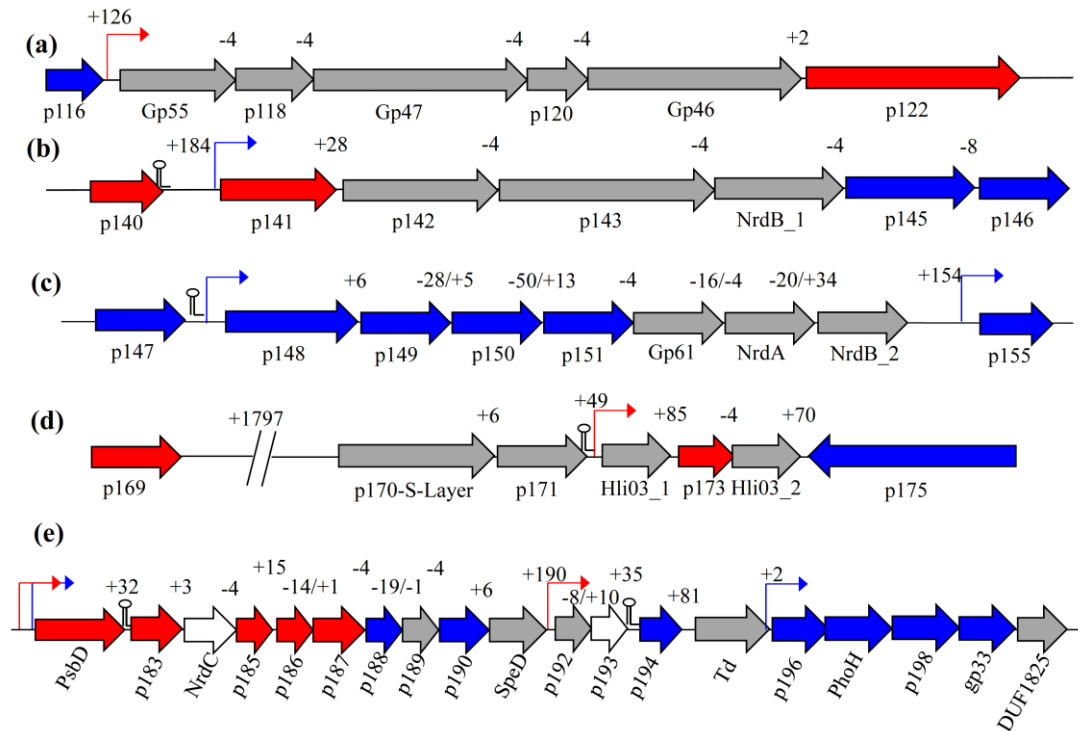
Figure 4.14: Genomic context of unclustered ORFs. Arrows are coloured according to the cluster assignment. Red is early, blue is late, grey is unclustered and white indicates genes whose expression fell below the expression threshold and was not considered for analysis. Red right-angled arrows indicate early promoters, blue indicate late promoters and stem-loops indicate predicted rho independent terminators. ORF annotations are shown below the arrows and the intergenic length is shown above the arrows.

In Fig. 4.14c, an apparent operon is present starting with ORF p148 and ending in NrdB_2, despite no rho-independent termination signal being present. This operon is driven by a late promoter and indeed the first four genes of this operon display a late mode. However, the last three genes display a 'continuous' mode, characterised by an early onset of expression followed by high levels of relative expression throughout infection. Whilst this apparent transcript also retains a highly structured 3' end, it is unlikely that this participates in the expression profile of these genes. Therefore, the factors that may contribute to the irregular expression profile of this apparent operon remain to be elucidated.
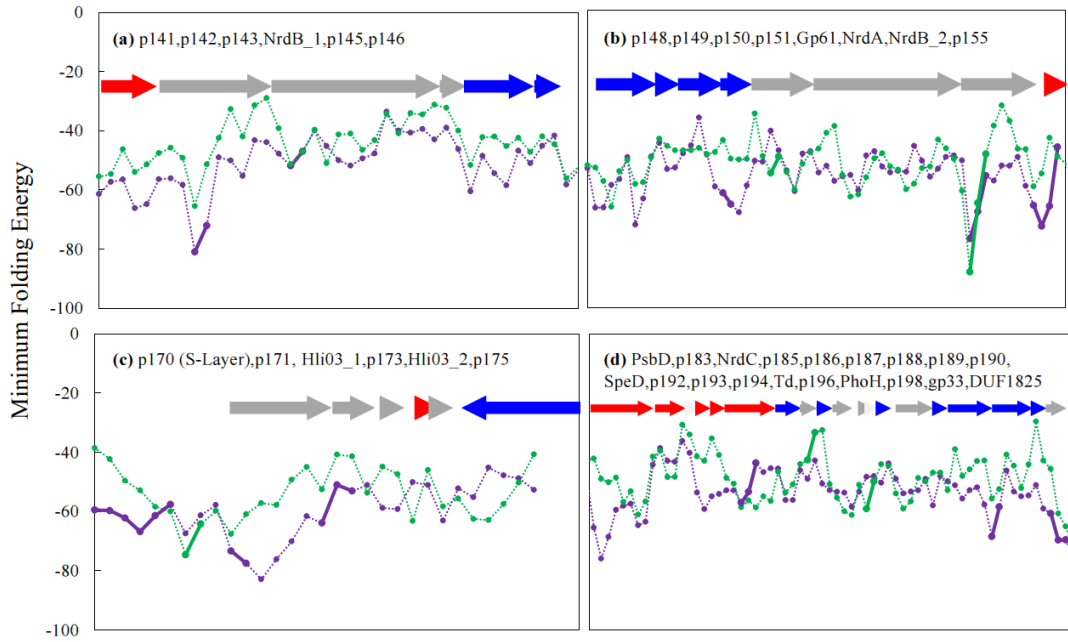
Figure 4.15: Unclustered ORFs with regions of significant predicted RNA secondary structure. ORFs are shown as arrows with early ORFs shown in red, late ORFs shown in blue, unclustered ORFs shown in grey and ORFs with low read counts shown in light grey. Hashed line graphs show the minimum folding energy of the RNA window. Emboldened lines indicate windows whose secondary structure was significant when compared to 100 scrambles of the RNA sequence that retains the nucleotide and dinucleotide composition. Here, the window size was 125 nt with no overlap between windows. Purple lines show the Watson strand and green shows the Crick strand.

Upstream of ORF *p*170, encoding a protein with an S-Layer domain, is a long stretch of non-coding nucleotides. It is likely that this region encodes at least one if not two unannotated ORFs due to the absence of stop codons in this region. Further work is required to validate these ORFs. The ORFs downstream of *p*170 are largely unclustered with respect to expression with the exception of the small ORF *p*173, which has a well-supported late expression mode (combined silhouette score=2.161) (Fig. 4.14, Appendix 5). Downstream of Hli03_2 is ORF *p*175 which has a well-supported late mode of expression (combined silhouette score=2.696) and is found on the Crick strand. It is possible that a lack of termination of the transcript encoding *p*175 may lead to antisense expression of ORFs *p*170-*hli03* that, through complimentary base pairing, acts to stabilise the transcripts encoding ORFs *p*170-*hli03*. This hypothesis remains to be tested in greater detail.

In Fig 4.14e, the unclustered ORFs seem to form a boundary between an early cluster and a late ORF cluster. The early cluster appears to be driven by the early promoter upstream of *psbD*. *psbD* is followed by an obvious rho-independent terminator. Interestingly the last 'U' of the rho-independent terminator is located -1 nt relative to the start codon of the downstream ORF *p*183. Therefore, if this terminator was fully functional it is unclear how *p*183 would be expressed. Therefore, it is likely that the terminator does not function with 100% efficiency, a hypothesis that is supported by numerous reads that overlap the terminator (Fig. 4.16). Interestingly, the terminator proceeding *psbA* shows a strong termination signal in RNA-Seq reads and thus differential termination may play a key role in driving the transcriptional patterns observed in cyanophage S-PM2. An interesting junction is between ORF *p*187 and *p*188 (Fig 4.14e). *p*187 is a strong scoring early gene (silhouette score= 2.469) whilst *p*188 clusters within the late group, albeit with silhouette score that just passes the applied threshold (1.780). These two ORFs overlap by four nucleotides and therefore it is unclear how these ORFs may be differentially regulated. An asRNA is predicted antisense to *p*188 due to a high folding energy (Fig. 4.15d) and therefore may act to regulate *p*188.
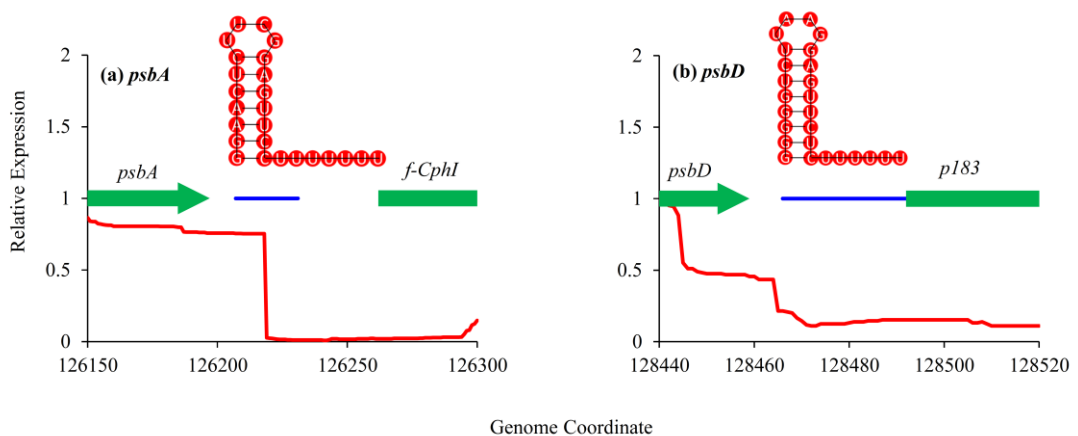
Figure 4.16: Comparison of termination signals downstream of the cyanophage S-PM2 *psbA* **(a)** and *psbD* **(b)** genes. The relative expression of each nucleotide is shown on the y-axis and is relative to the maximum of the window. Green arrows show locations of ORFs and the blue lines show locations of the terminator helices. A cartoon of the terminator structure and sequence is shown in the inset of each plot.

The boundary between *p*190 and *speD* is also extremely interesting. There are six nucleotides in between *p*190 and *speD*. However, a putative rho-independent terminator links these two genes (Fig. 4.17a). Indeed, during early infection approximately half of the reads that cover the termination signal end within the termination signal (Fig. 4.17b). At 9 hr after infection this ratio is increased to 4 (Fig. 4.17b). It is possible that transcriptional readthrough, due to incomplete termination, from p190 into SpeD during the late period facilitates its 'continuous' expression mode.
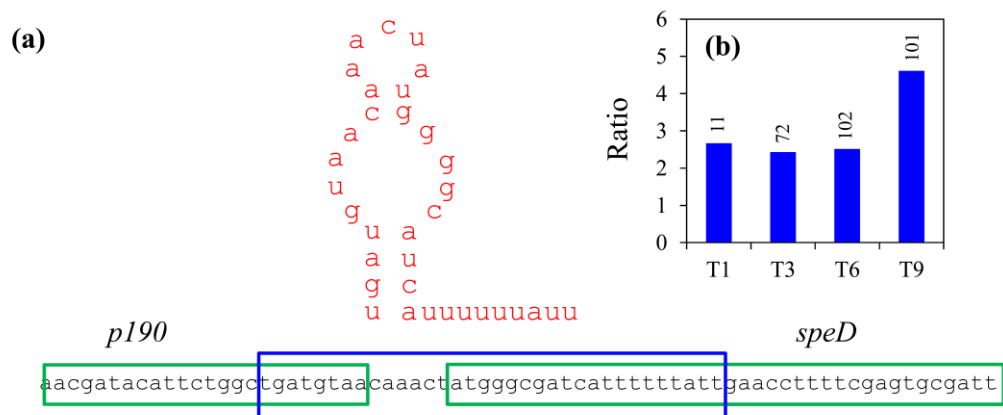
Figure 4.17: Possible anti-termination between the *p190* and *speD* genes. (a) The two genes are shown by green boxes and the predicted terminator is shown by the blue box (b) Ratio of the region coverage to the number of reads that terminate in the region at different time points. The number of reads that cover the region are shown above the bars.

There exists one unclustered ORF whose expression profile is distinct from all other unclustered ORFs (Fig. 4.13). Instead of the 'continuous' or 'pseudo-middle' mode, ORF p162 appears to be expressed early but then undergoes a rapid decrease in expression from T1-T3, where it's expression remains low for the rest of the infection period (Fig. 4.12). However, this ORF had an extremely low number of mapped reads (14 across all time points) and therefore this expression profile generated is probably erroneous.

4.3.6. Validation of the S-PM2 transcriptional model

Cyanophage S-PM2 has been shown to contain T4-like early and late promoters (Mann et al. 2005). Indeed, cyanophage S-PM2 gene expression profiles largely fall into early and late

clusters as discussed above. Therefore, to validate the transcriptional model proposed for cyanophage S-PM2 the expression profiles of the downstream ORFs from those identified early or late promoters (Mann et al. 2005) were compared (Fig. 4.18). Only 50% of genes with a predicted early promoter displayed an early mode of expression. Of the remainder, the majority were unassigned, with 3 ORFs possessing an early promoter having a late expression profile (Fig 4.18). In comparison, the vast majority (83.3%) of ORFs with a putative late promoter display a late expression profile (Fig. 4.18). 4 ORFs containing an upstream late promoter displayed an early profile. Interestingly, 5 cyanophage S-PM2 ORFs have been shown to retain both an early and late promoter upstream. In these analyses, expression profiles of these ORFs are split fairly evenly between the early and late clusters (Fig. 4.18).
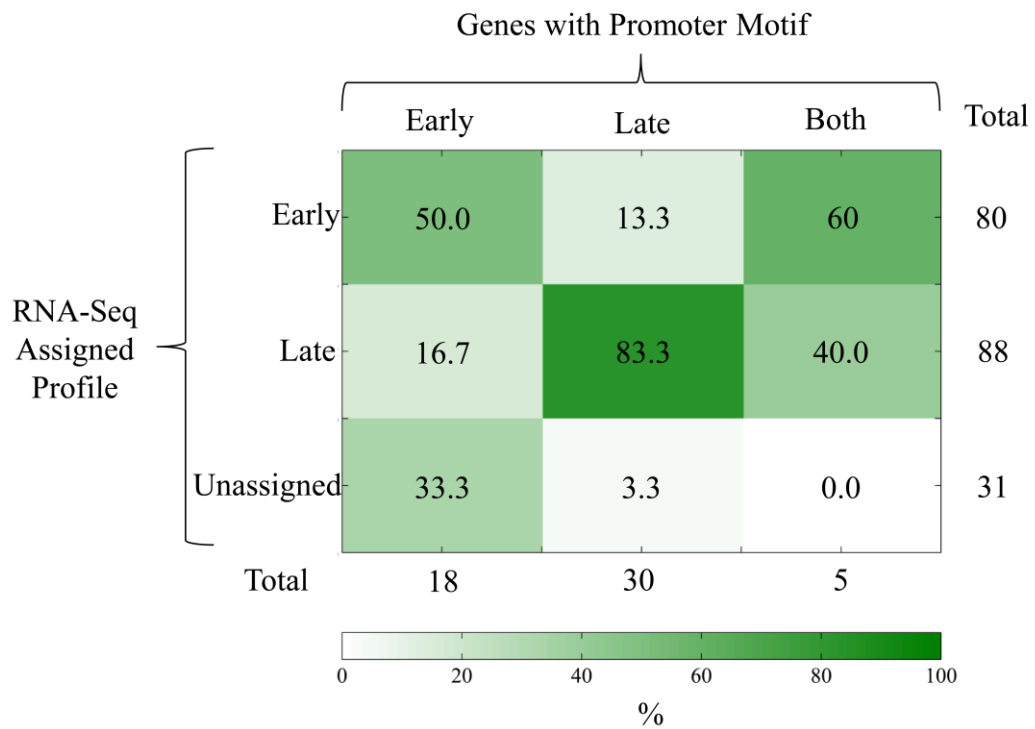
Figure 4.18: Relationship between the clustered expression profiles of ORFs immediately downstream of predicted early and late promoters.

4.3.7. Non-conforming ORFs

In total, the expression profiles of 14 ORFs did not conform to their predicted upstream promoter (Table 4.1). Of these, 8 had expression profiles that were unclustered, instead displaying a 'continuous'/'pseudo-middle' mode or their expression estimate was below the threshold set.

ORFs that encode an upstream late promoter that instead displayed an early profile are all hypothetical (p013, p141 and p216). p013 is located within the early cluster between genomic coordinate 181,422 and 14,606 (Fig. 4.13) and therefore the cluster assignment is likely correct. It is unclear why the existence of a late promoter upstream of this gene does not contribute to a late profile. Interestingly, p013 is extremely rare amongst cyanophages only found in the coastal isolate S-CBM2 and the freshwater isolate S-CRM01.

p141 is also found within an early gene cluster. Interestingly, this ORF has a low silhouette score, just above the threshold and therefore transcriptional read-through from the upstream early genes may contribute to the early profile whilst the late promoter may extend expression later during infection that causes a decrease in the silhouette score.

ORF No	Name	Annotation	Start	Stop	Promoter Type	Profile	Silhouette Score
p011	p011	Hypothetical	3550	3678	Early	n.d	n.d
p013	p013	Hypothetical	4201	4605	Late	Early	2.554
p117	gp55	Sigma factor for late transcription	74110	74604	Early	UC	0.697
p139	Maz G	Pyrophosphatase	91875	92282	Early	UC	1.283
p141	p141	Hypothetical	92603	93181	Late	Early	1.574
p142	p142	Cytitidyl-transferase	93210	94382	Early	UC	0.142
p152	gp61	DNA primase subunit	100153	101142	Early	UC	0.624
p157	p157	Lysin	106077	106295	Early	Late	2.719
p165	p165	Hypothetical	109393	109677	Early	UC	0.586
p172	Hli03_1	High light inducible polypeptide	113940	114134	Early	UC	1.038
p189	p189	Hypothetical	130947	131234	Late	UC	0.328
p199	gp33	Late promoter	134876	135124	Early	Late	2.309

		transcription accessory protein					
p214	p214	Hypothetical	152403	152747	Early	Late	2.557
p216	p216	Hypothetical	153265	153537	Late	Early	2.363

Table 4.1: ORFs whose expression profile does not conform to their relevant upstream promoter.

p216 is found upstream of the AMG *cpeT*, encoding a putative phycobilin lyase (Bretaudeau et al. 2013). The two ORFs overlap by 8 nucleotides. *p216* contains an uncharacteristically long upstream intergenic region (361 bp). This region is characterised by high MFED scores (Fig 4.19a), which suggests conservation of RNA secondary structure. Structural prediction for this region suggests that the late promoter may form a paired helix with an upstream sequence. Such paired helices may cause transcriptional termination or stalling of the elongating RNAP. It would be tempting to speculate that certain environmental conditions may cause a refolding of the RNA to allow access of the RNAP through the late promoter to drive late expression of this gene and the downstream *cpeT*, reminiscent of a prokaryotic riboswitch. Interestingly *cpeT* was one ORF whose expression was unclustered (Appendix 5) and displayed a 'continuous' mode.
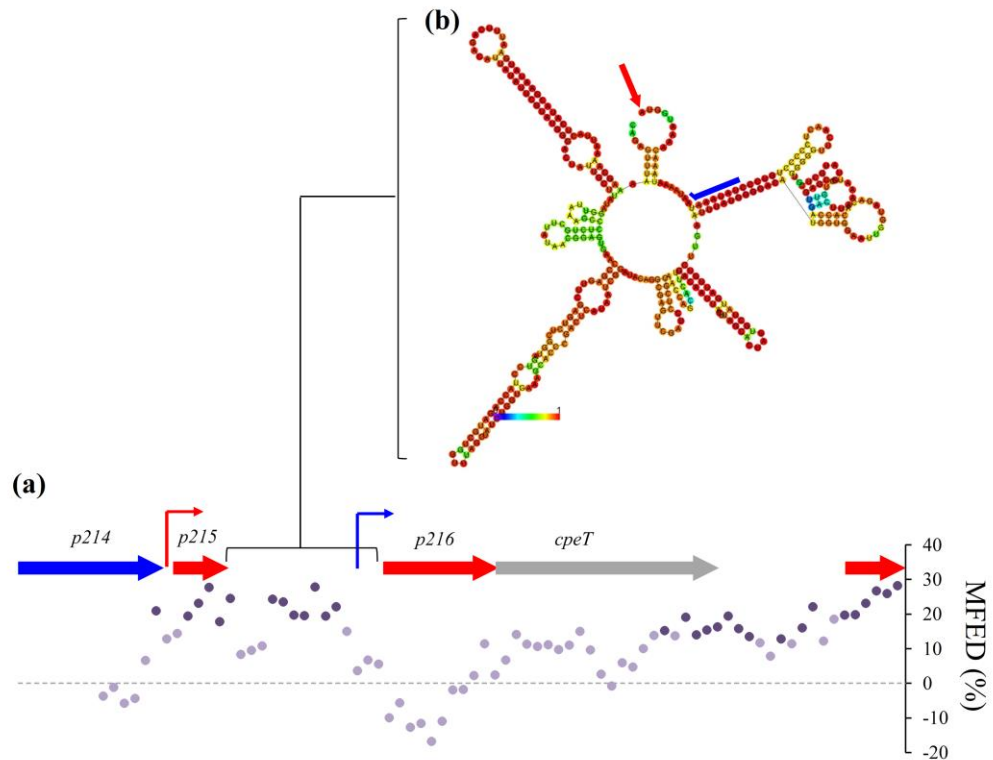
Figure 4.19: (a) Genomic context of the *p216* region. ORF *p216* displays a late expression profile despite retaining an upstream early promoter. The minimum folding energy difference is symbolised by purple data points. The minimum folding energy difference is defined as the difference in minimal folding energy between the native window and the mean of 99 scrambles that retain nucleotide and dinucleotide frequency. Here, the window size was 250 nt and the overlap between windows was 225 nt. Emboldened data points show windows whose MFED was greater than 99% of the scrambles **(b)** Predicted secondary structure of the intragenic region between *p215* and *p216* as predicted by RNAfold. The free energy of the thermodynamic ensemble was -121.78 kcal/mol and the fragment was 358 nt in length. Colours correspond to the probability of the base being paired (paired bases) or unpaired (unpaired bases). The late promoter region is shown highlighted with the blue line and the start codon of *p216* is shown by the red arrow.

ORFs with an upstream early promoter which instead display a late profile are p157, p199, p214. p214 contains an uncharacteristically long upstream intergenic region between itself and p213 (160 nt), but unlike the region upstream of p216 displays no conserved secondary structure. There is, however, a late promoter like motif with the sequence 'CTAAATA' that differs from the late consensus by one nucleotide which has been suggested to function in a freshwater cyanophage (Dreher et al. 2011). *p199*, encodes a protein required for late transcription so it is surprising to find this gene with a late expression profile, especially given the presence an early promoter. It is, however, found in a late gene cluster. ORF *p157*

encodes a putative lysin and so one may expect it to have a late expression profile. The transcriptional organisation of this region is discussed below.

4.3.8. Transcriptional organisation of the lysin region

The genomic and transcriptional organisation of the putative lysin region is shown in Fig. 4.20. The region begins with expression of the late gene *p155,* which contains a late promoter upstream. A strong termination signal is observed downstream of *p155* in RNA-Seq reads (Fig. 4.26b). However, this termination is incomplete and allows partial read-through at approximately 20% (Fig. 4.26b). A possible rho-independent terminator is observed at the C-terminal region of the p155 ORF. This termination signal contains a poly U tract but the identified paired helix is extremely weak (Fig 4.21). The annotated ORF boundaries of the downstream lysin *p156* may be incorrect. Currently the start codon is annotated to genomic coordinate 105,112 resulting in an overlap of 14 bp between ORFs p155 and p156 (Fig. 4.21). However, a much more likely start codon is found downstream at position 105,227 which retains an obvious RBS upstream (Fig. 4.21). Relocation of the start codon opens up the possibility of a late promoter downstream of the poly U tract of the *p155* transcript (Fig 4.21). This late promoter may drive expression of p156 leading to the observed late expression profile for this ORF despite the presence of an early promoter upstream of *p155* (Fig. 4.21).
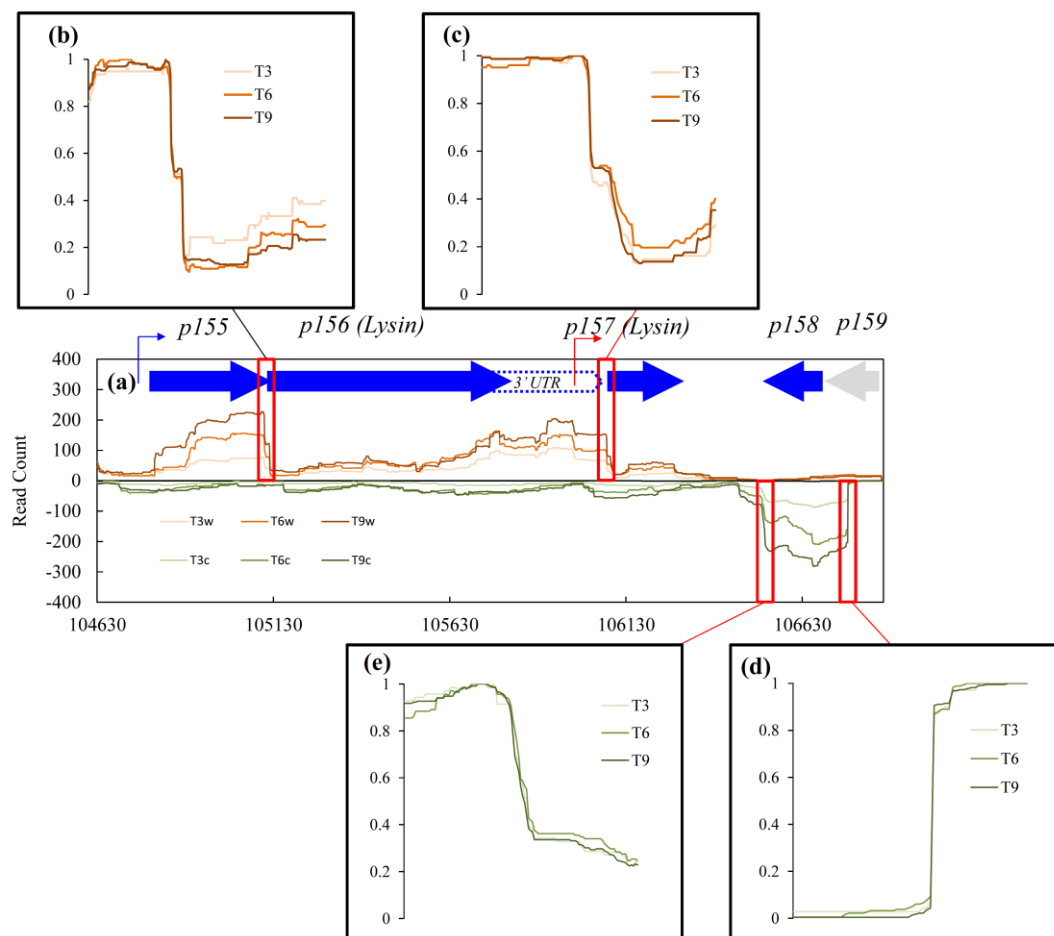
Figure 4.20: Genomic and transcriptional organisation of the lysin region. **(a)** Coverage plots of the Watson strand (orange shades) and Crick strand (green strands). Dashed box shows the position of the putative 3' UTR. Late promoters are shown in blue and early in red. Red boxes show regions where transcription termination or initiation is thought to occur from RNA-Seq data. **(b-d)** Close-up view of transcription termination/ initiation regions. The x-axis shows the expression relative to the maximum in the window.
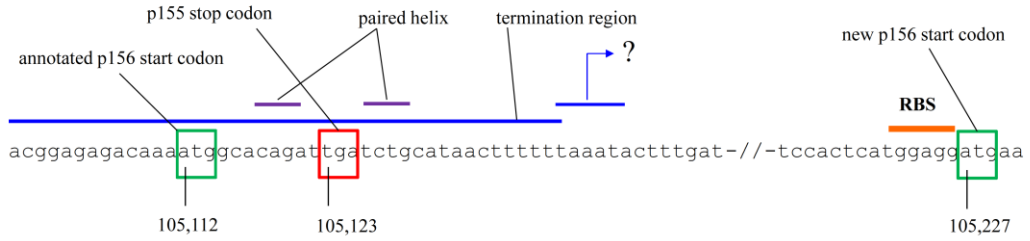
Figure 4.21: Re-annotation of the *p156* gene. Features of this region are discussed in the text

The annotated stop codon of *p156* maps to position 105,807. Here, it is shown that *p156* encodes an extended 3' UTR with a strong termination signal in the RNA-seq data observed between genomic coordinates 106,075 and 106,090 (Fig. 4.20). No rho-independent termination signal is observed in the genomic sequence. Interestingly the 3'UTR encodes an early promoter sequence. Folding of the 3'UTR may act to limit availability of the promoter to the RNA polymerase. The downstream gene *p157* also encodes a protein with sequence similarity to a lysin. There are two ORFs present on the Crick strand, p158 and p159. p158 displays a strong TSS (Fig. 4.22). This region appears to be unique in the cyanophage S-PM2 transcriptome where obvious TSSs are lacking. Analysis of the genomic sequence upstream reveals a putative promoter sequence that is distinct from the late promoter sequence (Fig. 4.22). Therefore, there appears to be greater diversity in the late promoter sequence than was previously thought (Mann et al. 2005).
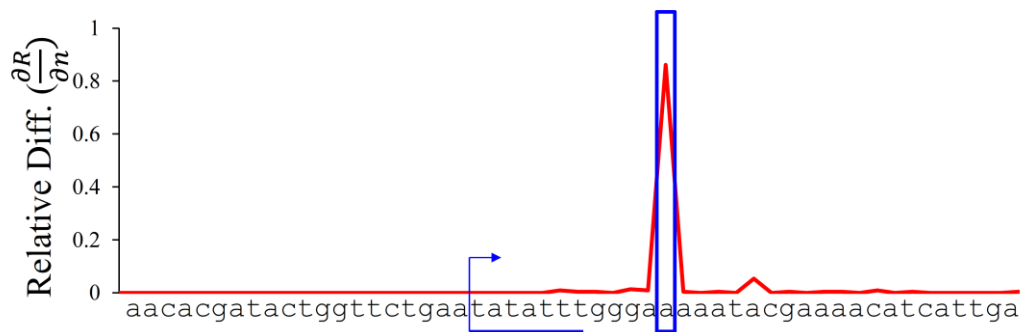
Figure 4.22: Mapped TSS of the p158 gene. The y-axis shows the relative differential in read coverage per nucleotide. The putative promoter is underlined in blue.

p158 displays a termination signal as evidenced by a sharp decline in the RNA-Seq coverage (Fig. 4.20). No obvious termination sequence could be found. Interestingly, the ORF also shows incomplete termination that leads to read-through at about 30-40%. It would be tempting to speculate that downstream expression from this terminator may interact with the antisense lysin mRNA. Regulation of p158 termination could therefore play a role in regulating lysis

4.3.9. Antisense transcription

The presence of low-level genome-wide antisense transcription in cyanophage S-PM2 can be seen in Fig. 4.7. Fig. 4.23 shows the relative expression of every cyanophage S-PM2 ORF in relation to its antisense expression, i.e. the relative expression of the genomic region antisense to that ORF. In all cases a positive linear relationship exists between the sense expression and antisense expression. This may be an artefact of the DNA dependent DNA polymerase activity of reverse transcriptase (Perocchi et al. 2007). However, this phenomena is only estimated to occur in a small fraction of cases (~1-3%) (Wolfgang Hess, *pers. comm.*). In Fig. 4.23 reference lines are shown at 1, 10 and 100% intervals and show that the antisense expression is greater than this amount in most instances. Further, the antisense RNA *CfrI* has been validated in cyanophage S-PM2 (Millard et al. 2010) and occurs at a ratio that is significantly less than the majority of ORFs (Fig. 4.23). Thus, much of this antisense expression is thought to be *bona fide*. In many examples, the relative antisense expression exceeds that of the sense, often by at least an order of magnitude (Fig. 4.23).
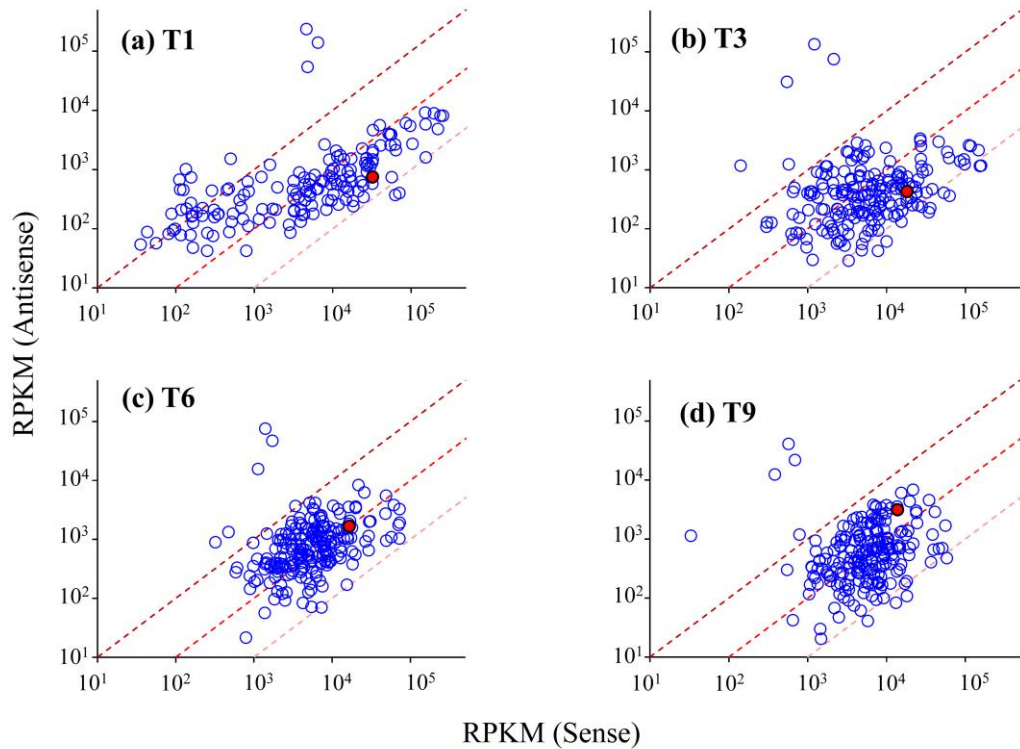
Figure 4.23: Relationship between sense expression and antisense expression for every cyanophage S-PM2 ORF. Reference lines correspond to 1% (light red), 10% (red) and 100% (dark red) antisense: sense expression. The red data point corresponds to the validated asRNA *CfrI* (Millard et al., 2010)

The genomic context of the top 10 ORFs with the greatest antisense: sense ratio are shown in Fig. 4.24. 5 of the top 10 ORFs are located on the Crick strand. This represents a significant enrichment given that only 13/208 ORFs are located on the Crick strand. Thus, it appears that the massive levels of antisense 'read-through' transcription overwhelms the modest expression of Crick strand ORFs.

Of the remaining top scoring ORFs, 4/5 (p213, *g23*, *g5* and *g14*) have a late transcriptional mode. The other, p163 was unclustered. g*23*, *g5* and *g14* all encode structural proteins specifically the major head protein, the baseplate hub and tail lysozyme and the neck protein respectively.

The SeqMonk package was used to test for ORFs that contained a significant amount of antisense expression (see section 4.2.8). The results from these tests are shown in Table 4.2. These reveal a range of potential asRNAs that were not in the top 10 scoring elements. All

of these appear to be late genes and all are found within the late genomic cluster discussed in section 4.3.4. Transcript assembly by cufflinks suggests the presence of one long asRNA that extends antisense to this genomic cluster (Appendix 6). The relevance and mechanisms regarding this asRNA are discussed in section 4.4.

ORF No	Name	Annotation	Strand	T1	T3	T6	T9
p063	p063	-	-	<0.0001	<0.0001	<0.0001	<0.0001
p067	p067	-	-	0	0	0	0
p070	p070	-	-	0	0	0	0
p087	p087	-	+	n.d.	n.d.	<0.0001	<0.0001
p089	p089	-	+	n.d.	n.d.	<0.0001	<0.0001
p091	p091	-	+	n.d.	n.d.	<0.0001	<00001
p092	p092	-	+	n.d.	n.d.	n.d.	0.041
p093	p093	-	+	n.d.	n.d.	n.d.	0.031
p113	gp3	head-proximal tip of tail tube tail completion + sheath stabilizer protein	+	n.d.	n.d.	0.009	n.d.
p175	p176	-	-	n.d.	n.d.	<0.0001	<0.0001
p203	gp48	baseplate tail tube cap	-	n.d.	n.d.	<0.0001	<0.0001

Table 4.2: Results from statistical testing of antisense expression within the SeqMonk package
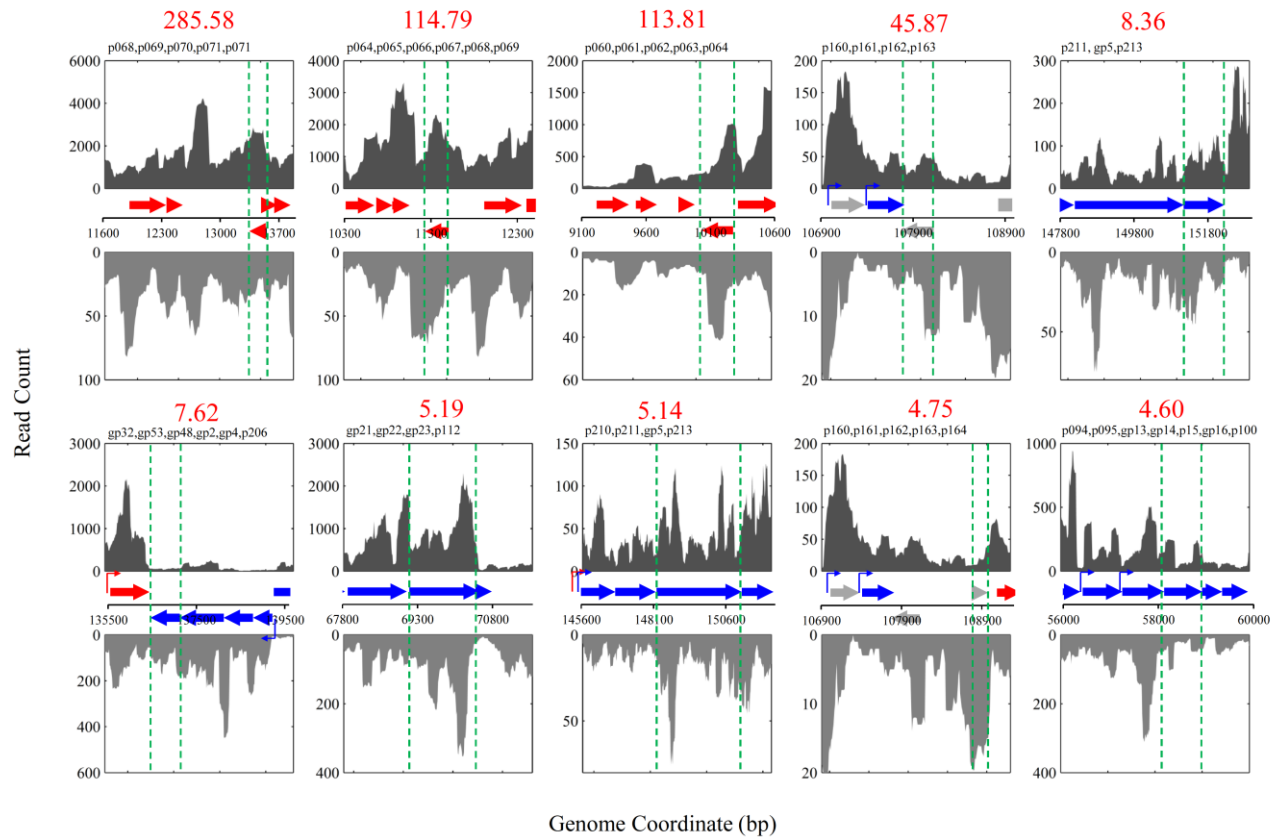
Figure 4.24: Genomic context and read coverage of top 10 scoring sense antisense pairs in cyanophage S-PM2. The ORF names are given above each plot in order left to right. The ORFs are coloured according to the expression cluster with red meaning early, blue meaning late and grey meaning unclustered. Early and late promoters are shown in red and blue respectively. In each case the top plot indicates the Watson strand expression and the bottom shows the Crick strand. Dashed green lines show the location of the top scoring ORF and the value above the plot in red is the ratio of antisense: sense expression. Note the differences in scales between Watson and Crick read coverages.

4.3.10. Attempts to map transcription start sites

According to the ScriptSeq™ library preparation, intact 5' transcript ends should be retained due to the random priming of the TTO to the 3' end of the cDNA fragment (Fig. 4.7). Given that transcription initiation is a directed process, i.e. it occurs at discrete TSSs then one may expect bias in the start site of reads to a specific base. Therefore, for each base in the genome the RSS score was calculated (see Section 4.2.5). Indeed, specific bases do have a bias for an increased RSS score (Fig. 4.31a). The upstream 50 bases from each high scoring RSS were retrieved from the cyanophage S-PM2 genome sequence. Figures 4.31b-d show sequence logos for each upstream region exceeding certain RSS cut-offs. What is immediately apparent is the sequence conservation at the -1:-5 region from the high scoring RSS. The conserved sequence has the consensus ' GATCT'.
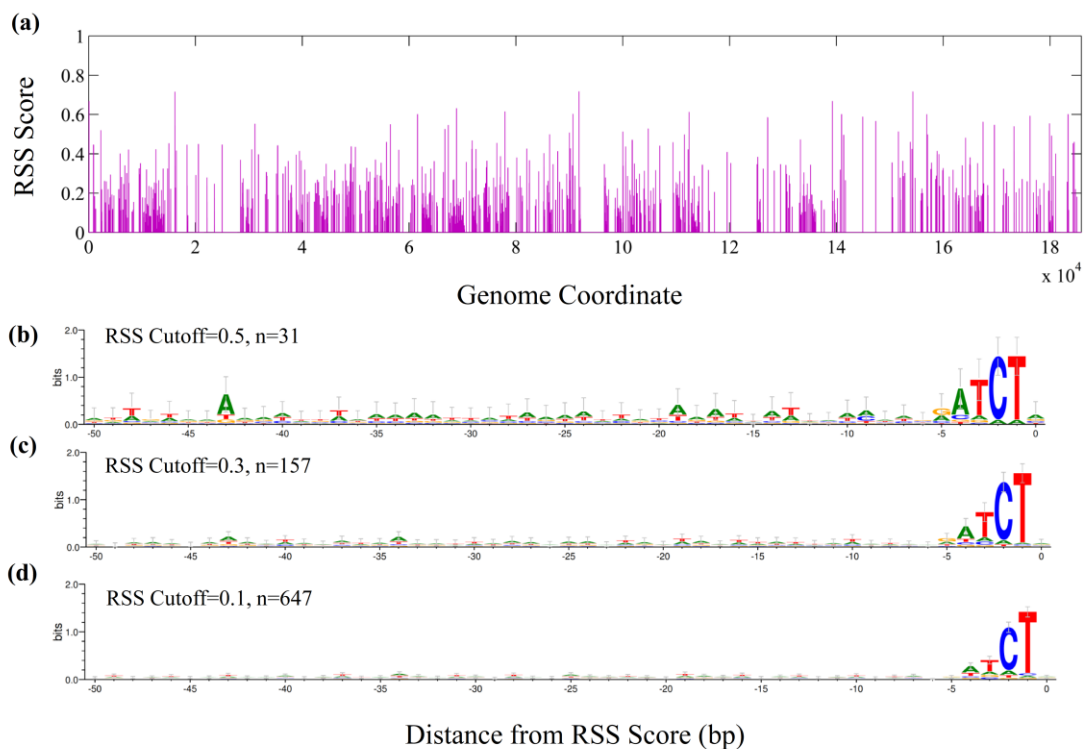
Figure 4.25: Results of global TSS mapping. **(a)** The distribution of RSS scores across the genome. **(b-d)** Sequence logos of the 50bp upstream regions of bases with a RSS score greater than 0.5 **(b)**, 0.3 **(c)**, 0.1 **(d).**

A closer inspection of the ScriptSeq^{TM} library preparation protocol reveals this conserved motif is an artefact of preferential annealing of the TTO to an RNA region that contain the 3' end of the TTO sequence. A thorough description of the bias is given in appendix 7. Thus, using the ScriptSeq^{TM} v2.0 protocol is was impossible to globally map TSSs.

4.4. Discussion

4.4.1. Deleted ORFs

The deletion of ~10kb (5.3% of the genome) observed in S-PM2 during routine propagation is unprecedented amongst bacteriophages. Due to the presence of a PCR product of one size on isolated plaques of S-PM2 it is likely that this deletion was spontaneous and involved deletion of 35 ORFs (14.3%). The deleted ORFs are between p017-p050 thus the deleted strain is named S-PM2$^{\Delta p017:050}$. It is suggested here that there is a significant fitness burden to carrying these ORFs during routine propagation, as S-PM2$^{\Delta p017:050}$ became dominant in the population as inferred by the number of RNA-Seq reads. No homologues of the deleted ORFs were found in the ncbi –nr database. Therefore the function and distribution of these ORFs is unknown. Homologues were found in the GOS database suggesting prevalence in marine organisms. Interestingly, the synteny of the deleted ORFs is not conserved amongst the GOS hits suggesting intense exchange in this genetic region. It is hypothesised here that the deleted ORFs may offer some fitness benefit to the phage during particular conditions encountered in the environment. The isolation of both S-PM2 and S-PM2$^{\Delta p017:050}$ will enable characterisation of such fitness benefits.

4.4.2. Global temporal regulation of S-PM2 gene expression

The data presented in this chapter suggest the temporal expression of S-PM2 ORFs can be divided into two distinct clusters. Thus ORFs can be broadly classified as to having an early or late mode of expression. This is consistent with the predicted choreography of transcription in S-PM2 (Mann et al. 2003; Clokie et al. 2006) as inferred through similarity

to the model myovirus T4 (Miller et al. 2003). In microarray analysis, T4 has been shown to possess 4 temporal clusters of gene expression: immediate early, delayed early, middle and late (Luke et al. 2002). The molecular machineries that regulate these clusters are well understood: Early transcripts possess an upstream promoter that broadly resembles the host σ^{70} factor binding site (Miller et al. 2003). That is, with the exception of an upstream extended -10 region and a slightly modified -35 region that contains an upstream A-rich tract (UP) (Miller et al. 2003). These modifications are thought to enhance interaction with the host σ^{70} factor by direct binding and also through the effect on DNA curvature (Ross et al. 1998; Estrem et al. 1999). Acting in concert with the RNAP modifying protein, GpAlt, these factors contribute to greater expression of the T4 early genes compared with the host's (Koch et al. 1995). The middle profile is controlled by two T4 encoded proteins, AsiA and MotA. AsiA is an anti-σ factor that when complexed with σ^{70} interrupts interaction with the -35 element and redirects binding of RNAP to the middle promoter sequences (Colland et al. 1998; Hinton and Vuthoori 2000; Urbauer et al. 2002; Minakhin et al. 2003). MotA is a transcriptional activator that binds to the 'MotA box' located at -30 relative to the TSS. Finally, late transcription is accomplished through the activity of three T4 encoded proteins: Gp55, Gp33 and Gp45 (Miller et al. 2003). Gp55 and Gp33 both bind RNAP (Horvitz 1973; Ratner 1974) with Gp55 encoding a small σ factor. Gp45 encodes the sliding-clamp processivity factor of T4 DNA polymerase (Wu and Geiduschek 1975) and hence T4 late transcription relies on concurrent DNA replication (Riva et al. 1970). S-PM2 shares all three genes required for late transcription but is lacking the middle regulators a*siA* and *motA* (Mann et al. 2005). This may explain the apparent lack of a robust middle profile in the cluster analyses. Interestingly, a 'pseudo-middle' profile was identified from those ORFs that did not cluster into the early or late modes. In addition a 'continuous' profile was also found that is characterised by early onset of expression followed by maintained expression throughout infection. The existence of these profiles suggests selection for expression of genes that are outside the 'umbrella' of early/late global regulation. Many of these genes share homology with genes of known function. Of these, many cases for necessity of

'pseudo-middle' or 'continuous' expression can be made. For instance, Gp55, the σ-factor for late expression, requires an early onset of expression to allow for transition from early to late transcriptional phases. Indeed this is the case in T4, where Gp55 clusters within the early group but is characterised by a reduction in expression after the initial peak (Luke et al. 2002). This is not observed in the S-PM2 Gp55, where expression continues throughout infection.

Many of the other unclustered ORFs are components of nucleotide metabolism (Td, NrdA, NrdB_1 and NrdB_2) and therefore may be required early during infection to provide the building blocks for DNA synthesis. Continued expression may be required to fuel DNA synthesis that extends further into the middle and late period. Similarly, two subunits of the putative T4-like replisome are unclustered (Gp61 and Gp62) as well as two components of the join-copy recombination-replication pathway (Gp46 and Gp47) as well as the Dam methylase, p129.

Interestingly multiple AMGs are also found in the unclustered group. These include the NTP pyrophosphohydrolase MazG, the S-adenosylmethionine decarboxylase SpeD, the putative phycobilin lyase CpeT and both copies of the high light inducible polypeptide Hli03 (Fig 4.12). S-PM2 is relatively depauperate in AMGs (Millard et al. 2009; Clokie et al. 2010; Sullivan et al. 2010). Thus the only AMGs present in S-PM2 that were not unclustered are *psbA* and *psbD,* encoding the core subunits of the PSII reaction centre (early cluster), the phosphate-induced stress gene *phoH* (late cluster), the gene *cobS* involved in cobalamin synthesis (early cluster), and the 3 copies of the 2-OG-Fe(II) dependent oxygenases (1 early, 2 late). Thus there appears to be a propensity for horizontally acquired AMGs to display an irregular profile. It is shown that many of these AMGs are found in genomic locations between clusters of early or late genes and a wide diversity of mechanisms could act to produce such expression profiles. For instance it is speculated that transcriptional read through from upstream ORFs, highly structured RNA sequences and antisense expression

could all act to modify the expression of these ORFs. A major future goal will be to provide experimental evidence for these mechanisms.

4.4.3. Complexities of the S-PM2 transcriptional landscape

The analyses presented here reveal several features of the S-PM2 transcriptional landscape that appear to be anomalous. In particular, the late period of infection is characterised by genome-wide pervasive expression on the Watson strand. This is evinced in Figs. 4.8 and 4.9 where expression of intergenic nucleotides is considerable, but not apparent in WH7803. This feature precludes designation of transcriptional boundaries and instead one or two continuous transcripts are predicted for the entire S-PM2 Watson strand during the late period (Appendix 6). It is suggested that two separate events can lead to such a landscape. The first is pervasive transcription initiation (PTI) and the latter is incomplete or inefficient transcription termination (Singh et al. 2014). Genome-wide PTI has been shown to be a general feature of prokaryotic transcriptomes (Dornenburg et al. 2010; Sharma et al. 2010a; Mitschke et al. 2011; Voss et al. 2013). Recent studies suggest that AT rich regions are particularly susceptible to non-canonical PTI (Singh and Grainger 2013; Singh et al. 2014). In *E. coli,* the histone-like nucleoid structuring (H-NS) protein acts to silence genome wide PTI, which was once thought to solely silence transcription from horizontally acquired AT rich DNA (Ueda et al. 2013). H-NS appears to be absent form cyanobacteria (Tendeng and Bertin 2003) and it is not clear whether a separate mechanism has evolved to silence PTI.

In T4, the onset of the late period of transcription is marked by increased specificity of RNAP to the late promoter motif with consensus TATAAATA, occurring between -13→-6 bp from the TSS (Kassavetis et al. 1986; Miller et al. 2003; Geiduschek and Kassavetis 2010). Whilst the conservation of this promoter sequence is remarkable compared with either the early or middle consensuses (Miller et al. 2003; Geiduschek and Kassavetis 2010), there remains several exceptions to the consensus. In all, only the A at the -9 site is conserved at 100% across all identified T4 late promoters (Geiduschek and Kassavetis

2010). In addition, *in vitro* transcription by T4 modified RNA polymerase occurs at several variants of the consensus, including the conserved -9 nucleotide and including variants in multiple positions (Kassavetis et al. 1986). Further, three important caveats to the use of the T4 consensus have been raised by Geiduschek and Kassavetis (2010). 1. Many sites have not been experimentally validated and thus are biased because they are being searched for. 2. Most mapping of promoters has been by primer extension analysis that cannot differentiate between *bona fide* transcription initiation and RNA processing and 3. The relative activity of the late consensus and its variants have not been established *in vivo*. Finally initial binding of Gp55 complexed RNAP does not occur with any great site specificity (Kolesky et al. 2002; Nechaev and Geiduschek 2008), yet upon opening of the RNAP complex, site specific transcription initiation is observed. In comparison to the T4 consensus, the S-PM2 consensus shows no conservation of the T at position -13 with the consensus ATAAATA. Moreover, in the freshwater cyanophage S-CRM01 it has been suggested that the CTAAATA motif plays a prominent role in late transcription initiation. It is suggested here that such degeneracy must cause a 'problem' for phages with propensity for AT rich genomes (T4=64.7%, S-PM2=62.2%). This is emphatically demonstrated in Fig 4.25. Here, the frequency of the ATAAATA promoter sequence is shown in the S-PM2 genome. With only 1 mismatch in the consensus, the frequency of the occurrence of the consensus is dramatically increased from 69 to 997. Moreover, there is no bias against any other nucleotide in any position in the consensus (Fig. 4.26b). Interestingly, the only positively identified TSS in the genome was that upstream of the high scoring late gene *p158* (Fig. 4.22). Here, the -12:-6 region has the sequence AATATAT, which diverges from the predicted S-PM2 consensus at 5/7 sites. Thus it may be the case that low-to mid-level transcription initiation is pervasive throughout the genome at such AT rich sites. Certainly Gp55 appears to be expressed continuously during infection which is not observed in T4 (Luke et al. 2002). It is likely therefore that Gp55 will be in abundance during the late period and therefore likely outcompetes any other sigma factor for RNAP. Unfortunately the ScriptSeq™ v2.0 protocol contained a bias that contributed to annealing of the TTO to specific sequences that can be found within a

transcript rather than to the 5' end of the transcript as predicted. This precluded the experimental identification of TSSs to add strength to the above hypothesis. Instead, the dRNA-Seq protocol suggested in (Sharma et al. 2010a; Mitschke et al. 2011) could be used to map genome-wide TSSs. Alternatively 5' RACE with TAP treatment would allow for identification of a selection of TSSs that can be differentiated from processed transcripts (Vogel 2003).

It is curious why T4-like phages encode such a divergent σ-factor (Geiduschek and Kassavetis 2010) with such low specificity in promoter sequence. As described below it is suggested here that this may have been selected for during evolution and that spurious transcription initiation may provide a fitness benefit for T4-like phages (section 4.4.4).

.
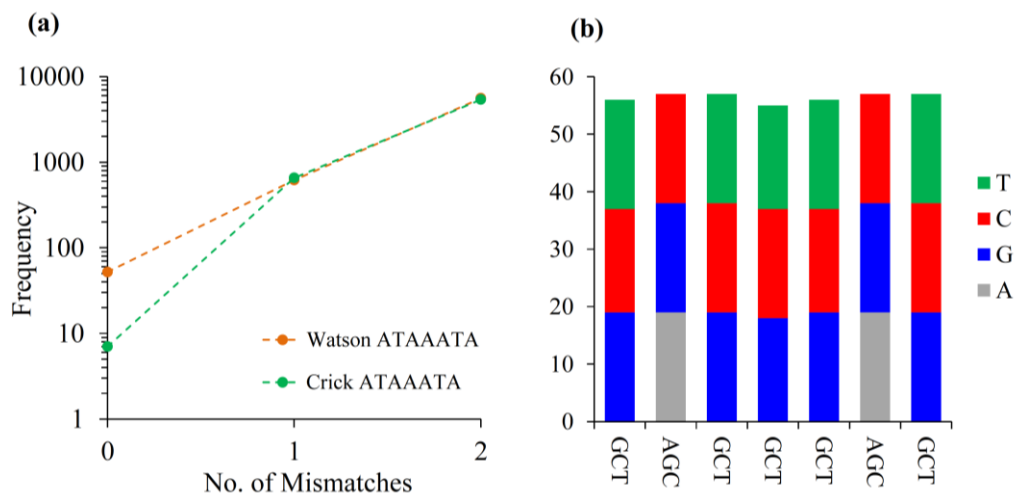
Figure 4.26: (a) Frequency of late promoter motifs in the S-PM2 genome. **(b)** Nucleotide bias in late promoter mismatches. Each bar represents each nucleotide of the late promoter consensus ATAAATA. The label corresponds to each variant other than the consensual nucleotide and the y-axis indicates the frequency of those variants.

The second hypothesis is that inefficient transcription termination between neighbouring ORFs contributes to the observed transcriptional landscape. No rho-independent terminators have been experimentally validated in S-PM2, however, several programs exist for their prediction. Predictions of rho-independent terminators for the Watson strand of S-PM2 were

therefore made using ARNold (http://rna.igmors.u-psud.fr/toolbox/arnold/). This predicted 47 Watson strand terminators of which 25 were intergenic (Appendix 3). The remainders were either antisense to a Crick strand ORF or were intragenic. Of the 25 intergenic, 13 were found in what are referred to here as boundary loci. These are defined as contiguous pairs of ORFs where the downstream ORF either has a different expression profile (Appendix 5) or has a different orientation. It was calculated that there are 44 boundary loci on the Watson strand. In comparison, 12 of the terminators were found in non-boundary intergenic loci. This is out of a possible 140 Watson strand non-boundary loci. Thus there is a clear enrichment for a rho independent termination signal in boundary loci as is observed for T4 (Miller et al. 2003). Figure 4.27b shows the observed mean frequency of termination of these predicted terminators as described in section 4.2.9. Only 4 termination signals displayed termination frequencies above 80%. These are the termination region in between the early *psbA* gene and the late downstream homing endonuclease *f-CphI*, the region between hypothetical late gene *p095* and late clustering head completion protein *gp13*, the region between the early expressed *psbD* and the early hypothetical *p183* and finally the region between the unclassified *hli03_2* gene and the late Crick strand encoded *p175* gene. ¾ of these genes are horizontally acquired AMGs and rho independent terminators can be identified downstream of host orthologues.

Most other terminators appear to be between 50 and 80% 'efficient' using this method and indeed many show no termination activity whatsoever. Therefore it is highly likely that the lack of transcriptional termination and by consequence read-through transcription contributes to the observed transcriptional landscape of S-PM2. Recently, it has been shown that 44% of intrinsic terminators from *E. coli*, *Bacillus subtilis,* T-odd phages, M13 phage and λ lambda phage were <80% efficient in *E. coli* cell expression systems (Cambray et al. 2013) and therefore inefficient intrinsic termination is not unprecedented. The efficiency of termination appears to be predominantly related to the characteristics of the 3' polyU tract as

well as the relationship between the thermodynamic stability and length of the hairpin structure (Cambray et al. 2013).
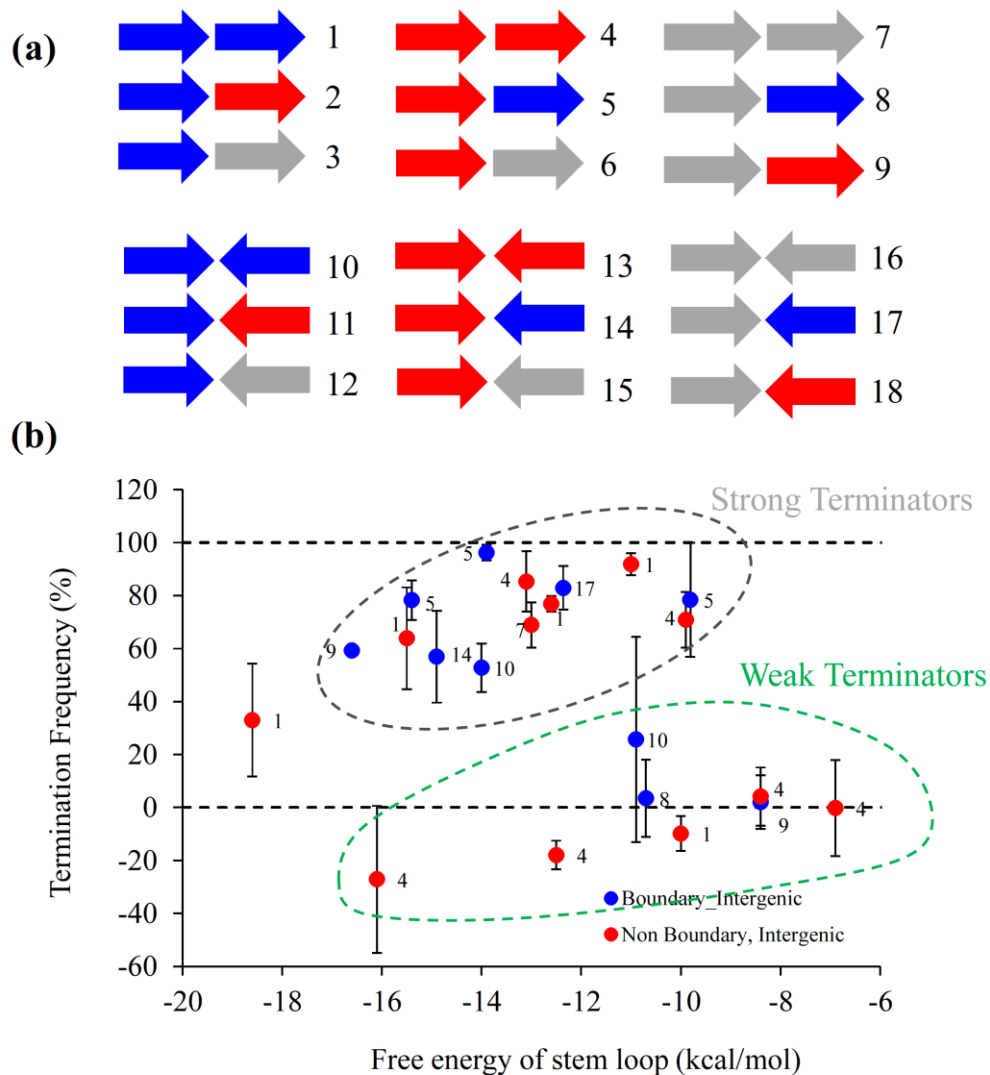
Figure 4.27: Termination frequency of bioinformatically predicted terminators **(a)** Schematic showing the possible locations of bioinformatically predicted terminators. In each case the terminator is located between two ORFs. The direction of the arrow indicates the direction of the ORF. The ORFs are then coloured by their expression profile, red is early, blue is late and grey is unclustered. **(b)** For each ORF the termination frequency is shown as described in section 4.2.10. Data points are coloured according to whether they are in a boundary loci or not as described in the text.

4.4.4. Cyanophage transcriptional landscape in the context of AMG acquisition and strand bias

It is suggested above that a combination of inefficient transcription termination and genome-wide pervasive transcription at AT rich sites as a function of low late promoter specificity

contributes to the observed transcriptional landscape of S-PM2. Moreover, cyanophages are characterised by frequent acquisition and maintenance of horizontally acquired AMGs (Millard et al. 2009; Sullivan et al. 2010). Thus it is hypothesised here that this transcriptional landscape has been selected for to allow for expression of recently acquired genes. It remains largely unclear how AMGs are acquired (Sullivan et al. 2010), but it has been speculated that many factors may be responsible. For instance, site-specific recombination with nearby tRNA genes (Williams 2002; Campbell 2003; Sullivan et al. 2010), whole-module shuffling facilitated by mobile genetic elements (Zeng et al. 2009; Sullivan et al. 2010) and illegitimate recombination events with low sequence similarity (Hendrix et al. 1999; Hendrix 2002; Arbiol et al. 2010; Hatfull et al. 2011) all play a role. Indeed, many cyanophages, including S-PM2, encode homologues of the T4 recombination genes; *uvsW, uvsX* and *uvsY* (Mann et al. 2005). Regardless of the mode of acquisition, recently acquired genes would probably lack the transcriptional regulatory requirements (e.g. late promoter) to be expressed during late infection. Therefore, retention of such genes may be selected against if they cannot be expressed. It is suggested that the presence of read-through transcription, particularly on the Watson strand facilitates late period transcription of recently acquired genes. Subsequently, during evolution, AMGs gain the sequence requirements for regulated transcription. This scenario is best witnessed in the S-PM2 copy of *psbA* as will be discussed in chapters 5 and 6. This hypothesis also explains the remarkable propensity for ORF strand bias in cyanophage genomes (Fig. 4.28), which has received little attention in the literature. For almost every cyanomyovirus >90% of ORFs are found on one strand. This is with the exception of S-TIM5 which belongs to a divergent lineage of myoviruses with very few cultures representatives (Sabehi et al. 2012) and the dwarf myovirus MED4-213, whose genomic signatures resemble a podovirus (section 1.4). T4 also displays a strand bias yet it appears that this bias is related to expression profiles of the ORFs. In particular early and middle promoters are solely found on the Crick strand whilst only 22/50 late promoters are found on this strand (Miller et al. 2003). This is not the

case for S-PM2 where only 1 early and 2 late promoters are found on the Crick strand. Indeed, the expression profiles presented here to not seem to be correlated with polarity.
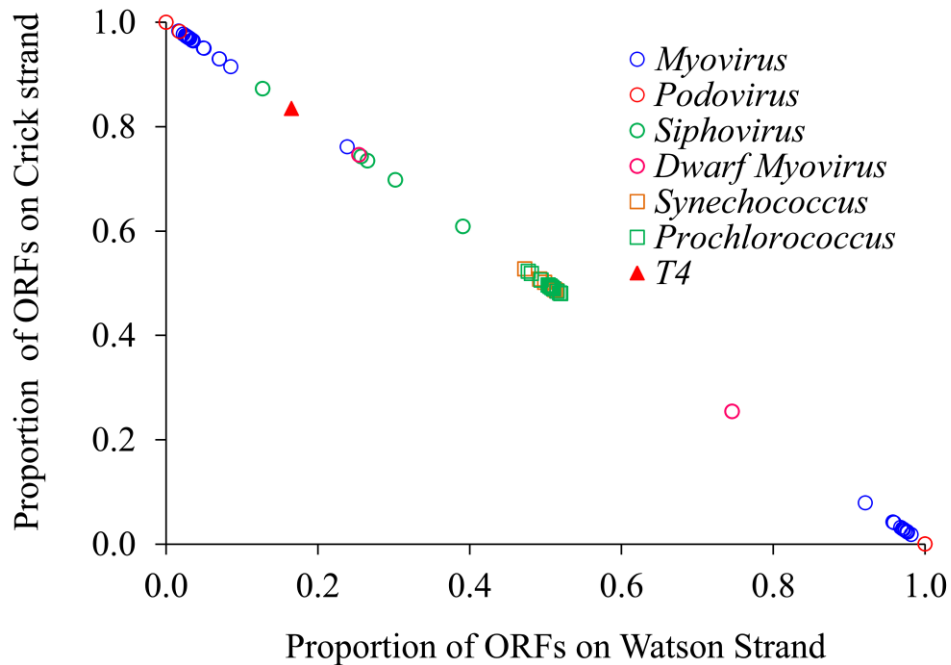
Figure 4.28: Strand bias of ORFs in cyanophage and cyanobacterial genomes.

In closely related coliphages, genome divergence often occurs at discrete loci in the genome known as hyper variable regions (HVRs) (Arbiol et al. 2010). Such HVRs are also found in cyanophages (Millard et al. 2009; Sullivan et al. 2010; Ignacio-Espinoza and Sullivan 2012). Recently Arbiol et al. (2010) has revealed a fascinating mechanism by which such HVR can originate. It is shown in the closely related coliphages RB49 and phi1 that HVRs are more likely to contain Promoter early Stem Loop motifs (PeSLs). These regions contain host σ^{70} like -35 and -10 regions with an AT rich upstream sequence and a paired helix (stem-loop) either upstream, within or downstream of the promoter. It is elegantly demonstrated that these regions act as recombination substrates that cross-over with the transcriptional regulator sequences (promoters and terminators) from other phages during co-infection leading to both gene gain and loss. Such a mechanism means that horizontally acquired genes retain self-regulation. There is some support for the distribution of PeSLs in

cyanophages, but greater genome divergence, reduced conservation of synteny and the fact that PeSLs are, by their nature, similar to endogenous regulatory sequences, means that PeSL designation is difficult (Ignacio-Espinoza and Sullivan 2012). In addition it is likely that PeSL may only provide expression for the early period. It is argued here that most selection for AMGs may occur in the uncharacteristically long late periods in cyanophages. Certainly this has been suggested to be the case for the photosynthesis gene *psbA* (Bailey et al. 2004; Hellweger 2009). Therefore in addition to PeSLs, late period genome-wide pervasive transcription may drive the selection of AMGs.

4.4.5. The role of antisense transcription in cyanophages

Antisense transcription is a prominent feature of bacterial transcriptomes (Georg et al. 2009; Lasa et al. 2011; Georg and Hess 2011; Courtney and Chatterjee 2014). Yet the distribution of asRNAs in bacteriophages is less well known. To date only 2 asRNAs have been identified; the OOP asRNA of bacteriophage λ (section 1.5.1; Krinke and Wulff 1990) and the asRNA of S-PM2 *psbA, CfrI* (Millard et al. 2010). Here asRNAs were detected to virtually every ORF with *CfrI* being shown to be low scoring asRNA: mRNA pair. Therefore this strongly suggests that other higher scoring asRNA: mRNA pairs are *bona fide.* It is suggested here that the wide abundance of asRNAs act to tweak expression levels of their cognate sense mRNAs. asRNAs have suggested to be fast response regulators of gene expression where protein transcription factors may take time to elicit a response (Steglich et al. 2008). This may be crucial for a phage, whose life cycle is much reduced compared with that of its host and where synthesis of protein transcription factors may be energetically costly (see section 3.4.1.). Therefore asRNAs may explain the diversity in the levels in gene expression despite the existence of only two clusters of regulatory proteins. This is especially true for the late period where the late promoter is almost always completely conserved.

There was a propensity for genes with high scoring asRNAs to be located on the Crick strand. This is driven by the genome –wide pervasive transcription on the opposing Watson strand. How these Crick strand mRNAs avoid base pairing with their cognate antisense RNAs remains to be tested? Perhaps Crick strand ORFs maintain their orientation such that their expression is silenced at crucial times during infection by the Watson strand expression. Unfortunately the majority of Crick strand ORFs are hypothetical. This is with the exception of the late cluster of Crick genes encoding Gp53, Gp48, Gp2, Gp4. In T4 these structural proteins are often required at the extreme late period of infection (Miller et al. 2003) where they encode the baseplate wedge protein, the tail tube assembly protein, a protein that is packaged into the virion and protects the phage DNA from endonuclease cleavage and the head completion protein respectively (Miller et al. 2003). Therefore the expression of these genes may be silenced earlier during infection by asRNA interaction.

These hypotheses remain to be tested but future work should aim to validate the existence of the asRNAs by direct methods such as northern blotting and also seek to validate their interaction with target mRNAs.

Chapter Five: Differential expression of genes in response to light intensity

5.1. Introduction

Marine environments are extremely dynamic, providing a rapidly changing light environment for marine cyanobacteria (MacIntyre et al. 2000). This is superimposed on the diurnal and seasonal changes in the solar flux which have been shown to drive niche partitioning in marine cyanobacteria (Johnson et al. 2006; Six et al. 2007). The ability for organisms to sense and respond to environmental stimuli determines their success (Mella-Flores et al. 2012). Here, this concept is extended to cyanophages, where, especially in the *Myoviridae* family, latent periods have shown to be relatively long (~12 hrs). There have been very few reports of environmental sensing in lytic cyanophages. Zeng and Chisholm (2012) demonstrate increased expression of the phage encoded *pstS* and *phoA* genes in response to phosphate limitation. These encode a high-affinity phosphate binding protein and an alkaline phosphatase respectively. In the host, the phosphate stress response is controlled by the Pho-regulon: a two-component system comprised of a sensor histidine kinase (PhoR) which is autophosphorylated in P_i limiting conditions (Tetu et al., 2009). PhoR activates its cognate response regulator, PhoB. Phosphorylated PhoB acts as a transcriptional activator of many P_i stress response genes by binding to the Pho-box which occupies the -35 region. A Pho-box is found upstream of *pstS* and *phoA* in cyanophage S-SM1, which form an apparent operon with gene *g172*. Recombinant host PhoB binds to this Pho-box *in vitro* and is therefore thought to activate the phage genes. Many other phages can 'sense' the metabolic state of the cell with the archetypal example being the regulation of the lytic/lysogenic switch of bacteriophage λ, which is largely regulated by the metabolic state of the cell through the concentration of intracellular proteases (Oppenheim et al. 2005).

Cyanophage S-PM2 was isolated from station L4 in the Western English Channel (Wilson et al., 1993, 1996), which undergoes large and rapid fluctuations in the light environment due to daily and seasonal fluxes as well as short term mixing events and water column biomass accumulation (blooms) (Groom et al. 2009). The results from Chapter three demonstrated that light strongly influenced the infection dynamics of cyanophage S-PM2. In particular,

the maturation period was significantly reduced following infection of the *Synechococcus* sp. WH7803 host under HL conditions. It was hypothesised that this may in-fact be a regulated event and that regular oscillations in light flux in the environment would have selected for a light responsive transcriptional regulatory system in cyanophage S-PM2. To address this RNA-Seq was used to quantify expression of all cyanophage genes in HL and LL.

The specific aims of this chapter were:

1. To identify cyanophage S-PM2 genes whose expression was regulated by light availability
2. To determine the regulatory circuit controlling the light dependent response
3. To identify *Synechococcus* sp. WH7803 host genes that were differentially expressed in response to light.

5.2. Methods

5.2.1. Infection

Synechococcus sp. WH7803 was grown in ASW medium (section 2.1.2) to 1×10^8 cells ml^{-1} at 23˚C under continuous illumination of 15 μmol photons m^{-2} s^{-1}. Infection was carried out with cyanophage S-PM2 at a VBR of 10. Samples were incubated at LL (15 μmol photons m^{-2} s^{-1}) for 1 hr to allow adsorption. Three biological replicates were then shifted to 210 μmol photons m^{-2} s^{-1} and three were maintained at 15 μmol photons m^{-2} s^{-1}. 50 ml samples were taken at 1 (immediately after the shift), 3, 6 and 9 hr after infection for total RNA extraction.

5.2.2. RNA extraction, rRNA removal, library construction and sequencing

RNA extraction followed the protocol described in section 2.1.7. Samples were DNase treated using a TURBO DNA-*free*™ kit (Ambion®-Life Technologies, Carlsbad, U.S.A) following the manufacturer's instructions with 5μg RNA. Samples were tested for gDNA

contamination using primers phoH_F and phoH_R (Appendix 1). Any samples shown to yield a PCR product were further DNase treated. In total, samples were DNase treated twice. For each biological replicate, total RNA from time points 1, 3, 6 and 9 hr after infection were pooled in equimolar quantities. rRNA removal was achieved using Terminator^{TM} 5'-monophosphate dependent nuclease as described in section 4.2.2. RNA Seq libraries were prepared using the ScriptSeq^{TM} v2.0 kit as described in section 4.7. Libraries were prepared by the Centre for Genomic Research, University of Liverpool (http://www.liv.ac.uk/genomic-research/). Sequencing was also performed by the Centre for Genomic Research, University of Liverpool on an Illumina GAIIx platform.

5.2.3. Sequence mapping

Paired end reads were mapped to the S-PM2$^{\Delta p017:050}$ genome using Bowtie 2 (Langmead and Salzberg 2012) as described in section 4.2.4. For counting and differential expression analyses presented in this chapter paired reads were not joined to create fragments as described in section 4.2.4.

5.2.4. Read counting

Differential expression analyses by DE-Seq (Anders and Huber 2010) requires raw counts that map to a given locus. Thus, HTSeq-count (Anders et al. 2014) was used to count reads that map to every ORF, tRNA and RNA sequence in the S-PM2 and *Synechococcus* sp. WH7803 genome. Counts were made using –union –intersection-strict and intersection-nonempty methods (Anders et al. 2014).

5.2.5. Differential expression

DE-Seq v. 1.16 (Anders and Huber 2010) was used for testing of differential expression. Counts were normalised to the apparent library size factor within the DE-Seq package. Multiple hypothesis testing was corrected for using the Benjamini-Hochberg procedure as computed by DE-Seq and p-values less than 0.05 were considered significant. Only genes

that were significantly differentially expressed using all three counting methods within the HT-Seq package were accepted for conservatism.

5.2.6. Sequence analysis

MEME v.4.9.1 (Bailey et al. 2009) was used to detect conserved motifs in upstream regions of cyanophage *psbA*s. A maximum of 10 motifs were searched for with a minimum size of 5 and a maximum of 50. A minimum occurrence of 70% and a maximum of 100% were used. Zero or one occurrences of the motif per sequence were predicted. A cut-off of 1e-5 was used for significant motifs.

Significant motifs reported by MEME were searched for in fasta files of all sequenced *Synechococcus, Prochlorococcus* and S-PM2 genomes using FIMO (Bailey et al. 2009), with a *p*-value cut-off of 1e-9.

5.2.7. Conservation of RNA secondary structure

To predict whether conserved secondary structure existed in the upstream regions of myoviral *psbA* sequences LocARNA was used (Will et al. 2012). Global alignments were made using LocARNA-P with a structure weight of 200, an indel opening score of -500, and indel score of -350.

5.2.8. asRNA: RNA interaction prediction

Sites of interaction between *psbA* mRNA and the asRNA *CfrI* were predicted using IntaRNA (Busch et al. 2008) and computed within the Freiberg RNA Tools suite (Smith et al. 2010). A seed length of 7 was used with 0 mismatches in the seed allowed. Energy computation was made for 23°C with a folding windows size target of 150 and a maximum bp distance of 100.

5.3. Results

5.3.1. High light increased yield of mRNA

Mapping statistics and coverage estimation for these sequenced libraries are shown in appendix 4b. The fraction of non-rRNA reads that mapped to the cyanophage S-PM2 genome was significantly increased in HL ($\bar{x}$=2.34 fold increase, t_3=3.98, p=0.014, Fig. 5.1a). In HL, 18.90±4.42 % of reads mapped to cyanophage S-PM2 compared with only 8.09±1.61% in LL. This means a significant fraction of the reads still mapped to *Synechococcus* sp. WH7803. Due to pooling of time points it is likely that the majority of these came from samples early during infection. The estimated RAW coverage for S-PM2 and *Synechococcus* sp. WH7803 in HL and LL is shown in Fig. 5.1b. It shows that the *Synechococcus* sp. WH7803 coverage was unaffected by light intensity, whereas cyanophage S-PM2 coverage was increased in HL. This strengthens the idea that most *Synechococcus* sp. WH7803 reads come from the early time points.
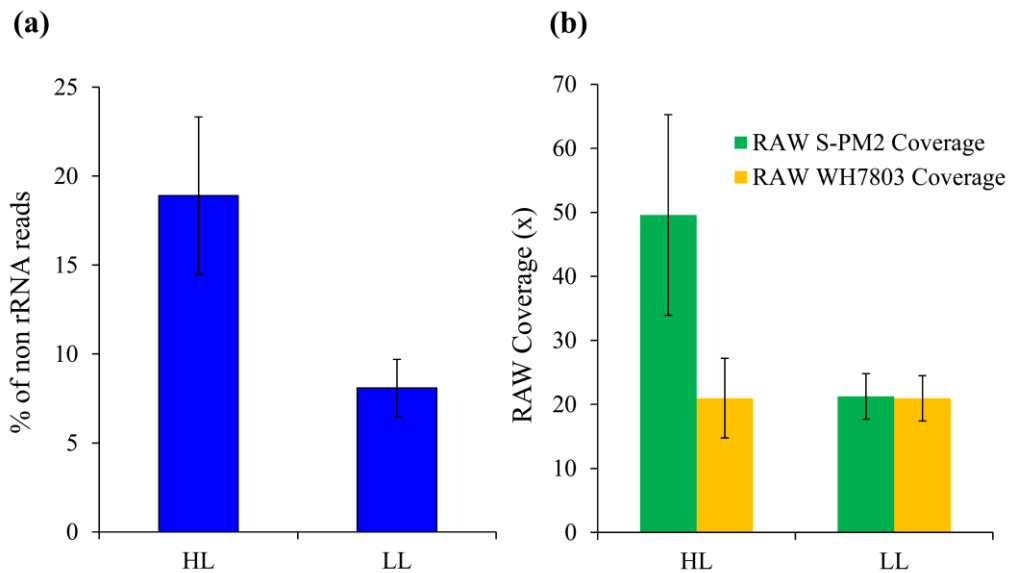
Figure 5.1: **(a)** Percentage of reads mapping to the cyanophage S-PM2 genome from HL and LL samples. **(b)** Raw coverage of S-PM2 and Synechococcus sp. WH7803 genomes at HL and LL. Coverage was calculated as the sum of lengths of all fragments (paired reads) divided by the genome size.

5.3.2. S-PM2 genes responsive to light

Results from combined counting with HTSeq-count and differential expression testing with DE-Seq revealed only one S-PM2 ORF that was statistically differentially expressed in response to increased light intensity. This was the horizontally acquired photosynthesis gene *psbA*. S-PM2 *psbA* was up-regulated 1.72±0.33 fold (*p*=0.007±0.004) in HL. Here, the mean fold change and *p* value is of the 3 counting methods employed within HTSeq-count.

RNA-Seq allows for base-pair resolution of gene expression data. Therefore the coverage per nucleotide was calculated for the *psbA* region as described in section 4.2.6. Fig. 5.2a shows there is a general increase in RPKM across the entire length of the S-PM2 *psbA* ORF in HL. This suggests that this is a *bona fide* increase in expression in response to high light and not due to spurious read-through into the *psbA* ORF. In general, the coverage of the *psbA* ORF is quite unusual. At the 5' end there is low relative expression. There is a bias for reads to start approximately 80bp into the ORF (Fig. 5.2b). Inspection of the sequence upstream reveals a sequence similar to the 5' of the terminal tagging oligo (TTO) which was the cause of read start bias shown in section 4.3.9. Hence, these reads probably extend into the 5'UTR of *psbA* but the presence of the bias precludes examination of the TSS. In addition, an interesting feature exists in all samples approximately half way into the ORF. This feature is characterised by a large reduction in the read coverage (Fig. 5.2a) driven by a propensity for reads to stop at genomic co-ordinate 125,627 (Fig. 5.2b). An inspection of the genomic sequence of this region reveals no obvious site of termination, nor any significant RNA secondary structure. Therefore, it is hypothesised that this could be the site of complimentary base pairing by the antisense *CfrI,* whose 3' termination site was unknown (see Chapter 6). Such a dsRNA site may be subject to endonuclease cleavage and may therefore manifest as a reduction in the reads mapping to this loci. This hypothesis is further discussed in chapter 6. To provide evidence for this site being the site of interaction, the program IntaRNA was used. This program predicts sites of interaction based on similarity (identical in this case) and based on the availability of unpaired nucleotides in the predicted

secondary structure (Busch et al. 2008). IntaRNA predicted that the most likely site of interaction was downstream of the observed decline in read coverage (Fig. 5.2a) and thus does not support the hypothesis that this is indeed the site of interaction.
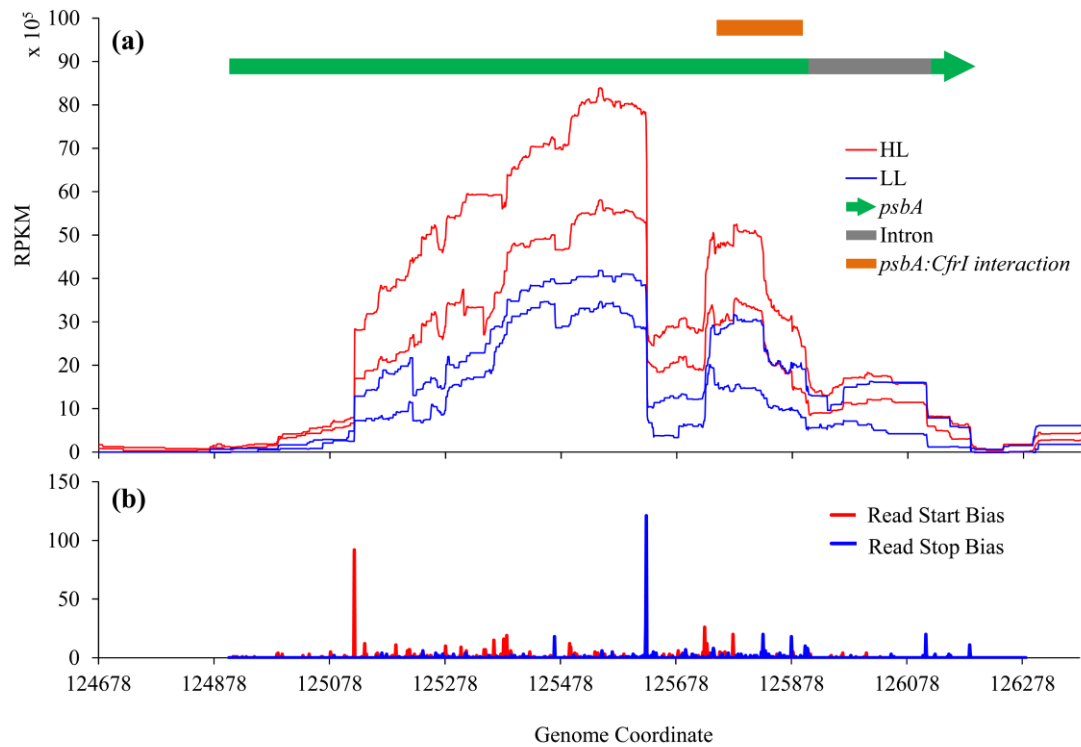
Figure 5.2: (a) RPKM coverage plot of the S-PM2 psbA region. The two red lines indicate the maximum and minimum coverage from HL and the blue represent the maximum and minimum coverage from LL. The orange bar represents the predicted sites of interaction between psbA and CfrI as predicted by IntaRNA. **(b)** Read start (red line) and read stop (blue line) bias of the psbA region.

5.3.3. The upstream region of psbA

The upstream region of the cyanophage S-PM2 *psbA* gene is unusually long (232bp, Fig. 5.3) and contains both a T4-like early and late promoter (Fig. 5.3). Both promoters are located distantly from the S-PM2 start codon (85-112bp, Fig. 5.3). Therefore, it was hypothesised that this extensive region could encode transcriptional regulatory sequences. To provide evidence for this, the regions upstream of *psbA* were scanned in sequenced cyanomyoviruses that contain *psbA* upstream regions greater than 50bp in length using MEME software. 29/41 sequenced myoviruses contain *psbA* upstream lengths greater than 50bp.

133
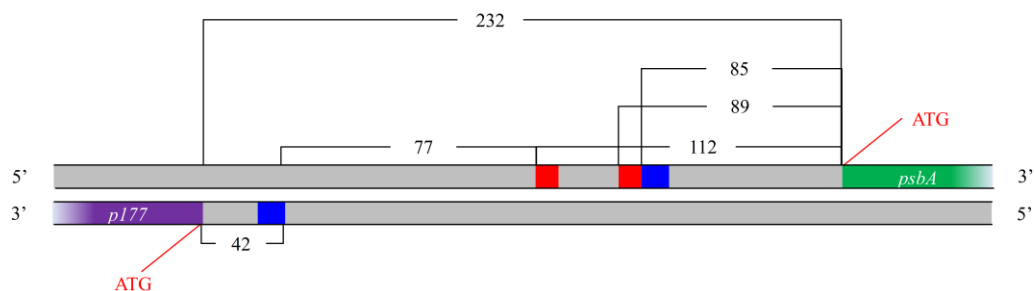
Figure 5.3: Genomic context of the cyanophage S-PM2 *psbA* upstream region. Start codons of *psbA* and ORF *p177* are shown in red. Blue boxes correspond to sites of late promoters and red boxes indicate -35 and -10 regions of the early promoter. The lengths of the various regions are indicated in the line intersections and are in bp.

5.3.4. Conserved motifs predicted by MEME

MEME predicted 6 significant motifs in the upstream regions of cyanophage *psbA* genes (Fig. 5.4). The locations of these motifs in each of the 29 selected *psbA* upstream regions are shown in Fig. 5.5. MEME succesfully detected the sequence conservation conferred by the T4-like early and late promoter motifs (Fig. 5.4). 20/29 sequences contain motif 3 that contains the late promoter sequence. Of the nine that do not, four encode the late promoter sequence in a different location and thus contribute the weak late promoter signal shown at the 3' end of motif 1 (blue square bracket). Therefore, 5 *psbA* upstream regions do not encode a late promoter. These are found in phages P-SSM3, S-TIM5, S-SSM4, S-SSM6a and S-SSM6b. Of these, all encode early promoter -35 and -10 regions that contribute to motifs 6 and 3 respectively. These motifs (3 & 6) are referred to as 'known regulatory'. An exception is S-TIM5 that does not contain either motif 6 or 3 or a late promoter. However, S-TIM5 does contain the sequence TTTAC-N_{17}-TATAAT that differs from the *psbA* early consensus TTGAC-N_{18}-TATAAT by only 1 nucleotide. Interestingly, there exists greater sequence conservation in regions downstream of the early and late promoters in cyanophage *psbA* upstream regions. Specifically, the -35 pentamer appears to be extended to an octomer with the consensus TTGACGGA (Fig 5.4). In addition downstream of the late promoter, there is extensive conservation with the consensus ATAAATAGGTAAACAAAT (Fig 5.4).

The S-PM2 sequence matches this consensus identically. The S-PM2 genome was searched with the position weighted score matrix (PWSM) for motif 3. Returned hits were screened for possession of the exact sequence TAAATA, corresponding to the late promoter. Of those that contained this sequence none contained the aditional nucleotides and indeed FIMO did not return any other significant hits in the S-PM2 genome for motif 3.

Motifs 1, 2, 4 and 5 are referred to here as 'accessory motifs' since they contain no sequence of known regulatory function. Motif 1 occurs in 26/29 cyanophages and has the 38bp consensus YMMCWSCCCTYWAACCRAGACCWMKAGGGTGTHWAAWY. Three phages do not contain this motif: METAG-MbCM1, P-HM2 and S-PM2. In fact S-PM2 is rather depauperate in conserved motifs, only possessing the known regulatory motifs 6 and 3 corresponding to the early -35 and the early -10/late promoter respectively (Fig. 5.4 & 5.5). Motif 1 is almost exclusively (22/26) found downstream of motifs 6 and 3, suggesting it occurs within the 5'UTRs of *psbA* mRNAs. Motif 1 is also present in sequences which lack an apparent early or late promoter, specifically in S-TIM5 and S-SSM6a. In cyanophages MED4-213 and P-HM1 motif 1 appears to be found upstream of a late and early promoters respectively. FIMO was used to scan genomes of all sequenced marine *Synechococcus* and *Prochlorococcus* for motif 1. The motif was not found in any marine *Synechococcus* or *Prochlorococcus* with an E-value greater than 1e-9.
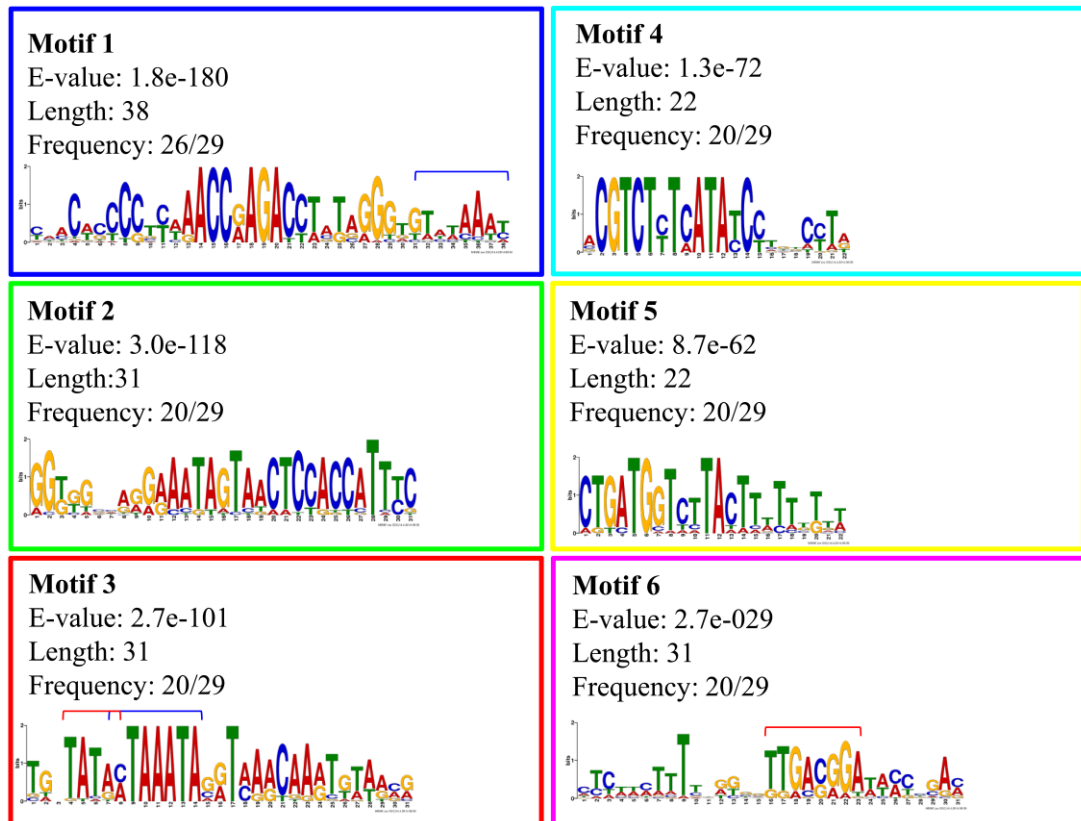
Figure 5.4: Sequence logos of the 5 significant motifs predicted by MEME. Length of the motif and e-value are shown. Blue bracket indicates sequence conservation corresponding to a T4-like late promoter and the red bracket indicates the -35 or -10 regions of the early promoter.

Canonicaly the order of the motifs is 5'-6-3-1-4-2-3-3', ocurring in 14/29 phage. All of these are *Synechcoccoccus* phages isolated from diverse oceanic regions (e.g. Sargasso Sea, Red Sea, Indian Ocean, California Upwelling and Western Atlantic contintental shelf seas). *Prochlorococcus* phages are lacking in many of these motifs, though motif 1 is found in all apart from P-HM2. 3 *Prochlorococcus* phages (P-SSM5, P-SSM2 and P-RSM4) contain motif 3 containing the T4 late promoter sequence. Those that do not contain motif 1, instead contain motif 6 encoding the early -35 hexamer (Fig 5.4). This suggests there is a dichotomy amongst *Prochlorococcus* phages to either select for the early or late promoter upstream of *psbA*. Motifs 2 and 4 can be found in two *Prochlorococcus* phages (P-SSM5 and P-SSM2). In fact, the *psbA* upstream region of these phages are identical, suggesting they are closely related.
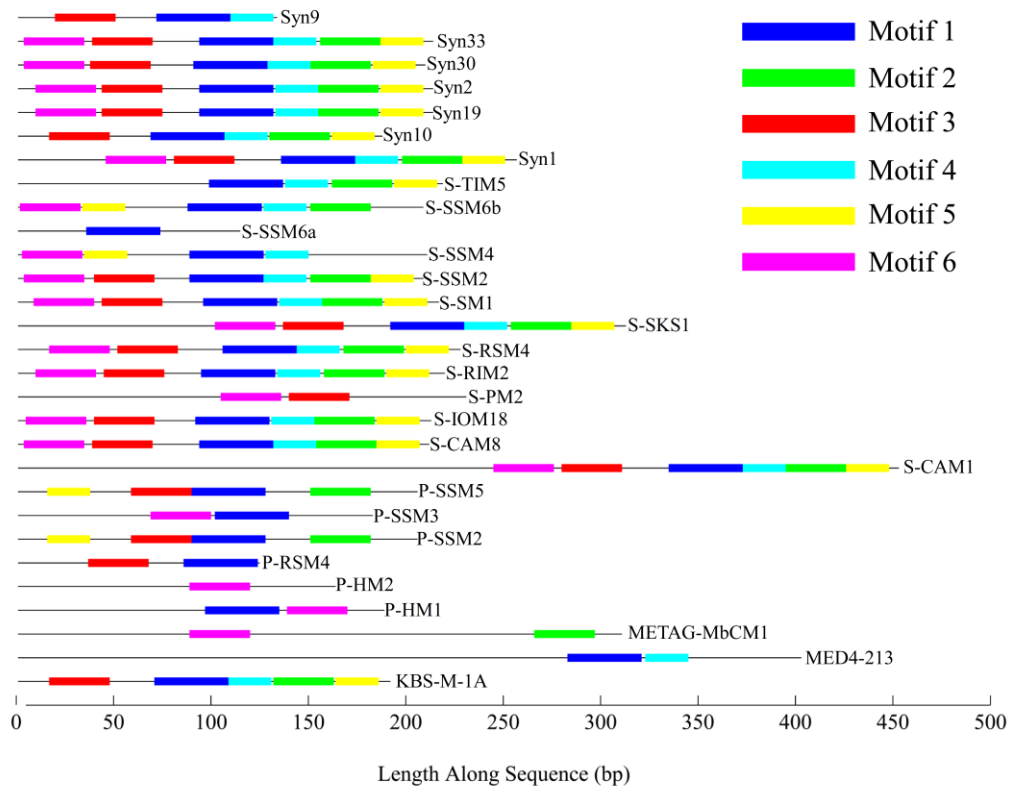
Figure 5.5: Position of high scoring motifs as predicted by MEME on upstream regions of cyanophage encoded *psbAs*

In general these motifs are all found downstream of the early or late promoter sequences, suggesting they occur in the 5'UTRs of *psbA* mRNAs. This is with the exception of motif 5 which is found upstream of the early promoter in phages P-SSM5 and P-SSM2. Indeed, motif 5 shows greater flexibility in its location relative to other motifs. When present, motif 4 is always found upstream of motif 2 and downstream of motif 1. In comparison, motif 5 can be found upstream of the early or late promoter and upstream of motif 1 (Fig. 5.5). Again, none of the sequences were found in any marine *Synechococcus* or *Prochlorococcus* as reported by FIMO.

Due to the putative presence of these motifs in 5'UTRs, it was hypothesised that they may form some conserved secondary structure that may be related to stability of the *psbA* transcript or may encode some regulatory function, for instance a riboswitch. The program LocARNA (Will et al. 2012) was used to align these sequences based on predicted

secondary structures. This analysis suggested no conserved secondary structure existed despite the level of sequence conservation.

5.3.5. Synechococcus sp. WH7803 genes responsive to light

Differential expression analyses revealed some 52 genes in cyanophage S-PM2 infected *Synechococcus* sp. WH7803 that were differentially expressed in response to light intensity (Table 5.1). 26 were up-regulated following transfer to HL and 26 were down-regulated.

Of the 26 genes up-regulated in HL, 19 were either in the functional category 'translation' or 'post-translational modification, protein turnover and chaperones'. These include 12 subunits of the 50S ribosomal complex, the molecular chaperone genes *dnaJ* and *dnaK,* both copies of the chaperonin *groEL,* the *msrB* gene encoding a peptide methionine sulfoxide reductase and *ahpC,* encoding a peroxiredoxin. In addition, a protein required for cytochrome oxidase assembly was also up-regulated in HL.

Other HL up-regulated genes included three copies of the host photosynthesis gene *psbA* (0366, 0790, 2084), all of which encode a high light isoform (Garczarek et al. 2008) and of which the phage orthologue was also up-regulated in HL (section 5.3.2). In addition the alpha subunit of the F_0F_1 ATP synthase was also up-regulated (*1870*).

26 genes were down-regulated in HL. 2 genes were involved in carbon metabolism (*gap1* and *0756*). In the 'transcription' functional category 3 genes were down-regulated. This included two copies of the host σ^{70} factor *rpoD*, of which 7 copies exist in *Synechococcus* sp. WH7803 and a gene encoding a protein responsible for elongation of the period of the circadian rhythm in freshwater *Synechococcus* (*pex*) (Kutsuna et al. 1998). In addition, the ribozyme RNase P appeared to be down-regulated.

Many genes involved in light harvesting and photosynthesis were down-regulated in HL. This included both genes encoding the core heterodimer of PSI, *psaA* and *psaB.* In addition, the expression of the PsaL subunit of PSI was down-regulated. This protein is thought to be

responsible for trimerisation of PSI reaction centres (Chitnis and Chitnis 1993). Subunit IV of the cytochrome b_6f complex, encoded by the *petD* gene was also down-regulated in HL. This protein functions with cytochrome b_6 as the docking site for plastoquinol during electron transfer (Kurisu et al. 2003). The genes *mpeA* and *mpeB* were also down-regulated in HL. These genes are an apparent operon and together encode the α and β subunits of the class II phycoerythrin proteins (Glazer and Cohen-Bazire 1971). In addition the *chlL* gene involved in chlorophyll biosynthesis was down-regulated.

8 down-regulated genes were hypothetical (*69, 463, 934, 940, 1800, 2356, 2423* and *543*) the latter of which contains a LysM repeat domain (Buist et al. 2008).

Interestingly, 5 HL down-regulated genes are thought to be involved in twitching motility. These include the *pilT* gene thought to be involved in providing ATP for motility. Further, a cluster of genes annotated as involved in pilus assembly were down-regulated (Fig. 5.6). The genomic context of these clusters reveals they are found downstream of a putative two component system with *1793* encoding the response regulator and *1794* encoding the histidine kinase (Fig. 5.6). All the genes in the cluster were shown to be down-regulated in HL apart from *1798* whose fold change was approximately equal to the rest of the genes in the cluster but was not significant. This gene cluster is found in many other sequenced marine *Synechococcus* strains from diverse environments. These include the Red Sea isolates RS9917 and RS9916 that belong to *Synechococcus* clades VIII and IX respectively. It is also found in *Synechococcus* sp. CC911, a coastal strain belonging to clade I. The cluster is also found in *Synechococcus* sp. RCC307 but the two genes encoding the two component system regulator are not homologous (Fig. 5.6).

ORF No.	Name	Annotation	CyanoBase[*] GO	Direction	Start	Stop	Base Mean HL	Base Mean LL	Fold change from LL	*p*-value (adj.)
2514	*dnaK*	Molecular chaperone DnaK	Posttranslational modification, protein turnover, chaperones	Up	2344278	2346194	1473.859	794.878	1.860	1.58E-03
2374	*rplK*	50S ribosomal protein L11	Translation	Up	2193738	2194163	84.673	38.520	2.225	8.16E-04
2372	*rplJ*	50S ribosomal protein L10	Translation	Up	2192181	2192708	454.389	189.862	2.402	1.18E-10
2371	*rplL*	50S ribosomal protein L7/L12	Translation	Up	2191741	2192133	260.348	139.603	1.858	6.44E-04
2084	*psbA*	Photosystem II protein D1	Photosynthesis	Up	1914146	1915225	830.324	260.471	3.187	9.24E-04
2069	-	Carbohydrate-binding protein; modular; contains a central CBM2 module	Not in COGs	Up	1898360	1900210	905.012	227.461	3.978	1.67E-12
2017	*atpA*	F0F1 ATP synthase subunit alpha	Energy production and conversion	Up	1844242	1845762	166.451	101.071	1.647	1.45E-02
1998	*groEL*	chaperonin GroEL	Posttranslational modification, protein turnover, chaperones	Up	1826304	1827938	1408.840	621.005	2.270	8.86E-13
1930	-	Alpha-glycosidase of family GH13	Carbohydrate transport and metabolism	Up	1763437	1765419	33.475	9.672	3.456	5.32E-03
1870	-	Uncharacterized protein required for cytochrome oxidase assembly	Posttranslational modification, protein turnover, chaperones	Up	1712400	1713314	34.480	10.661	3.256	1.23E-03
1863	*groEL*	chaperonin GroEL	Posttranslational modification, protein turnover, chaperones	Up	1706431	1708092	356.307	123.142	2.896	2.00E-11
1329	*rpsB*	30S ribosomal protein S2	Translation	Up	1226333	1227052	110.025	49.117	2.220	1.91E-03
1118	*ahpC*	Peroxiredoxin, AhpC/TSA family	Posttranslational modification, protein turnover, chaperones	Up	1024460	1025062	81.567	33.750	2.399	7.81E-03
790	*psbA*	Photosystem II protein D1	Photosynthesis	Up	776911	777990	224.503	62.625	3.594	6.92E-04
434	*rplC*	50S ribosomal protein L3	Translation	Up	446621	447277	147.692	62.019	2.412	1.38E-03

433	*rplD*	50S ribosomal protein L4	Translation	Up	445986	446621	69.100	33.287	2.076	9.94E-03
431	*rplB*	50S ribosomal protein L2	Translation	Up	444812	445675	104.801	46.339	2.260	1.75E-04
428	*rpsC*	30S ribosomal protein S3	Translation	Up	443380	444111	84.335	41.690	2.022	4.73E-03
427	*rplP*	50S ribosomal protein L16	Translation	Up	442890	443363	55.361	25.461	2.177	1.93E-02
422	*rplE*	50S ribosomal protein L5	Translation	Up	441084	441623	110.013	49.668	2.217	1.75E-04
420	*rplF*	50S ribosomal protein L6	Translation	Up	440108	440647	101.087	53.207	1.895	1.04E-02
418	*rpsE*	30S ribosomal protein S5	Translation	Up	439045	439692	99.528	49.825	1.997	6.09E-03
366	*psbA*	Photosystem II protein D1	Photosynthesis	Up	384540	385619	610.568	204.901	2.980	6.65E-04
169	*ahcY*	S-adenosyl-L-homocysteine hydrolase	Coenzyme transport and metabolism	Up	187422	188852	99.561	43.246	2.301	3.37E-03
23	-	DnaJ-class molecular chaperone	Posttranslational modification, protein turnover, chaperones	Up	29438	30568	55.527	26.503	2.097	2.48E-02
16	*msrB*	Peptide methionine sulfoxide reductase	Posttranslational modification, protein turnover, chaperones	Up	18790	19308	209.423	122.141	1.722	1.21E-03
2501	*rpoD*	Alternative RNA polymerase sigma factor, sigma-70 family	Transcription	Down	2330046	2331065	50.111	101.013	0.499	1.23E-02
611	*rpoD*	RNA polymerase sigma factor RpoD, sigma-70 family	Transcription	Down	612192	613544	287.679	488.757	0.587	9.77E-04
464	*pex*	Possible Pex protein (Period-extender gene product)	Transcription	Down	477913	478314	21.909	99.498	0.220	2.75E-04
RNA_3	*rnpB*	bacterial ribonuclease P RNA component of ribozyme	RNA	Down	221709	222102	3527.547	6354.044	0.555	5.56E-08
534	*petD*	Cytochrome b6-f complex subunit 4	Photosynthesis	Down	530577	531059	137.894	247.289	0.565	3.90E-03
396	*psaL*	Photosystem I reaction centre subunit XI	Photosynthesis	Down	424189	424680	50.408	103.036	0.487	9.84E-04
392	*psaB*	photosystem I P700 chlorophyll a apoprotein A2	Photosynthesis	Down	417251	419467	1626.355	3441.585	0.472	1.34E-03
391	*psaA*	photosystem I P700 chlorophyll a apoprotein A1	Photosynthesis	Down	414925	417228	855.180	1570.171	0.546	5.88E-04
2423	-	hypothetical protein	Not in COGs	Down	2255527	2255955	73.155	166.277	0.439	4.91E-05

2356	-	hypothetical protein	Not in COGs	Down	2176215	2178122	155.492	272.235	0.570	5.47E-04
1800	-	Uncharacterised conserved membrane protein	Not in COGs	Down	1648950	1650587	260.300	826.146	0.315	5.39E-08
940	-	hypothetical protein	Not in COGs	Down	878481	878789	17.585	85.210	0.198	1.81E-02
934	-	hypothetical protein	Not in COGs	Down	875397	875651	283.272	2333.681	0.127	5.88E-04
463	-	hypothetical protein	Not in COGs	Down	477328	477882	9.930	37.131	0.266	4.28E-04
69	-	hypothetical protein	Not in COGs	Down	70114	70749	31.375	67.678	0.465	2.44E-02
493	*mpeB*	C-phycoerythrin class II beta chain	Light Harvesting	Down	500108	500653	413.895	816.737	0.518	1.06E-05
492	*mpeA*	C-phycoerythrin class II alpha chain	Light Harvesting	Down	499571	500068	164.702	408.999	0.403	8.30E-04
670	*chlL*	protochlorophyllide reductase iron-sulfur ATP-binding protein	Inorganic ion transport and metabolism	Down	667390	668280	132.112	356.969	0.375	1.59E-04
543	-	LysM-repeat protein	Cell wall/membrane biogenesis	Down	545118	546173	368.624	542.378	0.676	1.13E-02
1841	*pilT*	Twitching motility protein	Cell motility	Down	1687892	1689184	136.572	377.005	0.362	2.77E-03
1798	-	Uncharacterized conserved secreted protein, pili subunit superfamily	Cell motility	Down	1647606	1648121	35.457	98.804	0.361	1.25E-05
1797	-	Uncharacterised conserved secreted protein	Cell motility	Down	1645501	1647477	53.570	155.638	0.344	3.96E-04
1796	-	Uncharacterized conserved secreted protein, pili subunit superfamily	Cell motility	Down	1644983	1645486	78.030	227.037	0.344	2.76E-04
1795	-	Uncharacterized conserved secreted protein, pili subunit superfamily	Cell motility	Down	1644150	1644614	812.139	2612.330	0.310	5.88E-04
756	-	glycogen branching enzyme	Carbohydrate transport and metabolism	Down	738615	740900	41.237	87.293	0.472	1.67E-03
29	*gap1*	Glyceraldehyde-3-phosphate dehydrogenase	Carbohydrate transport and metabolism	Down	34586	35611	70.088	134.549	0.521	9.88E-04

Table 5.1: Differentially expressed genes in cyanophage S-PM2 infected *Synechococcu*s sp. WH7803 following a shift to HL. Gene ontology is based on CyanoBase GOs available at (http://genome.microbedb.jp/cyanobase/)
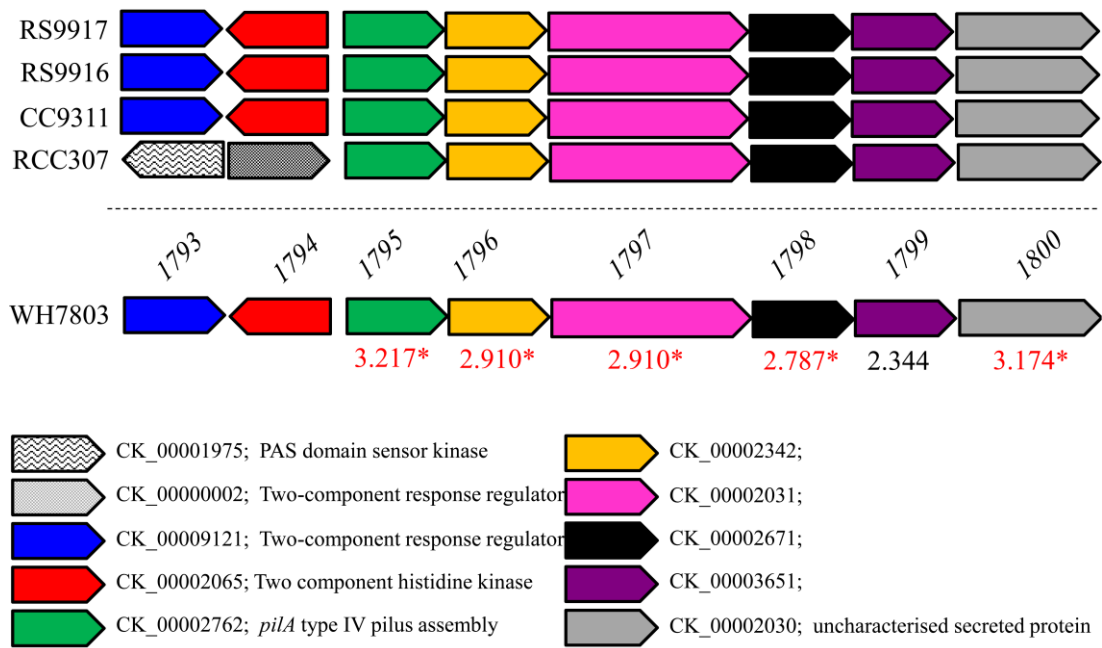
Figure 5.6: Genomic context of the *Synechococcus* sp. WH7803 pilin-like cluster down-regulated in HL and relationship with *Synechococcus* genomes. The *Synechococcus* sp. WH7803 ORF numbers are shown above the genes. The fold change in repression is shown beneath the genes. Significantly down-regulated genes are shown in red with an asterisk.

5.4. Discussion

Despite the dramatic changes in infection dynamics in different light conditions (section 3.3.2) there were comparatively few changes in cyanophage S-PM2 gene expression. Overall, there was an increase in reads that mapped to cyanophage S-PM2 following a shift to HL (~2 fold). This suggests a general increase in transcription across the S-PM2 genome; thus, every gene was approximately 2 fold up-regulated. The cause of this general increase in expression is unclear. In photoautotrophs, the availability of light affects the rate of ATP synthesis through photophosphorylation (Bornefeld and Simonis 1974; Kallas and Castenholz 1982). Hence, light influences the global physiological state of the cell. It is well known that the physiological state of the cell affects overall gene expression (Klumpp et al. 2009; Berthoumieux et al. 2013). Mechanistically, this has been explained by various factors. In particular, the concentration of RNAP and ribosomes, the gene copy number, and the size of the amino acid and nucleotide pools are suggested to play a role (Klumpp et al. 2009; Berthoumieux et al. 2013). Both HL and LL *Synechococcus* were both acclimated to

LL before infection, thus prior to infection the physiological states of each of the cultures would have been similar and therefore it is unlikely that the concentration of RNAP would be different between the light treatments. Berthoumieux et al. (2013) suggests the concentration of the ribosomes plays a role in changes in transcription as a result of global cell physiology. Yet, this is only likely to affect results obtained from promoter-reporter type assays of expression, where expression is actually a function of transcriptional and translational processes rather than the RNA-Seq analysis presented here. Interestingly, many subunits of the host's 50S ribosomal complex were up-regulated in HL. As discussed below, this may contribute to the changes in infection dynamics but is unlikely to affect the rate of transcription.

Other factors have also been shown to affect the genome-wide expression in bacteria. In particular, cellular concentrations of specific alarmones have been shown to be global regulators of gene expression (Magnusson et al. 2005). The "magic spot" nucleotide ppGpp has been shown to bind the β and β' subunits of RNAP and behave as a global regulator of transcription initiation (Toulokhonov 2000; Artsimovitch et al. 2004; Magnusson et al. 2005). A shift of the freshwater cyanobacterium *Synechococcus* PCC7942 to LL causes a rapid increase in ppGpp (Mann et al. 1975). This is complemented with a reduction in RNA synthesis. Interestingly, S-PM2 and many other cyanophages (Bryan et al. 2008) contain a divergent form of the gene *mazG* encoding a ppGpp hydrolase. It has been hypothesised that upon infection, cyanophage MazG prevents the accumulation of ppGpp which would otherwise accumulate as part of the cellular stress response to cyanophage infection. This has been observed in phage infected *Synechococcus* PCC7942 (Borbely et al. 1980). Thus, it is suggested here that this alarmone is unlikely to play a role in the increase in genome-wide transcription at HL.

The energy for transcription comes from hydrolysis of the triphosphate bond of the nucleotide precursor NTP, thus it is conceivable that a reduction in ATP synthesis under a LL regime may limit the energy availability for transcription. However, this could also be

true for DNA replication, yet the rate of phage genome replication is identical in HL and LL (section 3.3.2). Thus, cyanophage S-PM2 may partition energy into DNA replication as a priority over gene expression.

The single phage gene that was differentially expressed above the overall background level was that of *psbA*. As previously discussed, the gene product of *psbA*, D1, is particularly susceptible to light damage (Bailey et al. 2004). A dedicated repair cycle continuously replaces the damaged D1 with *de novo* synthesised polypeptides (Tyystjärvi and Aro 1996; Bailey et al. 2004). Therefore, damage to D1 is dependent on light flux. Up-regulation of phage *psbA* at HL is expected to maintain the supply of D1 to PSII, hence allowing continuing functioning of PSII to support a PMF for ATP synthesis. The enhanced operation of PSII in HL compared with uninfected cells is clearly demonstrated in chapter 3. Thus, it is suggested here that this improved functioning is facilitated by increased HL transcription of the phage *psbA*.

In cyanobacteria the regulation of delivery of D1 is largely thought to occur at the level of transcription (Mulo et al. 2012). D1 is a small gene family with each cyanobacterial strain encoding between 2-6 isoforms, (Mulo et al. 2009). The growth light condition is known to select for each particular isoform (Garczarek et al. 2008; Mulo et al. 2009; Mulo et al. 2012). In the host, *Synechococcus* sp. WH7803, cells adapted to low (15 μmol photons m^{-2} s^{-1}) or medium (75 μmol photons m^{-2} s^{-1}) growth irradiance show a rapid increase (~6 fold) in the *psbA* mRNA pool upon a shift to high light (350 μmol photons m^{-2} s^{-1}) (Garczarek et al. 2008). This is a transient increase that occurs approximately 1 hr after the light shift, after which the *psbA* mRNA pool is reduced back to pre-shift conditions (Garczarek et al. 2008). This shift is coupled with a change in the dominant isoform of *psbA*, from the *psbAI* to the *psbAII* isoform (Garczarek et al. 2008). A similar situation is observed in model freshwater cyanobacteria (Kulkarni and Golden 1994; Soitamo et al. 1996; Tichý et al. 2003; Sander et al. 2010). Cyclostatic growth coupled with analysis of gene expression in *Synechococcus* sp. WH7803 have shown that this isoform-swapping is relevant for variations in the light regime

faced in the environment, occurring in a temporal fashion correlated with diurnal variations in intensity (Mella-Flores et al. 2012). Whilst these global changes in *psbA* expression in response to light are well established, the exact regulatory network remains controversial and incomplete (Mulo et al. 2012). Most work has been in freshwater *Synechocystis* sp. PCC6803 and *Synechococcus* sp. PCC7942. Here, all *psbAs* contain long 5'UTRs (49-88bps) (Mulo et al. 2012). Upstream of the TSS, *psbAs* of the D1:2 isoform contain the hexameric -35 and -10 regions (Mohamed et al. 1993; Li and Golden 1993) that are bound by the principal σ^{70} factor (Shibato et al. 1998). However, light dependent expression has been shown to have a requirement for the alternative σ factors SigB, SigD and SigE (Imamura et al. 2003; Imamura et al. 2004; Yoshimura et al. 2007; Pollari et al. 2009). The *psbA* upstream region is also characterised by AT rich tracts which have been suggested to affect the tertiary structure of the DNA and are required for light responsive expression (Asayama et al. 2002). In addition, several regulatory motifs have been identified in these upstream regions (Li and Golden 1993). However, it is clear that significant regulatory differences exist between *psbA* isoforms and between cyanobacterial strains (mainly *Synechocystis* sp. PCC6803 and *Synechococcus* sp. PCC7942). For instance, *Synechococcus* sp. PCC7942 is thought to sense light quantity and quality through its impact on thiol reductants (Mulo et al. 2012). Addition of a thiol reductant (DTT) under LL conditions induces expression of the HL induced *psbAII* and *psbAIII*, whilst thiol oxidants reversed this result (Sippola and Aro 1999). In *Synechocystis* the thiol reductant pool is thought to be less important for *psbA* regulation (Mulo et al. 2012), though evidence for this is yet to be seen. In addition to differences in light sensing, there appears to be differences in the transcription factors required for light dependent expression. In particular, the protein PsfR has been shown to be a transcriptional enhancer of *psbAI* expression but not *psbAII* or *psbAIII* in *Synechococcus* sp. PCC7942 (Thomas 2004). In contrast, the LysR family transcriptional regulator CmpR has been shown to be a positive regulator of *psbAII* and *psbAIII* expression, binding to enhancer elements upstream (Takahashi et al. 2004), yet is not required for light dependent expression (Takahashi et al. 2004). Interestingly, CmpR also appears to be responsible for

regulating the increase in expression of *psbAII* and *psbAIII* in response to low CO_2 conditions (Takahashi et al. 2004), which may be a result of increased photosynthetic capacity leading to inorganic carbon depletion. In addition, it appears that breakdown products of the D1 polypeptide may specifically bind to the upstream regions of *psbA* and enhance transcription (Stelljes and Koenig 2007). In this way *psbA* may be able to auto-regulate its expression; coupling transcription to light through light dependent damage to D1. Indeed, D1 degradation correlates with *psbA* expression in *Synechocystis* sp. PCC6803 (Tyystjärvi et al. 1994; Mulo et al. 1997; Mulo et al. 1998; Komenda et al. 2000; Li and Sherman 2000). In *Synechocystis* sp. PCC6803, DNA binding proteins have been shown to repress expression of *psbAII* and *psbAIII* in low light and darkness (Eriksson et al. 2000; Herranen et al. 2001).

It is clear from these studies that many environmental stimuli affect expression of the *psbA* gene family. Further, the transcriptional network in response to the light environment is poorly understood. Moreover, there appears to be cross-talk between various environmental stimuli, and finally the mechanisms regulating the transcriptional response may not be conserved between taxa. This makes the study of *psbA* transcription exceptionally difficult and underlines the importance of a systems biology approach to study transcriptional networks, a factor that is only being realised for far better characterised enteric bacteria (Berthoumieux et al. 2013).

Therefore, the exact mechanism that regulates light responsive transcription in cyanophage S-PM2 *psbA* is difficult to predict. However, it shown here that there are sequences upstream of the *psbA* start codon that are conserved between many cyanophages. Firstly, the analysis correctly identified the presence of the putative cyanophage early and late promoter motifs that are found in the vast majority of cyanophages and are referred to here as known regulatory motifs. The early and late promoters are frequently found upstream of *psbA* in the same phage, yet there are instances where only one exists. This suggests there is diversity in the expression patterns of *psbA* during infection between cyanophages. This may be related

to specific changes in the phage's infection dynamics or particular environmental forcing in the isolation site of that phage. The analyses also highlighted a range of other 'accessory motifs'. These appeared to be completely lacking in cyanophage S-PM2 and as such may not participate in light dependent expression. Interestingly, both the early and late promoters appear to have extended conservation surrounding the -35 hexamer and the late heptamer respectively. Such extended regions are not found in any other instances of the early or late promoters throughout cyanophage genomes. Therefore, it is postulated that specific transcription factors may bind to these extended regions and regulate *psbA* expression.

Many of the accessory motifs occur downstream of the early and late promoters and thus may play a role in turnover of the *psbA* transcript or alternatively transcription factors may bind to such motifs and prevent RNA polymerase elongation.

It could be argued that the sequence conservation upstream of *psbA* is merely a reflection of the ancestral host's *psbA* from where the cyanophage acquired the gene. The phylogeny of the cyanobacterial and cyanophage *psbAs* has been contentious (Lindell et al. 2004a; Zeidner et al. 2005; Sullivan et al. 2006; Chénard and Suttle 2008). These variations arise from systematic errors related to variations in G+C% between cyanobacteria and cyanophages and potential intergenic recombination between host to phage (Zeidner et al. 2005; Sullivan et al. 2006) and even between phages (Sullivan et al. 2006). Perhaps the most robust analysis of *psbA* phylogeny comes from Sullivan et al. (2006) and thus the conclusions presented here will follow this phylogenetic hypothesis. It is proposed that *Synechococcus* myoviruses acquired *psbA* at least once from a *Synechcoccus* strain. In comparison, *Prochlorococcus* myoviruses have acquired *psbA* at least twice from a *Prochlorococcus* host, whilst *Prochlorococcus* podoviruses acquired their *psbA* earlier during evolutionary time from an ancestor to all *Prochlorococcus* and *Prochlorococcus* myoviruses. Therefore, this could apply to the motifs identified, given that the canonical motif order (5'-6,3,1,4,2,5-3') is frequently found in *Synechcococcus* phages (14/21) but is not encountered in *Prochlorococcus* phages, though this may also just reflect the low number of

Prochlorococcus phages analysed (6). However, 3 *Prochlorococcus* phages contain at least 1 accessory motif.

In addition, none of the identified motifs could be found in upstream regions of *Synechococcus* host *psbAs*. If these motifs were apparent in the last common ancestor (LCA) and are not selected for in cyanophages, one might expect them to have been lost by drift. This strongly suggests independent selection of cyanophage *psbA* upstream regions compared with the host. Alternatively, the upstream region may not have come from the host but rather was present in the LCA prior to acquisition of *psbA*. Subsequently, divergence of the LCA or mobility of the *psbA* including its upstream region has led to proliferation throughout the population. In either case the conservation of these motifs suggests some purifying selection and therefore functionality. Further work is required to elucidate the biological significance of these motifs but it is proposed here that such sites could be binding sites for *trans*-acting DNA binding proteins that may act as transcription factors to regulate *psbA* in response to a changing light environment.

The analyses presented here revealed a number of host genes that were differentially expressed in response to light intensity. However, the interpretation of these data requires some caution. As reported by Lindell et al. (2007), changes in host expression may reflect the host's response to stress before the phage has taken over the cell's regulatory network. In addition, despite the high VBR (10), some cells may remain that were not infected. Deng et al. (2012) have shown that in the myoviruses S-SM1 and Syn33 a maximum of 50-60% of *Synechcocccus* sp. cells are infected regardless of the VBR used (0.05-10). Thus, the responses observed may be the uninfected host's response and therefore the lack of a no phage control precludes the conclusion that the changes are elicited by phage infection.

There are few reports of the transcriptional response of *Synechococcus* sp. WH7803 to changes in the light environment (Garczarek et al. 2008; Mella-Flores et al. 2012). *Synechococccus* sp. WH7803 encodes 4 copies of the *psbA* gene. 3 are identical at the DNA

level and encode a HL induced isoform D1:2 (Garczarek et al. 2008). The other encodes the D1:1 isoform that is constitutively expressed under LL. It is unclear whether there is a reduction in expression of this isoform when shifted to HL but certainly the D1:2 isoform dominates the *psbA* mRNA pool at 1 hr after the light shift. The results presented here largely agree with the data of Garczarek et al. (2008). Here, all three D1:2 isoforms were significantly up-regulated upon shift to HL, whereas the D1:1 isoform showed no change in expression. Given that all three D1:2 isoforms are identical at the DNA level it is impossible to map reads to each specific ORF (unless the reads extend into the 5'/3' end of the gene). Therefore, given equal probabilities of mapping to all three loci, Bowtie2 aligns the read pseudo-randomly. Thus, the specific fold changes for each D1:2 isoform may be misleading. Clokie et al. (2006) showed an increase in expression of the D1:2 isoform upon infection with S-PM2. This occurred rapidly after infection (1hr) and was reduced thereafter. Thus, the D1:2 isoform appears to be responsive to both HL shift and cyanophage infection.

In addition to *psbA* many other *Synechococcus* host genes were up-regulated under HL. Interestingly, many of these genes were involved in translation and post translational modification. In particular, 12 subunits of the 50S ribosomal complex were up-regulated. The 50S subunit catalyses peptide bond formation by an induced fit mechanism (Schmeing et al. 2005). In chapter 3, it was demonstrated that HL caused early lysis of cyanophage S-PM2. It was discussed that perhaps cyanophage development was limited by the energy required for polymerisation of the components of the virion and that increased rates of photophosphorylation under HL provided energy for an increased rate of translation. It would be tempting to postulate that under HL, the increase in expression of the 50S ribosome causes crowding of ribosomes around the higher abundance phage mRNAs and promotes an increased rate of translation of the structural components of the virion. Whether this is regulated specifically by S-PM2 or whether it is an inherent cellular response to increased light remains to be tested.

Whilst host HL up-regulated genes related to photosynthesis were mainly constrained to the *psbA* family, many other photosynthesis related genes were down-regulated. In particular, the operon consisting of *mpeA* and *mpeB* was down-regulated in HL. Shan et al. (2008) reported an increase in expression of this operon when *Synechococcus* sp. WH7803 was infected with S-PM2. These experiments were conducted at LL and thus it appears that S-PM2 can co-ordinately regulate this operon in response to a wide range of irradiances. *mpeA/B* encodes the α and β subunits of phycoerythrin apoproteins classes I and II, respectively (Bretaudeau et al. 2013). These pigment-protein complexes form part of the phycobilisome structure required for light harvesting in cyanobacteria. Mella-Flores et al. (2012) demonstrate that this operon is up-regulated during the day in uninfected *Synechococcus* sp. WH7803 in cyclostatic growth, showing maximal expression when light was highest. Thus, this appears to be the opposite response to what was seen in cells infected with cyanophage S-PM2 in this study. Moreover, the genes encoding the core complex of PSI (*psaA* and *psaB*) were also down-regulated in HL. Again, the expression of these genes has been shown to be maximal during the light period in uninfected cells (Mella-Flores et al. 2012). Due to these discrepancies between the observed changes in expression here with that reported in the literature it is likely the responses observed here are specific to phage infected cells, though this would require further examination.

Of particular interest are the HL down-regulated genes encoding the putative pilin cluster. This apparent operon is preceded by a putative histidine kinase and response regulator and thus probably regulates expression of this cluster. This pilin cluster is frequently annotated as containing components of the type IV pilus required for twitching motility. *Synechocystis* sp. PCC6803 displays positive phototaxis that is controlled by a type IV pilus complex (Bhaya et al. 2000). Light responsive expression of these genes has not been established to date. However, phototaxis has not been established in *Synechococcus* sp. WH7803, which is largely thought to be immotile (Brahamsha 1996). Moreover, electron micrographs of *Synechococcus* sp. WH7803 (Kana and Glibert 1987) do not reveal the pili structures

observed in *Synechocystis* (Bhaya et al. 2000). In addition, none of the genes in the pilin cluster of *Synechococcus* sp. WH7803 show any conserved motifs that belong specifically to the pilin biogenesis family. For instance, in *Synechocystis* the *hofG* gene (*sll1694*) contains the Pilin_PilA superfamily domain (pfam: 14245) and has been shown to encode the 'thick' pilin (Bhaya et al. 2000). However, such domains are lacking in the *Synechococcus* sp. WH7803 pilin cluster. Indeed, the annotations appear to have arisen from the presence of the N_methyl domain (Pfam:13544) found in ORFs *1795-1799*. Whilst this domain is found in proteins required for the biogenesis of the type IV pilus it is also found in components of the type II secretion system of bacteria (Cianciotto 2005), which itself is used to target the type IV pilin to the outer membrane (Russel 1998; Bhaya et al. 1999).

This is particularly relevant to the *Synechococcus* sp. WH7803 pilin cluster given that that last gene in the cluster (*1800*) belongs to the cyanorak cluster CK_00002030 containing putative secreted proteins. Interestingly, the protein contains von Willebrand Factor (vWF) and metal-ion dependent adhesion site (MIDAS) domains at the C-terminus. The vWF family is widely distributed in all domains of life (Whittaker and Hynes 2004) and in humans is often found in extracellular multiprotein complexes and is linked to many bleeding diseases (Ruggeri 1993). Yet in bacteria it has been completely unstudied (Whittaker and Hynes 2004). Here, it is suggested that *1800* could be a secreted protein that is exported through the type II secretion pathway and may a play a role in the stress response to high light. Unfortunately, whether this response is specific to phage infected cells is unclear from the current dataset.

5.5. Concluding remarks

The data from this chapter suggests that whilst the rate of transcription by cyanophage S-PM2 was approximately 2 fold greater in HL, there was only 1 gene that was specifically differentially expressed (*psbA*). In addition, an apparent paucity of light responsive genetic elements disproves the hypothesis that variations in light intensity have selected for a light

responsive regulatory circuit in S-PM2. This may suggest that the differences observed in the timing of lysis reported in chapter 3 are not a regulatory event or are regulated at the post-transcriptional level. Instead, it is hypothesised that the observed variations in infection dynamics are the result of differences in the energy state of the phage infected cell resulting from increased rates of photophosphorylation under HL. A rate limiting step of the rate of photophosphorylation is the continued generation of a transmembrane potential. This transmembrane potential is derived from charge separations in the PSII and PSI complexes leading to linear and cyclic photophosphorylation. HL can inactivate PSII (through damage to the D1 polypeptide) and hence the rate of photophosphorylation would be limited unless the supply of D1 could be increased at HL. Here, it is shown that the only differentially expressed gene in HL was *psbA*, indicating a clear mechanism to increase the supply of D1 under HL. Thus, the acquisition of *psbA* and in particular light responsive expression of *psbA* by the cyanophage has overcome a major metabolic bottleneck in the cell which provides energy for genome wide transcription and translation of the virion.

Chapter Six:Transcriptional regulation of the cyanophage S-PM2 *psbA* gene

6.1. Introduction

In chapter 4 it was shown that the temporal expression of cyanophage S-PM2 genes during infection can be separated into two distinct clusters. Further, it was suggested that these global changes in expression are caused by modifications of the host RNA polymerase in a manner similar to bacteriophage T4. *psbA* formed one of few genes whose affiliation to either cluster was not strongly supported statistically. Thus, expression of *psbA* appears to be distinct from global regulation. In addition, in chapter 5 it was shown that cyanophage S-PM2 *psbA* is the only S-PM2 gene whose expression is responsive to HL. Curiously, as discussed in section 1.5, the genetic organisation of cyanophage S-PM2 *psbA* is most unusual (Millard et al. 2004; Zeng et al. 2009; Millard et al. 2010; Bonocora et al. 2011). S-PM2 *psbA* is interrupted by a 212 nt group I self-splicing intron which protects *psbA* from cleavage by a downstream homing endonuclease (Zeng et al. 2009). Further an asRNA overlaps the 5' end of the homing endonuclease and the 3' end of *psbA* and therefore links these two genetic elements. Thus, it was suspected that either the intron or the asRNA may contribute to the peculiarities surrounding *psbA* expression.

The specific aims of this chapter were:

1) To examine the kinetics of unspliced and spliced *psbA* transcript accumulation during infection under LL and HL by use of RT-qPCR. This was to test the hypothesis that the cyanophage S-PM2 *psbA* intron may respond to light and thus have a role in gene regulation.

2) To examine the kinetics of *CfrI* transcript accumulation during infection under HL and LL with RT-qPCR.

3) To create an artificial repressor mutant in *Synechococcus* sp. WH7803, such that upon infection *CfrI* expression would be repressed and therefore the exact role in regulation of phage (and perhaps host) *psbA* could be elucidated.

6.2. Methods

6.2.1. Infection of Synechococcus sp. WH7803 with cyanophage S-PM2

To compare the accumulation of spliced and unspliced *psbA* at HL and LL, infection was carried out with cyanophage S-PM2 infecting *Synechococcus* sp. WH7803. *Synechococcus* sp. WH7803 was grown in ASW medium (see section 2.1.2) to 1×10^8 cells ml^{-1} at 23˚C under continuous illumination of 15 μE m^{-2} s^{-1}. Infection was carried out with cyanophage S-PM2 at a VBR of 10. Phages were left for 1 hour to adsorb at LL before three biological replicates were shifted to HL (210 μE m^{-2} s^{-1}) and three were maintained at LL. 50 ml samples were taken at 0, 1, 3, 6, 9 hrs after infection for total RNA extraction.

6.2.2. Total RNA extraction and cDNA synthesis

RNA extraction followed that described in section 2.1.6. RNA was extracted and cDNA synthesized. cDNA synthesis was carried out with SuperScript III reverse transcriptase (Life Technologies, Carlsbad, USA) in 20 μl volumes with 2 μg of total RNA. Each reaction contained 0.5 mM each dNTP, 250 ng random hexamers or 2 pM of gene-specific primer (GSP), 1x SuperScript III buffer, 5 mM DTT and 200U SuperScript III. The RNA, dNTP and random hexamers/GSP were mixed and heated to 65˚C for 5 mins before being cooled on ice for 5 mins. SuperScript III, DTT and buffer were then added, mixed and incubated at 50˚C for 60 mins. When using a GSP, the incubation temperature was increased to 55˚C. When using random primers the reaction was pre-incubated at 25˚C for 5mins. The reaction was inactivated by heating to 70˚C for 15 mins.

6.2.3. Cloning of psbA spliced and unspliced transcripts

Validation of qPCR assays required the cloning of both spliced and unspliced *psbA* transcripts. Total RNA extracted from infection of *Synechococcus* sp. WH7803 by cyanophage S-PM2 was subjected to cDNA synthesis as described above with random hexamers. cDNA was amplified using primers psbA_F/SPM2psbA_R (see Appendix 1)

which were directed at cyanophage S-PM2 *psbA*. PCRs were in 50 μl and contained 1 μl cDNA synthesis reaction, 5 μl 10x Platinum® Taq PCR buffer (Life Technologies, Carlsbad, USA), 1.5 mM MgCl$_2$, 0.4 μM each primer, 0.2 mM each dNTP, 1 unit Platinum® Taq (Life Technologies, Carlsbad, USA) and water. Annealing was at 55°C and ran for 35 cycles.10 μl of PCR product was run on a 1% (w/v) agarose gel for confirmation of splicing (i.e. 2 bands corresponding to the correct size). The PCR product was then subcloned into pCR2.1® TOPO® TA (Life Technologies, Carlsbad, USA) following the manufacturer's instructions and transformed into chemically competent *E. coli* DH5α as described in section 2.1.9. 20 white colonies were picked and grown overnight in LB medium (see section 2.1.2) +50 μg ml⁻¹kanamycin. Plasmids were purified from these clones using a miniprep kit (Qiagen, Venlo, Netherlands). These plasmids were then subject to PCR using the primers SPM2psbAF/SPM2psbAR using the same conditions as described above. 5 μl PCR product was analysed on a 1% (w/v) agarose gel to detect cloning of either the spliced or unspliced *psbA*. Plasmid inserts were sequenced using M13F/M13R primers to confirm cloning of the correct transcript. Purified plasmids containing the unspliced *psbA* were named pCR2.1_psbAUns and those containing the spliced *psbA* were called pCR2.1_psbASpl. These purified plasmids were used for template DNA when validating *psbA* spliced and unspliced qPCR assays.

6.2.4. Validation of qPCR Assays

qPCR assays were developed to quantify 4 targets in this study: 1. Cyanophage S-PM2 unspliced *psbA,* 2. Cyanophage S-PM2 spliced *psbA,* 3. *Synechococcus* sp. WH7803 16S rRNA gene, 4. *CfrI*. The *Synechococcus* sp. WH7803 16S rRNA gene qPCR assay was as reported in Clokie et al. (2006). The other primer/probe sequences are reported in Appendix 1 (psbA_spl_F, psbA_spl_R, psbA_spl_probe, psbA_unspl_F, psbA_unspl_R psbA_unspl_probe). To employ relative quantification by the efficiency corrected ΔΔCT method (Pfaffl 2004), validation of these assays was required. This involved amplification of a dilution series of the relevant template such that a standard curve can be generated (Bustin

et al. 2009). Parameters of validated assays are reported in Table 6.1 and conformed to MIQE guidelines (Bustin et al. 2009).

Parameter	16S	*psbA*_Unspliced	*psbA*_Spliced	*CfrI*
Amplification Efficiency (E)	1.06	1.12	1.13	1.09
R^2	0.9983	0.9986	0.9981	0.9992
Dynamic Range (CTs)	9.115-18.735	9.354-35.876	9.192-36.019	10.236-35.978

Table 6.1: Parameters of validated qPCR assays used in this study

6.2.5. qPCR assays

Total RNA was extracted as described in section 2.1.7. cDNA synthesis was as described above. In the case of the 16S rRNA gene, random hexamers were used to prime cDNA synthesis. In the case of unspliced *psbA*, spliced *psbA and CfrI,* GSPs were used (see Appendix 1). This was because random hexamers would detect the opposing strand (i.e. *CfrI* or the 3' end of *psbA* respectively). 2 µg of total RNA was used as template. Reaction volumes were 20 µl containing 1x Brilliant III Ultra-Fast qPCR Master Mix (Agilent, Santa Clara, USA), 1x relevant primer/probe assay (500 nM primers and 250nM probe), 1 µl cDNA and nuclease free water to 20 µl. An Applied Biosystems® 7500 Fast Real Time PCR System (Life Technologies, Carlsbad, USA) was used for quantification. Cycling parameters were 95˚C for 1min followed by 40 cycles of 15 s at 95˚C and 15 s at 62˚C. Raw C_T values were exported and analysed using the efficiency corrected $^{\Delta\Delta}$CT method (eq. 6.1; Pfaffl 2004).

$$R = \frac{E_{Tg}^{\Delta C_T Tg}}{E_{Rf}^{\Delta C_T Rf}} \quad \text{[Eq. 6.1]}$$

Where E_{Tg} is the amplification efficiency of the target gene, $\Delta C_T Tg$ is the change in C_T between the control and sample. E_{Rf} is the amplification efficiency of the reference gene, and $\Delta C_T Rf$ is the change in C_T between the control and the reference gene. If the gene expression value is relative to T0, then the control is the C_T at time point 0 and the 16S

rRNA gene is the reference. If the gene expression is relative to another gene (i.e. spliced to unspliced or relative to *CfrI*), then that C_T is the reference and the 16S rRNA gene is the control. Relative expression values were tested for significance using REST (Pfaffl et al. 2002).

6.2.6. 'Transcript Walking'

The 5' end(s) of *CfrI* have previously been mapped by 5' RACE (Millard et al. 2010). Based on northern blots a prediction was made as to the 3' terminus of the transcript (Millard et al. 2010). To validate this, attempts were made to map the 3' end of the transcript using variations of 3' RACE. Unfortunately, none of these attempts were successful (data not shown) and therefore an alternative method was employed involving RT-PCR to 'walk' down the transcript. Total RNA from infection of *Synechococcus* sp. WH7803 with cyanophage S-PM2 was isolated as previously described (see section 2.1.7). cDNA synthesis was carried out as previously described (see section 6.2.2) but with 17 different GSPs (Fig. 6.1; Appendix 1, cfrI_Walk_R_A-Q) and with 2 µg total RNA. The GSPs were located at varying points downstream of the mapped TSS of *CfrI* and are shown in Fig. 6.1. PCR was then conducted using primer CfrI_F and the same GSP used for cDNA synthesis. PCR reactions included 1 µl cDNA reaction, 1x MyTaq master mix (BioLine, London, UK), 0.4 µM each primer and water to 50 µl. Absence of a band at a given GSP loci was taken to be indicative of transcription termination upstream of that GSP. No reverse transcriptase controls and positive cyanophage S-PM2 gDNA controls were used to avoid false positives from gDNA contamination or false negatives from failed amplification, respectively.
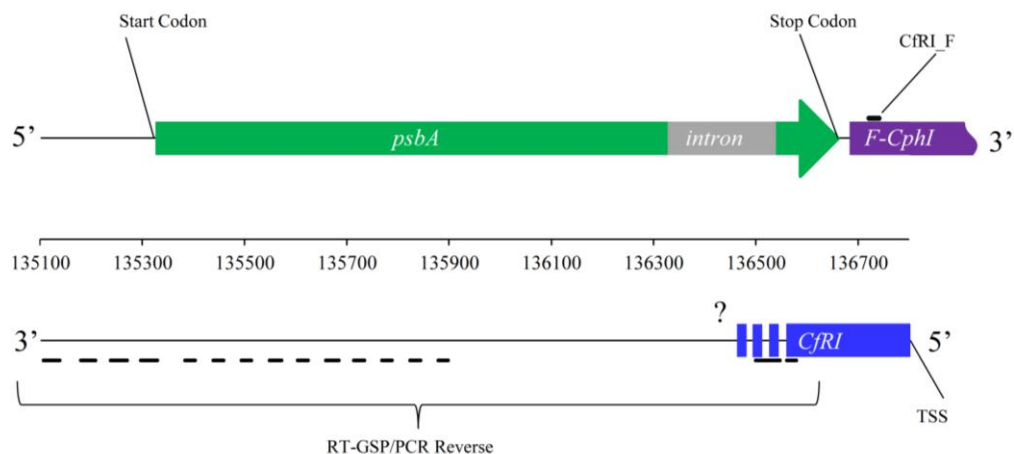
Figure 6.1: Strategy for determination of the 3' termini of *CfrI*. The X axis gives the cyanophage S-PM2 genome co-ordinates with Watson strand features above and Crick strand features below. ? symbolises the unknown 3' terminus of *CfrI*, TSS is the CfrI transcription start site as mapped by 5' RACE (Millard et al., 2010). RT GSP stands for reverse transcription gene specific primer. Black lines indicates the exact positions of the RT-GSPs.

6.2.7. Attempts to express CfrI in Synechococcus sp. WH7803

To test the function of *CfrI* during infection, a strain of *Synechococcus* sp. WH7803 was sought that over-expressed this gene. This over-expression approach has been successful in other asRNAs in cyanobacteria (Dühring et al. 2006) To date there are only two reports of successful heterologous expression of genes in marine cyanobacteria (Tolonen et al., 2006; Shan, 2008). These studies relied on expression from RSF1010 based plasmids that have been shown to replicate in cyanobacteria (Elhai and Wolk 1988; Brahamsha 1996). Thus, the RSF1010 derivative pRL153 (Elhai and Wolk 1988; Brahamsha 1996) was used as the expression backbone. This plasmid has two unique restriction sites that allow for directional cloning (Fig. 6.2). To allow expression, two cyanophage S-PM2 putative promoter regions were fused upstream of *CfrI*. The two putative promoter regions were that of the cyanophage S-PM2 *cpeT* gene (YP_195250.1) that has been used previously to express the cyanophage S-PM2 *cpeT* gene in *Synechococcus* (Shan 2008) and is referred to here as *cpeTp*. The second was the upstream region of the gene encoding a putative surface layer protein in cyanophage S-PM2 (YP_195203.1), which had been shown to be strongly expressed (Clokie et al. 2006) and in a temporal fashion that correlates with cyanophage S-PM2 *psbA* (Clokie

et al. 2006). This DNA fragment is referred to as *slp* here. Termination of transcription was accomplished through use of the bacteriophage λ oop terminator incorporated onto the 5' end of the reverse primer used to amplify *CfrI* (Burr et al. 2000).
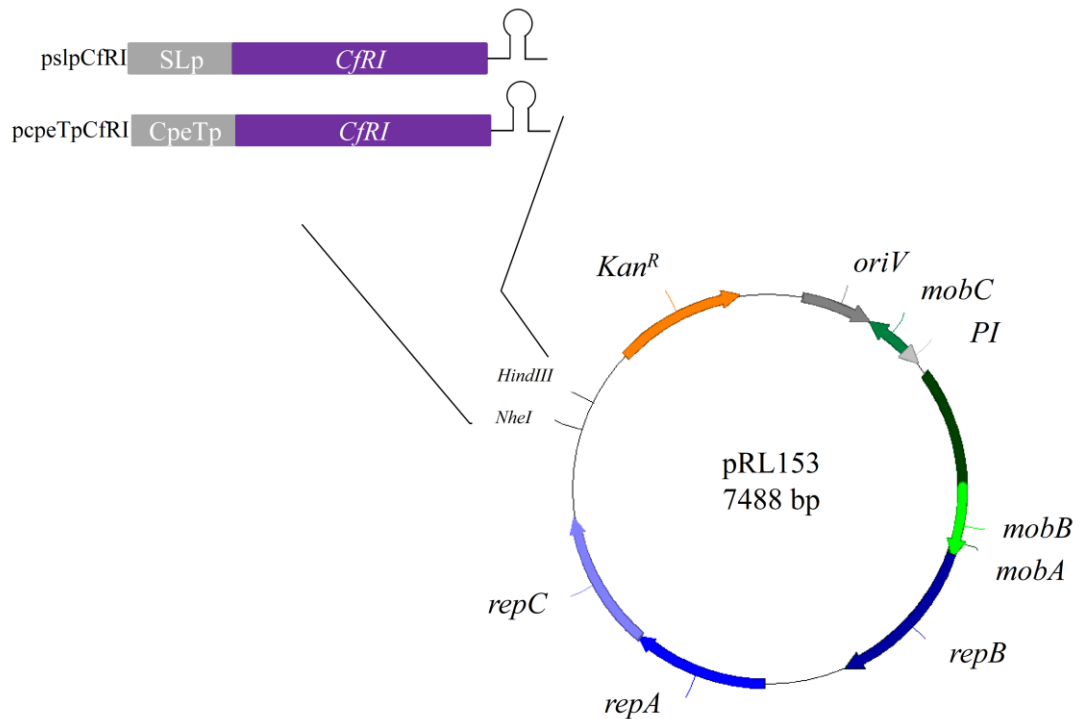
Figure 6.2: Cloning strategy used to produce *CfrI* expression constructs, pslpCfrI and pcpeTpCfrI. The fused PCR products were cloned into the *Nhe*I/*Hind*III sites of the pRL153 plasmid backbone.

Fusion PCR was used to join the promoter fragments to two *CfrI* variants, *CfrI_sl* and *CfrI_cpeT*. These had 5' overlaps at the 3' end of the promoter fragment to allow fusion during PCR. Figure 6.3 shows initial PCRs of the *slp, cpeTp, CfrI_sl* and *CfrI_cpeT* fragments. PCRs were in 50 μl volumes containing 40-100 ng of template, 0.4 μM each primer, 1x KAPA HiFi buffer (KAPA Biosystems, Wilmington, USA), 0.3mM each dNTPs, 1U KAPA HiFi polymerase (KAPA Biosystems, Wilmington, USA) and water to 50μl. Thermal cycling parameters were 95°C for 3mins, followed by 35 cycles of denaturation at 98 °C for 20 s, annealing at 65°C for 15 s, followed by extension at 72°C for 1min. A final extension at 72°C for 1min was applied. Only amplification of the *cpeTp* fragment was

successful at 65°C (Fig. 6.3). Thus a 'touch-up' gradient PCR was used to amplify the remaining fragments.
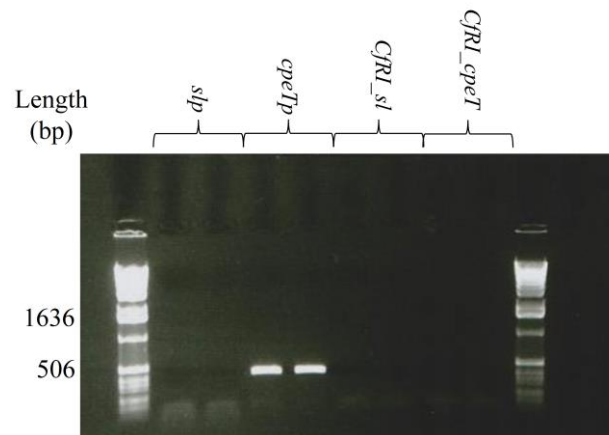
Figure 6.3: Initial PCRs to construct pslpCfrI and pcpeTpCfrI. Lane 1 = ladder. Lane 2&3 = Failed amplification of the upstream region of the putative S-Layer gene of S-PM2 (*slp*). Lane 4&5 = Amplification of the upstream region of the S-PM2 cpeT gene (cpeTp). Lane 6&7 = Failed amplification of the *CfrI* gene with 5' overlap with 3' of slp (*CfrI_sl*). Lane 8&9 = Failed amplification of the *CfrI* gene with 5' overlap with 3' of cpeTp (*CfrI_cpeT*).

'Touch-up' gradient PCR reactions were the same as applied above except the cycling parameters varied as follows: Initial 95°C for 3 mins, followed by 10 cycles of denaturation at 98°C for 20 s, annealing at 50-70°C for 15 s, extension at 72°C for 1 min, 15 cycles of annealing at 50°-70°C with increments of 0.5°C per cycle, followed by extension at 72°C for 1min. In this case amplification of all three remaining fragments was successful (Fig. 6.4). Products which showed specific amplification were used for fusion PCR.
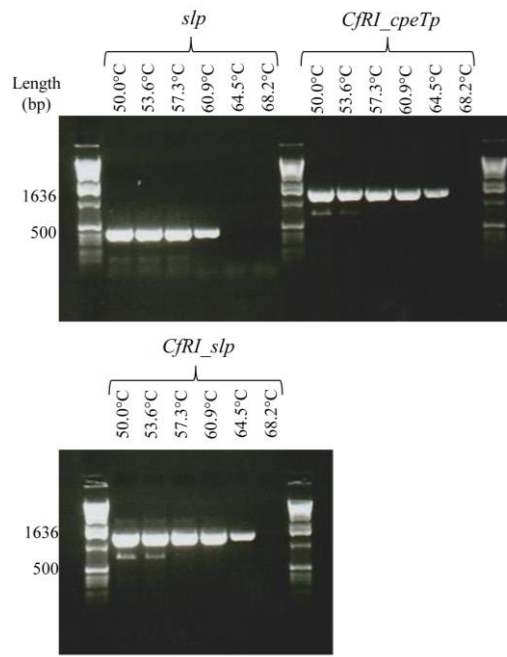
Figure 6.4: Gradient PCR of slp, *CfrI_cpeTp* and *CfrI_slp* fragments. The description of gene targets is given in the text.

Fragments from the promoter regions were fused to the *CfrI* variants using fusion PCR. PCRs were in a 50 μl volume, containing 1μl of promoter region PCR reaction and 1μl of *CfrI* variant PCR reaction, 0.4 μM of each primer, 1x KAPA HiFi buffer, 0.3mM each dNTPs, 1U KAPA HiFi polymerase and water to 50 μl. In each case the primers used were the forward primer used to amplify the promoter fragment and the reverse primer used to amplify the *CfrI* fragment. Fig. 6.5 shows fragments were successfully fused resulting in a band that corresponded in size to the promoter plus the *CfrI* fragment. However, further spurious bands were present and thus the fragment was purified by gel extraction (Qiagen, Venlo, Netherlands) following the manufacturer's instructions.
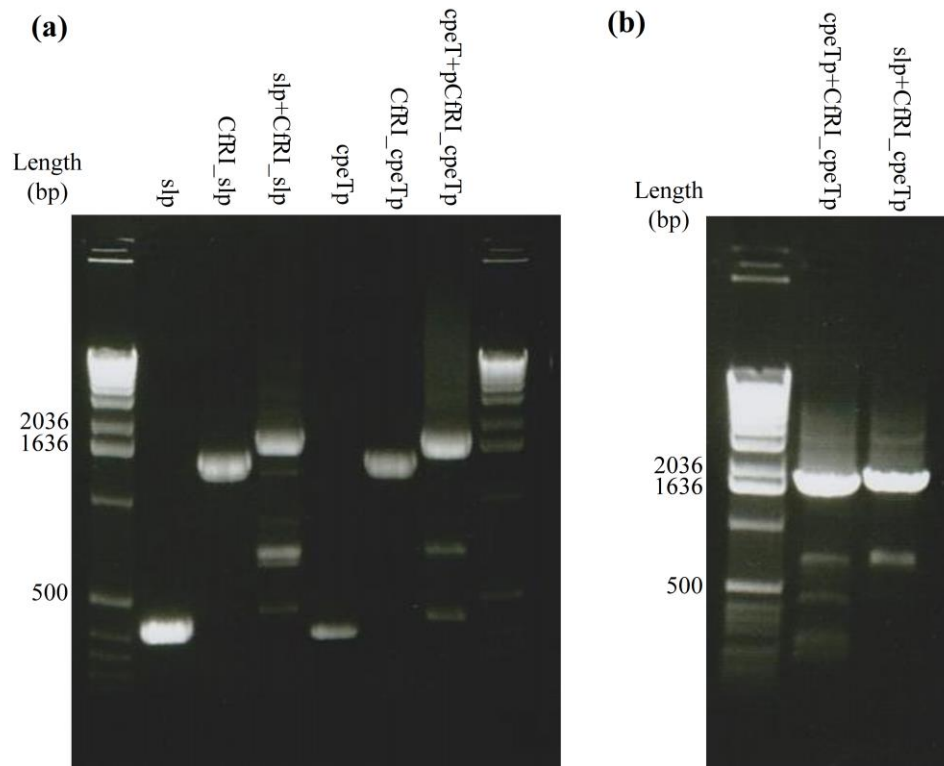
Figure 6.5: (a) Fusion of promoter regions to the *CfrI* gene by fusion PCR. Lane 1 = ladder. Lane 2 = slp fragment. Lane 3 = CfrI_slp fragment. Lane 4 = fusion of slp to CfrI_slp. Lane 5 = cpeTp fragment. Lane 6 = CfrI_cpeT fragment. Lane 7 = fusion of the cpeTp to CfrI_cpeT. Lane 8= ladder **(b)** re-amplification of gel purified fusion PCR fragments. Lane 1=ladder. Lane 2 fusion of the cpeTp to CfrI_cpeT. Lane 3= fusion of slp to CfrI_slp.

Gel purified fragments were further subject to PCR as described above resulting in a more pronounced band of the correct size (Figure 6.5b). The product was subcloned into pCR2.1® TOPO® TA (Life Technologies, Carlsbad, U.S.A.), yielding plasmids pCRcpeTpCfrI and pCRslpCfrI. Sequencing of these plasmids revealed no mutations in the sequence. These plasmids as well as plasmid pRL153 were digested with *Nhe*I and *Hin*dIII. Digestions were in 50μl containing 5μg plasmid, 1 x NEB buffer 4 (New England Biolabs, Ipswich, USA), 100μg ml⁻¹ BSA, 2U *Hin*dIII-HF® (New England Biolabs, Ipswich, USA), 2U *Nhe*I-HF® (New England Biolabs, Ipswich, USA) and water to 50 μl. Digestion of pRL153 further contained 2U FastAP thermosensitive alkaline phosphatase (ThermoFisher Scientific, Waltham, USA) to prevent vector re-ligation. Digestions were incubated at 37°C for 2 hrs, followed by inactivation at 80°C for 20mins. DNA from pRL153 digestion was precipitated

by ethanol precipitation as described in section 2.1.10 and resuspended in 10 µl H$_2$O. The cpeTp*CfrI* and slp*CfrI* fragments from digestion of pCRcpeTpCfrI and pCRslpCfrI, respectively were gel purified (Qiagen, Venlo, Netherlands) following the manufacturer's instructions and subsequently precipitated using ethanol precipitation. Ligation of digested pRL153 to cpeTp*CfrI* and slp*CfrI* was done in 20 µl volumes containing 1x T4 ligation buffer (Promega, Madison, USA), 1 µl T4 DNA ligase, 8.5 µl digested pRL153 (unquantified) and 8.5 µl of cpeTp*CfrI* or slp*CfrI* (unquantified). Ligation reactions were left at 4°C overnight before being transformed into chemically competent *E. coli* DH5α (see section 2.1.9) and plated onto LB+50 µg ml⁻¹ kanamycin. Transformants were picked into LB+50 µg ml⁻¹ kanamycin and grown overnight. Plasmids were then purified by miniprep (Qiagen, Venlo, Netherlands). PCR was used to check for cloning of the insert. PCRs were performed in a 50 µl volume containing ~10ng plasmid template, 1x MyTaq master mix (BioLine, London, UK), 0.4 µM each primer and water to 50 µl. Primers were pRL153_FSeq and pRL153_RSeq (Appendix 1) that flanked the cloning site shown in Fig. 6.2. All transformants contained inserts of the correct length (Fig. 6.6).
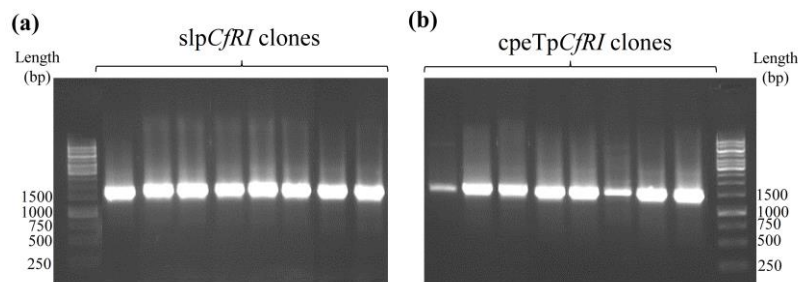
Figure 6.6: Confirmation of cloned inserts into pRL153. Products are the result of PCR amplification with primers pRL153_FSeq and pRL153_RSeq.

Expression constructs were transferred to the *E. coli* conjugal strain MC10161 containing the plasmid pRK24 (Brahamsha 1996). This was accomplished through chemical transformation and selection on LB + ampicillin (100 µg ml⁻¹) and kanamycin (50 µg ml⁻¹). The expression constructs were transferred to *Synechococcus* sp. WH7803 using conjugation as described in (Brahamsha 1996).

6.2.8. In vitro splicing

Group I self-splicing introns have the propensity to splice *in-vitro* given an exogenous source of GTP (Cech 1990). This property was used to test the hypothesis that *in-vitro* expression of *CfrI* could inhibit the splicing reaction through complimentary base pairing. The cyanophage S-PM2 unspliced *psbA* gene was amplified by PCR using primers psbA_FT7/psbA_R (see Appendix 1). PCRs composed 45 ng cyanophage S-PM2 gDNA, 1x MyTaq master mix (BioLine, London, UK), 0.4 μM each primer and water to 50 μl. Annealing of primers took place at 55°C. The spliced *psbA* gene was also amplified using the same PCR conditions except the pCR2.1_psbASpl was used as the template. 4 different variants of the antisense ncRNA *CfrI* were amplified. These variants differed in length, with three terminating before overlapping complimentary to the *psbA* intron/exon2 boundary and the other representing the full length of *CfrI* (Fig. 6.12a). These fragments were amplified from S-PM2 gDNA using primers CfrI_FT7 and primers cfrI_trunc1_R, CfrI_trunc2_R, CfrI_trunc3_R and CfrI_full_R (see Appendix 1). PCR conditions were the same as that for the amplified cyanophage S-PM2 *psbA*. All PCR products were purified using the Wizard ® SV PCR clean up kit (Promega, Madison, USA) following the manufacturer's instructions. psbA_FT7 and CfrI_FT7 contain the T7 promoter sequence that is recognised by T7 RNA polymerase used in *in-vitro* transcription. *In-vitro* transcription was thus carried out on spliced and unspliced *psbA* templates as well as all *CfrI* variants. Reactions were in 100 μl and contained 1x T7 transcription buffer (Thermo Fisher Scientific, Waltham, USA), 2mM each NTPs, 50U RiboLock™ RNase Inhibitor (Thermo Fisher Scientific, Waltham, USA), 30U T7 RNA polymerase (Thermo Fisher Scientific, Waltham, USA), 10 pmol of PCR product as template and nuclease-free water up to 100 μl. These data revealed that the cyanophage S-PM2 *psbA* intron spliced efficiently. Thus, to determine whether *CfrI* affected the splicing of the *psbA* intron, both templates were mixed in equimolar quantities and subject to *in-vitro* transcription. Conditions were as described above except 10 pmol of both unspliced *psbA* and *CfrI* variant PCR product was used as template. *In-vitro* transcriptions

ran for two hours followed by DNase digestion by TURBO DNA-*free*[TM] Kit (Life Technologies, Carlsbad, USA) following the manufacturer's instructions. RNA was then precipitated by NaAc/Ethanol precipitation. At no point was the RNA heated, which would have allowed denaturing of potential RNA duplexes. The precipitated RNAs were quantified by Nanodrop and 200 ng run on a Bioanalyser Total RNA Nano Chip (Agilent, Santa Clara, USA) to detect inhibition of *psbA* intron splicing.

6.2.9. Construction of a splicing reporter strain

To show that anti-intron expression could inhibit *psbA* intron splicing *in vivo,* an *E. coli* based reporter strain was constructed. The *psbA* intron contains several in-frame stop codons such that should the unspliced transcript be translated, the resulting polypeptide would probably be non-functional. This property was exploited by introducing the *psbA* intron into the UV excitable *gfpuv* gene (Lee and Keasling 2005), such that splicing would give a green phenotype under UV light (Fig. 6.7). Further, the use of inducible promoters would allow for fine tuning of expression.
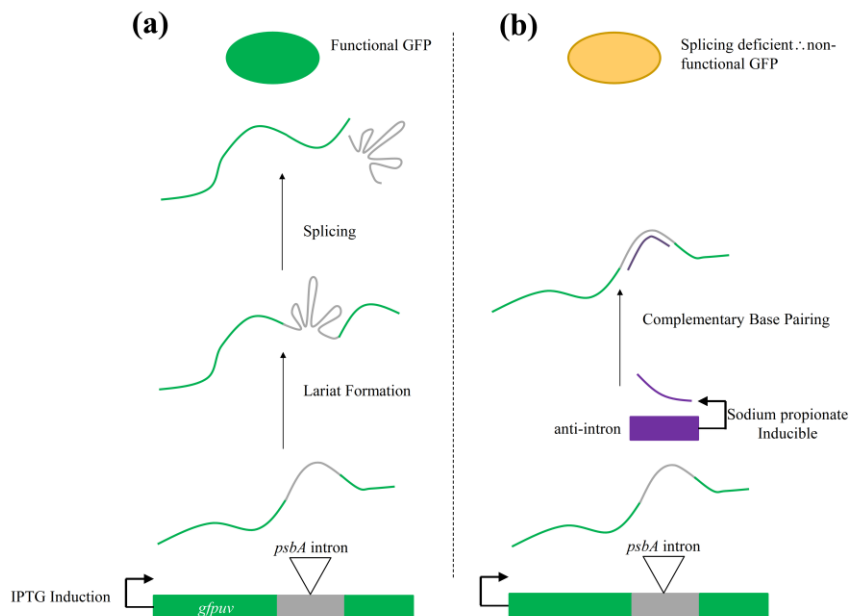
Figure 6.7: Strategy for the construction of the splicing reporter strain. **(a)** Without induction of the anti-intron **(b)** with sodium propionate induced transcription of the anti-intron.

For induction of *gfp* expression, the T7 promoter was used and genes were expressed from the pET151-D backbone (Life Technologies, Carlsbad, USA). The intron was inserted into two distinct loci in *gfpuv* (after codons 17 and codons 143) using fusion PCR. These two variants were thus named *gfp*int_L1 and *gfp*int_L2. In addition, a control strain was constructed that placed *gfpuv* under the transcriptional control of the T7 promoter and is expressed from the same pET-151D backbone. This plasmid is referred to here as pET*gfp*. Initial PCRs (Figure 6.8a) were in a 50 µl volume, containing 20-150ng template, 0.4µM each primer, 1x KAPA HiFi buffer, 0.3 mM each dNTP, 1U KAPA HiFi polymerase and water to 50µl. Primer sequences and templates are shown in Table 6.2. Thermal cycling parameters were 95°C for 3mins, followed by 35 cycles of denaturation at 98°C for 20 s, annealing at 65°C for 15 s, followed by extension at 72°C for 1 min. For subsequent fusion of fragments, 1µl of the products *gfpuv*_exon1_L1, Intron_L1 and *gfpuv*_exon2_L1 were mixed and PCR was conducted under the same conditions but with primers gfpuv_F and gfpuv_R yielding the product *gfp*intL1 (Figure 6.8b). Similarly 1µl of the products *gfpuv*_exon1_L2, Intron_L2 and *gfpuv*_exon2_L2 were mixed and PCR was conducted yielding the product *gfp*intL2 (Figure 6.8b).

Desired PCR product	Forward primer name	Forward Primer Sequence	Reverse Primer Name	Reverse Primer Sequence	Template
pET*gfp uv*	gfpuv_F	5'-CACCTAGATATAAGGAGGCTCACAGCTATGAGTAAAGGAGAAGAACTTTTC-3'	gfpuv_R	5'-TTATTTGTAGAGCTCATCCATG-3'	pPro24*gfpu v*
*gfpuv*_exon1_L1	gfpuv_ex1_L1_F	5'-CACCTAGATATAAGGAGGCTCACAGCTATGAGTAAAGGAGAAGAACTTTTC-3'	gfpuv_ex1_L1_R	5'-cacaacttgAAAGCATTGAACACCATAAG-3'	pPro24*gfpu v*
*gfpuv*_exon1_L2	gfpuv_ex1_L2_F	5'-CACCTAGATATAAGGAGGCTCACAGCTATGAGTAAAGGAGAAGAACTTTTC-3'	gfpuv_ex1_L2_R	5'-cacaacttgGTACTCGAGTTTGTGTCC-3'	pPro24*gfpu v*
*gfpuv*_exon2_L1	gfpuv_ex2_L1_F	5'-gaagtttaagTCCCGTTATCCGGATCATATG-3'	gfpuv_ex2_L1_R	5'-TTATTTGTAGAGCTCATCCATG-3'	pPro24*gfpu v*
*gfpuv*_exon2_L2	gfpuv_ex2_L2_F	5'-gaagtttaagAACTATAACTCACACAATGTATAC-3'	gfpuv_ex2_L2_R	5'-TTATTTGTAGAGCTCATCCATG-3'	pPro25*gfpu v*
Intron_L1	Intron_L1_F	5'caatgctttCAAGTTGTGCTTTGCGCTC-3'	Intron_L1_R	5'-taacgggaCTTAAACTTCCCTTAACTATCTGGTATTAC-3'	S-PM2 gDNA
Intron_	Intron_	5'-	Intron_	5'-	S-PM2

L2	L2_F	tcgagtacCAAGTTGTGCTTTGCGCTC-3'	L2_R	agttatagttCTTAAACTTCCCTTAACTATCTGGTATTAC-3'	gDNA
*gfp*intL1	gfpuv_F	5'-CACCTAGATATAAGGAGGCTCACAGCTATGAGTAAAGGAGAAGAACTTTTC-3'	gfpuv_R	5'-TTATTTGTAGAGCTCATCCATG-3'	*gfpuv*_exon1_L1, Intron_L1, *gfpuv*_exon2_L1
*gfp*intL2	gfpuv_F	5'-CACCTAGATATAAGGAGGCTCACAGCTATGAGTAAAGGAGAAGAACTTTTC-3'	gfpuv_R	5'-TTATTTGTAGAGCTCATCCATG-3'	*gfpuv*_exon1_L2, Intron_L2, *gfpuv*_exon2_L2

Table 6.2: Primer sequences and templates used in the fusion PCR construction of pET*gfpuv*, *gfp*intL1 and *gfp*intL2. Underlined sequences are those required for directional cloning into pET-151D. Those in blue are stop codons used to stop translation of the N-terminal signal peptide encoded on pET-151D. Those in red are the start codon for translation of the GFP variant. Emboldened sequences are RBSs. Those sequences in lower case are required for fusion to the upstream/downstream fragment. The pPro24*gfpuv* plasmid was a kind gift from Dr. Jay Keasling, University of California, Berkeley.
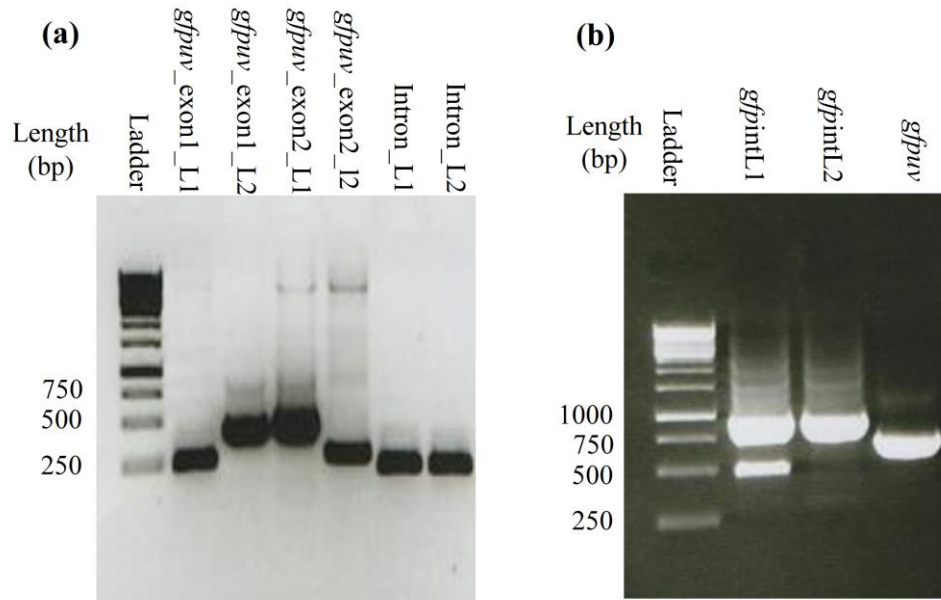
Figure 6.8: Fusion PCRs of the cyanophage S-PM2 psbA intron into *gfpuv*. **(a)** PCR amplification of the two gfpuv variant exons as well the two variant introns. **(b)** Fusion of the intron/exon products to form gfpintL1, *gfp*intL2. Also shown is the amplification of *gfpuv* of the control strain.

The fused PCR products were directionally cloned into pET151-D (Life Technologies, Carlsbad, USA) and transformed into *E. coli* TOP10 following the manufacturer's instructions. This yielded plasmids pET*gfpuv*, pET*gfp*intL1 and pET*gfp*intL2. These plasmids were purified by miniprep (Qiagen, Venlo, Netherlands) and sequenced. No mutations were detected in any clones. Thus, these plasmids were transformed into the *E.*

coli expression strain BL21 (DE3). For induction of expression of GFP from pET vectors, clones were plated onto LB Agar + 1mM IPTG and grown at 30°C overnight. For induction of GFP from pPro24, clones were plated onto LB Agar +5 mM sodium propionate and incubated overnight at 30°C.

6.3. Results

6.3.1. Light responsive splicing of a group I intron

To quantify the amount of unspliced and spliced cyanophage S-PM2 *psbA* during infection of *Synechococcus* sp. WH7803 under HL and LL conditions(see section 6.2.1), RT-qPCR was used with primer/probe assays specific for the unspliced and spliced transcripts (Fig. 6.9a). These data show an increase in expression of both spliced and unspliced cyanophage S-PM2 *psbA* in HL compared with LL conditions. This increase was statistically significant at all time points as analysed by REST analysis. If one considers that total *psbA* expression is the sum of both unspliced and spliced transcripts, then HL expression of *psbA* is between 2.56 and 10.36 greater than LL depending on the time point. Beyond this, the patterns in accumulation of the unspliced and spliced *psbA* transcripts are very different. In both cases there is an increase in expression during progression through infection. However, during infection under LL both the unspliced and spliced transcripts seem to reach saturation approximately 6hrs after infection (Fig. 6.9b). In comparison, at HL, both the unspliced and spliced transcripts continue to increase past 6hrs (Fig. 6.9c). Relative quantification allows for determination of the ratio of spliced: unspliced *psbA* at HL and LL (Fig 6.9d). During early infection (0-3hrs), the ratio increases in both light conditions such that the spliced transcript abundance is approximately twice that of the unspliced. In LL, after 3hrs the ratio drops rapidly such that by 6 hrs after infection the unspliced and spliced transcripts are approximately equimolar. From 6-9 hrs the molar ratio drops even further such that the unspliced transcript is in excess of the spliced transcript. In comparison, at HL, after 3hrs the ratio is maintained at approximately 2 until 9 hrs after infection. REST analysis shows that

the spliced: unspliced ratio is significantly greater at time points 6 and 9 hrs past infection in HL compared with LL.
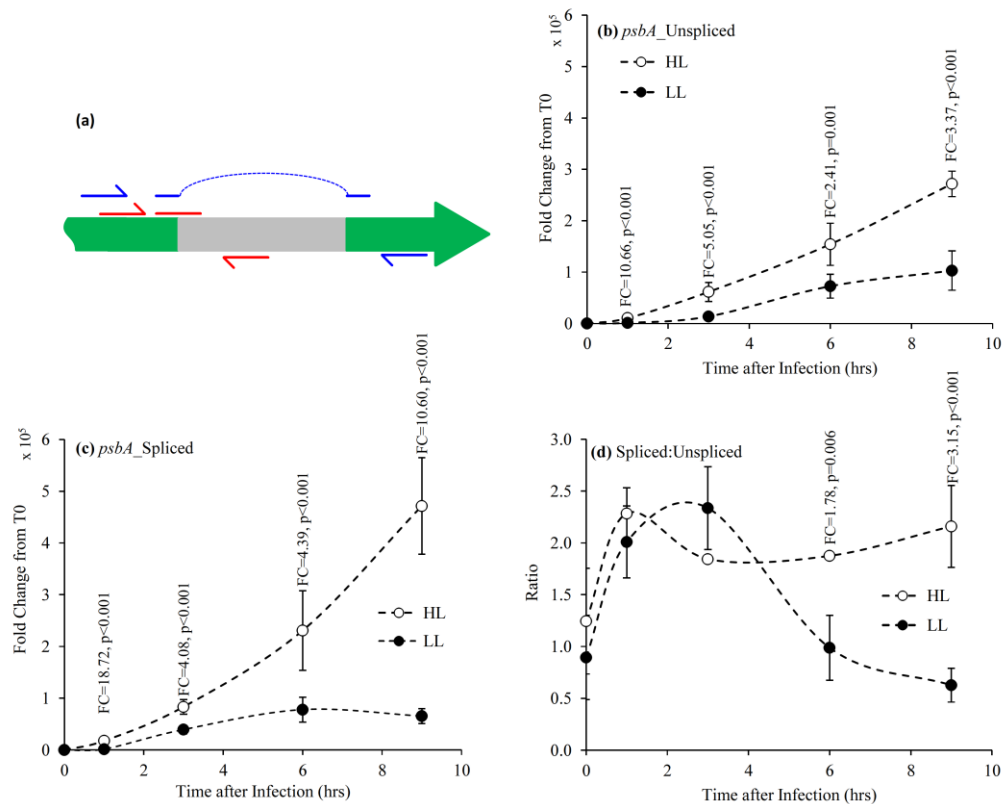
Figure 6.9: Expression patterns of unspliced and spliced cyanophage S-PM2 *psbA* following infection of *Synechococcus* sp. WH7803 under LL and HL growth conditions. **(a)** Position of unspliced and spliced primer/probes. The blue colour symbolises primer/probe assay to detect the spliced transcript whilst red shows the primer/probe assay used to detect the unspliced transcript. **(b)** Temporal expression of the unspliced *psbA* relative to time 0 in LL and HL. **(c)** Temporal expression of the spliced *psbA* relative to time 0 in LL and HL. **(d)** Temporal expression of spliced *psbA* relative to unspliced *psbA* in LL and HL. FC stands for fold change as measured by REST analysis and only shown for significantly different time points. P values are shown for significance.

6.3.2. An antisense ncRNA bridges the psbA exon/exon boundary

During attempts to map the 3' end of the antisense ncRNA, *CfrI,* it was shown that this transcript extends past the putative terminator helix as identified in Millard et al. (2010). Fig. 6.10 shows the results of 'transcript walking'. Using this method, the length of *CfrI* is predicted to be between 1306-1342 nt terminating in the region between the S-PM2 genomic coordinates 135,513-135,549 (NC_006820.1). This observation is strengthened by the reads mapped to this genomic region from RNA-Seq data presented in chapter 4 (Fig. 6.10).
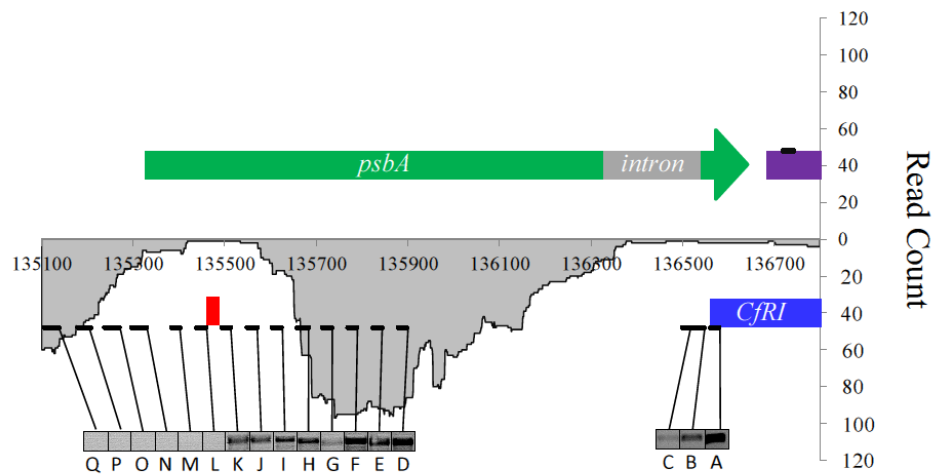
Figure 6.10: Results of transcript walking overlain on the number of crick strand reads derived from RNA Seq (see Chapter 4). Watson strand features are shown above the x-axis and crick strand features are shown below. Black lines show absolute positions of GSPs/ reverse RT-PCR primers. Bands corresponding to the presence of transcripts by RT-PCR and agarose gel electrophoresis are shown below. The red line shows the approximate region of termination

6.3.3. Attempts to express CfrI in Synechococcus sp. WH7803

With the observation that *CfrI* extended through the region antisense to the intron, it was hypothesised that *CfrI* could bind through complementary base pairing to the intron RNA. Thus it was further hypothesised that this may prevent splicing of the intron due to inability to form the lariat required for splicing (Cech 1990). To test this a *Synechococcus* sp. WH7803 mutant was sought that overexpressed *CfrI* (WH7803::S-PM2_*CfrI*). Thus, upon infection with cyanophage S-PM2 splicing of S-PM2 *psbA* would be inhibited. Constructs were made that used two cyanophage S-PM2 putative promoter regions for expression (see section 6.2.7). Figure 6.11 shows the efficiency of conjugal transfer of these constructs from *E. coli* to *Synechococcus* sp. WH7803.
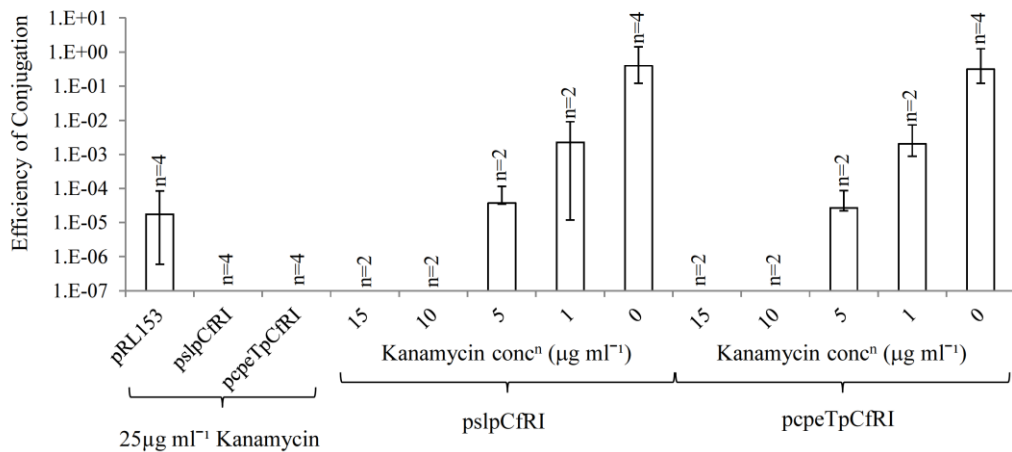
Figure 6.11: Conjugal transfer efficiency of pslpCfrI and pcpeTCfrI relative to the control pRL153.

The conjugal transfer efficiency to Synechococcus sp. WH7803 was 0 for the two expression constructs (i.e. no clones were obtained that were able to maintain these constructs). Decreasing the concentration of the antibiotic resulted in colony formation at $5\mu g$ ml^{-1} or lower. These colonies were subsequently picked but were unable to grow in liquid media plus the antibiotic. In contrast the pRL153 control plasmid could transfer to Synechococcus at high frequency (6.0x10-7-1.2x10-5).

6.3.4. Intron splicing is inhibited by CfrI expression in vitro

Due to inability to construct a *Synechococcus* sp. WH7803::*CfrI* strain, it was tested whether the presence of *CfrI* could inhibit *psbA* intron splicing *in-vitro*. Therefore, *in-vitro* transcription took place with *psbA*_unspliced co-expressed with 4 *CfrI* variants. The first three (*CfrI*_trunc1, *CfrI*_trunc2 and *CfrI*_trunc3) are truncated forms of the full length *CfrI*. They terminate in various positions downstream (antisense) of the 5' of *psbA* exon 2 (dashed line in Fig. 6.12a). The fourth product is the full length *CfrI* and extends through the intron antisense. *In vitro* transcription of unspliced *psbA* yields three bands (Fig. 6.12b). The band lengths (~1300 bp, ~1100 bp and ~210 bp) corresponding to: unspliced *psbA* (1292 nt), spliced *psbA* (1080 nt) and the spliced intron (212 nt), respectively. Expression of the 4 *CfrI* variants yields bands of roughly correct sizes (Fig. 6.12b). Co-expression of unspliced *psbA*

with the first three variants yields bands of various sizes. In each case the band corresponding to the intron is present (green box Fig. 6.12b). Larger bands are present that correspond to duplex RNAs between the *CfrI* variants and the spliced transcript and unspliced transcript. As such these products are slightly larger than that of the *psbA* unspliced and spliced transcripts when expressed from unspliced *psbA*. When unspliced *psbA* is co-expressed with the full length *CfrI,* there is a disappearance of the band corresponding to the intron. Further, only one larger band is present that corresponds to a duplex RNA between *CfrI* and unspliced *psbA*. Should splicing have occurred one may expect two bands corresponding to a *CfrI:* unspliced and *CfrI:* spliced duplexes. Therefore, it is clear that the kinetics of splicing *in-vitro* are slower than that of complementary base pairing.
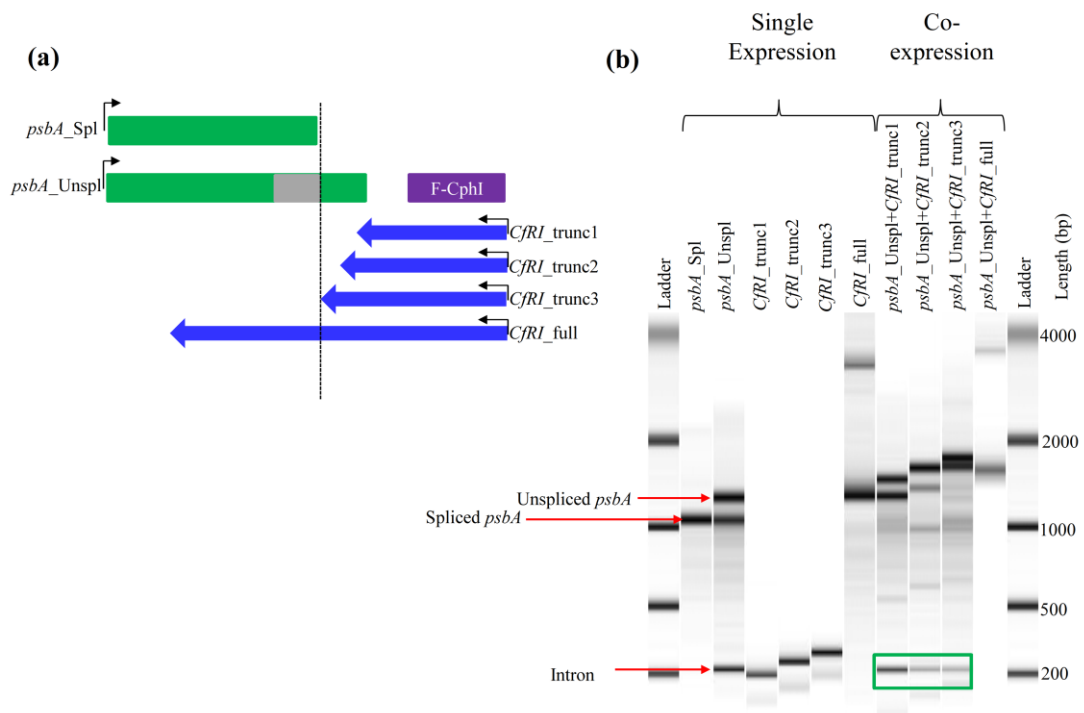
Figure 6.12: Inhibition of cyanophage S-PM2 *psbA* in-vitro splicing by co-expression with *CfrI*. **(a)** Strategy for co-expression assay. 4 *CfrI* variants were made by PCR. 3 truncated variants (trunc1, trunc2, trunc3) terminated before the intron antisense. The fourth was the full length *CfrI*. **(b)** Bioanalyser gel of in vitro transcriptions.

6.3.5. Expression of CfrI

To further elucidate the role of *CfrI* in the splicing phenomena, its expression was monitored during infection of *Synechococcus* sp. WH7803 under LL and HL conditions. *CfrI* appears to have an extremely late mode of expression. In the case of HL, there is no observed plateau in relative expression as is observed for all other S-PM2 genes studied so far (Clokie et al. 2006). Instead, there is a linear increase in expression throughout infection. In comparison, there appears to be a repression in *CfrI* expression in LL during the first 3hrs of infection relative to HL. From 3-6hrs there is a rapid rise in *CfrI* expression, thus by 6hrs *CfrI* expression is equal in HL and LL. In LL, a plateau is observed after 6hrs whereas at HL *CfrI* expression continues to accumulate (Fig. 6.13a). Relative quantification allows one to calculate the ratio of both spliced and unspliced *psbA* to *CfrI* expression (Fig. 6.13b). At HL both spliced and unspliced *psbA* appear to be in excess of *CfrI*. There is no overall trend throughout infection with spliced *psbA* remaining in ~4 fold excess and unspliced *psbA* remaining in ~2 fold excess of *CfrI*. In contrast, at LL unspliced *psbA* remains approximately equimolar to *CfrI* throughout infection. Also during LL, there appears to be an initial increase in the ratio of spliced *psbA: CfrI* to approximately 3 fold excess by 3 hrs. After this there is a dramatic decrease in this ratio such that by 6hrs the two are approximately equimolar. This trend continues to 9 hrs where *CfrI* is actually found in excess of spliced *psbA*. The reduction in the ratio appears to be driven by a large increase in expression of *CfrI* between 3-6 hrs.
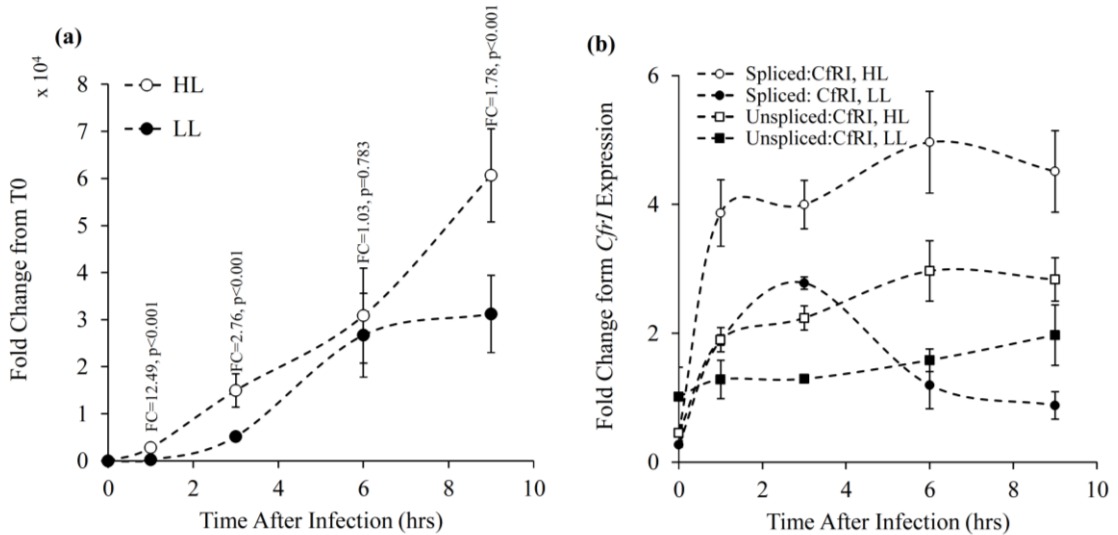
Figure 6.13: (a) Expression of CfrI during infection under HL and LL. FC is the fold change as reported by REST analysis with *p* values given. **(b)** Expression of spliced and unspliced psbA relative to CfrI

6.3.6. Inhibition of splicing in the E. coli reporter strain

To test whether *CfrI* could indeed alleviate splicing of the cyanophage S-PM2 intron, in the absence of obtaining a WH7803::*CfrI* strain, a heterologous splicing reporter strain was constructed in *E. coli* (see section 6.2.9). The reporter strain functioned such that splicing of the intron would create a functional GFP and thus the fluorescence intensity would be dependent on the rate of splicing. Fig. 6.14a shows that whilst the control strain (pET*gfpuv*) containing no intron insertion was capable of GFP expression when induced with IPTG, neither GFP variants containing introns in two distinct loci expressed GFP when induced with IPTG. This was tested under several temperatures (18, 23, 30, 37˚C) yet GFP expression could not be detected Thus, it was clear that the cyanophage S-PM2 *psbA* intron was unable to splice in this background. As such the reporter strain was non-functional. The reasons for this and possible solutions are discussed in section 6.4.3.
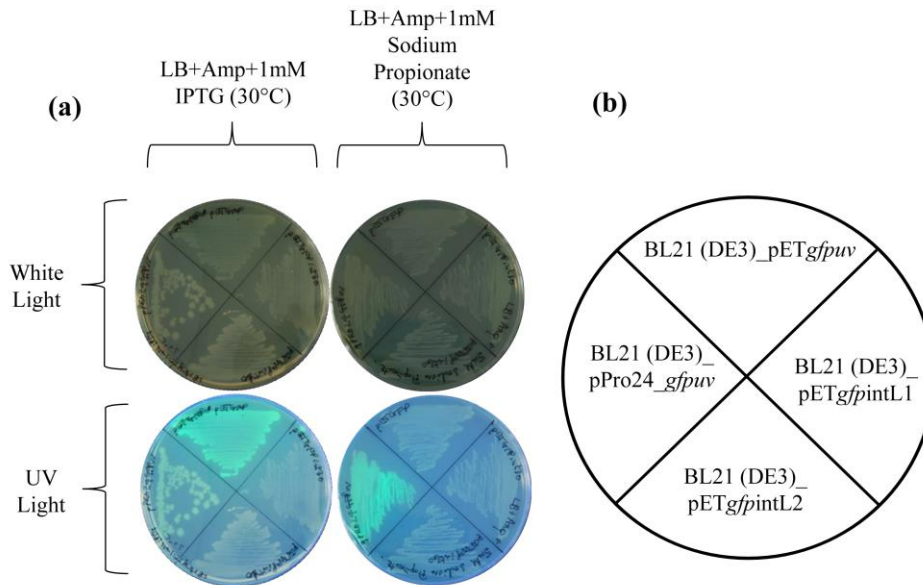
Figure 6.14: Inhibition of splicing in the *E. coli* reporter strain. (a) Colonies induced with either IPTG (T7 induction in the lambda lysogen strain) or sodium propionate (pPro plasmids) (b) Description of strains in (a).

6.4. Discussion

6.4.1. Splicing of the psbA intron: Implications for regulation of cyanophage photosynthesis

Both unspliced and spliced cyanophage *psbA* were shown to be up-regulated in HL compared with LL conditions. The fold change in gene expression was estimated to be between 2.41 and 10.66 for unspliced *psbA* and between 4.08 and 18.72 for spliced *psbA*. However, these estimates must be treated with some caution due to the observed changes in the global gene expression reported in section 5.3.1. It was shown that at HL there was an increase in the reads mapped to cyanophage S-PM2 by approximately 2-fold. Thus, S-PM2 *psbA* shows large HL up-regulation when normalised to host 16S rRNA gene expression but this difference is dampened relative to all other phage genes at HL. Light dependent differential expression of S-PM2 *psbA* was observed in RNA-Seq experiments (section 5.3.2) and thus it is extremely likely that S-PM2 *psbA* is differentially expressed, it is just the fold-change that is under contention. The observation of global changes in gene expression under varying physiological conditions casts doubt on the use of RT-qPCR to determine

gene expression changes in phages. Specifically, it highlights the importance of the reference gene used for normalisation of expression between samples. Future studies may seek a phage encoded gene as the reference, but given diversity in temporal expression of phage genes, discovery of such a 'housekeeping' gene is an almost impossible task. RNA-Seq circumvents this issue by assessing expression of all genes simultaneously.

In contrast, the use of relative expression of spliced: unspliced *psbA* is unaffected by the large global changes in gene expression between HL and LL. Thus, here it is demonstrated that splicing of the cyanophage S-PM2 *psbA* group I intron is responsive to light intensity. Increased light intensity caused an increase in splicing such that the *in vivo* ratio of spliced: unspliced *psbA* is much greater in HL compared with LL conditions. During HL, the rate of damage to the D1 polypeptide, encoded by the *psbA* gene, is significantly higher (see chapter 3), and therefore a higher rate of delivery of D1 to PSII is required to maintain stable PSII photochemistry (Tyystjärvi and Aro 1996; Bailey et al. 2004). Conversely, under LL, unnecessary accumulation of the D1 polypeptide is a waste of cellular energy given the amount of ATP equivalents required for amino acid polymerisation (see section 3.4.1). The presence of several in frame stop codons within the intron suggests that unspliced *psbA* does not encode a functional protein. Therefore, splicing of the *psbA* intron could be a mechanism for regulating the amount of D1 polypeptide delivered to the photosystem. In the eukaryotic algae *Chlamydomonas reinhardtii*, the *psbA* gene contains four group I introns (Erickson et al. 1984). In the dark, splicing of each of these introns is inhibited, whilst upon the onset of the light period efficient splicing of each intron occurs (Deshpande et al. 1997). Moreover, the process appears to be regulated by electron transport given that electron transport inhibitors and mutants deficient in components of the PET chain cease splicing. In this example a nuclear encoded gene, *css1*, has been implicated in regulation of light dependent splicing (Li et al. 2002). Thus light dependent splicing is not unprecedented.
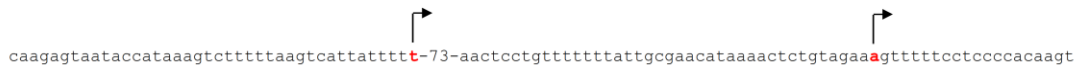
The transcriptional regulation of the cyanobacterial *psbA* family and comparison to cyanophage S-PM2 is discussed in detail in section 5.4. During the late period of infection,

gene expression was thought (Mann et al. 2005; Sullivan et al. 2010) and now validated to be driven by the T4-like late promoter (section 4.3.6). Since this period and indeed the infection cycle in general is short relative to the life cycle of the host, it is suggested that many of the host's transcriptional strategies to regulate *psbA* would be of negligible benefit. Instead, here it is suggested that cyanophage S-PM2 maximally expresses the unspliced *psbA* through interaction of the putative phage modified RNA polymerase with the upstream T4-like late promoter of cyanophage S-PM2 *psbA*. This produces transcripts that are in excess of what is required for stable PSII photochemistry at LL and thus inhibition of intron splicing could prevent wasteful accumulation of the D1 polypeptide. When the phage infected cell is exposed to increased quanta the inhibition of splicing is alleviated and the production of functional D1 can match the rate of photoinduced damage. The consequences of this regulation are that NTPs are sacrificed at the expense of translational amino acid polymerisation, which in section 3.4.1 is demonstrated to be a greater energetic demand. Thus the splicing of the cyanophage S-PM2 *psbA* intron may act as a fast response regulator to changing light levels. The mechanism that regulates the splicing is unclear and would require further experimental work. An alternative hypothesis is that light dependent splicing is not an active process and instead is a bi-product of expression of *CfrI* as will be discussed below.

6.4.2. Expression of CfrI

The expression dynamics of *CfrI* during infection are somewhat enigmatic. Under LL there is an initial delay in *CfrI* expression relative to HL, followed by a rapid increase from 3hrs to 6hrs, where expression plateaus. This is indicative of a late transcriptional mode. The TSS of *CfrI* have been mapped to two distinct loci (Millard et al. 2010). However, neither of these loci shows the presence of an upstream T4 late promoter (Fig. 6.15). Indeed, there is also an absence of any σ_{70}-like promoter. Thus, it is unclear how initiation of transcription progresses in *CfrI* and that *CfrI* may be driven by a 'rogue' promoter. However, the presence of two TSSs suggests variability in transcriptional regulation that may contribute to the

expression changes in HL and LL. REST analysis shows a 12.49 fold increase in *CfrI* expression at HL at 1 hr post infection This increase reduces towards the end of infection to ~2 fold increase at 3, 6 and 9 hrs post infection. However, it is argued here that given the global increase in expression observed under HL, there is likely to be no net increase in *CfrI* expression at HL and perhaps even a slight increase at LL. This is evidenced in that no significantly different expression for *CfrI* was found in the RNA-Seq analysis.

caagagtaataccataaagtctttttaagtcattatttt**t**-73-aactcctgtttttttattgcgaacataaaactctgtagaa**a**gtttttcctccccacaagt

Figure 6.15: Nucleotide sequences of upstream regions of the two *CfrI* TSSs. TSSs are shown in red with an arrow above.

6.4.3. The role of CfrI in regulation of psbA expression

The inability to generate WH7803::*CfrI* precludes examination of the functional role of this asRNA. Conjugal transfer of the *CfrI* expression constructs were unsuccessful yet the control plasmid transferred with relatively high frequency, albeit slightly lower than reported elsewhere (Brahamsha 1996). Given that successful expression from this plasmid has been achieved previously (Shan 2008), it is likely that the *CfrI* product is toxic to *Synechococcus* sp. WH7803. Indeed, the reverse complement of *CfrI* shows between 70.19-70.80% nucleotide identity to the 4 copies of host *psbA* in the region that overlaps them (917bp). This suggests that *CfrI* interacts with the host *psbA* mRNA when introduced causing an artificial 'knockdown' and given that *psbA* is essential for phototrophic growth is therefore lethal (Debus et al. 1988). Thus, there is potential to use this system to generate 'knockdowns' in non-lethal genes of *Synechococcus* where considerable difficulty in knockout mutant generation exists. It is unlikely that *CfrI* interacts with the host *psbA in-vivo* due to the late expression of *CfrI* and the fact that host *psbA* mRNA is decreased dramatically 1hr after infection (Clokie et al. 2006).

In spite of this, elucidation of the 3' terminus of *CfrI* demonstrates that this gene extends antisense through the intron. Therefore, upon expression of *CfrI* there is the possibility of

interaction between the *CfrI* RNA and the RNA of the unspliced *psbA* intron. Indeed, in northern blot analysis, *CfrI* displays a major band of ~225bp, which equates to the approximate coordinates of the experimentally validated TSS to just downstream of the 5' end of *psbA* exon 2 (Millard et al. 2010). However, a range of higher molecular weight (HMW) products also exist in this analysis (Millard et al. 2010). Here, it is suggested that these HMW products correspond to the full length *CfrI* transcript that is significantly longer (up to 1342bp). The existence of a prominent band of ~225 bp suggests therefore that this may be a site of processing of the transcript (Millard et al. 2010). Explicitly, it is proposed that downstream of this site, *CfrI* specifically interacts with the *psbA* intron. The duplex formed may therefore be rapidly degraded by the host RNase III complex and therefore the longer transcript is absent in northern blot analysis. *In-vitro* it is shown here that duplex formation between full length *CfrI* and unspliced *psbA* prevents self-splicing, whilst expression of *CfrI* variants that terminate before the intron allow splicing to occur. These experiments demonstrate that the kinetics of splicing are slow, comparative to that of antisense complimentary base pairing *in-vitro*. How these kinetics relate to splicing *in-vivo* are unclear. Indeed the large ribosomal RNA intron of *Tetrahymena* shows a splicing rate that is 20-50 times greater *in-vivo* that *in-vitro* (Zhang et al. 1995).

Light dependent intron splicing may be regulated by *CfrI* expression but not in an active mode. Indeed, in Fig. 6.9d the reduction in the spliced: unspliced ratio at LL between 3-6hrs after infection is concomitant with a rapid increase in *CfrI* expression (Fig. 6.13a) such that by 6hrs the ratio of spliced *psbA: CfrI* is approximately 1. Coupled with the knowledge that *CfrI* can prevent splicing *in-vitro* suggests that the expression of *CfrI* may be sequestering the *psbA* transcript by complementary base pairing. This effect is not present under HL due to the HL inducible transcription of *psbA* that dominates that of *CfrI*. In this way splicing may not be the dominant factor that regulates *psbA* expression but instead may be a bi-product of interactions of *CfrI* with the low abundance *psbA*. Millard et al. (2010) suggested that *CfrI* may 'protect' the *psbA* transcript from RNase degradation, a situation analogous to

the asRNA Yfr15 of *Prochlorococcus* sp. MED4 (Steglich et al. 2008; Stazic et al. 2011). In this latter situation, upon cyanophage infection there is an increased abundance of a 3-kb transcript containing three small ORFs of unknown function, a ncRNA also of unknown function, the 6S RNA and partially an ORF encoding *purK* involved in purine metabolism. In addition, an increased abundance of a long (~3.5kb) antisense transcript that completely overlaps the sense transcript is observed upon cyanophage infection (Steglich et al. 2008). It is shown that duplex RNAs produced from *in-vitro* transcription of both transcripts are protected from cleavage by the ssRNA endonuclease, RNase E *in-vitro* (Stazic et al. 2011) and hence *in-vivo,* results in a decrease in the degradation rate of this transcript. However, it is unclear what the net effect of this mechanism is in the cell during infection. Presumably, duplex RNAs are physically occluded from being translated into proteins (Brantl 2002; Georg and Hess 2011). Thus, an equally probable scenario is that cyanophage infection increases the expression of the antisense RNA to prevent translation of the duplex and that the RNase E protection afforded is simply a bi-product. The 'protection' of sense transcripts as a regulatory mechanism is much clearer in the case of the *Synechocystis* sp. PCC6803 *psbA2*. Here, an asRNA overlaps the 5'UTR of the *psbA2* gene. Over-expression of the asRNA leads to a decrease in the rate of degradation of the *psbA2* transcript, whilst repression of the asRNA leads to an increase the degradation rate (Sakurai et al. 2012). Further expression of the asRNA is correlated with that of *psbA2*. Thus, this asRNA acts as a positive regulator. This is conceivable as the asRNA only overlaps a fraction of the 5'UTR antisense (interestingly not including the RBS) and therefore translation of the transcript can still occur. Therefore, it is suggested here that *CfrI* does not act as a positive regulator as once suspected (Millard et al. 2010), but rather acts to sequester low abundance *psbA* transcripts at LL. One may conceivably argue why cyanophage S-PM2 would invest in such an elaborate method of regulation when repression of *psbA* transcription initiation would yield the same effect. The reason may be due to the need to retain a late promoter upstream of *psbA*. The late promoter motif is an absolute requirement for transcription of late genes in T4 (Miller et al. 2003). The late motif is an 8bp AT rich sequence with the consensus

TATAAATA. In comparison with early transcription, this is the only sequence required for binding of the late-infection modified RNA polymerase complex. Cyanophage S-PM2 *psbA* retains this motif and therefore is thought to be responsible for continued transcription throughout infection (Mann et al. 2005). Therefore, repression of transcription initiation by a hypothetical repressor can only occur by binding to this motif. This in turn would repress the transcription of all other late genes. Thus, under LL a degree of *psbA* expression occurs. *CfrI* then post-transcriptionally represses transcription by sequestration of the transcript at the 3' end. Increases in expression of *psbA* at HL, by the activity of an activator or by decreases in *psbA* transcript degradation (Sakurai et al. 2012) overwhelms the low level constitutive expression of *CfrI* in late infection. Thus, the regulation is analogous to the threshold-linear response (Dühring et al. 2006; Legewie et al. 2008; Courtney and Chatterjee 2014). The classical example is the asRNA *isrR* that represses expression of the iron stress induced gene *isiA* in *Synechocystis* sp. PCC6803 (Dühring et al. 2006). Here, *isrR* shows constitutive low level expression, whereas its antisense partner, *isiA* exhibits strict iron deficiency dependent transcription initiation. *isrR* binds to *isiA* mRNA and causes degradation of the duplex RNA. Thus, for the *isiA* transcript to be effectively translated, the antisense *isrR* must first be 'titrated' out. This manifests as a delay in the iron stress response which is alleviated upon artificial repression of *isrR* (Dühring et al. 2006). Thus, this paradigm is extended to explain the transcriptional regulation of cyanophage S-PM2 *psbA* and is summarised in Fig. 6.16. This type of regulation would cause the light dependent splicing phenotype observed and thus it is concluded that *CfrI* represents the dominant regulator of *psbA* expression and that light dependent splicing is simply a bi-product.
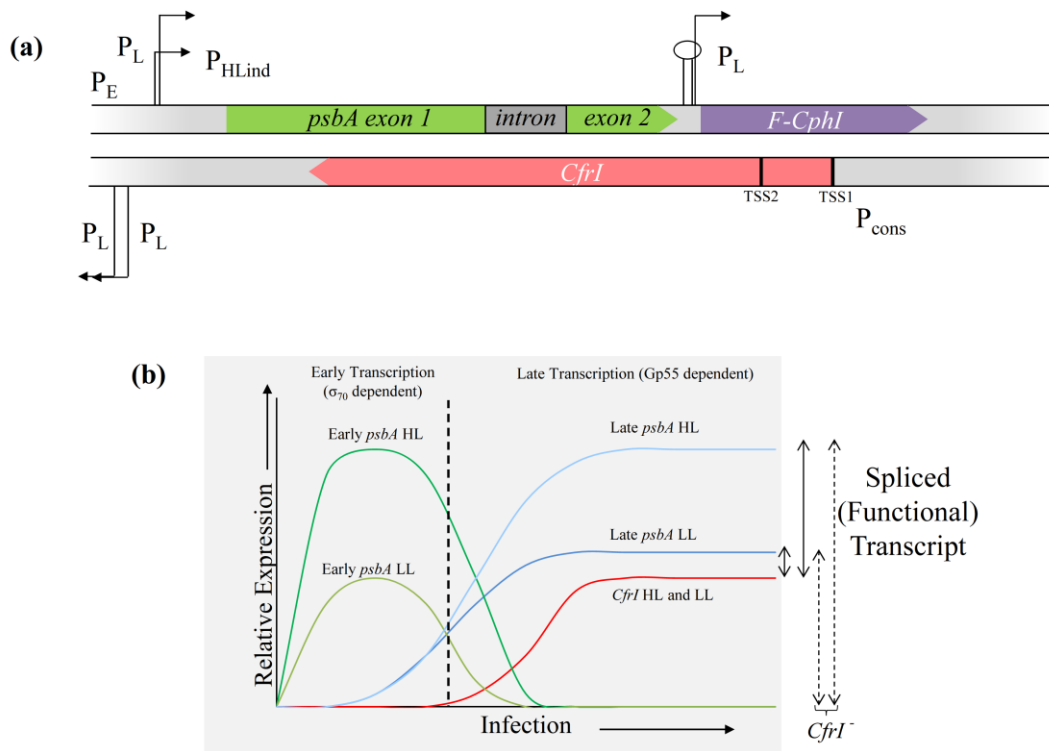
Figure 6.16: Model of transcriptional regulation of cyanophage S-PM2 *psbA*. **(a)** Transcriptional organisation of the *psbA* region. P_L represents the late promoter, P_E is the early promoter, P_{HLind} represents a HL inducible promoter, P_{cons} represents a constitutive promoter. TSS1 and 2 represent the mapped transcriptional start sites of *CfrI*. The stem loop structure represents a rho-independent terminator downstream of *psbA*. **(b)** Kinetics of transcripts originating from the *psbA* region. Early expression of *psbA* is greater at HL than LL through some yet undefined regulation. This is also true for late expression of *psbA*. In the late period accumulation of *CfrI* sequesters the low abundance psbA leading to the light dependent splicing phenotype. The amount of 'functional' transcript during late infection is shown by the arrows, where the dashed arrows represent the abundance in a hypothetical *CfrI* null mutant.

Unfortunately the inability to generate a *CfrI* repressor mutant limits the testing of this hypothesis rigorously. Moreover the development of genetic manipulation of cyanophages has not yet been achieved. The most powerful tool to unravel this transcriptional regulation may be heterologous re-constitution of the two regulatory elements (the intron and *CfrI*). Attempts to achieve this in *E. coli* were limited by the inability of the intron to self-splice *in-vivo* (Fig. 6.14a). It is suggested here that this may be due to the lack of the formation of paired helix 1, which is formed by interaction of the 5' region of the intron with the 3' of *psbA* exon 1 (Fig. 6.17). Therefore, splicing requires a certain sequence of *psbA* that is lacking in *gfpuv*. As an alternative approach *gfpuv* could be fused to the C-terminus of the full *psbA*, As such, it is expected that self-splicing of the intron will occur as all sequence

requirements for splicing are present within *psbA*. Then the effect of *CfrI* on both *psbA* and the *psbA* intron could be assessed.

6.4.4. Impacts on the evolution of mobile group I introns

The genetic organisation of the *psbA* region in cyanophage S-PM2 is certainly unusual (Millard et al. 2004; Zeng et al. 2009; Millard et al. 2010; Bonocora et al. 2011). The *psbA* intron insertion site 'masks' the 20bp endonuclease cleavage site of the downstream F-CphI (Zeng et al. 2009). This prevents self-cleavage of *psbA* by F-CphI which would almost certainly create a non-functional protein. Thus, the intron has been suspected to offer protection from self-cleavage in a process called collaborative homing (Zeng et al. 2009). However, a more common arrangement is for the homing endonuclease to be located within the intron (Bonocora and Shub 2009), thus creating a mobile group I intron. Therefore, it has been proposed that the situation in *psbA* will eventually form a mobile group I intron and thus represents a snapshot in evolution before the development of mobility (Zeng et al. 2009). Millard et al. (2010) have shown, through metagenomic surveys, that the arrangement in cyanophage S-PM2 is common and F-CphI is yet to be found to have invaded the *psbA* intron. This is somewhat strange considering that that the terminal loop of the intron paired helix 1 contains a sequence that is very similar to the F-CphI cleavage site (Fig. 6.17). 14/20 nucleotides are present in the cyanophage S-PM2 intron. Of the remaining six, four are found at sites of loop formation or unduplexed RNA, thus are not thought to be essential for splicing and hence sequence conservation is not required. The other two are found at sites that correspond to the formation of paired helix 1 of the cyanophage S-PM2 *psbA* intron (Fig. 6.17). However, these two sites are shared in the *psbA* of cyanophage P-SSM2 which can be cleaved by F-CphI (Zeng et al. 2009). Therefore, some other selective forcing must be preventing inclusion of F-CphI into the *psbA* intron.
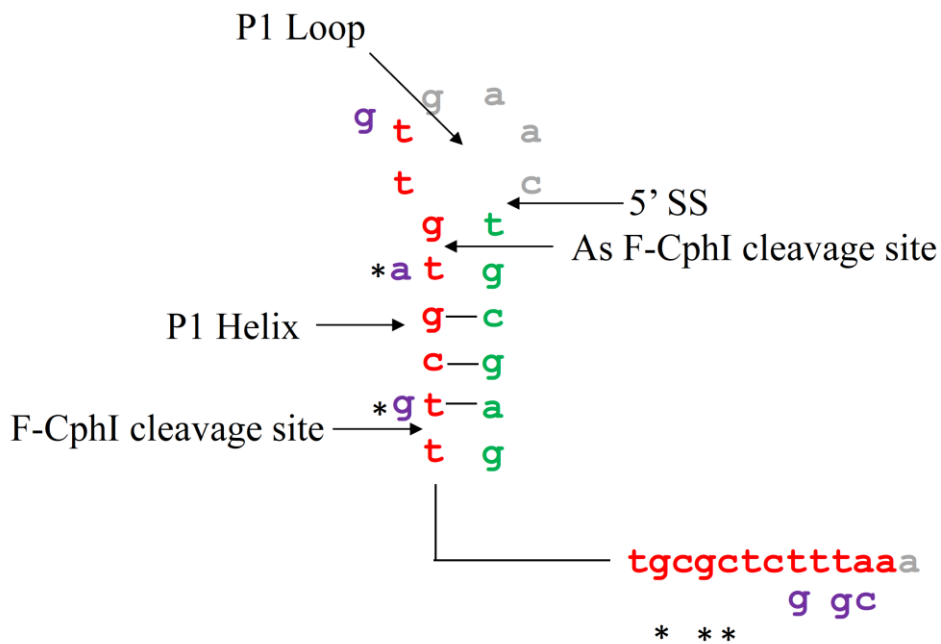
Figure 6.17: Schematic of the putative paired helix 1 from the *psbA* intron. The paired helix is formed from the 3' end of the *psbA* exon 1 (green letters) and the 5' end of the intron (grey and red letters). The formation of this loop is suggested to be essential for splicing and hence why the intron could not splice when inserted into *gfpuv*. The red letters highlight the 20bp region that is strikingly similar to the cleavage site of F-CphI. Differences in nucleotides compared with the mapped cleavage site in phage S-BM1 are shown in purple. Asterisks correspond to sites that are not conserved in the F-*CphI* cleavage recognition site.

Here, we provide evidence that *CfrI* may function in regulating LL expression of its sense partner *psbA*. Therefore, invasion of the *psbA* intron will be concomitant with interruption of *CfrI*. As such the function of *CfrI* may offer enough selective pressure to conserve the separated arrangement of the intron and the homing endonuclease, such that a mobile intron may never be formed.

Chapter Seven: Conclusions and future directions

Here, the results of this thesis are discussed in the context of the original aims of the study highlighted in section 1.7. Further, future approaches are considered that will allow for understanding of interactions between cyanophages and their hosts.

Firstly (Aim 1), an understanding of the role of light in regulating the infection dynamics of cyanophage S-PM2 was sought. It was hypothesised that increased light intensity, through an increased rate of photophosphorylation, would increase the available energy for genome replication. This was not observed in S-PM2 where at both HL and LL the rate of genome replication was identical. However, the latent period was significantly reduced in HL, marked by a reduced maturation period. It was hypothesised that two mechanisms may contribute to this reduced maturation period at HL. The first is that there is a regulated LL lysis delay ('LL lysis delay hypothesis'). It was argued that this may benefit viral development in a diurnal light regime as experienced in the environment given that phage development is often contingent on photophosphorylation for development and that many cyanophages are dependent on light for adsorption to their host. Thus it was speculated that lysis inhibition may occur during the night as regulated by LL lysis delay in the dusk period. However, it was shown that light had very little effect on differential gene expression of S-PM2 genes (Aim 5). Thus for this to be case the regulation of lysis must be occurring post-transcriptionally.

The other explanation, which is favoured here, is that light intensity plays a direct role in controlling lysis through limiting the availability of cellular energy for translation ('Energy limited translation hypothesis'). Crude estimates suggest that the energy required for synthesis of the structural components of the virion far outweigh those required for genome replication even if each nucleotide was synthesised *de novo*, which is unlikely in S-PM2. Thus protein synthesis appears to be the main energy consuming process of phage development, which has largely remained unnoticed in cyanophages. Here, gene acquisitions are often discussed in terms of the fitness benefit afforded to genome replication (Sullivan et al. 2010; Thompson et al. 2011). This is particularly relevant to this study where the only S-

PM2 gene that was differentially expressed in response to light intensity was the horizontally acquired *psbA* (Aim 5). Under the 'energy limited translation hypothesis', an increased rate of translation is afforded by increased rate of photophosphorylation in HL. However, this mechanism is contingent on a continually functioning PSII, which becomes inactivated during periods of HL in the host. Thus by acquiring *psbA* and by co-ordinating light responsive expression, cyanophages may overcome the metabolic bottleneck of translation through making use of excess quanta. Whilst this has long been the hypothesised function for cyanophage-encoded *psbA*, little experimental evidence has been gathered to support this. In addition it shown here, that during infection with S-PM2, PSII photochemistry is altered to favour electron flow at HL (Aim 2). It is suggested that either cyanophage specific residues of the D1 polypeptide or increased expression of cyanophage *psbA* act to increase the turnover rate of D1 and sustain PSII photochemistry at increased irradiances.

This study represents the first approach to understanding the functioning of cyanophage 'photosynthesis' during infection and its specific effect on the developmental program of the phage. In so doing, the importance of energetics during infection is revealed.

Future studies should seek to unravel the mechanism of lysis in S-PM2 and cyanophages in general. Whilst a putative lysin can be identified by homology, a possible holin has not been identified due to extreme sequence divergence in this family. Identification of the holin will be challenging due to the lack of a genetic system in these phages, but an understanding of its functioning and regulation will provide insights into the factors that control this important process within the cyanophage developmental cycle.

In addition, to resolve the exact mechanism of improved PSII functioning under HL during infection, cyanophage *psbAs* could be expressed heterologously in infection free backgrounds. Whilst this may be impossible to do in the host strain due to lack of genetic tractability, freshwater strains of cyanobacteria offer an alternative. Such an approach has

recently revealed modifications to PSI by the cyanophage encoded *psaJF* fusion protein (Mazor et al. 2014).

An additional aim of this study was to provide experimental proof that cyanophage infection inhibits CO_2 fixation while maintaining PET (Aim 3). This is clearly demonstrated in both cyanophages S-PM2 and S-RSM4. S-RSM4 contains the genes *cp12, talC, zwf* and *gnd* required for redirection of host carbon metabolism, whereas S-PM2 does not. It is shown that infection with S-RSM4 'shuts-off' CO_2 fixation much earlier during the infection period compared with S-PM2, yet maintains PET throughout. It is suggested that this process prevents the waste of cellular energy into carbon fixation, whilst at the same time providing NADPH and nucleotide precursors to facilitate genome replication. This data provides clear experimental support of this hypothesis derived from genomic mining and in so doing raises an important facet to the functioning of photosynthesis in the global ocean. In particular, cyanophage infection acts to decouple photochemistry from CO_2 fixation and thus may represent an important alternative electron sink when photochemical based estimates of photosynthesis are used.

Future studies may wish to assess the magnitude of this electron sink in the environment. Whilst extremely challenging, Raman and or Raman-FISH spectroscopy may reveal novel biomarkers of cyanophage infection which can be coupled to rate measurement of CO_2 fixation in uncultured organisms (Huang et al. 2007).

In addition to the effect of light, the power of RNA-Seq was exploited to validate the transcriptional model proposed for S-PM2 (Mann et al. 2005) (Aim 4). It is revealed that only two clusters of genes exist corresponding to early and late modes of expression. The assignment of ORFs to these clusters will benefit future studies that seek functions for many of these ORFs. In so doing, a number of ORFs were discovered that did not conform to this 'umbrella' of regulation. Many of these genes have predicted functions and strong cases can be made for a necessity of 'continued' or 'middle-period' expression evinced in their

profiles. A number of hypotheses are highlighted which may act to modify the expression of these ORFs to suit these modes. A major goal of future work will be the elucidation of the mechanisms that act to regulate expression of these genes.

Moreover, RNA-Seq revealed an unprecedented transcriptional landscape in bacteriophages. The data suggest that genome-wide read-through transcription occurs, particularly on the Watson strand of S-PM2 and particularly during the late period. It is shown that this is not the case for the host WH7803 which therefore acts as a control. This may result from late period pervasive transcription initiation and or by large scale inefficient transcription termination. The divergent late σ-factor of S-PM2 is thought to recognise low sequence information promoters that are AT rich in a genomic background that is also AT rich. Two questions therefore remain. How the σ-factor directs expression to these promoters and why encode such low sequence information for the promoter? It is suggested here, perhaps speculatively, that T4-like cyanophages have evolved to initiate transcription in a shotgun fashion across the genome. Whilst in general this may produce a misappropriation of cellular resources, it may also allow for expression of horizontally acquired genes and therefore selection of these genes if a fitness-benefit is afforded. Subsequently, the full regulatory capacity of such genes is evolved. Therefore it is hypothesised that gene expression places a major barrier to gene flow amongst phages and that the evolution of the T4-like late σ-factor overcomes this barrier. This mechanism results in the strand bias that is frequently observed in cyanophage genomes. This hypothesis requires further attention. Future studies should seek transcriptomes of many other T4-like phages to understand if this is a common phenomenon. Moreover, the high-throughput sequencing of 5' ends of primary transcripts will provide further evidence to support or reject the findings of this thesis.

Previous studies revealed the presence of at least one and the prediction of numerous asRNAs in S-PM2 (Millard et al. 2010). It was therefore hypothesised that asRNA-RNA interactions may be a dominant form of gene regulation in bacteriophages. Thus RNA-Seq was further used to validate these predictions. The analysis suggests that virtually every

ORF in the S-PM2 genome contains significant antisense expression. Such expression is often greater in magnitude than the already validated asRNA *CfrI* and thus these asRNAs are likely to be *bona fide*. A major goal of the near future work will be to identify whether these RNAs are indeed present by direct methods such as northern blotting and by mapping transcript termini by RACE. Moreover, the functional relevance of these RNAs requires attention. If the case of pervasive transcription initiation at AT rich loci produces such transcripts, their regulatory role may be unimportant during infection. However, asRNA-RNA interactions may act to modify the magnitude of expression of a multitude of genes to account for limited global regulatory capacity in S-PM2. The antisense-expression approach highlighted in Chapter 6 may prove a valuable tool for assessment of asRNA functionality. In addition the new method of double-stranded transcripotome sequencing (Lybecker et al. 2014) will detect whether these asRNAs pair with their cognate sense RNA.

Finally, aim 6 sought to determine the role of the asRNA, *CfrI* in regulating cyanophage encoded *psbA* expression in response to changes in light intensity. The aim to overexpress *CfrI* in the host in a bid to 'knock-down' S-PM2 *psbA* expression was unsuccessful, most likely due to interaction with the essential host *psbA* mRNA. However, it is demonstrated that *CfrI* extends further into the *psbA* gene than previously thought (Millard et al. 2010) and therefore overlaps antisense the *psbA* exon/exon boundary. It is shown, through qPCR, that the *psbA* intron is differentially spliced in response to light intensity. Therefore it is suggested that interaction between *CfrI* and *psbA* unspliced mRNA causes this light dependent splicing. A regulatory model is hypothesised that suggests that *CfrI* functions to limit *psbA* expression at LL where production of the D1 polypeptide would be unnecessary due to low levels of photodamage to PSII. Light dependent transcription of *psbA* has to overcome base pairing of *CfrI* in a manner similar to the model suggested by (Dühring et al. 2006). Validation of this exact regulatory mechanism could be achieved by re-constitution of both elements in a heterologous strain that can sustain metabolism through heterotrophy, for example the freshwater cyanobacterium *Synechocystis* PCC6803.

In conclusion, this thesis makes a significant contribution to the study of cyanophage-cyanobacteria interactions through experimental validation of a range of previously hypothesised interactions (e.g. cyanophage modified PSII electron transport, cyanophage redirection of CO_2 fixation and cyanophage temporal regulation of transcription). It also provides novel hypotheses that may explain, in part, the acquisition and maintenance of AMGs in cyanobacteria and how these acquisitions influence the evolution of the genome. Moreover, a range of putative transcription regulation scenarios are presented, whose study are expected to reveal a new level of regulatory complexity in these ecologically important viruses.

References

Ackermann HW, Krisch, HM, and Comeau AM (2011). Morphology and genome sequence of phage Φ1402. *Bacteriophage*, **1** (3), 138–142.

Aiba H and Mizuno T (1994). A novel gene whose expression is regulated by the response-regulator, SphR, in response to phosphate limitation in *Synechococcus* species PCC7942. *Molecular Microbiology*, **13** (1), 25–34.

Aiba H, Nagaya M and Mizuno T (1993). Sensor and regulator proteins from the cyanobacterium *Synechococcus* species PCC7942 that belong to the bacterial signal-transduction protein families: implication in the adaptive response to phosphate limitation. *Molecular Microbiology*, **8** (1), 81–91.

Aller RC (1998). Mobile deltaic and continental shelf muds as suboxic, fluidized bed reactors. *Marine Chemistry*, **61** (3-4), 143–155. doi:10.1016/S0304-4203(98)00024-3

Anders S and Huber W (2010). Differential expression analysis for sequence count data. *Genome Biology*, **11** (10), R106. doi:10.1186/gb-2010-11-10-r106

Anders S, Pyl PT and Huber W (2014). HTSeq A Python framework to work with high-throughput sequencing data. BioRxiv preprint, 1–5. doi:10.1101/002824

Arbiol C, Comeau AM, Kutateladze M, Adamia R and Krisch HM. (2010) Mobile regulatory cassettes mediate modular shuffling in T4-type phage genomes. *Genome Biology and Evolution* **2**, 140–52. doi: 10.1093/gbe/evq006

Artsimovitch I, Patlan V, Sekine S, Vassylyeva MN, Hosaka T, Ochi K, Yokoyama S and Vassylyev DG (2004). Structural basis for transcription regulation by alarmone ppGpp. *Cell*, **117**, 299–310.

Asayama M, Kato H, Shibato J, Shirai M and Ohyama T (2002). The curved DNA structure in the 5'-upstream region of the light-responsive genes: its universality, binding factor and function for cyanobacterial *psbA* transcription. *Nucleic Acids Research*, **30** (21), 4658–4666.

Avrani S, Wurtzel O, Sharon I, Sorek R, and Lindell D (2011). Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature*, **474**, 604–608. doi:10.1038/nature10172

Bailey S, Clokie MR J, Millard A, and Mann NH (2004). Cyanophage infection and photoinhibition in marine cyanobacteria. *Research in Microbiology*, **155** (9), 720–725. doi:http://dx.doi.org/10.1016/j.resmic.2004.06.002

Bailey S, Melis A, Mackey KRM, Cardol P, Finazzi G, van Dijken G, Berg GM, Arrigo K, Shrager J and Grossman A (2008). Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. *Biochimica et Biophysica Acta*, **1777** (3), 269–76. doi:10.1016/j.bbabio.2008.01.002

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li W and Noble WS (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, **37** (2), 202–208. doi:10.1093/nar/gkp335

Bailleul B, Cardol P, Breyton C, and Finazzi G (2010). Electrochromism: a useful probe to study algal photosynthesis. *Photosynthesis Research*, **106** (1-2), 179–189. doi:10.1007/s11120-010-9579-z

Bailly-Bechet M, Vergassola M and Rocha E (2007). Causes for the intriguing presence of tRNAs in phages. *Genome Research*, **17** (10), 1486–1495. doi:10.1101/gr.6649807

Barber J (2008). Photosynthetic generation of oxygen. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **363** (1504), 2665–74. doi:10.1098/rstb.2008.0047

Barik S, Ghosh B, Whalen W, Lazinski D and Das A (1987). An antitermination protein engages the elongating transcription apparatus at a promoter-proximal recognition site. *Cell*, **50** (6), 885–899.

Barker M, de Vries R, Nield J, Komenda J and Nixon PJ (2006). The deg proteases protect *Synechocystis* sp. PCC 6803 during heat and light stresses but are not essential for removal of damaged D1 protein during the photosystem two repair cycle. *The Journal of Biological Chemistry*, **281** (41), 30347–30355. doi:10.1074/jbc.M601064200

Battchikova N, Eisenhut M and Aro, E-M (2011). Cyanobacterial NDH-1 complexes: Novel insights and remaining puzzles. *Biochimica et Biophysica Acta-Bioenergetics*, **1807** (8), 935–944.

Behrenfeld MJ (1999). Widespread iron limitation of phytoplankton in the South Pacific Ocean. *Science*, **283**, 840–843. doi:10.1126/science.283.5403.840

Behrenfeld MJ, Boss E, Siegel DA, Shea DM (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, **19** (1), 1-14. doi:10.1029/2004GB002299

Behrenfeld MJ, O'Malley RT, Siegel DA, McClain CR, Sarmiento JL, Feldman GC, Milligan AJ, Falkowski PG, Letelier RM and Boss ES (2006). Climate-driven trends in contemporary ocean productivity. *Nature*, **444***, 752–755. doi:10.1038/nature05317

Behrenfeld MJ, Worthington K, Sherrell RM, Chavez FP, Strutton P, McPhaden M and Shea DM (2006). Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics. *Nature*, **442**, 1025–1028. doi:10.1038/nature05083

Béjà O, Fridman S and Glaser F (2012). Viral clones from the GOS expedition with an unusual photosystem-I gene cassette organization. *The ISME Journal*, **6** (8), 1617–20. doi:10.1038/ismej.2012.23

Belfort M (1989). Bacteriophage introns: parasites within parasites? *Trends in Genetics*, **5**, 209–213.

Belfort M. (1990). Phage T4 introns: Self-splicing and mobility. *Annual Review of Genetics*, **24**, 363–385.

Belfort M, Derbyshire V, Parker MM, Cousineau B, and Lambowitz AM (2002). Mobile introns: pathways and proteins. *In* Craig NL, Craigie R, Gellert M and Lambowitz AM (Eds.), *Mobile DNA II,* Washington, ASM Press, 761–783.

Bender M, and Sowers T (1994). The Dole effect and its variations during the last 130,000 years as measured in the Vostok ice core. *Global Biogeochemical Cycles*, **8** (3), 363–376.

Ben-Shem A, Frolow F, and Nelson N (2003). Crystal structure of plant photosystem I. *Nature*, **426** (6967), 630–635. doi:10.1038/nature02200

Berthoumieux S, de Jong H, Baptist G, Pinel C, Ranquet C, Ropers D, and Geiselmann J (2013). Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Molecular Systems Biology*, **9** (634), 634. doi:10.1038/msb.2012.70

Bhaya D, Bianco NR, Bryant D, and Grossman A (2000). Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. PCC6803. *Molecular Microbiology*, **37** (4), 941–951.

Bhaya D, Watanabe N, Ogawa T and Grossman AR (1999). The role of an alternative sigma factor in motility and pilus formation in the cyanobacterium *Synechocystis* sp. strain PCC6803. *Proceedings of the National Academy of Sciences of the United States of America*, **96** (6), 3188–3193.

Bidle KD and Vardi A (2011). A chemical arms race at sea mediates algal host-virus interactions. *Current Opinion in Microbiology*, **14** (4), 449–457. doi:10.1016/j.mib.2011.07.013

Blankenship RE (2001). Molecular evidence for the evolution of photosynthesis. *Trends in Plant Science*, **6** (1), 4–6.

Blankenship RE (2010). Early evolution of photosynthesis. *Plant Physiology*, **154** (2), 434–438. doi:10.1104/pp.110.161687

Blot N, Mella-Flores D, Six C, Le Corguillé G, Boutte C, Peyrat A, Monnier A, Ratin M, Gourvil P, Campbell DA and Garczarek L (2011). Light history influences the response of the marine cyanobacterium *Synechococcus* sp. WH7803 to oxidative stress. *Plant Physiology*, **156** (4), 1934–1954. doi:10.1104/pp.111.174714

Blower TR, Pei XY, Short FL, Fineran PC, Humphreys DP, Luisi BF and Salmond GPC (2011). A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nature Structural & Molecular Biology*, **18** (2), 185–190. doi:10.1038/nsmb.1981

Boehme J, Frischer ME, Jiang SC, Kellogg CA, Pichard S, Rose JB, Steinway C and Paul JH (1993). Viruses, bacterioplankton, and phytoplankton in the Southeastern Gulf of Mexico : Distribution and contribution to oceanic DNA pools. *Marine Ecology Progress Series*, **97**, 1–10.

Bonocora RP and Shub DA (2009). A likely pathway for formation of mobile group I introns. *Current Biology*, **19** (3), 223–228. doi:10.1016/j.cub.2009.01.033

Bonocora RP, Zeng Q, Abel EV and Shub DA (2011). A homing endonuclease and the 50-nt ribosomal bypass sequence of phage T4 constitute a mobile DNA cassette. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (39), 16351–16356. doi:10.1073/pnas.1107633108

Borbely G, Kaki C, Gulyas A and Farkas GL (1980). Bacteriophage infection interferes with guanosine accumulation induced by energy and nitrogen starvation in the cyanobacterium *Anacystis nidulans*. *Journal of Bacteriology*, **144** (3), 859–864.

Bornefeld T and Simonis W (1974). Effects of light, temperature, pH, and inhibitors on the ATP level of the blue-green alga *Anacystis nidulans*. *Planta*, **115**, 309-318.

Bragg JG and Chisholm SW (2008). Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PloS One*, **3** (10), e3550. doi:10.1371/journal.pone.0003550

Brahamsha B (1996). A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Applied and Environmental Microbiology*, **62** (5), 1741–1751.

Brantl S (2002). Antisense-RNA regulation and RNA interference. *Biochimica et Biophysica Acta*, **1575** (1-3), 15–25.

Breitbart M (2012). Marine viruses: truth or dare. *Annual Review of Marine Science*, **4**, 425–48. doi:10.1146/annurev-marine-120709-142805

Breitbart M, Thompson LR, Suttle CA and Sullivan MB (2007). Exploring the vast diversity of marine viruses. *Oceanography*, **20** (2), 135–139.

Bretaudeau A, Coste F, Humily F, Garczarek L, Le Corguillé G, Six C, Ratin M, Collin O, Schuluchter WM and Partensky F (2013). CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. *Nucleic Acids Research*, **41** (D1), D396–401. doi:10.1093/nar/gks1091

Bruyant F, Babin M, Sciandra A, Marie D, Genty B, Claustre H, Blanchot J, Bricaud A, Rippka R, Boulben S, Louis F and Partensky F (2001). An axenic cyclostat of *Prochlorococcus* PCC 9511 with a simulator of natural light regimes. *Journal of Applied Phycology*, **13**, 135–142.

Bruyant F, Babin M and Sud P (2005). Diel variations in the photosynthetic parameters of *Prochlorococcus* strain PCC 9511 : Combined effects of light and cell cycle. *Limnology and Oceanography,* **50** (3), 850–863.

Bryan MJ, Burroughs NJ, Spence EM, Clokie MRJ, Mann NH and Bryan SJ (2008). Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. *PloS One*, **3** (4), e2048. doi:10.1371/journal.pone.0002048

Buist G, Steen A, Kok J and Kuipers OP (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. *Molecular Microbiology*, **68** (4), 838–847. doi:10.1111/j.1365-2958.2008.06211.x

Burr T, Mitchell J, Kolb A, Minchin S and Busby S (2000). DNA sequence elements located immediately upstream of the − 10 hexamer in *Escherichia coli* promoters : a systematic study. *Nucleic Acids Research*, **28** (9), 1864–1870.

Busch AWU, Reijerse EJ, Lubitz W, Hofmann E and Frankenberg-Dinkel N (2011). Radical mechanism of cyanophage phycoerythrobilin synthase (PebS). *The Biochemical Journal*, **433** (3), 469–476. doi:10.1042/BJ20101642

Busch A, Richter AS and Backofen R (2008). IntaRNA : efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, **24** (24), 2849–2856. doi:10.1093/bioinformatics/btn544

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley G, Vandesompele J and Wittwer CT (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, **55** (4), 611–22. doi:10.1373/clinchem.2008.112797

Callen BP, Shearwin KE and Egan JB (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. *Molecular Cell*, **14** (5), 647–656. doi:10.1016/j.molcel.2004.05.010

Cambray G, Guimaraes JC, Mutalik VK, Lam C, Mai Q-A, Thimmaiah T, Carothers JM, Arkin AP and Endy D. (2013) Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Research,* **41**, 5139–4148. doi: 10.1093/nar/gkt163

Catalão MJ, Gil F, Moniz-Pereira J, São-José C and Pimentel M (2013). Diversity in bacterial lysis systems: bacteriophages show the way. *FEMS Microbiology Reviews*, **37** (4), 554–571. doi:10.1111/1574-6976.12006

Cech TR (1990). Self-splicing of group I introns. *Annual Review of Biochemistry.* **59**, 543-568.

Chang C, Nam K and Young R (1995). S gene expression and the timing of lysis by bacteriophage lambda. *Journal of Bacteriology*, **177** (11), 3283-3294.

Chang FN, Sih CJ and Wesiblum B (1966). Lincomycin, An inhibitor of aminoacyl sRNA binding to ribosomes. *Proceedings of the National Academy of Sciences of the USA* , **55**, 431–438.

Chatterjee A, Johnson CM, Shu C-C, Kaznessis YN, Ramkrishna D, Dunny GM and Hu W-S (2011). Convergent transcription confers a bistable switch in *Enterococcus faecalis* conjugation. *Proceedings of the National Academy of Sciences of the USA*, **108** (23), 9721–9726. doi:10.1073/pnas.1101569108

Chen Z and Schneider TD (2005). Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases. *Nucleic Acids Research*, **33** (19), 6172–6187. doi:10.1093/nar/gki915

Chénard C, and Suttle CA (2008). Phylogenetic diversity of sequences of cyanophage photosynthetic gene *psbA* in marine and freshwaters. *Applied and Environmental Microbiology*, **74** (17), 5317–5324. doi:10.1128/AEM.02480-07

Chitnis VP and Chitnis PR (1993). PsaL subunit is required for the formation of photosystem I trimers in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS Letters*, **336** (2), 330–334.

Chow C-ET and Fuhrman JA (2012). Seasonality and monthly dynamics of marine myovirus communities. *Environmental Microbiology*, **14** (8), 2171–2183. doi:10.1111/j.1462-2920.2012.02744.x

Cianciotto NP (2005). Type II secretion: a protein secretion system for all seasons. *Trends in Microbiology*, **13** (12), 581–588. doi:10.1016/j.tim.2005.09.005

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** (1), 17–31. doi:10.3354/ame01613

Clokie MRJ, Shan J, Bailey S, Jia Y, Krisch HM, West S and Mann NH (2006). Transcription of a "photosynthetic" T4-type phage during infection of a marine cyanobacterium. *Environmental Microbiology*, **8** (5), 827–35. doi:10.1111/j.1462-2920.2005.00969.x

Clokie MRJ, Thalassinos K, Boulanger P, Slade SE, Stoilova-McPhie S, Cane M, Scrivens JH and Mann NH (2008). A proteomic approach to the identification of the major virion structural proteins of the marine cyanomyovirus S-PM2. *Microbiology*, **154** (6), 1775–1782. doi:10.1099/mic.0.2007/016261-0

Cochlan WP, Wikner J, Steward GF, Smith DC and Azam F (1993). Spatial distribution of viruses, bacteria and chlorophyll *a* in neritic, oceanic and estuarine environments. *Marine Ecology Progress Series,* **92**, 77–87.

Comeau AM, Tremblay D, Moineau S, Rattei T, Kushkina AI, Tovkach FI, Krisch HM and Ackermann HW (2012). Phage morphology recapitulates phylogeny: the comparative genomics of a new group of myoviruses. *PloS One*, **7** (7), e40102. doi:10.1371/journal.pone.0040102

Courtney CM and Chatterjee A (2014). *cis*-Antisense RNA and transcriptional interference : Coupled layers of gene regulation. *Journal of Gene Therapy*, **2** (1), 1–9.

Crampton N, Bonass WA, Kirkham J, Rivetti C and Thomson NH (2006). Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Research*, **34** (19), 5416–5425. doi:10.1093/nar/gkl668

Culley AI and Welschmeyer NA (2002). The abundance, distribution, and correlation of viruses, phytoplankton, and prokaryotes along a Pacific Ocean transect. *Limnology and Oceanography*, **47** (5), 1508–1513.

Dammeyer T, Bagby SC, Sullivan MB, Chisholm SW and Frankenberg-Dinkel N (2008). Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Current Biology*, **18** (6), 442–448. doi:10.1016/j.cub.2008.02.067

Danovaro R, Corinaldesi C, Dell'anno A, Fuhrman JA, Middelburg JJ, Noble RT and Suttle CA (2011). Marine viruses and global climate change. *FEMS Microbiology Reviews*, **35** (6), 993–1034. doi:10.1111/j.1574-6976.2010.00258.x

Debus RJ, Barry BA, Sithole I, Babcock GT and McIntosh L (1988). Directed mutagenesis indicates the donor to P+680 in photosystem II Is Tyrosine-161 of the D1 polypeptide. *Biochemistry*, **27** (26), 9071–9074.

Dedrick RM, Marinelli LJ, Newton GL, Pogliano K, Pogliano J and Hatfull GF (2013). Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. *Molecular Microbiology*, **88** (3), 577–89. doi:10.1111/mmi.12210

Del Giorgio PA, Cole JJ and Cimbleris A (1997). Respiration rates in bacteria exceed phytoplankton production in unproductive aquatic systems. *Nature*, **385**, 148–150.

Deng L, Gregory A, Yilmaz S, Poulos BT, Hugenholtz P and Sullivan MB (2012). Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria , as revealed by viral tagging. *American Society for Microbiology*, **3** (6), e00373-12. doi:10.1128/mBio.00373-12

Deshpande NN, Bao Y and Herrin DL (1997). Evidence for light/redox-regulated splicing of *psbA* pre-RNAs in *Chlamydomonas* chloroplasts. *RNA*, **3** (1), 37–48.

Dornenburg JE, Devita AM, Palumbo MJ and Wade JT (2010). Widespread antisense transcription in *Escherichia coli*. *MBio*, **1** (1), 00024-10. doi:10.1128/mBio.00024-10

Dreher TW, Brown N, Bozarth CS, Schwartz AD, Riscoe E, Thrash C, Bennett E, Shin-Cheng T and Maier CS (2011). A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. *Environmental Microbiology*, **13** (7), 1858–74. doi:10.1111/j.1462-2920.2011.02502.x

Duarte CM, Regaudie-de-Gioux A, Arrieta JM, Delgado-Huertas A and Agustí S (2013). The oligotrophic ocean is heterotrophic. *Annual Review of Marine Science*, **5**, 551–69. doi:10.1146/annurev-marine-121211-172337

Ducklow HW and Doney SC (2013). What is the metabolic state of the oligotrophic ocean? A debate. *Annual Review of Marine Science*, **5**, 525–33. doi:10.1146/annurev-marine-121211-172331

Dühring U, Axmann IM, Hess WR and Wilde A (2006). An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proceedings of the National Academy of Sciences of the United States of America*, **103** (18), 7054–8. doi:10.1073/pnas.0600927103

Elhai J and Wolk CP (1988). Conjugal transfer of DNA to cyanobacteria. *Methods in Ezymology*, **167**, 747–754.

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** (3), 619–628. doi:10.1038/ismej.2011.146

Enav H, Mandel-Gutfreund Y and Béjà O (2014). Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome*, **2** (1), 9. doi:10.1186/2049-2618-2-9

Erickson JM, Rahire M and Rochaix J (1984). *Chlamydomonas reinhardii* gene for the 32,000 mol. Wt. protein of photosystem II contains four large introns and is located entirely within the chloroplast inverted repeat. *The EMBO Journal*, **3** (12), 2753–2762.

Eriksson J, Salih GF, Ghebramedhin H and Jansson C (2000). Deletion mutagenesis of the 5' *psbA2* region in *Synechocystis* 6803: identification of a putative *cis* element involved in photoregulation. *Molecular Cell Biology Research Communications*, **3** (5), 292–298. doi:10.1006/mcbr.2000.0227

Falkowski PG, Fenchel T and Delong EF (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, **320**, 1034–1039. doi:10.1126/science.1153213

Falkowski PG and Godfrey LV (2008). Electrons, life and the evolution of Earth's oxygen cycle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **363** (1504), 2705–2716. doi:10.1098/rstb.2008.0054

Falkowski PG, Katz ME, Milligan AJ, Fennel K, Cramer BS, Aubry MP, Berner RA Novacek MJ and Zapol WM (2005). The rise of oxygen over the past 205 million years and the evolution of large placental mammals. *Science*, **309**, 2202–2204. doi:10.1126/science.1116047

Falkowski PG and Raven JA (2007). *Aquatic Photosynthesis* (2nd ed.). Princeton: Princeton University Press.

Fassioli F, Dinshaw R, Arpin PC and Scholes GD (2014). Photosynthetic light harvesting: excitons and coherence. *Journal of the Royal Society Interface.* **11** (92), 20130901.

Field CB, Behrenfeld MJ, Randerson JT and Falkowski PG (1998). Primary production of the biosphere: Integrating terrestrial and Oceanic components. *Science*, **281**, 237–240. doi:10.1126/science.281.5374.237

Filiatrault MJ, Stodghill PV, Bronstein PA, Moll S, Lindeberg M, Grills G, Schweitzer P, Wang W, Scroth GP, Luo S, Khrebtukova I, Yang Y, Thannhauser T, Butcher BG, Cartinhour S, et al. (2010). Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. *Journal of Bacteriology*, **192** (9), 2359–2372. doi:10.1128/JB.01445-09

Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS and Salmond GPC (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (3), 894–899. doi:10.1073/pnas.0808832106

Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, and Jiao N (2013). Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences*, **110** (24). doi:10.1073/pnas.1307701110

Frada M, Probert I, Allen MJ, Wilson WH and de Vargas C (2008). The "Cheshire Cat" escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (41), 15944–15999. doi:10.1073/pnas.0807707105

Fuhrman JA (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, **399** (6736), 541–548. doi:10.1038/21119

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

Geiduschek EP and Kassavetis GA (2010). Transcription of the T4 late genes. *Virology Journal*, **7** (1), 288-300. doi:10.1186/1743-422X-7-288

Genty B, Briantais J-M and Baker NR (1989). The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochimica et Biophysica Acta (BBA) - General Subjects*, **990** (1), 87–92. doi:10.1016/S0304-4165(89)80016-9

Georg J and Hess WR (2011). *cis*-antisense RNA, another level of gene regulation in bacteria. *Microbiology and Molecular Biology Reviews*, **75** (2), 286–300. doi:10.1128/MMBR.00032-10

Georg J, Voss B, Scholz I, Mitschke J, Wilde A and Hess WR (2009). Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Molecular Systems Biology*, **5** (1), 1-17. doi:10.1038/msb.2009.63

Glazer AN and Cohen-Bazire G (1971). Subunit structure of the phycobiliproteins of blue-green algae. *Proceedings of the National Academy of Sciences of the United States of America*, **68** (7), 1398–1401.

Groom S, Martinez-Vicente V, Fishwick J, Tilstone G, Moore G, Smyth T, and Harbour D (2009). The Western English Channel observatory: Optical characteristics of station L4. *Journal of Marine Systems*, **77** (3), 278–295. doi:10.1016/j.jmarsys.2007.12.015

Güell M, van Noort V, Yus E, Chen W-H, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, Rode M, Suyama M, Schmidt S, Gavin A-C, Bork P, et al. (2009). Transcriptome complexity in a genome-reduced bacterium. *Science*, **326**, 1268–1271. doi:10.1126/science.1176951

Hausner G, Hafez M and Edgell DR (2014). Bacterial group I introns: mobile RNA catalysts. *Mobile DNA*, **5** (1), 8. doi:10.1186/1759-8753-5-8

Haussühl 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** (4), 713–22. doi:10.1093/emboj/20.4.713

Havaux M (2003). Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll a/b-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1557**, 21–33. doi:10.1016/S0005-2728(02)00391-2

He Q, Dolganov N, Bjorkman O and Grossman AR (2001). The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *The Journal of Biological Chemistry*, **276** (1), 306–14. doi:10.1074/jbc.M008686200

Hedges JI and Keil RG (1995). Sedimentary organic matter preservation: an assessment and speculative synthesis. *Marine Chemistry*, **49** (2-3), 81–115. doi:10.1016/0304-4203(95)00008-F

Hellweger FL (2009). Carrying photosynthesis genes increases ecological fitness of cyanophage *in silico*. *Environmental Microbiology*, **11** (6), 1386–1394. doi:10.1111/j.1462-2920.2009.01866.x

Hernández JA, Muro-Pastor AM, Flores E, Bes MT, Peleato ML and Fillat MF (2006). Identification of a *furA cis* antisense RNA in the cyanobacterium *Anabaena* sp. PCC 7120. *Journal of Molecular Biology*, **355** (3), 325–334. doi:10.1016/j.jmb.2005.10.079

Hernandez-Prieto MA, Tibiletti T, Abasova L, Kirilovsky D, Vass I and Funk C (2011). The small CAB-like proteins of the cyanobacterium *Synechocystis* sp. PCC 6803: their involvement in chlorophyll biogenesis for Photosystem II. *Biochimica et Biophysica Acta*, **1807** (9), 1143–1151. doi:10.1016/j.bbabio.2011.05.002

Herranen M, Aro E-M and Tyystjärvi T (2001). Two distinct mechanisms regulate the transcription of photosystem II genes in *Synechocystis* sp. PCC 6803. *Physiologia Plantarum*, **112** (4), 531–539.

Hirani TA, Suzuki I, Murata N, Hayashi H, and Eaton-Rye JJ (2001). Characterization of a two-component signal transduction system involved in the induction of alkaline phosphatase under phosphate-limiting conditions in *Synechocystis* sp. PCC 6803. *Plant Molecular Biology*, **45** (2), 133–144.

Holland HD (2006). The oxygenation of the atmosphere and oceans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **361** (1470), 903–915. doi:10.1098/rstb.2006.1838

Huang S, Wang K, Jiao N, and Chen F (2012). Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environmental Microbiology*, **14** (2), 540–58. doi:10.1111/j.1462-2920.2011.02667.x

Huang WE, Stoecker K, Griffiths R, Newbold L, Daims H, Whiteley AS and Wagner M (2007). Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence *in situ* hybridization for the single cell analysis of identity and function. *Environmental Microbiology* **9**, 1878–1889. doi: 10.1111/j.1462-2920.2007.01352.x

Hurwitz BL, Hallam SJ, and Sullivan MB (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biology*, **14** (11), R123. doi:10.1186/gb-2013-14-11-r123

Hyman P and Abedon ST (2009). Practical methods for determining phage growth parameters. *In* Clokie MRJ and Kropinski AM (Eds.), *Bacteriophages Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions*. 1st ed., New York, Humana Press, 175–202.

Ignacio-Espinoza JC and Sullivan MB (2012). Phylogenomics of T4 cyanophages: lateral gene transfer in the "core" and origins of host genes. *Environmental Microbiology*, **14** (8), 2113–2126. doi:10.1111/j.1462-2920.2012.02704.x

Imamura S, Asayama M, and Shirai M (2004). *In vitro* transcription analysis by reconstituted cyanobacterial RNA polymerase: roles of group 1 and 2 sigma factors and a core subunit, RpoC2. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms*, **9** (12), 1175–1187. doi:10.1111/j.1365-2443.2004.00808.x

Imamura S, Yoshihara S, Nakano S, Shiozaki N, Yamada A, Tanaka K, Takahashi H, Asayama M and Shirai M (2003). Purification, characterization, and gene expression of all sigma factors of RNA Polymerase in a cyanobacterium. *Journal of Molecular Biology*, **325** (5), 857–872. doi:10.1016/S0022-2836(02)01242-1

Jacquet S, Heldal M, Iglesias-Rodriguez D, Larsen A, Wilson W and Bratbak G (2002). Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. *Aquatic Microbial Ecology*, **27**, 111–124. doi:10.3354/ame027111

Jia Y, Shan J, Millard A, Clokie MRJ and Mann NH (2010). Light-dependent adsorption of photosynthetic cyanophages to *Synechococcus* sp. WH7803. *FEMS Microbiology Letters*, **310** (2), 120–126. doi:10.1111/j.1574-6968.2010.02054.x

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, and Chisholm SW (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*, **311**, 1737–1740. doi:10.1126/science.1118052

Jordan P, Fromme P, Witt HT, Klukas O, Saenger W and Krauss N (2001). Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature*, **411** (6840), 909–917. doi:10.1038/35082000

Kallas T and Castenholz R W (1982). Internal pH and ATP-ADP pools in the cyanobacterium *Synechococcus* sp. during exposure to growth-inhibiting low pH. *Journal of Bacteriology*, **149** (1), 229–236.

Kana TM and Glibert PM (1987). Effect of irradiances up to 2000 μE m^{-2} s^{-1} on marine *Synechococcus* WH7803-I. Growth, pigmentation and cell composition. *Deep Sea Research A*, **34** (4), 479-495.

Kawano M, Aravind L and Storz G (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Molecular Microbiology*, **64** (3), 738–754. doi:10.1111/j.1365-2958.2007.05688.x

Klumpp S, Zhang Z and Hwa T (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cell*, **139** (7), 1366–1375. doi:10.1016/j.cell.2009.12.001

Komenda J, Hassan HAG, Diner BA, Debus RJ, Barber J and Nixon PJ (2000). Degradation of the Photosystem II D1 and D2 proteins in different strains of the cyanobacterium *Synechocystis* PCC6803 varying with respect to the type and level of *psbA* transcript. *Plant Molecular Biology*, **42**, 635–645.

Komenda J, Tichy M, Prásil O, Knoppová J, Kuviková S, de Vries R and Nixon PJ (2007). The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during photosystem II repair in *Synechocystis* sp PCC 6803. *The Plant Cell*, **19** (9), 2839–2854. doi:10.1105/tpc.107.053868

Kozloff LM, Knowlton K, Putnam FW and Evans Jnr EA (1951). Biochemical studies of virus reproduction: V. The origin of bacteriophage nitrogen. *Journal of Biological Chemistry,* **188**, 101-116.

Krinke L and Wulff DL (1990). RNase III-dependent hydrolysis of lambda cII-O gene mRNA mediated by lambda OOP antisense RNA. *Genes & Development*, **4** (12a), 2223–2233. doi:10.1101/gad.4.12a.2223

Kulkarni RD and Golden SS (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** (4), 959–965.

Kurisu G, Zhang H, Smith JL and Cramer WA (2003). Structure of the cytochrome *b6f* complex of oxygenic photosynthesis: tuning the cavity. *Science*, **302**, 1009–1014. doi:10.1126/science.1090165

Kutsuna S, Kondo T and Aoki S (1998). A Period-Extender gene, *pex,* that extends the period of the circadian clock in the cyanobacterium *Synechococcus* sp. Strain PCC7942. *Journal of Bacteriology*, **180** (8), 2167-2175.

Kuty GF, Xu M, Struck DK, Summer EJ and Young R (2010). Regulation of a phage endolysin by disulfide caging. *Journal of Bacteriology*, **192** (21), 5682–5687. doi:10.1128/JB.00674-10

Labrie SJ, Frois-Moniz K, Osburne MS, Kelly L, Roggensack SE, Sullivan MB, Gearin G, Zeng Q, Fitzgerald M, Henn MR and Chisholm SW (2013). Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environmental Microbiology*, **15** (5), 1356–1376. doi:10.1111/1462-2920.12053

Langmead B and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Lasa I, Toledo-Arana A, Dobin A, Villanueva M, Ruiz I, Mozos DL, Vergara-Irigaray M, Segura V, Fagegaltier D, Penandes JR, Valle J, Solano C and Gingeras TR (2011). Genome-wide antisense transcription drives mRNA processing in bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* **108** (50), 20172-20177 doi:10.1073/pnas.1113521108

Lasa I, Toledo-Arana A, and Gingeras TR (2012). An effort to make sense of antisense transcription in bacteria. *RNA Biology*, **9** (8), 1039–1044. doi:10.4161/rna.21167

Lawrenz E, Silsbe G, Capuzzo E, Ylöstalo P, Forster RM, Simis SGH, Prášil O, Kromkamp JC, Hickman AE, Moore M, Forget M-H, Geider R and Suggett DJ (2013). Predicting the electron requirement for carbon fixation in seas and oceans. *PloS One*, **8** (3), e58137. doi:10.1371/journal.pone.0058137

Lee SK and Keasling JD (2005). A propionate-inducible expression system for enteric bacteria. *Applied and Environmental Microbiology*, **71** (11), 6856-6862. doi:10.1128/AEM.71.11.6856

Legewie S, Dienst D, Wilde A, Herzel H and Axmann IM (2008). Small RNAs establish delays and temporal thresholds in gene expression. *Biophysical Journal*, **95** (7), 3232–3238. doi:10.1529/biophysj.108.133819

Lennon JT, Khatana SAM, Marston MF and Martiny JBH (2007). Is there a cost of virus resistance in marine cyanobacteria? *The ISME Journal*, **1** (4), 300–312. doi:10.1038/ismej.2007.37

Li F, Holloway SP, Lee J and Herrin DL (2002). Nuclear genes that promote splicing of group I introns in the chloroplast 23S rRNA and *psbA* genes in *Chlamydomonas reinhardtii*. *The Plant Journal : For Cell and Molecular Biology*, **32** (4), 467–480.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25** (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li H and Sherman LA (2000). A redox-responsive regulator of photosynthesis gene expression in the cyanobacterium *Synechocystis* sp. Strain PCC 6803. *Journal of Bacteriology*, **182** (15), 4268–4277.

Li R and Golden SS (1993). Enhancer activity of light-responsive regulatory elements in the untranslated leader regions of cyanobacterial *psbA* genes. *Proceedings of the National Academy of Sciences of the United States of America*, **90** (24), 11678–11682.

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Wolfgang RH, Church GM & Chisholm SW (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*, **449** , 83–86. doi:10.1038/nature06130

Lindell D, Jaffe JD, Johnson ZI, Church GM and Chisholm SW (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*, **438**, 86–89. doi:10.1038/nature04111

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F and Chisholm SW (2004a). Transfer of photosynthesis genes to and from P*rochlorococcus* viruses. *Proceedings of the National Academy of Sciences of the United States of America*, **101** (30), 11013–11018. doi:10.1073/pnas.0401526101

Logemann J, Schell J and Willmitzer L (1987). Improved method for the isolation of RNA from plant tissues. *Analytical Biochemistry*, **163**, 16–20.

Lonetto M Gribskov M and Gross CA (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology*, **174** (12), 3843–3849.

Longhurst A, Sathyendranath S, Platt T and Caverhill C (1995). An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research*, **17** (6), 1245–1271. doi:10.1093/plankt/17.6.1245

Lubberding HJ and Schroten W (1984). The ATP level in the thermophilic cyanobacterium *Synechococcus* 6716 during light-dark transition and in the presence of some specific inhibitors. *FEMS Microbiology Letters*, **22** (2), 93–96.

Luke K, Radek A, Liu X, Campbell J, Uzan M, Haselkorn R and Kogan Y (2002). Microarray analysis of gene expression during bacteriophage T4 infection. *Virology*, **299** (2), 182–191. doi:10.1006/viro.2002.1409

Luz P, Lavigne R, Lecoutere E, Wagemans J, Cenens W, Aertsen A, Schoofs L, Landuyt B, Paeshuyse J, Scheer M, Schobert M and Ceyssens P-J (2013). A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent Podovirus LUZ19. *MBio*, **4** (2), e0061-13 doi:10.1128/mBio.00061-13

Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N and Schroeder R (2014) The double-stranded transcriptome of *Escherichia coli*. *Proceeding of the National Academy of Sciences of the U.S.A.*, **111**, 3134–3139. doi: 10.1073/pnas.1315974111

MacIntyre HL, Kana TM and Geider RJ (2000). The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. *Trends in Plant Science*, **5** (1), 12–17.

Magnusson LU, Farewell A and Nyström T (2005). ppGpp: a global regulator in *Escherichia coli. Trends in Microbiology*, **13** (5), 236–242. doi:10.1016/j.tim.2005.03.008

Mann NH, Carr NG and Midgley JEM (1975). RNA synthesis and accumulation of guanine nucleotides during growth shift down in the blue-green alga *Anacystis nidulans. Biochimica et Biophysica Acta*, **402**, 41–50.

Mann NH, Clokie MRJ, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A and Krisch HM (2005). The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *Journal of Bacteriology*, **187** (9), 3188–3200. doi:10.1128/JB.187.9.3188-3200.2005

Mann NH, Cook A, Millard A, Bailey S and Clokie M (2003). Bacterial photosynthesis genes in a virus. *Nature*, **424**, 741–742.

Marchant H, Davidson A, Wright S and Glazebrook J (2000). The distribution and abundance of viruses in the Southern Ocean during spring, **12** (4), 414–417.

Marston MF, Pierciey FJ, Shepard A, Gearin G, Qi J, Yandava C, Schuster SC, Henn MR and Martiny JBH (2012). Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proceedings of the National Academy of Sciences of the USA*, **109** (12), 4544–4549. doi:10.1073/pnas.1120310109

Marston MF and Sallee JL (2003). Genetic diversity and temporal variation in the Cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Applied and Environmental Microbiology*, **69** (8). doi:10.1128/AEM.69.8.4639

Martiny JBH, Riemann L, Marston MF and Middelboe M (2014). Antagonistic coevolution of marine planktonic viruses and their hosts. *Annual Review of Marine Science*, **6**, 393–414. doi:10.1146/annurev-marine-010213-135108

Mazor Y, Greenberg I, Toporik H, Beja O and Nelson N (2012). The evolution of photosystem I in light of phage-encoded reaction centres. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **367** (1608), 3400–3405. doi:10.1098/rstb.2012.0057

Mazor Y, Nataf D, Toporik H and Nelson N (2014). Crystal structures of virus-like photosystem I complexes from the mesophilic cyanobacterium *Synechocystis* PCC 6803. *eLife*, **3**, e01496. doi:10.7554/eLife.01496

McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, Vanderpool CK and Tjaden B. (2013). Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research*, **41** (14), e140. doi:10.1093/nar/gkt444

Mella-Flores D, Six C, Ratin M, Partensky F, Boutte C, Le Corguillé G, Marie D, Blot N, Gourvil P, Kolowrat C and Garczarek L (2012). *Prochlorococcus* and *Synechococcus* have evolved different adaptive mechanisms to cope with light and UV stress. *Frontiers in Microbiology*, **3**, 285, 1-20. doi:10.3389/fmicb.2012.00285

Mendoza-Vargas A, Olvera L, Olvera M, Grande R, Vega-Alvarado L, Taboada B, Jiminez-Jacinto V, Salgado H, Juárez K, Contreas-Moreira B, Huerta AM, Collado-Vides J and Morett, E (2009). Genome-wide identification of transcription start sites, promoters

and transcription factor binding sites in *E. coli*. *PloS One*, **4** (10), e7526. doi:10.1371/journal.pone.0007526

Millard A, Clokie MRJ, Shub DA and Mann NH (2004). Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus strains*. *Proceedings of the National Academy of Sciences of the USA,* **101** (30), 11007–11012. doi:10.1073/pnas.0401478101

Millard AD (2009). Isolation of cyanophages from aquatic environments. *In* Clokie MRJ and Kropinski AM (Eds.), *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions*. 1st ed., New York, Humana Press, 33-42. doi:10.1007/978-1-60327-164-6

Millard AD, Gierga G, Clokie MRJ, Evans DJ, Hess WR and Scanlan DJ (2010). An antisense RNA in a lytic cyanophage links *psbA* to a gene encoding a homing endonuclease. *The ISME Journal*, **4** (9), 1121–1135. doi:10.1038/ismej.2010.43

Millard AD and Mann NH (2006). A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba , Red Sea. *Journal of the Marine Biological Association U.K,* **86**, 507-515.

Millard AD, Zwirglmaier K, Downey MJ, Mann NH, and Scanlan DJ (2009). Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environmental Microbiology*, **11** (9), 2370–2387. doi:10.1111/j.1462-2920.2009.01966.x

Miller ES, Kutter E, Mosig G, Kunisawa T, Rüger W, Arisaka F and Ru W (2003). Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews*, **67** (1), 86–156. doi:10.1128/MMBR.67.1.86

Mitchell P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature*, **191**, 144–148.

Mitschke J, Georg J, Scholz I, Sharma, CM, Dienst D, Bantscheff J, Voss B, Steglich C, Wilde A, Vogel J and Hess WR (2011). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proceedings of the National Academy of Sciences*, **108** (5), 2124–2129. doi:10.1073/pnas.1015154108

Mohamed A, Eriksson J, Osiewacz HD and Jansson C (1993). Differential expression of the *psbA* genes in the cyanobacterium *Synechocystis* 6803. *Molecular & General Genetics*, **238** (1-2), 161–168.

Ackermann HW, Krisch HM, Comeau AM (2011) Morphology and genome sequence of phage Φ1402. Bacteriophage 1:3:138–142.

Aiba H, Mizuno T (1994) A novel gene whose expression is regulated by the response-regulator, SphR, in response to phosphate limitation in Synechococcus species PCC7942. Mol Microbiol 13:25–34.

Aiba H, Nagaya M, Mizuno T (1993) Sensor and regulator proteins from the cyanobacterium Synechococcus species PCC7942 that belong to the bacterial signal-

transduction protein families: implication in the adaptive response to phosphate limitation. Mol Microbiol 8:81–91.

Aller RC (1998) Mobile deltaic and continental shelf muds as suboxic, fluidized bed reactors. Mar Chem 61:143–155. doi: 10.1016/S0304-4203(98)00024-3

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. doi: 10.1186/gb-2010-11-10-r106

Anders S, Pyl PT, Huber W (2014) HTSeq A Python framework to work with high-throughput sequencing data. 0–5. doi: 10.1101/002824

Arbiol C, Comeau AM, Kutateladze M, et al. (2010) Mobile regulatory cassettes mediate modular shuffling in T4-type phage genomes. Genome Biol Evol 2:140–52. doi: 10.1093/gbe/evq006

Artsimovitch I, Patlan V, Sekine S, et al. (2004) Structural Basis for Transcription Regulation by Alarmone ppGpp. 117:299–310.

Asayama M, Kato H, Shibato J, et al. (2002) The curved DNA structure in the 5'-upstream region of the light-responsive genes: its universality, binding factor and function for cyanobacterial psbA transcription. Nucleic Acids Res 30:4658–66.

Avrani S, Wurtzel O, Sharon I, et al. (2011) Genomic island variability facilitates Prochlorococcus-virus coexistence. Nature 474:604–8. doi: 10.1038/nature10172

Bailey S, Clokie MRJ, Millard A, Mann NH (2004) Cyanophage infection and photoinhibition in marine cyanobacteria. Res Microbiol 155:720–725. doi: http://dx.doi.org/10.1016/j.resmic.2004.06.002

Bailey S, Melis A, Mackey KRM, et al. (2008) Alternative photosynthetic electron flow to oxygen in marine Synechococcus. Biochim Biophys Acta 1777:269–76. doi: 10.1016/j.bbabio.2008.01.002

Bailey TL, Boden M, Buske F a, et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–8. doi: 10.1093/nar/gkp335

Bailleul B, Cardol P, Breyton C, Finazzi G (2010) Electrochromism: a useful probe to study algal photosynthesis. Photosynth Res 106:179–89. doi: 10.1007/s11120-010-9579-z

Bailly-Bechet M, Vergassola M, Rocha E (2007) Causes for the intriguing presence of tRNAs in phages. Genome Res 17:1486–95. doi: 10.1101/gr.6649807

Barber J (2008) Photosynthetic generation of oxygen. Philos Trans R Soc Lond B Biol Sci 363:2665–74. doi: 10.1098/rstb.2008.0047

Barik S, Ghosh B, Whalen W, et al. (1987) An antitermination protein engages the elongating transcription apparatus at a promoter-proximal recognition site. Cell 50:885–99.

Barker M, de Vries R, Nield J, et al. (2006) The deg proteases protect Synechocystis sp. PCC 6803 during heat and light stresses but are not essential for removal of damaged

D1 protein during the photosystem two repair cycle. J Biol Chem 281:30347–55. doi: 10.1074/jbc.M601064200

Battchikova N, Eisenhut M, Aro E-M (2011) Cyanobacterial NDH-1 complexes: Novel insights and remaining puzzles. Biochim Biophys Acta -Bioenergetics 1807:935–944.

Baudoux A, Veldhuis M, Noordeloos A, et al. (2008) Estimates of virus- vs. grazing induced mortality of picophytoplankton in the North Sea during summer. Aquat Microb Ecol 52:69–82. doi: 10.3354/ame01207

Behrenfeld MJ (1999) Widespread Iron Limitation of Phytoplankton in the South Pacific Ocean. Science (80-) 283:840–843. doi: 10.1126/science.283.5403.840

Behrenfeld MJ, Boss E, Siegel D a., Shea DM (2005) Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochem Cycles 19:n/a–n/a. doi: 10.1029/2004GB002299

Behrenfeld MJ, O'Malley RT, Siegel D a, et al. (2006a) Climate-driven trends in contemporary ocean productivity. Nature 444:752–5. doi: 10.1038/nature05317

Behrenfeld MJ, Worthington K, Sherrell RM, et al. (2006b) Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics. Nature 442:1025–8. doi: 10.1038/nature05083

Béjà O, Fridman S, Glaser F (2012) Viral clones from the GOS expedition with an unusual photosystem-I gene cassette organization. ISME J 6:1617–20. doi: 10.1038/ismej.2012.23

Belfort M (1990) Phage T4 Introns: Sefl-Spling and Mobility. Annu Rev Genet 24:363–385.

Belfort M (1989) Bacteriophage introns:parasites within parasites? Trends Genet 5:209–213.

Belfort M, Derbyshire V, Parker MM, et al. (2002) Mobile introns: pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mob. DNA II. pp 761–783

Bender M, Sowers T (1994) The Dole effect and its variations during the last 130,000 years as measured in the Vostok ice core. Global Biogeochem Cycles 8:363–376.

Ben-Shem A, Frolow F, Nelson N (2003) Crystal structure of plant photosystem I. Nature 426:630–5. doi: 10.1038/nature02200

Berthoumieux S, de Jong H, Baptist G, et al. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. Mol Syst Biol 9:634. doi: 10.1038/msb.2012.70

Bhaya D, Bianco NR, Bryant D, Grossman a (2000) Type IV pilus biogenesis and motility in the cyanobacterium Synechocystis sp. PCC6803. Mol Microbiol 37:941–51.

Bhaya D, Watanabe N, Ogawa T, Grossman a R (1999) The role of an alternative sigma factor in motility and pilus formation in the cyanobacterium Synechocystis sp. strain PCC6803. Proc Natl Acad Sci U S A 96:3188–93.

Bidle KD, Vardi A (2011) A chemical arms race at sea mediates algal host-virus interactions. Curr Opin Microbiol 14:449–57. doi: 10.1016/j.mib.2011.07.013

Blankenship RE (2001) Molecular evidence for the evolution of photosynthesis. Trends Plant Sci 6:4–6.

Blankenship RE (2010) Early evolution of photosynthesis. Plant Physiol 154:434–8. doi: 10.1104/pp.110.161687

Blot N, Mella-Flores D, Six C, et al. (2011) Light history influences the response of the marine cyanobacterium Synechococcus sp. WH7803 to oxidative stress. Plant Physiol 156:1934–54. doi: 10.1104/pp.111.174714

Blower TR, Pei XY, Short FL, et al. (2011) A processed noncoding RNA regulates an altruistic bacterial antiviral system. Nat Struct Mol Biol 18:185–90. doi: 10.1038/nsmb.1981

Boehme J, Frischer ME, Jiang SC, et al. (1993) Viruses , Bacterioplankton , and Phytoplankton in the Southeastern Gulf of Mexico : Distribution and Contribution to Oceanic DNA Pools. Mar Ecol Prog Ser 97:1–10.

Bonocora RP, Shub D a (2009) A likely pathway for formation of mobile group I introns. Curr Biol 19:223–8. doi: 10.1016/j.cub.2009.01.033

Bonocora RP, Zeng Q, Abel E V, Shub D a (2011) A homing endonuclease and the 50-nt ribosomal bypass sequence of phage T4 constitute a mobile DNA cassette. Proc Natl Acad Sci U S A 108:16351–6. doi: 10.1073/pnas.1107633108

Borbely G, Kaki C, Gulyas A, Farkas GL (1980) Bacteriophage infection interferes with guanosine accumulation induced by energy and nitrogen starvation in the cyanobacterium Bacteriophage Infection Interferes with Guanosine 3 ' - Diphosphate-5 ' -Diphosphate Accumulation Induced by Energy and Nitrogen. J Bacteriol 144:859–864.

Bornefeld T, Simonis W (1974) Effects of Light, Temperature, pH, and Inhibitors on the ATP Level of the Blue-Green Alga Anacystis nidulans. Planta 115:

Bouman H, Ulloa O, Scanlan DJ, et al. (2006) Oceanographic basis of the global surface distribution of Prochlorococcus ecotypes. Science 312:918–21. doi: 10.1126/science.1122692

Bragg JG, Chisholm SW (2008) Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. PLoS One 3:e3550. doi: 10.1371/journal.pone.0003550

Brahamsha B (1996) A genetic manipulation system for oceanic cyanobacteria of the genus Synechococcus . Appl Environ Microbiol 62:1741–1751.

Brantl S (2002) Antisense-RNA regulation and RNA interference. Biochim Biophys Acta 1575:15–25.

Breitbart M (2012) Marine viruses: truth or dare. Ann Rev Mar Sci 4:425–48. doi: 10.1146/annurev-marine-120709-142805

Breitbart M, Thompson LR, Suttle CA, Sullivan MB (2007) Exploring the Vast Diversity of Marine Viruses. 135–139.

Bretaudeau A, Coste F, Humily F, et al. (2013) CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. Nucleic Acids Res 41:D396–401. doi: 10.1093/nar/gks1091

Bruyant F, Babin M, Sciandra A, et al. (2001) An axenic cyclostat of Prochlorococcus PCC 9511 with a simulator of natural light regimes. 135–142.

Bruyant F, Babin M, Sud P (2005) Diel variations in the photosynthetic parameters of Prochlorococcus strain PCC 9511 : Combined effects of light and cell cycle. 50:850–863.

Bryan MJ, Burroughs NJ, Spence EM, et al. (2008) Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. PLoS One 3:e2048. doi: 10.1371/journal.pone.0002048

Buist G, Steen A, Kok J, Kuipers OP (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol 68:838–47. doi: 10.1111/j.1365-2958.2008.06211.x

Burr T, Mitchell J, Kolb A, et al. (2000) DNA sequence elements located immediately upstream of the – 10 hexamer in Escherichia coli promoters : a systematic study. Nucleic Acids Res 28:1864–1870.

Busch A, Richter AS, Backofen R, et al. (2008) IntaRNA : efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. 24:2849–2856. doi: 10.1093/bioinformatics/btn544

Busch AWU, Reijerse EJ, Lubitz W, et al. (2011) Radical mechanism of cyanophage phycoerythrobilin synthase (PebS). Biochem J 433:469–76. doi: 10.1042/BJ20101642

Bustin S a, Benes V, Garson J a, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55:611–22. doi: 10.1373/clinchem.2008.112797

Callen BP, Shearwin KE, Egan JB (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol Cell 14:647–56. doi: 10.1016/j.molcel.2004.05.010

Cambray G, Guimaraes JC, Mutalik VK, et al. (2013) Measurement and modeling of intrinsic transcription terminators. Nucleic Acids Res 41:5139–48. doi: 10.1093/nar/gkt163

Campbell A (2003) Prophage insertion sites. Res Microbiol 154:277–82. doi: 10.1016/S0923-2508(03)00071-8

Catalão MJ, Gil F, Moniz-Pereira J, et al. (2013) Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiol Rev 37:554–71. doi: 10.1111/1574-6976.12006

Cech TR (1990) E � F.

Chang C, Nam K (1995) S gene expression and the timing of lysis by bacteriophage lambda . These include : S Gene Expression and the Timing of Lysis by Bacteriophage . 177:

Chang FN, Sih CJ, Wesiblum B (1966) Lincomycin, An Inhibitor of Aminoacyl sRNA Binding to Ribosomes. Proc Natl Acad Sci 55:431–438.

Chatterjee A, Johnson CM, Shu C-C, et al. (2011) Convergent transcription confers a bistable switch in Enterococcus faecalis conjugation. Proc Natl Acad Sci U S A 108:9721–6. doi: 10.1073/pnas.1101569108

Chen Z, Schneider TD (2005) Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases. Nucleic Acids Res 33:6172–87. doi: 10.1093/nar/gki915

Chénard C, Suttle C a (2008) Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. Appl Environ Microbiol 74:5317–24. doi: 10.1128/AEM.02480-07

Chitnis VP, Chitnis PR (1993) PsaL subunit is required for the formation of photosystem I trimers in the cyanobacterium Synechocystis sp. PCC 6803. FEBS Lett 336:330–4.

Chow C-ET, Fuhrman J a (2012) Seasonality and monthly dynamics of marine myovirus communities. Environ Microbiol 14:2171–83. doi: 10.1111/j.1462-2920.2012.02744.x

Cianciotto NP (2005) Type II secretion: a protein secretion system for all seasons. Trends Microbiol 13:581–8. doi: 10.1016/j.tim.2005.09.005

Clasen J, Hanson C, Ibrahim Y, et al. (2013) Diversity and temporal dynamics of Southern California coastal marine cyanophage isolates. Aquat Microb Ecol 69:17–31. doi: 10.3354/ame01613

Clokie MRJ, Millard AD, Mann NH (2010) T4 genes in the marine ecosystem: studies of the T4-like cyanophages and their role in marine ecology. Virol J 7:291. doi: 10.1186/1743-422X-7-291

Clokie MRJ, Millard AD, Wilson WH, Mann NH (2003) Encapsidation of host DNA by bacteriophages infecting marine Synechococcus strains. FEMS Microbiol Ecol 46:349–52. doi: 10.1016/S0168-6496(03)00247-2

Clokie MRJ, Shan J, Bailey S, et al. (2006) Transcription of a "photosynthetic" T4-type phage during infection of a marine cyanobacterium. Environ Microbiol 8:827–35. doi: 10.1111/j.1462-2920.2005.00969.x

Clokie MRJ, Thalassinos K, Boulanger P, et al. (2008) A proteomic approach to the identification of the major virion structural proteins of the marine cyanomyovirus S-PM2. Microbiology 154:1775–82. doi: 10.1099/mic.0.2007/016261-0

Cochlan WP, Wikner J, Steward GF, Smith DC (1993) environments. 92:77–87.

Colland F, Orsini G, Brody EN, et al. (1998) The bacteriophage T4 AsiA protein: a molecular switch for sigma 70-dependent promoters. Mol Microbiol 27:819–29.

Comeau AM, Tremblay D, Moineau S, et al. (2012) Phage morphology recapitulates phylogeny: the comparative genomics of a new group of myoviruses. PLoS One 7:e40102. doi: 10.1371/journal.pone.0040102

Courtney CM, Chatterjee A (2014) cis-Antisense RNA and Transcriptional Interference : Coupled Layers of Gene Regulation. 2:1–9.

Crampton N, Bonass W a, Kirkham J, et al. (2006) Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. Nucleic Acids Res 34:5416–25. doi: 10.1093/nar/gkl668

Culley AI, Welschmeyer NA (2002) The abundance , distribution , and correlation of viruses , phytoplankton , and prokaryotes along a Pacific Ocean transect. Limnol Oceanogr 47:1508–1513.

Dammeyer T, Bagby SC, Sullivan MB, et al. (2008) Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. Curr Biol 18:442–8. doi: 10.1016/j.cub.2008.02.067

Danovaro R, Corinaldesi C, Dell'anno A, et al. (2011) Marine viruses and global climate change. FEMS Microbiol Rev 35:993–1034. doi: 10.1111/j.1574-6976.2010.00258.x

Debus RJ, Barry BA, Sithole I, et al. (1988) Directed Mutagenesis Indicates the Donor to P+680 in Photosystem II Is Tyrosine-161 of the D1 Polypeptide. 27:9071–9074.

Dedrick RM, Marinelli LJ, Newton GL, et al. (2013) Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. Mol Microbiol 88:577–89. doi: 10.1111/mmi.12210

Deng L, Gregory A, Yilmaz S, et al. (2012) Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria , as Revealed by Viral Tagging. doi: 10.1128/mBio.00373-12.Editor

Deshpande NN, Bao Y, Herrin DL (1997) Evidence for light/redox-regulated splicing of psbA pre-RNAs in Chlamydomonas chloroplasts. RNA 3:37–48.

Dornenburg JE, Devita AM, Palumbo MJ, Wade JT (2010) Widespread Antisense Transcription in Escherichia coli. doi: 10.1128/mBio.00024-10.Updated

Dreher TW, Brown N, Bozarth CS, et al. (2011) A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. Environ Microbiol 13:1858–74. doi: 10.1111/j.1462-2920.2011.02502.x

Duarte CM, Regaudie-de-Gioux A, Arrieta JM, et al. (2013) The oligotrophic ocean is heterotrophic. Ann Rev Mar Sci 5:551–69. doi: 10.1146/annurev-marine-121211-172337

Ducklow HW, Doney SC (2013) What is the metabolic state of the oligotrophic ocean? A debate. Ann Rev Mar Sci 5:525–33. doi: 10.1146/annurev-marine-121211-172331

Dühring U, Axmann IM, Hess WR, Wilde A (2006) An internal antisense RNA regulates expression of the photosynthesis gene isiA. Proc Natl Acad Sci U S A 103:7054–8. doi: 10.1073/pnas.0600927103

Elhai J, Wolk CP (1988) Conjugal transfer of DNA to cyanobacteria. Methods in Ezymology 167:747–754.

Enav H, Béjà O, Mandel-Gutfreund Y (2012) Cyanophage tRNAs may have a role in cross-infectivity of oceanic Prochlorococcus and Synechococcus hosts. ISME J 6:619–28. doi: 10.1038/ismej.2011.146

Enav H, Mandel-Gutfreund Y, Béjà O (2014) Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. Microbiome 2:9. doi: 10.1186/2049-2618-2-9

Erickson JM, Rahire M, Rochaix J (1984) entirely within the chloroplast inverted repeat. 3:2753–2762.

Eriksson J, Salih GF, Ghebramedhin H, Jansson C (2000) Deletion mutagenesis of the 5' psbA2 region in Synechocystis 6803: identification of a putative cis element involved in photoregulation. Mol Cell Biol Res Commun 3:292–8. doi: 10.1006/mcbr.2000.0227

Estrem ST, Ross W, Gaal T, et al. (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes Dev 13:2134–47.

Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320:1034–9. doi: 10.1126/science.1153213

Falkowski PG, Godfrey L V (2008) Electrons, life and the evolution of Earth's oxygen cycle. Philos Trans R Soc Lond B Biol Sci 363:2705–16. doi: 10.1098/rstb.2008.0054

Falkowski PG, Katz ME, Milligan AJ, et al. (2005) The rise of oxygen over the past 205 million years and the evolution of large placental mammals. Science 309:2202–4. doi: 10.1126/science.1116047

Falkowski PG, Raven JA (2007) Aquatic Photosynthesis, 2nd ed. Princeton University Press, Princeton

Fassioli F, Dinshaw R, Arpin PC, Scholes GD (2014) Photosynthetic light harvesting : excitons and coherence Photosynthetic light harvesting : excitons and coherence.

Field CB (1998) Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science (80-) 281:237–240. doi: 10.1126/science.281.5374.237

Filiatrault MJ, Stodghill P V, Bronstein P a, et al. (2010) Transcriptome analysis of Pseudomonas syringae identifies new genes, noncoding RNAs, and antisense activity. J Bacteriol 192:2359–72. doi: 10.1128/JB.01445-09

Fineran PC, Blower TR, Foulds IJ, et al. (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A 106:894–9. doi: 10.1073/pnas.0808832106

Flombaum P, Gallegos JL, Gordillo RA, et al. (2013) Present and future global distributions of the marine. Proc Natl Acad Sci. doi: 10.1073/pnas.1307701110

Frada M, Probert I, Allen MJ, et al. (2008) The "Cheshire Cat" escape strategy of the coccolithophore Emiliania huxleyi in response to viral infection. Proc Natl Acad Sci U S A 105:15944–9. doi: 10.1073/pnas.0807707105

Fuhrman J a (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399:541–8. doi: 10.1038/21119

Garczarek L, Dufresne A, Blot N, et al. (2008) Function and evolution of the psbA gene family in marine Synechococcus: Synechococcus sp. WH7803 as a case study. ISME J 2:937–953.

Geiduschek EP, Kassavetis G a (2010) Transcription of the T4 late genes. Virol J 7:288. doi: 10.1186/1743-422X-7-288

Genty B, Briantais J-M, Baker NR (1989) The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochim Biophys Acta - Gen Subj 990:87–92. doi: 10.1016/S0304-4165(89)80016-9

Georg J, Hess WR (2011) cis-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev 75:286–300. doi: 10.1128/MMBR.00032-10

Georg J, Voss B, Scholz I, et al. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. Mol Syst Biol 5:305. doi: 10.1038/msb.2009.63

Del Giorgio PA, Cole JJ, Cimbleris A (1997) Respiration rates in bacteria exceed phytoplankton production in unporductive aquatic systems. Nature 385:148–150.

Glazer a N, Cohen-Bazire G (1971) Subunit structure of the phycobiliproteins of blue-green algae. Proc Natl Acad Sci U S A 68:1398–401.

Groom S, Martinez-Vicente V, Fishwick J, et al. (2009) The Western English Channel observatory: Optical characteristics of station L4. J Mar Syst 77:278–295. doi: 10.1016/j.jmarsys.2007.12.015

Güell M, van Noort V, Yus E, et al. (2009) Transcriptome complexity in a genome-reduced bacterium. Science 326:1268–71. doi: 10.1126/science.1176951

Hatfull GF, Jacobs-sera D, Lawrence JG, et al. (2011) Comparative genomic analysis of sixty mycobacteriophage genomes: Genome clustering, gene acquisition and gene size. J mo 397:119–143. doi: 10.1016/j.jmb.2010.01.011.Comparative

Hausner G, Hafez M, Edgell DR (2014) Bacterial group I introns: mobile RNA catalysts. Mob DNA 5:8. doi: 10.1186/1759-8753-5-8

Haussühl K, Andersson B, Adamska I (2001) A chloroplast DegP2 protease performs the primary cleavage of the photodamaged D1 protein in plant photosystem II. EMBO J 20:713–22. doi: 10.1093/emboj/20.4.713

Havaux M (2003) Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll a/b-binding proteins results in aberrant photoacclimation in Synechocystis PCC6803. Biochim Biophys Acta - Bioenerg 1557:21–33. doi: 10.1016/S0005-2728(02)00391-2

He Q, Dolganov N, Bjorkman O, Grossman AR (2001) The high light-inducible polypeptides in Synechocystis PCC6803. Expression and function in high light. J Biol Chem 276:306–14. doi: 10.1074/jbc.M008686200

Hedges JI, Keil RG (1995) Sedimentary organic matter preservation: an assessment and speculative synthesis. Mar Chem 49:81–115. doi: 10.1016/0304-4203(95)00008-F

Hellweger FL (2009) Carrying photosynthesis genes increases ecological fitness of cyanophage in silico. Environ Microbiol 11:1386–94. doi: 10.1111/j.1462-2920.2009.01866.x

Hendrix RW (2002) Bacteriophages: Evolution of the Majority. Theor Popul Biol 61:471–480. doi: 10.1006/tpbi.2002.1590

Hendrix RW, M.C.M. S, Burns RN, et al. (1999) Evolutionary relationships among diverse bacteriophages and prophages : All the world ' s a phage. Proc Natl Acad Sci 96:2192–2197.

Hernández J a, Muro-Pastor AM, Flores E, et al. (2006) Identification of a furA cis antisense RNA in the cyanobacterium Anabaena sp. PCC 7120. J Mol Biol 355:325–34. doi: 10.1016/j.jmb.2005.10.079

Hernandez-Prieto MA, Tibiletti T, Abasova L, et al. (2011) The small CAB-like proteins of the cyanobacterium Synechocystis sp. PCC 6803: their involvement in chlorophyll biogenesis for Photosystem II. Biochim Biophys Acta 1807:1143–51. doi: 10.1016/j.bbabio.2011.05.002

Herranen M, Aro E-M, Tyystjärvi T (2001) Two distinct mechanisms regulate the transcription of photosystem II genes in Synechocystis sp. PCC 6803. Physiol Plant 112:531–539.

Hinton DM, Vuthoori S (2000) Efficient inhibition of Escherichia coli RNA polymerase by the bacteriophage T4 AsiA protein requires that AsiA binds first to free sigma70. J Mol Biol 304:731–739.

Hirani T a, Suzuki I, Murata N, et al. (2001) Characterization of a two-component signal transduction system involved in the induction of alkaline phosphatase under phosphate-limiting conditions in Synechocystis sp. PCC 6803. Plant Mol Biol 45:133–44.

Holland HD (2006) The oxygenation of the atmosphere and oceans. Philos Trans R Soc Lond B Biol Sci 361:903–15. doi: 10.1098/rstb.2006.1838

Horvitz HR (1973) Polypeptide bound to the host RNA polymerase is specified by T4 control gene 33. Nat New Biol 244:137–140.

Huang S, Wang K, Jiao N, Chen F (2012) Genome sequences of siphoviruses infecting marine Synechococcus unveil a diverse cyanophage group and extensive phage-host genetic exchanges. Environ Microbiol 14:540–58. doi: 10.1111/j.1462-2920.2011.02667.x

Huang WE, Stoecker K, Griffiths R, et al. (2007) Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis

of identity and function. Environ Microbiol 9:1878–89. doi: 10.1111/j.1462-2920.2007.01352.x

Hurwitz BL, Hallam SJ, Sullivan MB (2013) Metabolic reprogramming by viruses in the sunlit and dark ocean. Genome Biol 14:R123. doi: 10.1186/gb-2013-14-11-r123

Hyman P, Abedon ST (2009) Practical Methods for Determining Phage Growth Parameters. In: Clokie MRJ, Kropinski AM (eds) Bacteriophages Methods Protoc. Vol. 1 Isol. Charact. Interact., 1st ed. Humana Press, pp 175–202

Ignacio-Espinoza JC, Sullivan MB (2012) Phylogenomics of T4 cyanophages: lateral gene transfer in the "core" and origins of host genes. Environ Microbiol 14:2113–26. doi: 10.1111/j.1462-2920.2012.02704.x

Imamura S, Asayama M, Shirai M (2004) In vitro transcription analysis by reconstituted cyanobacterial RNA polymerase: roles of group 1 and 2 sigma factors and a core subunit, RpoC2. Genes Cells 9:1175–87. doi: 10.1111/j.1365-2443.2004.00808.x

Imamura S, Yoshihara S, Nakano S, et al. (2003) Purification, Characterization, and Gene Expression of All Sigma Factors of RNA Polymerase in a Cyanobacterium. J Mol Biol 325:857–872. doi: 10.1016/S0022-2836(02)01242-1

Jacquet S, Heldal M, Iglesias-Rodriguez D, et al. (2002) Flow cytometric analysis of an Emiliana huxleyi bloom terminated by viral infection. Aquat Microb Ecol 27:111–124. doi: 10.3354/ame027111

Jardillier L, Zubkov M V, Pearman J, Scanlan DJ (2010) Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. ISME J 4:1180–92. doi: 10.1038/ismej.2010.36

Jia Y, Shan J, Millard A, et al. (2010) Light-dependent adsorption of photosynthetic cyanophages to Synechococcus sp. WH7803. FEMS Microbiol Lett 310:120–6. doi: 10.1111/j.1574-6968.2010.02054.x

Johnson ZI, Zinser ER, Coe A, et al. (2006) Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. Science 311:1737–40. doi: 10.1126/science.1118052

Jordan P, Fromme P, Witt HT, et al. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. Nature 411:909–17. doi: 10.1038/35082000

Kallas T, Castenholz RW (1982) Internal pH and ATP-ADP pools in the cyanobacterium Synechococcus sp. during exposure to growth-inhibiting low pH. J Bacteriol 149:229–36.

Kana TM, Glibert PM (2000) RAPID RESPONSE PAPER Effect of irradiances up to 2000 laE m -2 s -l on marine. 34:

Kassavetis GA, Zentner PG, Geiduschek EP (1986) Transcription at Bacteriophage T4 Variant Late Promoters. 14256–14265.

Kawano M, Aravind L, Storz G (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol Microbiol 64:738–54. doi: 10.1111/j.1365-2958.2007.05688.x

Kimura S, Yoshida T, Hosoda N, et al. (2012) Diurnal infection patterns and impact of Microcystis cyanophages in a Japanese pond. Appl Environ Microbiol 78:5805–11. doi: 10.1128/AEM.00571-12

Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. Cell 139:1366–75. doi: 10.1016/j.cell.2009.12.001

Koch T, Raudonikiene A, Wilkens K, Rüger W (1995) Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of E. coli RNA polymerase modulates T4 "early" transcription. Gene Expr 4:253–264.

Kolesky SE, Ouhammouch M, Peter Geiduschek E (2002) The Mechanism of Transcriptional Activation by the Topologically DNA-linked Sliding Clamp of Bacteriophage T4. J Mol Biol 321:767–784. doi: 10.1016/S0022-2836(02)00732-5

Komenda J, Hassan HAG, Diner BA, et al. (2000) Degradation of the Photosystem II D1 and D2 proteins in different strains of the cyanobacterium Synechocystis PCC 6803 varying with respect to the type and level of psbA transcript. 635–645.

Komenda J, Tichy M, Prásil O, et al. (2007) The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during photosystem II repair in Synechocystis sp PCC 6803. Plant Cell 19:2839–54. doi: 10.1105/tpc.107.053868

Kozloff LM, Knowlton K, Putnam W, Evans EA (1951) ARTICLE : BIOCHEMICAL STUDIES OF VIRUS REPRODUCTION : V . THE ORIGIN OF BACTERIOPHAGE NITROGEN.

Krinke L, Wulff DL (1990) RNase III-dependent hydrolysis of lambda cII-O gene mRNA mediated by lambda OOP antisense RNA. Genes Dev 4:2223–2233. doi: 10.1101/gad.4.12a.2223

Kulkarni RD, Golden SS (1994) Adaptation to high light intensity in Synechococcus sp. strain PCC 7942: regulation of three psbA genes and two forms of the D1 protein. J Bacteriol 176:959–65.

Kurisu G, Zhang H, Smith JL, Cramer W a (2003) Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. Science 302:1009–14. doi: 10.1126/science.1090165

Kutsuna S, Kondo T, Aoki S (1998) A Period-Extender Gene , pex , That Extends the Period of the Circadian Clock in the Cyanobacterium Synechococcus sp . Strain A Period-Extender Gene , pex , That Extends the Period of the Circadian Clock in the Cyanobacterium Synechococcus sp . Strain PCC.

Kuty GF, Xu M, Struck DK, et al. (2010) Regulation of a phage endolysin by disulfide caging. J Bacteriol 192:5682–7. doi: 10.1128/JB.00674-10

Labrie SJ, Frois-Moniz K, Osburne MS, et al. (2013) Genomes of marine cyanopodoviruses reveal multiple origins of diversity. Environ Microbiol 15:1356–76. doi: 10.1111/1462-2920.12053

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

Lasa I, Toledo-arana A, Dobin A, et al. (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. doi: 10.1073/pnas.1113521108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1113521108

Lasa I, Toledo-Arana A, Gingeras TR (2012) An effort to make sense of antisense transcription in bacteria. RNA Biol 9:1039–44. doi: 10.4161/rna.21167

Lawrenz E, Silsbe G, Capuzzo E, et al. (2013) Predicting the electron requirement for carbon fixation in seas and oceans. PLoS One 8:e58137. doi: 10.1371/journal.pone.0058137

Lee SK, Keasling JD (2005) A Propionate-Inducible Expression System for Enteric Bacteria A Propionate-Inducible Expression System for Enteric Bacteria. doi: 10.1128/AEM.71.11.6856

Legewie S, Dienst D, Wilde A, et al. (2008) Small RNAs establish delays and temporal thresholds in gene expression. Biophys J 95:3232–8. doi: 10.1529/biophysj.108.133819

Lennon JT, Khatana SAM, Marston MF, Martiny JBH (2007) Is there a cost of virus resistance in marine cyanobacteria? ISME J 1:300–12. doi: 10.1038/ismej.2007.37

Li F, Holloway SP, Lee J, Herrin DL (2002) Nuclear genes that promote splicing of group I introns in the chloroplast 23S rRNA and psbA genes in Chlamydomonas reinhardtii. Plant J 32:467–80.

Li H, Handsaker B, Wysoker A, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–9. doi: 10.1093/bioinformatics/btp352

Li H, Sherman L a (2000) A redox-responsive regulator of photosynthesis gene expression in the cyanobacterium Synechocystis sp. Strain PCC 6803. J Bacteriol 182:4268–77.

Li R, Golden SS (1993) Enhancer activity of light-responsive regulatory elements in the untranslated leader regions of cyanobacterial psbA genes. Proc Natl Acad Sci U S A 90:11678–82.

Lindell D, Jaffe JD, Coleman ML, et al. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449:83–6. doi: 10.1038/nature06130

Lindell D, Jaffe JD, Johnson ZI, et al. (2005) Photosynthesis genes in marine viruses yield proteins during host infection. Nature 438:86–9. doi: 10.1038/nature04111

Lindell D, Sullivan MB, Johnson ZI, et al. (2004a) Transfer of photosynthesis genes to and from Prochlorococcus viruses. Proc Natl Acad Sci U S A 101:11013–8. doi: 10.1073/pnas.0401526101

Lindell D, Sullivan MB, Johnson ZI, et al. (2004b) Transfer of photosynthesis genes to and from Prochlorococcus viruses. Proc Natl Acad Sci U S A 101:11013–8. doi: 10.1073/pnas.0401526101

Liu H, Campbell L, Landry MR (1994) Growth and mortality rates of Prochlorococcus and Synechococcus measured with a selective inhibitor techniaue.

Logemann J, Schell J, Willmitzer L (1987) Improved Method for the Isolation of RNA from Plant Tissues. Anal Biochem 163:16–20.

Lonetto M, Gribskov M, Gross C a (1992) The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol 174:3843–9.

Longhurst A, Sathyendranath S, Platt T, Caverhill C (1995) An estimate of global primary production in the ocean from satellite radiometer data. J Plankton Res 17:1245–1271. doi: 10.1093/plankt/17.6.1245

Lubberding HJ, Schroten W (1984) The ATP level in the thermophilic cyanobacterium Synechococcus 6716 during light-dark transition and in the presence of some specific inhibitors. FEMS Microbiol Lett 22:93–96.

Luke K, Radek A, Liu X, et al. (2002) Microarray Analysis of Gene Expression during Bacteriophage T4 Infection. Virology 299:182–191. doi: 10.1006/viro.2002.1409

Luz P, Lavigne R, Lecoutere E, Wagemans J (2013) A Multifaceted Study of Pseudomonas aeruginosa Shutdown by. doi: 10.1128/mBio.00061-13.Editor

Lybecker M, Zimmermann B, Bilusic I, et al. (2014) The double-stranded transcriptome of Escherichia coli. Proc Natl Acad Sci U S A 111:3134–9. doi: 10.1073/pnas.1315974111

MacIntyre HL, Kana TM, Geider RJ (2000) The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. Trends Plant Sci 5:12–7.

Magnusson LU, Farewell A, Nyström T (2005) ppGpp: a global regulator in Escherichia coli. Trends Microbiol 13:236–42. doi: 10.1016/j.tim.2005.03.008

Mann NH, Carr NG, Midgley JEM (1975) RNA synthesis and accumulation of guanine nucleotides during growth shift down in the blue-green alga Anacystis nidulans. Biochim Biophys Acta 402:41–50.

Mann NH, Clokie MRJ, Millard A, et al. (2005) The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. J Bacteriol 187:3188–200. doi: 10.1128/JB.187.9.3188-3200.2005

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

Marchant H, Davidsonl A, Wright S, Glazebrook J (2000) The distribution and abundance of viruses in the Southern Ocean during spring. 12:4–7.

Marston MF, Pierciey FJ, Shepard A, et al. (2012) Rapid diversification of coevolving marine Synechococcus and a virus. Proc Natl Acad Sci U S A 109:4544–9. doi: 10.1073/pnas.1120310109

Marston MF, Sallee JL (2003) Genetic Diversity and Temporal Variation in the Cyanophage Community Infecting Marine Synechococcus Species in Rhode Island ' s Coastal Waters Genetic Diversity and Temporal Variation in the Cyanophage Community Infecting Marine Synechococcus Species in R. doi: 10.1128/AEM.69.8.4639

Martin AP, Zubkov M V, Burkill PH, Holland RJ (2005) Extreme spatial variability in marine picoplankton and its consequences for interpreting Eulerian time-series. Biol Lett 1:366–9. doi: 10.1098/rsbl.2005.0316

Martiny JBH, Riemann L, Marston MF, Middelboe M (2014) Antagonistic coevolution of marine planktonic viruses and their hosts. Ann Rev Mar Sci 6:393–414. doi: 10.1146/annurev-marine-010213-135108

Mazor Y, Greenberg I, Toporik H, et al. (2012) The evolution of photosystem I in light of phage-encoded reaction centres. Philos Trans R Soc Lond B Biol Sci 367:3400–5. doi: 10.1098/rstb.2012.0057

Mazor Y, Nataf D, Toporik H, Nelson N (2014) Crystal structures of virus-like photosystem I complexes from the mesophilic cyanobacterium Synechocystis PCC 6803. Elife 3:e01496. doi: 10.7554/eLife.01496

McClure R, Balasubramanian D, Sun Y, et al. (2013) Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res 41:e140. doi: 10.1093/nar/gkt444

Mella-Flores D, Six C, Ratin M, et al. (2012) Prochlorococcus and Synechococcus have Evolved Different Adaptive Mechanisms to Cope with Light and UV Stress. Front Microbiol 3:285. doi: 10.3389/fmicb.2012.00285

Mendoza-Vargas A, Olvera L, Olvera M, et al. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. PLoS One 4:e7526. doi: 10.1371/journal.pone.0007526

Millard A, Clokie MRJ, Shub DA, Mann NH (2004) Genetic organization of the psbAD region in phages infecting marine Synechococcus strains. Proc Natl Acad Sci U S A 101:11007–12. doi: 10.1073/pnas.0401478101

Millard AD (2009) Bacteriophages. 501:33–42. doi: 10.1007/978-1-60327-164-6

Millard AD, Gierga G, Clokie MRJ, et al. (2010) An antisense RNA in a lytic cyanophage links psbA to a gene encoding a homing endonuclease. ISME J 4:1121–35. doi: 10.1038/ismej.2010.43

Millard AD, Mann NH, O ADM (2006) Publisher statement : None A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba , Red Sea.

Millard AD, Zwirglmaier K, Downey MJ, et al. (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. Environ Microbiol 11:2370–87. doi: 10.1111/j.1462-2920.2009.01966.x

Miller ES, Kutter E, Mosig G, et al. (2003) Bacteriophage T4 genome. Microbiol Mol Biol Rev 67:86–156. doi: 10.1128/MMBR.67.1.86

Minakhin L, Niedziela-Majka A, Kuznedelov K, et al. (2003) Interaction of T4 AsiA with its Target Sites in the RNA Polymerase σ70 Subunit Leads to Distinct and Opposite Effects on Transcription. J Mol Biol 326:679–690. doi: 10.1016/S0022-2836(02)01442-0

Mitchell P (1961) Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic type of Mechanism. Nature 191:144–148.

Mitschke J, Georg J, Scholz I, et al. (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803. Proc Natl Acad Sci 108 :2124–2129. doi: 10.1073/pnas.1015154108

Mohamed a, Eriksson J, Osiewacz HD, Jansson C (1993) Differential expression of the psbA genes in the cyanobacterium Synechocystis 6803. Mol Gen Genet 238:161–8.

Møller T, Franch T, Højrup P, et al. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 9:23–30.

Moore CM, Mills MM, Arrigo KR, et al. (2013) Processes and patterns of oceanic nutrient limitation. Nat Geosci 6:701–710. doi: 10.1038/ngeo1765

Moussa SH, Kuznetsov V, Tran TAT, et al. (2012) Protein determinants of phage T4 lysis inhibition. Protein Sci 21:571–82. doi: 10.1002/pro.2042

Mühling M, Fuller NJ, Somerfield PJ, et al. (2005) Genetic diversity of marine Synechococcus and co-occurring cyanophage communities : evidence for viral control of phytoplankton. 7:499–508. doi: 10.1111/j.1462-2920.2004.00713.x

Mulo P, Laakso S, Mäenpää P, Aro EM (1998) Stepwise photoinhibition of photosystem II. Studies with Synechocystis species PCC 6803 mutants with a modified D-E loop of the reaction center polypeptide D1. Plant Physiol 117:483–90.

Mulo P, Sakurai I, Aro E-M (2012) Strategies for psbA gene expression in cyanobacteria, green algae and higher plants: From transcription to PSII repair. Biochim Biophys Acta - Bioenerg 1817:247–257. doi: 10.1016/j.bbabio.2011.04.011

Mulo P, Sicora C, Aro E-M (2009) Cyanobacterial psbA gene family: optimization of oxygenic photosynthesis. Cell Mol Life Sci 66:3697–710. doi: 10.1007/s00018-009-0103-6

Mulo P, Tyystjärvi T, Tyystjärvi E, et al. (1997) Mutagenesis of the D-E loop of photosystem II reaction centre protein D1. Function and assembly of photosystem II. Plant Mol Biol 33:1059–71.

Naithani S, Hou JM, Chitnis PR (2000) Targeted inactivation of the psaK1, psaK2 and psaM genes encoding subunits of Photosystem I in the cyanobacterium Synechocystis sp. PCC 6803. Photosynth Res 63:225–36. doi: 10.1023/A:1006463932538

Navarro JA, Hervás M, Rosa MA De (1997) Co-evolution of cytochrome c 6 and plastocyanin , mobile proteins transferring electrons from cytochrome b 6 f to photosystem I. 11–22.

Nechaev S, Geiduschek EP (2008) Dissection of the bacteriophage T4 late promoter complex. J Mol Biol 379:402–13. doi: 10.1016/j.jmb.2008.03.071

Nicolas P, Mäder U, Dervyn E, et al. (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. Science 335:1103–6. doi: 10.1126/science.1206848

Nixon PJ, Barker M, Boehm M, et al. (2005) FtsH-mediated repair of the photosystem II complex in response to light stress. J Exp Bot 56:357–63. doi: 10.1093/jxb/eri021

Nixon PJ, Michoux F, Yu J, et al. (2010) Recent advances in understanding the assembly and repair of photosystem II. Ann Bot 106:1–16. doi: 10.1093/aob/mcq059

Nomura M, Matsubara K, Okamoto K, Fujimura R (1962) Inhibition of host nucleic acid and protein synthesis by bacteriophage T4: Its relation to the physical and functional integrity of host chromosome. J Mol Biol 5:535–549. doi: 10.1016/S0022-2836(62)80127-2

Oppenheim AB, Kobiler O, Stavans J, et al. (2005) Switches in bacteriophage lambda development. Annu Rev Genet 39:409–29. doi: 10.1146/annurev.genet.39.073003.113656

Palmer AC, Ahlgren-Berg A, Egan JB, et al. (2009) Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter. Mol Cell 34:545–55. doi: 10.1016/j.molcel.2009.04.018

Parsons RJ, Breitbart M, Lomas MW, Carlson CA (2012) Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. ISME J 6:273–84. doi: 10.1038/ismej.2011.101

Parsons TR, Maita Y, Lalli CM (1984) A Manual of Chemical and Biological Methods for Seawater Analysis.

Partensky F, Hess WR, Vaulot D (1999) Prochlorococcus, a Marine Photosynthetic Prokaryote of Global Significance. Microbiol Mol Biol Rev 63:106–127.

Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Res 35:e128. doi: 10.1093/nar/gkm683

Pfaffl MW (2004) Relative quantification. 63–82.

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30:e36.

Philosof A, Battchikova N, Aro E-M, Béjà O (2011) Marine cyanophages: tinkering with the electron transport chain. ISME J 5:1568–70. doi: 10.1038/ismej.2011.43

Platt T, Gallegos CL, Harrison WG (1980) Photoinhibition of photosynthesis in natural assemblages of marine phytoplankton. J Mar Res 38:687–701.

Pollari M, Ruotsalainen V, Rantamäki S, et al. (2009) Simultaneous inactivation of sigma factors B and D interferes with light acclimation of the cyanobacterium Synechocystis sp. strain PCC 6803. J Bacteriol 191:3992–4001. doi: 10.1128/JB.00132-09

Polovina JJ, Howell E a., Abecassis M (2008) Ocean's least productive waters are expanding. Geophys Res Lett 35:L03618. doi: 10.1029/2007GL031745

Prasil O, Kolber Z, Berry JA, Falkowski PG (1996) Cyclic electron flow around Photosystem II in vivo. Photosynth Res 48:395–410.

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

Puxty RJ, Millard AD, Ph D (2014) Shedding new light on viral photosynthesis. Photosynth. Res.

Rabinovitch A, Hadas H, Einav M, et al. (1999) Model for Bacteriophage T4 Development in Escherichia coli. J Bacteriol 181:1677–1683.

Raghavan R, Minnick MF (2009) Group I Introns and Inteins: Disparate Origins but Convergent Parasitic Strategies. J Bacteriol 191:6193–6202. doi: 10.1128/JB.00675-09

Rajagopala S V, Casjens S, Uetz P (2011) The protein interaction map of bacteriophage lambda. BMC Microbiol 11:213. doi: 10.1186/1471-2180-11-213

Ratner D (1974) Bacteriophage T4 transcriptional control gene 55 codes for a protein bound to Escherichia coli RNA polymerase. J Mol Biol 89:803–807.

Riva S, Cascino A, Geiduschek EP (1970) Coupling of late transcription to viral replication in bacteriophage T4 development. J Mol Biol 54:85–102.

Ross W, Aiyar SE, Salomon J, Gourse RL (1998) Escherichia coli promoters with UP elements of different strengths: modular structure of bacterial promoters. J Bacteriol 180:5375–83.

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. doi: 10.1016/0377-0427(87)90125-7

Ruggeri M (1993) Iii.

Russel M (1998) Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. J Mol Biol 279:485–99. doi: 10.1006/jmbi.1998.1791

Sabehi G, Shaulov L, Silver DH, et al. (2012) A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. Proc Natl Acad Sci U S A 109:2037–42. doi: 10.1073/pnas.1115467109

Sakurai I, Stazic D, Eisenhut M, et al. (2012) Positive regulation of psbA gene expression by cis-encoded antisense RNAs in Synechocystis sp. PCC 6803. Plant Physiol 160:1000–10. doi: 10.1104/pp.112.202127

Sander J, Nowaczyk M, Buchta J, et al. (2010) Functional characterization and quantification of the alternative PsbA copies in Thermosynechococcus elongatus and their role in photoprotection. J Biol Chem 285:29851–6. doi: 10.1074/jbc.M110.127142

Scanlan DJ, Ostrowski M, Mazard S, et al. (2009) Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev 73:249–99. doi: 10.1128/MMBR.00035-08

Scanlan DJ, West NJ (2002) Molecular ecology of the marine cyanobacterial genera Prochlorococcus and Synechococcus. FEMS Microbiol Ecol 40:1–12. doi: 10.1111/j.1574-6941.2002.tb00930.x

Schmeing TM, Huang KS, Strobel S a, Steitz T a (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. Nature 438:520–4. doi: 10.1038/nature04152

Shan J (2008) An investigation into the effect of cyanophage infection on photosynthetic antenna. University of Warwick, UK

Shan J, Jia Y, Clokie MRJ, Mann NH (2008) Infection by the "photosynthetic" phage S-PM2 induces increased synthesis of phycoerythrin in Synechococcus sp. WH7803. FEMS Microbiol Lett 283:154–61. doi: 10.1111/j.1574-6968.2008.01148.x

Sharma CM, Hoffmann S, Darfeuille F, et al. (2010a) The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464:250–255.

Sharma CM, Hoffmann S, Darfeuille F, et al. (2010b) The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464:250–5. doi: 10.1038/nature08756

Sharon I, Alperovitch A, Rohwer F, et al. (2009) Photosystem I gene cassettes are present in marine virus genomes. Nature 461:258–62. doi: 10.1038/nature08284

Sharon I, Battchikova N, Aro E-M, et al. (2011) Comparative metagenomics of microbial traits within oceanic viral communities. ISME J 5:1178–90. doi: 10.1038/ismej.2011.2

Shibato J, Asayama M, Shirai M (1998) Promoter paper Speci ¢ c recognition of the cyanobacterial psbA promoter by RNA polymerases containing principal sigma factors. 1442:296–303.

Short FL, Pei XY, Blower TR, et al. (2013) Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. Proc Natl Acad Sci U S A 110:E241–9. doi: 10.1073/pnas.1216039110

Siegel DA, Buesseler KO, Doney SC, et al. (2014) Global Biogeochemical Cycles. Global Biogeochem Cycles 28:181–196. doi: 10.1002/2013GB004743.Received

Silva P, Thompson E, Bailey S, et al. (2003) FtsH Is Involved in the Early Stages of Repair of Photosystem II in Synechocystis sp PCC 6803. Plant Cell 15:2152–2164. doi: 10.1105/tpc.012609.tivity

Singh SS, Grainger DC (2013) H-NS can facilitate specific DNA-binding by RNA polymerase in AT-rich gene regulatory regions. PLoS Genet 9:e1003589. doi: 10.1371/journal.pgen.1003589

Singh SS, Singh N, Bonocora RP, et al. (2014) Widespread suppression of intragenic transcription initiation by H-NS. Genes Dev 28:214–9. doi: 10.1101/gad.234336.113

Sippola K, Aro EM (1999) Thiol redox state regulates expression of psbA genes in Synechococcus sp. PCC 7942. Plant Mol Biol 41:425–33.

Six C, Finkel Z V, Irwin AJ, Campbell D a (2007) Light variability illuminates niche-partitioning among marine Picocyanobacteria. PLoS One 2:e1341. doi: 10.1371/journal.pone.0001341

Smith C, Heyne S, Richter AS, et al. (2010) Freiburg RNA Tools : a web server integrating I NTA RNA , E XPA RNA and L OC ARNA. 38:373–377. doi: 10.1093/nar/gkq316

Soitamo a J, Zhou G, Clarke a K, et al. (1996) Over-production of the D1:2 protein makes Synechococcus cells more tolerant to photoinhibition of photosystem II. Plant Mol Biol 30:467–78.

Stazic D, Lindell D, Steglich C (2011) Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. Nucleic Acids Res 39:4890–9. doi: 10.1093/nar/gkr037

Steeman-Nielsen E (1952) The use of radioactive carbon (14C) for measuring organic production in the sea. J du Cons 18:117–140.

Steglich C, Futschik ME, Lindell D, et al. (2008) The challenge of regulation in a minimal photoautotroph: non-coding RNAs in Prochlorococcus. PLoS Genet 4:e1000173. doi: 10.1371/journal.pgen.1000173

Stelljes C, Koenig F (2007) Specific binding of D1 protein degradation products to the psbAI promoter in Synechococcus sp. strain PCC 7942. J Bacteriol 189:1722–6. doi: 10.1128/JB.01428-06

Stoddard LI, Martiny JBH, Marston MF (2007) Selection and characterization of cyanophage resistance in marine Synechococcus strains. Appl Environ Microbiol 73:5516–22. doi: 10.1128/AEM.00356-07

Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43:880–91. doi: 10.1016/j.molcel.2011.08.022

Stouthamer a H (1973) A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie Van Leeuwenhoek 39:545–65.

Suga M, Lai T-L, Sugiura M, et al. (2013) Crystal structure at 1.5Å resolution of the PsbV2 cytochrome from the cyanobacterium Thermosynechococcus elongatus. FEBS Lett 587:3267–72. doi: 10.1016/j.febslet.2013.08.023

Suggett DJ, Moore CM, Geider RJ (2010) Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications. doi: 10.1007/978-90-481-9268-7

Sullivan MB, Coleman ML, Quinlivan V, et al. (2008) Portal protein diversity and phage ecology. Environ Microbiol 10:2810–23. doi: 10.1111/j.1462-2920.2008.01702.x

Sullivan MB, Coleman ML, Weigele P, et al. (2005) Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol 3:e144. doi: 10.1371/journal.pbio.0030144

Sullivan MB, Huang KH, Ignacio-Espinoza JC, et al. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ Microbiol 12:3035–56. doi: 10.1111/j.1462-2920.2010.02280.x

Sullivan MB, Krastins B, Hughes JL, et al. (2009) The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial "mobilome". Environ Microbiol 11:2935–51. doi: 10.1111/j.1462-2920.2009.02081.x

Sullivan MB, Lindell D, Lee JA, et al. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol 4:e234. doi: 10.1371/journal.pbio.0040234

Sun Q, Kuty GF, Arockiasamy A, et al. (2009) Regulation of a muralytic enzyme by dynamic membrane topology. Nat Struct Mol Biol 16:1192–4. doi: 10.1038/nsmb.1681

Suttle C a (2007) Marine viruses--major players in the global ecosystem. Nat Rev Microbiol 5:801–12. doi: 10.1038/nrmicro1750

Suttle C a (2005) Viruses in the sea. Nature 437:356–61. doi: 10.1038/nature04160

Suttle CA, Chan AM (1993) Marine cyanophages infecting oceanic and coastal strains of Synechococcus: abundance, morphology, cross-infectivity and growth characteristics. Mar Ecol Prog Ser 92:99–109.

Suttle CA, Chan AM, Sutile CA, Chan AMYM (1994) Dynamics and Distribution of Cyanophages and Their Effect on Marine Synechococcus Dynamics and Distribution of Cyanophages and Their Effect on Marine Synechococcus spp . t. 60:

Tachibana Y, Vayssieres L, Durrant JR (2012) Artificial photosynthesis for solar water-splitting. doi: 10.1038/NPHOTON.2012.175

Takahashi Y, Yamaguchi O, Omata T (2004) Roles of CmpR, a LysR family transcriptional regulator, in acclimation of the cyanobacterium Synechococcus sp. strain PCC 7942 to low-CO(2) and high-light conditions. Mol Microbiol 52:837–45. doi: 10.1111/j.1365-2958.2004.04021.x

Tendeng C, Bertin PN (2003) H-NS in Gram-negative bacteria: a family of multifaceted proteins. Trends Microbiol 11:511–518. doi: 10.1016/j.tim.2003.09.005

Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 45:1320–1328. doi: 10.4319/lo.2000.45.6.1320

Thomas C (2004) PsfR, a factor that stimulates psbAI expression in the cyanobacterium Synechococcus elongatus PCC 7942. Microbiology 150:1031–1040. doi: 10.1099/mic.0.26915-0

Thompson LK, Brudvig GW (1988) Cytochrome b-559 May Function To Protect Photosystem II from Photoinhibition. Biochemistry 6653–6658.

Thompson LR, Zeng Q, Kelly L, et al. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc Natl Acad Sci 108:E757–E764. doi: 10.1073/pnas.1102164108

Tichý M, Lupínková L, Sicora C, et al. (2003) Synechocystis 6803 mutants expressing distinct forms of the Photosystem II D1 protein from Synechococcus 7942: relationship between the psbA coding region and sensitivity to visible and UV-B radiation. Biochim Biophys Acta - Bioenerg 1605:55–66. doi: 10.1016/S0005-2728(03)00064-1

Tolonen AC, Liszt GB, Hess WR (2006) Genetic manipulation of Prochlorococcus strain MIT9313: green fluorescent protein expression from an RSF1010 plasmid and Tn5 transposition. Appl Environ Microbiol 72:7607–13. doi: 10.1128/AEM.02034-06

Toulokhonov II (2000) Binding of the Transcription Effector ppGpp to Escherichia coli RNA Polymerase Is Allosteric, Modular, and Occurs Near the N Terminus of the beta '-Subunit. J Biol Chem 276:1220–1225. doi: 10.1074/jbc.M007184200

Tramonti A, De Canio M, De Biase D (2008) GadX/GadW-dependent regulation of the Escherichia coli acid fitness island: transcriptional control at the gadY-gadW divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. Mol Microbiol 70:965–82. doi: 10.1111/j.1365-2958.2008.06458.x

Trapnell C, Roberts A, Goff L, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–78. doi: 10.1038/nprot.2012.016

Traving SJ, Clokie MRJ, Middelboe M (2014) Increased acidification has a profound effect on the interactions between the cyanobacterium Synechococcus sp. WH7803 and its viruses. FEMS Microbiol Ecol 87:133–41. doi: 10.1111/1574-6941.12199

Tyystjärvi E, Aro EM (1996) The rate constant of photoinhibition, measured in lincomycin-treated leaves, is directly proportional to light intensity. Proc Natl Acad Sci U S A 93:2213–8.

Tyystjärvi T, Aro EM, Jansson C, Mäenpää P (1994) Changes of amino acid sequence in PEST-like area and QEEET motif affect degradation rate of D1 polypeptide in photosystem II. Plant Mol Biol 25:517–26.

Ueda T, Takahashi H, Uyar E, et al. (2013) Functions of the Hha and YdgT proteins in transcriptional silencing by the nucleoid proteins, H-NS and StpA, in Escherichia coli. DNA Res 20:263–71. doi: 10.1093/dnares/dst008

Umena Y, Kawakami K, Shen J-R, Kamiya N (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. Nature 473:55–60. doi: 10.1038/nature09913

Urbauer JL, Simeonov MF, Urbauer RJB, et al. (2002) Solution structure and stability of the anti-sigma factor AsiA: implications for novel functions. Proc Natl Acad Sci U S A 99:1831–5. doi: 10.1073/pnas.032464699

Vavilin D, Yao D, Vermaas W (2007) Small Cab-like proteins retard degradation of photosystem II-associated chlorophyll in Synechocystis sp. PCC 6803: kinetic analysis of pigment labeling with 15N and 13C. J Biol Chem 282:37660–8. doi: 10.1074/jbc.M707133200

Vermaas WFJ (2001) Photosynthesis and Respiration in Cyanobacteria.

Vijayan V, Jain IH, O'Shea EK (2011) A high resolution map of a cyanobacterial transcriptome. Genome Biol 12:R47. doi: 10.1186/gb-2011-12-5-r47

Vincent WF, Bowman JP, Rankin LM, Mcmeekin TA (2000) Phylogenetic diversity of picocyanobacteria in Arctic and Antarctic ecosystems.

Vogel J (2003) Experimental and computational analysis of transcriptional start sites in the cyanobacterium Prochlorococcus MED4. Nucleic Acids Res 31:2890–2899. doi: 10.1093/nar/gkg398

Voss B, Bolhuis H, Fewer DP, et al. (2013) Insights into the physiology and ecology of the brackish-water-adapted Cyanobacterium Nodularia spumigena CCY9414 based on a genome-transcriptome analysis. PLoS One 8:e60224. doi: 10.1371/journal.pone.0060224

Wang Q, Jantaro S, Lu B, et al. (2008) The high light-inducible polypeptides stabilize trimeric photosystem I complex under high light conditions in Synechocystis PCC 6803. Plant Physiol 147:1239–50. doi: 10.1104/pp.108.121087

Waterbury JB, Valois FW (1993) Resistance to Co-Occurring Phages Enables Marine Synechococcus Communities To Coexist with Cyanophages Abundant in Seawater Resistance to Co-Occurring Phages Enables Marine Synechococcus Communities To Coexist with Cyanophages Abundant in Seawatert. Appl Environ Microbiol 59:3393–3399.

Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136:615–28. doi: 10.1016/j.cell.2009.01.043

Weinbauer MG, Rassoulzadegan F (2003) Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1–11. doi: 10.1046/j.1462-2920.2003.00539.x

Westberry TK, Williams PJLB, Behrenfeld MJ (2012) Global net community production and the putative net heterotrophy of the oligotrophic oceans. Global Biogeochem Cycles 26:n/a–n/a. doi: 10.1029/2011GB004094

White R, Chiba S, Pang T, et al. (2011) Holin triggering in real time. Proc Natl Acad Sci U S A 108:798–803. doi: 10.1073/pnas.1011921108

Whittaker CA, Hynes RO (2004) Distribution and Evolution of von Willebrand / Integrin A Domains : Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere □. 13:3369–3387. doi: 10.1091/mbc.E02

Wikner J (1993) Nucleic acids from the host bacterium as a major source of nucleotides for three marine bacteriophages. FEMS Microbiol Ecol 12:237–248. doi: 10.1016/0168-6496(93)90047-B

Wilhelm SW, Suttle CA (1999) Viruses and Nutrient Cycles in the Sea aquatic food webs.

Will S, Joshi T, Hofacker IL, et al. (2012) LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs. Rna 18:900–914. doi: 10.1261/rna.029041.111

Williams KP (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res 30:866–75.

Williams PJ le B (1998) The balance of plankton respiration and photosynthesis in the open oceans. Nature 394:55–57.

Williams PJLB, Quay PD, Westberry TK, Behrenfeld MJ (2013) The oligotrophic ocean is autotrophic. Ann Rev Mar Sci 5:535–49. doi: 10.1146/annurev-marine-121211-172335

Wilms I, Nowrousian M, Sharma CM, Narberhaus F (2012) Do not distribute . © 2012 Landes Bioscience . 446–457.

Wilson JT (1965) A new class of faults and their bearing on Continental Drift. Nature 207:343–347.

Wilson WH, Carr NG, Mann NH (1994) The Effect of Phosphate Status on the Kinetics of Cyanophage Infection in the Ocenaic Cyanobacteium Synechococcus sp. WH7803. J P 32:506–516.

Wilson WH, Joint IR, Carr NG, Mann NH (1993a) Isolation and Molecular Characterization of Five Marine Cyanophages Propagated on Synechococcus sp. Strain WH7803. Appl Environ Microbiol 59 :3736–3743.

Wilson WH, Joint IR, Carr NG, Mann NH (1993b) Isolation and Molecular Characterization of Five Marine Cyanophages Propagated on Isolation and Molecular Characterization of Five Marine Cyanophages Propagated on Synechococcus sp . Appl Environ Microbiol 59:3736–3743.

Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 59:487–517. doi: 10.1146/annurev.micro.59.030804.121336

Winter C, Bouvier T, Weinbauer MG, Thingstad TF (2010) Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. Microbiol Mol Biol Rev 74:42–57. doi: 10.1128/MMBR.00034-09

Wommack KE, Colwell RR (2000) Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69–114.

Wright PR, Richter a. S, Papenfort K, et al. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci 110:E3487–E3496. doi: 10.1073/pnas.1303248110

Wu R, Geiduschek EP (1975) The role of replication proteins in the regulation of bacteriophage T4 transcription. II. Gene 45 and late transcription uncoupled from replication. J Mol Biol 96:539–562.

Xu H, Vavilin D, Funk C, Vermaas W (2002) Small Cab-like proteins regulating tetrapyrrole biosynthesis in the cyanobacterium Synechocystis sp. PCC 6803. Plant Mol Biol 49:149–60.

Xu H, Vavilin D, Funk C, Vermaas W (2004a) Multiple deletions of small Cab-like proteins in the cyanobacterium Synechocystis sp. PCC 6803: consequences for pigment biosynthesis and accumulation. J Biol Chem 279:27971–9. doi: 10.1074/jbc.M403307200

Xu M, Struck DK, Deaton J, et al. (2004b) A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. Proc Natl Acad Sci U S A 101:6415–20. doi: 10.1073/pnas.0400957101

Xu Q, Hoppe D, Chitnis VP, et al. (1995) Mutational Analysis of Photosystem I Polypeptides in the Cyanobacterium Synechocystis sp. PCC6803. J Biol Chem 270:16243–16250.

Yooseph S, Sutton G, Rusch DB, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 5:e16. doi: 10.1371/journal.pbio.0050016

Yoshimura T, Imamura S, Tanaka K, et al. (2007) Cooperation of group 2 sigma factors, SigD and SigE for light-induced transcription in the cyanobacterium Synechocystis sp. PCC 6803. FEBS Lett 581:1495–500. doi: 10.1016/j.febslet.2007.03.010

Young R (1992) Bacteriophage lysis: mechanism and regulation. Microbiol Rev 56:430–81.

Young R (2013) Phage lysis: do we have the hole story yet? Curr Opin Microbiol 16:790–7. doi: 10.1016/j.mib.2013.08.008

Young R (2002) Bacteriophage holins: deadly diversity. J Mol Microbiol Biotechnol 4:21–36.

Young R (2014) Phage lysis: Three steps, three choices, one outcome. J Microbiol 52:243–58. doi: 10.1007/s12275-014-4087-z

Zagotta MT, Wilson DB (1990) Oligomerization of the bacteriophage lambda S protein in the inner membrane of Escherichia coli. J Bacteriol 172:912–21.

Zeidner G, Bielawski JP, Shmoish M, et al. (2005) Potential photosynthesis gene recombination between Prochlorococcus and Synechococcus via viral intermediates. Environ Microbiol 7:1505–13. doi: 10.1111/j.1462-2920.2005.00833.x

Zeng Q, Bonocora RP, Shub D a (2009) A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages. Curr Biol 19:218–22. doi: 10.1016/j.cub.2008.11.069

Zeng Q, Chisholm SW (2012) Marine viruses exploit their host's two-component regulatory system in response to resource limitation. Curr Biol 22:124–8. doi: 10.1016/j.cub.2011.11.055

Zhang F, Ramsey E, Woodsen S (1995) In vivo facilitation of Tetrahymena group I intron splicing in Escherichia coli pre ribosomal RNA. RNA 1:284–292.

Zhu B, Tabor S, Raytcheva D a, et al. (2013) The RNA polymerase of marine cyanophage Syn5. J Biol Chem 288:3545–52. doi: 10.1074/jbc.M112.442350

Zinser ER, Johnson ZI, Coe A, et al. (2007) Influence of light and temperature on Prochlorococcus ecotype distributions in the Atlantic Ocean. Limnol Oceanogr 52:2205–2220. doi: 10.4319/lo.2007.52.5.2205

Zouni A, Witt HT, Kern J, et al. (2001) Crystal structure of photosystem II from Synechococcus elongatus at 3.8 A resolution. Nature 409:739–43. doi: 10.1038/35055589

Zwirglmaier K, Heywood JL, Chamberlain K, et al. (2007) Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. Environ Microbiol 9:1278–90. doi: 10.1111/j.1462-2920.2007.01246.x

Zwirglmaier K, Jardillier L, Ostrowski M, et al. (2008) Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. Environ Microbiol 10:147–61. doi: 10.1111/j.1462-2920.2007.01440.x

Zwirglmaier K, Spence E, Zubkov M V, et al. (2009) Differential grazing of two heterotrophic nanoflagellates on marine Synechococcus strains. Environ Microbiol 11:1767–76. doi: 10.1111/j.1462-2920.2009.01902.x

Appendices

Appendix 1: List of oligos used in this study

Primer Name	Sequence (5'-3')
pRL153_Fseq	AGGAGATTAACCCGCCCAAGG
pRL153_Rseq	GGACTGGCTTTCTACGTGTTC
cfrI_Walk_F	GTTTTTCCTCCCCACAAGTC
cfrI_Walk_R_A	GACCTTGCAGCAGCAGAAG
cfrI_Walk_R_B	AAGGGAAGTTTAAGAATGCACA
cfrI_Walk_R_C	AATCCTGGTAGTAATACCAGATAGTTAAGG
cfrI_Walk_R_D	TTTTCCAGGCAGAACACAAC
cfrI_Walk_R_E	TCTTTCAGTGATGGTATGCCG
cfrI_Walk_R_F	AGCCCCTGTTGCCGCTGCTA
cfrI_Walk_R_G	AGTGGGAACTCTCTTACCGAC
cfrI_Walk_R_H	TATCAACTAGTGGTCTTCCACTTCCT
cfrI_Walk_R_I	CTATCCCATCTGGGAAGCTGC
cfrI_Walk_R_J	ATAACATCATCTCTGGTGCTGTTA
cfrI_Walk_R_K	CCCGTCGATATTGACGGCAT
cfrI_Walk_R_L	TCCTACTCTTCTCGCAGCTGC
cfrI_Walk_R_M	GGGTTACCAGCACCGACAAC
cfrI_Walk_R_N	TACCTAGTAAACAACGTTCTAATAAAACTATGA
cfrI_Walk_R_O	ATAATAAATAGGTAAACAAATGTTAAGGAATCA
cfrI_Walk_R_P	GGACAAATGAGTACAAATACTCTACTTTG
cfrI_Walk_R_Q	TGCTTATTTTTATTTCTTTGATTATTTATAGG
psbA_F	ATGACTGCATCCATCGCTCA
psbA_R	TCAACCGATTGCGGGTG
psbA_F_T7	AGAGAGATAATACGACTCACTATAGGGAGAATGACTGCATCCATCGCTCA
psbA_intF	CTGGTCTGGGTATGGAGGTG
psbA_intR	TGTCGGACGCTTATTCCTGT
psbA_intProbe	5Cy5/ACGAGCGTCAAGTTGTGCTTTGCGC/3IAbRQSp
psbA_spl_F	CTGGTCTGGGTATGGAGGTG
psbA_spl_R	AGGCAACTGGTGTTGCTTCT
psbA_spl_probe	5TET/CACGAGCGTAATGCACACAACTTCCCTCTT/3IABkFQ
psbA_unspl_F	CTGGTCTGGGTATGGAGGTG
psbA_unspl_R	TGTCGGACGCTTATTCCTGT
psbA_unspl_probe	5Cy5/ACGAGCGTCAAGTTGTGCTTTGCGC/3IAbRQSp
16S_F	CCAAGGCATCGATCAGTAGCT
16S_R	CTGGGCCGTGTCTCAGT
16S_probe	CCAGTGTGGCTGATCAT
phoH_F	TCTTCTTGGCATCGAACCACT
phoH_R	GCGTTGGAAATCTAAGATCCCC
cfrI_T7	AGAGAGATAATACGACTCACTATAGGGAGAGTTTTTCCTCCCCACAAGTC
cfrI_trunc1_R	GACCTTGCAGCAGCAGAAG
cfrI_trunc2_R	AAGGGAAGTTTAAGAATGCACA

cfrI_trunc3_R	AATCCTGGTAGTAATACCAGATAGTTAAGG
cfrI_full_R	TCCTACTCTTCTCGCAGCTG

Appendix 2: Results of rRNA depletion with Terminator 5' monophosphate dependent nuclease. Total RNA was treated with 1 or 2U of Terminator nuclease in buffer A or B. Removal of RNA was assayed using a Bioanalyser and detecting the relative percentage of area under the 16S or 23S RNA peaks. Sample sizes were 8.
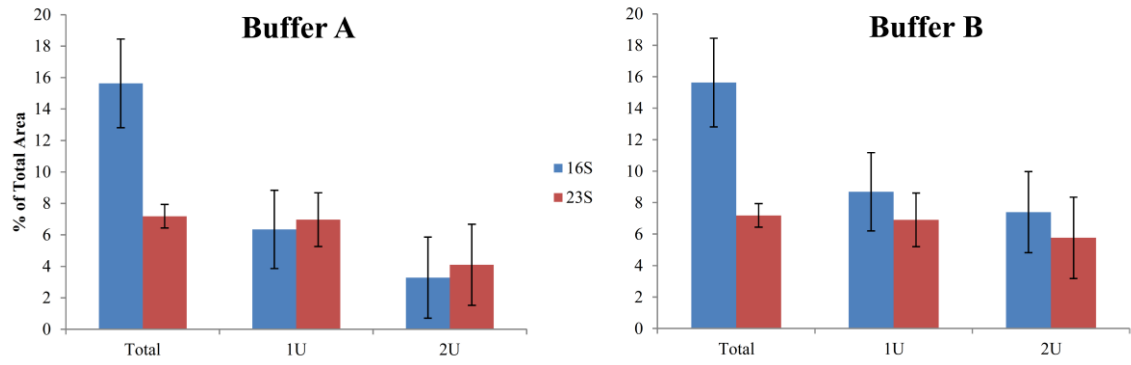
Appendix 3: Rho-independent terminators in the S-PM2$^{\Delta p017:050}$ genome as predicted by ARNold (rna.igmors.u-psud.fr/toolbox/ARNold/index.php). [1] Strand of terminator (+ is Watson, -is Crick). [2] Sequence of predicted terminator. Lower case indicates the spacer between the step loop and polyU tract. [3] Folding energy of the stem loop. [4] Either the ORF that is upstream of the terminator if the terminator is intergenic or the ORF that the terminator is found within if it is not intergenic. [5] Mean termination frequency as calculated using the method in section 4.2.9. The mean is of the 4 independent time points during infection (T1, T3, T6, and T9). [6] Classification of intergenic terminators. This is related to the orientation of the flanking ORFs and the expression cluster and is described in fig. 4.27.

Strand[1]	Sequence[2]	Folding energy (kcal/mol)[3]	Length (bp)	Start Coordinate	End Coordinate	Intergenic? (Y/N)	Upstream ORF/ORF[4]	Downstream ORF	Comments	Mean Termination Frequency[5]	Stdev	Classification[6]	Boundary/Non Boundary?
+	TATAATGTTAAGGAACTCTTCGGAGTTCCTTTT TTTATAAA	-13.9	42	126197	126239	Y	p178(psbA)	p180(f-CphI)		96.16	2.92	5	B
+	TAATCATAAATCCCTATCTGATAAATACAGGTA GGGaTTTTTAGTATCT	-11.0	50	56357	56407	Y	p095	p096(gp13)		91.85	4.18	1	N B
+	TGTGATATACTGGGGGTCTTAAGGGACCCTCTT TTTTATGATC	-13.1	44	128456	128500	Y	p182(psbD)	p183		85.29	11.41	4	N B
+	TCGGCGCTTAATGGGGTCTCTGGCAAAAACCA GAAGACCCCAtTTTTGTATCATGC	-12.4	57	114558	114615	Y	p174(hli03)	p175	Intergenic before antisense p175	82.87	8.18	17	B
+	TATTTCAACAAAGCGTGAGGAGCTGGGTGGTC ATTCCATCTGCCCTTACGCTTTTTCTGCATCTG	-9.8	66	15822	15888	Y	P079	P080	No intergenic nt between ORFs	78.38	21.53	5	B
+	AACTGATACCTGGGGGCAGTAATCTATGCCCC CTTTTTTATAGCT	-15.4	46	14602	14648	Y	p075	p076		78.31	7.49	5	B
+	CCGAACTTTAAAGTGGGGTCAGCATCACCCCA CTTTTTTTATGTCTT	-12.6	48	85624	85672	Y	p132 (regA)	p133 (hsp20)		76.87	2.87	1	N B
+	GAAGTGGCATAGGGACTTGCTAGTCCCTTTTTT CATGCC	-9.9	40	185228	185268	Y	p240	p241		70.87	10.50	4	N B
+	TCTTTAGTGGAGGGGGGTTGACTAATCCCCCTT TTTCCTATAT	-13.0	44	113871	113915	Y	p171	p172 (hli03_1)		68.90	8.59	7	N B
+	ATCTAGATTTGTGGGGATCCTAGGATCCCCATT TTTTATAAAT	-15.5	44	97632	97676	Y	p147	p148		63.84	19.22	1	N B
+	AAACTGGACTAGGGGTGCCCTCAAGGGGGTGC CCCTTTTGCTATGCT	-16.6	48	77854	77902	Y	p121(gp46)	p122		59.20	1.84	9	B
+	TAAGAAATTAGACGGGGGTTTAGACCCCCGTT	-14.9	45	136441	136486	Y	p201(gp32)	p202(gp53)	gp53 is on opposite strand	56.88	17.31	14	B

TTTTTGAGCCTT

	Sequence							Notes				
+	AAAGTTTCATACCCCCTTCTCTCGGATTCTTTG CAGTCCAGAGGAGGGGGaTTTTTTAGGTCTC	-14.0	65	106351	106416	Y	p157	p158	p158 on opposite strand	52.75	9.16	10	B
+	TATCTTTACGAGGGTGCCGAAAGGCACCCTTTT TTGTTGAT	-18.6	42	70554	70596	Y	p111	p112		32.99	21.26	1	N B
+	AGTCCAGAGGAGGGGGATTTTTTAGGTCTCCCT TTACATTAGCA	-10.9	45	106385	106430	Y	p157	p158		25.69	14.54	10	B
+	AAATCTATCACTATGCCTGCTATCATTAGTGGG CATAaTTTCTTATAACA	-8.4	51	6397	6448	Y	p055	p056		4.13	11.01	4	N B
+	AATAATTCTTAGCGGATCTAAGATCCGCTTTTT TTATAGG	-10.7	41	132297	132338	Y	p193	p194		3.43	36.07	8	B
+	TCCAAGTCAAGGTGTAGCGAGAGTTTGTTACA CTATTGTGTAATA	-8.4	46	107337	107383	Y	p160	p161	Termination before the stop codon	2.02	10.15	9	B
+	GTAGAACAATATCGTCCTAAGAGCGTGGATGA TTGTATTCTTCC	-6.9	45	82416	82461	Y	p124 (gp45)	p125 (gp44)		-0.23	18.14	4	N B
+	CCGATGAAGTATCGCCAGCAATGGCGGcTTATC TTCATGC	-10.0	41	69024	69065	Y	p110(gp22)	p111(gp23)		-9.85	6.59	1	N B
+	TTGAATATTCAGGGGGGCAACCCTCCCCCCTTT TCTATTCAT	-12.5	43	12350	12393	Y	p068	p069		-17.98	5.40	4	N B
+	ACCGACCACTAGGGGGCAGGGATGCCCCCaTTT TGCCCCTAT	-16.1	43	13638	13681	Y	p071	p072		-27.12	27.77	4	N B
+	AGCTGAGGAAGAGGCACAGGGGGCTTGACTCC CACCCCTGTGCCTgTTATTATGTCTAA	-13.7	60	135442	135502	Y	p200	p201	No poly U tract, p201 early promoter upstream of p200 terminator	-88.60	38.74	9	B
+	ATCTGTCACAGGGGGGCGTTGCTCCCTTTTGTT TTTTG	-9.1	39	86186	86225	Y	p133(hsp20)	p134	Late promoter of p134 Upstream of terminator of p133	-112.65	53.70	2	B
+	GTAAGATAAATACCTCCATATGGAGGTTTTTTA TTATGCG	-9.3	41	104744	104785	Y	p154(nrdB)	p155	Long 3'UTR, Late promoter of p155 upstream of terminator of p154	-248.95	131.75	8	B
+	ATAAAAGCAAGGGGAGGGAAACAAAAAGTTT CTCTCCCTTTTTTGTTGTCT	-12.0	52	108069	108121	N	p161	p163	Long 3'UTR antisense to p162	-14.71	11.95	OD D	N A
+	CATTAAATATTCCGTTTTTATAAGCGGTTATTT TTTTG	-5.6	40	118135	118175	N	p174(hli03)	p178 (psbA)	Long 3'UTR antisense to p175	-16.11	10.86	OD D	N A
+	GGATATATACTTGCTGGAACACCAGTATATTTT TCACGG	-6.8	40	118626	118666	N	p174(hli03)	p178 (psbA)	Long 3'UTR antisense to p176	-6.53	29.08	OD D	N A
+	TTATTGGTAACTTTGTTGGTCAACCACAACAGG TTTCTTATGATT	-5.4	46	168640	168686	N	p222	NA		-18.43	13.15	NA	N A
+	AGTAACTGAAGTTACCGTAACAGATGGTGGTA TTAATTATAC	-5.1	43	50161	50204	N	p091	NA		-2.68	14.03	NA	N A
+	CTAGAGGTTGTTGGTGGTATTAATTCACCACTA TTAGTTGTTGGT	-7.9	46	30802	30848	N	p087	NA		-10.50	1.55	NA	N A

+	AAGAAGACTGTTGATGCAGTTGGAGCGTCATA TTTCTTAGAT	-6.6	43	141760	141803	N	*p209*	NA	-83.77	50.05	NA	NA
+	GCAATTAATTGCTGGATTTGGTGGATCCAGTTC TATTGGTGG	-7.4	43	170971	171014	N	*p224*	NA	-51.96	0.94	NA	NA
+	GTATTCGATAGCGGTTATAAGTACGTTTATGAT CGTTTTACTGATAA	-5.5	48	63998	64046	N	*p105*	NA	-82.73	19.94	NA	NA
+	TTTATTCAACTCCTTCTAGTCAAGGAAGGTGTT ATTCAAAA	-6.6	42	31210	31252	N	*p087*	NA	-9.22	3.37	NA	NA
+	GGCGGTGGAGGTGGATCCAGATACAATGGATC TGTTTTGAGTTTGA	-8.6	47	173114	173161	N	*p225*	NA	-29.13	112.24	NA	NA
+	AAAATAAGTGAGGGTAACCAGAAGGTTGCCCT TATCCTTGCTA	-13.8	44	99831	99875	N	*p151*	NA	-9.90	4.06	NA	NA
+	ATGGTAAGAAGCACCTTTATATCGAAGGTGTTT TCTTGCAGT	-7.4	43	67261	67304	N	*p109*	NA	-128.90	49.70	NA	NA
+	TATCCTACTGCTTCTGCACATGCAGGAcaTTTTA AAGCGTT	-6.7	42	91305	91347	N	*gp41*	NA	25.61	13.00	NA	NA
+	TAAATTTAGAACATGTTTATGCAAATGGACATG TTTTTGGTGACA	-6.3	46	179910	179956	N	*p230*	NA	-163.95	30.13	NA	NA
+	TAGACCTGAAGGATCTTCTGAATTAGAAGAAG GTCaaTTTTGTAATAGT	-10.3	50	151220	151270	N	*p213*	NA	-125.41	60.33	NA	NA
+	GTTCTAAAAGATCGGTAATACGAGCCGATTTC AATTCTAT	-5.6	41	95563	95604	N	*p143*	NA	-85.35	86.00	NA	NA
+	GCACTTTACTGGTGGGGTGAGAATGAATACCC TGCaTTTTGGGTTGAA	-10.0	49	10897	10946	N	*p066*	NA	-47.07	26.15	NA	NA
+	AAAGTAAATAACATGATCGATCATGATCGTGaT TATCTATTTAC	-8.0	45	101539	101584	N	*nrdA*	NA	-22.98	12.07	NA	NA
+	TTCCTGCTACTGCAGACTTTATGTTTGCTTTGAT TAGTAC	-7.2	41	91597	91638	N	*gp41*	NA	-48.86	24.06	NA	NA
+	GATGAAGATCTTGCCAGTAAGTATTGGCGTTAC TTTTCTAA	-7.4	42	7458	7500	N	*p057*	NA	-57.15	38.50	NA	NA
+	TAATGTTGAGTTCCTCTCTTATGATAAGGGAGG AcaTTATGCTTGGCA	-10.5	49	8061	8110	N	*t004*	NA	-187.73	114.24	NA	NA

Appendix 4: RNA-Seq library mapping and coverage statistics **(a)** Samples from Chapter Four. **(b)** Samples from Chapter Five. [1]Reads are aligned if they are concordant and have a MAPQ>30. [2] Fold coverage is calculated as the sum of the length of all sequenced fragments divided by the genome length. 3Mapped to S-PM2$^{\Delta p017:p050}$

(a)

Sample	Library size (reads)	Reads aligned[1]			Computed fold coverage[2]	
		rRNA reads	S-PM2[3]	WH7803 (non-rRNA)	S-PM2	WH7803
T0	28,120,374	22,499,112 (80.01%)	0 (0%)	2,133,578 (7.59%)	0	112.67
T1	30,398,364	26,058,542 (85.72%)	27,110 (0.09%)	252,012 (0.83%)	18.23	13.31
T3	58,828,858	51,018,694 (86.72%)	119,910 (0.20%)	374,768 (0.64%)	80.65	19.79
T6	49,850,914	43,094,262 (86.45%)	165,728 (0.34%)	345,808 (0.69%)	111.46	18.26
T9	33,962,224	27,596,114 (81.26%)	169,664 (0.50%)	243,920 (0.72%)	114.12	12.88

(b)

Sample	Library size (reads)	Reads aligned[1]			Computed fold coverage[2]	
		rRNA reads	S-PM2[3]	WH7803 (non-rRNA)	S-PM2	WH7803
HL1	38,760,110	32,170,892 (83.00%)	97,749 (0.25%)	115,957 (0.30%)	65.74	22.21
HL2	43,396,178	35,541,470 (81.90%)	78,206 (0.18%)	129,072 (0.30%)	48.56	26.47
HL3	24,010,898	19,736,958 (82.20%)	61,240 (0.26%)	75,198 (0.31%)	34.47	14.21
LL1	34,175,164	27,647,706 (80.89%)	28,016 (0.08%)	86,535 (0.25%)	18.84	23.76
LL2	31,572,442	26,236,698 (83.01%)	29,019 (0.09%)	72,179 (0.23%)	19.52	16.97
LL3	36,823,838	30,637,434 (83.20%)	37,680 (0.10%)	83,906 (0.23%)	25.34	22.09

Appendix 5: Cluster assignments for all S-PM2$^{\Delta p017:050}$ ORFs, tRNAs and ncRNAs. [1] ORF numbers are as reported for S-PM2 WT as in Genbank accession NC_006820, (Downloaded on 25/1/2014). [2] Strand location of the feature (+ is Watson, - is Crick). [3] Cluster assignments using the 'Union' count method. Red is early, blue is late, n.d. is not determined. [4] Cluster assignments using the 'Intersection-strict' count method. Red is early, blue is late, n.d. is not determined. [5] Cluster assignments using the 'Union' count method. Red is early, blue is late, n.d. is not determined. [6] Final cluster assignments made according to the consensus between all three count methods.

ORF No.[1]	Annotation	Start codon coordinate	Stop codon coordinate	Strand[2]	Cluster (Union)[3]	Silhouette score	Cluster (Intersection-strict)[4]	Silhouette score	Cluster (Intersection-nonempty)[5]	Silhouette score	Sum of silhouette scores	Cluster[6]
S-PM2p001	*dam*	1	888	+	●	0.810	●	0.786	●	0.874	2.470	Early
S-PM2p002	*S-PM2p002*	925	1059	+	●	0.697	●	0.756	●	0.709	2.162	Early
S-PM2p003	*S-PM2p003*	1049	1258	+	●	0.764	●	0.752	●	0.833	2.349	Early
S-PM2p004	*S-PM2p004*	1255	1560	+	●	0.805	●	0.779	●	0.871	2.455	Early
S-PM2p005	*S-PM2p005*	1729	1932	+	●	0.815	●	0.805	●	0.877	2.497	Early
S-PM2p006	*S-PM2p006*	1913	2116	+	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p007	*S-PM2p007*	2113	2292	+	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p008	*S-PM2p008*	2289	2852	+	●	0.829	●	0.811	●	0.888	2.528	Early
S-PM2p009	*S-PM2p009*	2856	3251	+	●	0.824	●	0.816	●	0.889	2.529	Early
S-PM2p010	*S-PM2p010*	3248	3556	+	●	0.824	●	0.803	●	0.887	2.514	Early
S-PM2p011	*S-PM2p011*	3550	3678	+	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p012	*S-PM2p012*	3675	4079	+	●	0.815	●	0.816	●	0.882	2.512	Early
S-PM2p013	*S-PM2p013*	4201	4605	+	●	0.844	●	0.816	●	0.895	2.554	Early
S-PM2p014	*S-PM2p014*	4710	5090	+	●	0.649	●	0.642	●	0.677	1.968	Early
S-PM2p015	*S-PM2p015*	5160	5330	+	●	0.803	●	0.736	●	0.869	2.407	Early
S-PM2p016	*S-PM2p016*	5370	5780	+	●	0.831	●	0.805	●	0.889	2.525	Early
S-PM2p054	*S-PM2p054*	5888	6091	+	●	0.819	●	0.801	●	0.877	2.497	Early
S-PM2p055	*S-PM2p055*	6164	6388	+	●	0.837	●	0.823	●	0.892	2.553	Early

S-PM2p056	*S-PM2p056*	6608	7042	+	●	0.819	●	0.813	●	0.883	2.515	Early
S-PM2p057	*S-PM2p057*	7370	7597	+	●	0.791	●	0.778	●	0.864	2.434	Early
S-PM2p058	*S-PM2p058*	7594	7797	+	●	0.740	●	0.728	●	0.821	2.289	Early
S-PM2p059	*S-PM2p059*	8541	8774	+	●	0.782	●	0.780	●	0.863	2.425	Early
S-PM2p060	*S-PM2p060*	9215	9466	+	●	0.715	●	0.692	●	0.758	2.165	Early
S-PM2p061	*S-PM2p061*	9520	9684	+	●	0.836	●	0.827	●	0.862	2.525	Early
S-PM2p062	*S-PM2p062*	9860	9979	+	●	0.831	n.d	n.d	●	0.891	1.721	Early
S-PM2p063	*S-PM2p063*	10033	10278	-	●	0.692	●	0.651	●	0.746	2.088	Early
S-PM2p064	*S-PM2p064*	10316	10645	+	●	0.814	●	0.729	●	0.880	2.423	Early
S-PM2p065	*S-PM2p065*	10684	10857	+	●	0.809	●	0.790	●	0.884	2.484	Early
S-PM2p066	*S-PM2p066*	10854	11054	+	●	0.850	●	0.785	●	0.888	2.523	Early
S-PM2p067	*S-PM2p067*	11234	11506	-	●	0.771	●	0.693	●	0.842	2.306	Early
S-PM2p068	*S-PM2p068*	11918	12352	+	●	0.833	●	0.820	●	0.890	2.543	Early
S-PM2p069	*S-PM2p069*	12406	12552	+	●	0.850	n.d	n.d	●	0.888	1.738	Early
S-PM2p070	*S-PM2p070*	13346	13552	-	●	0.753	n.d	n.d	●	0.826	1.579	Early
S-PM2p071	*S-PM2p071*	13579	13680	+	●	0.829	n.d	n.d	●	0.886	1.714	Early
S-PM2p072	*S-PM2p072*	13658	13837	+	●	0.843	●	0.654	●	0.889	2.386	Early
S-PM2p073	*S-PM2p073*	13925	14125	+	●	0.846	●	0.836	●	0.888	2.570	Early
S-PM2p074	*S-PM2p074*	14195	14380	+	●	0.720	●	0.722	●	0.779	2.221	Early
S-PM2p075	*S-PM2p075*	14406	14606	+	●	0.740	●	0.744	●	0.807	2.291	Early
S-PM2p076	*S-PM2p076*	14846	15175	+	●	0.891	●	0.857	●	0.924	2.672	Late
S-PM2p077	*S-PM2p077*	15102	15365	+	n.d	n.d	●	0.001	●	-0.072	-0.071	UC
S-PM2p078	*S-PM2p078*	15293	15616	+	●	0.731	●	0.854	●	0.896	2.481	Late
S-PM2p079	*S-PM2p079*	15656	15799	+	●	0.784	●	0.672	●	0.823	2.279	Early
S-PM2p080	*S-PM2p080*	15783	16247	+	●	0.638	●	0.749	●	0.816	2.203	Late
S-PM2p081	*S-PM2p081*	16216	16410	+	n.d	n.d	●	0.552	●	0.621	1.173	UC
S-PM2p082	*gp25*	16444	16836	+	●	0.788	●	0.714	●	0.819	2.321	Late
S-PM2p083	*gp6*	16836	18644	+	●	0.887	●	0.893	●	0.917	2.697	Late
S-PM2p084	*S-PM2p084*	18650	27796	+	●	0.843	●	0.850	●	0.879	2.572	Late
S-PM2p085	*S-PM2p085*	27797	28324	+	●	0.870	●	0.837	●	0.893	2.600	Late

S-PM2p086	*gp8*	28387	30291	+	●	0.860	●	0.868	●	0.898	2.627	Late
S-PM2p087	*S-PM2p087*	30336	38579	+	●	0.885	●	0.890	●	0.912	2.687	Late
S-PM2p088	*S-PM2p088*	38576	38779	+	●	0.632	●	0.681	●	0.719	2.032	Late
S-PM2p089	*S-PM2p089*	38757	42512	+	●	0.878	●	0.884	●	0.910	2.673	Late
S-PM2p090	*S-PM2p090*	42514	43020	+	●	0.836	●	0.863	●	0.874	2.573	Late
S-PM2p091	*S-PM2p091*	43069	53085	+	●	0.869	●	0.876	●	0.900	2.646	Late
S-PM2p092	*S-PM2p092*	53095	54015	+	●	0.882	●	0.883	●	0.903	2.668	Late
S-PM2p093	*S-PM2p093*	54017	55000	+	●	0.903	●	0.904	●	0.926	2.732	Late
S-PM2p094	*S-PM2p094*	55002	56141	+	●	0.902	●	0.903	●	0.927	2.732	Late
S-PM2p095	*S-PM2p095*	56173	56358	+	●	0.709	●	0.638	●	0.727	2.074	Late
S-PM2p096	*gp13*	56411	57241	+	●	0.895	●	0.889	●	0.915	2.698	Late
S-PM2p097	*gp14*	57243	58121	+	●	0.875	●	0.874	●	0.902	2.651	Late
S-PM2p098	*gp15*	58121	58921	+	●	0.832	●	0.819	●	0.851	2.503	Late
S-PM2p099	*gp16*	58918	59337	+	●	0.835	●	0.834	●	0.873	2.542	Late
S-PM2p100	*S-PM2p100*	59337	59879	+	●	0.889	●	0.892	●	0.915	2.697	Late
S-PM2p101	*S-PM2p101*	59982	60212	+	●	0.366	●	0.044	●	0.399	0.810	UC
S-PM2p102	*S-PM2p102*	60190	60408	+	●	0.557	●	0.485	●	0.552	1.595	Late
S-PM2p103	*gp17*	60434	62080	+	●	0.826	●	0.807	●	0.858	2.490	Late
S-PM2p104	*S-PM2p104*	62068	62310	+	●	0.799	●	0.733	●	0.857	2.389	Late
S-PM2p105	*gp18*	62395	64626	+	●	0.846	●	0.855	●	0.886	2.587	Late
S-PM2p106	*gp19*	64665	65279	+	●	0.831	●	0.849	●	0.874	2.554	Late
S-PM2p107	*gp20*	65315	67009	+	●	0.850	●	0.849	●	0.887	2.586	Late
S-PM2p108	*S-PM2p108*	67032	67208	+	●	0.767	●	0.746	●	0.822	2.335	Late
S-PM2p109	*gp21*	67205	67849	+	●	0.897	●	0.895	●	0.922	2.713	Late
S-PM2p110	*gp22*	67906	69084	+	●	0.879	●	0.903	●	0.914	2.695	Late
S-PM2p111	*gp23*	69136	70542	+	●	0.891	●	0.897	●	0.916	2.705	Late
S-PM2p112	*S-PM2p112*	70618	70782	+	●	0.880	●	0.849	●	0.905	2.635	Late
S-PM2p113	*gp3*	71142	71651	+	●	0.888	●	0.870	●	0.912	2.670	Late
S-PM2p114	*UvsY*	71648	72121	+	●	0.896	●	0.890	●	0.920	2.706	Late
S-PM2p115	*UvsW*	72121	73584	+	●	0.844	●	0.840	●	0.877	2.562	Late

S-PM2p116	*S-PM2p116*	73584	73982	+	● (blue)	0.810	● (blue)	0.626	● (blue)	0.832	2.268	Late
S-PM2p117	*gp55*	74110	74604	+	● (blue)	0.271	● (red)	0.163	● (blue)	0.263	0.697	UC
S-PM2p118	*S-PM2p118*	74601	74819	+	● (red)	0.337	● (red)	0.349	● (red)	0.298	0.984	UC
S-PM2p119	*gp47*	74816	75865	+	● (red)	0.225	● (red)	0.304	● (red)	0.211	0.740	UC
S-PM2p120	*S-PM2p120*	75862	76137	+	● (blue)	0.418	● (blue)	0.483	● (blue)	0.556	1.456	UC
S-PM2p121	*gp46*	76134	77864	+	● (red)	0.093	● (red)	0.180	● (red)	0.068	0.340	UC
S-PM2p122	*S-PM2p122*	77867	80089	+	● (red)	0.838	● (red)	0.821	● (red)	0.890	2.549	Early
S-PM2p123	*CobS*	80163	81281	+	● (red)	0.816	● (red)	0.796	● (red)	0.882	2.494	Early
S-PM2p124	*gp45*	81729	82394	+	● (red)	0.851	● (red)	0.833	● (red)	0.887	2.572	Early
S-PM2p125	*gp44*	82394	83338	+	● (red)	0.808	● (red)	0.805	● (red)	0.834	2.447	Early
S-PM2p126	*S-PM2p126*	83335	83529	+	● (red)	0.744	● (red)	0.718	● (red)	0.803	2.265	Early
S-PM2p127	*S-PM2p127*	83522	83722	+	● (red)	0.843	● (red)	0.836	● (red)	0.893	2.572	Early
S-PM2p128	*S-PM2p128*	83754	83933	+	● (red)	0.808	● (red)	0.825	● (red)	0.877	2.510	Early
S-PM2p129	*Dam*	84134	84532	+	● (red)	0.503	● (red)	0.250	● (red)	0.504	1.257	UC
S-PM2p130	*S-PM2p130*	84522	84665	+	● (red)	0.810	n.d	n.d	● (red)	0.879	1.689	Early
S-PM2p131	*gp62*	84813	85199	+	● (red)	0.202	● (red)	0.434	● (red)	0.297	0.933	UC
S-PM2p132	*RegA*	85196	85621	+	● (blue)	0.539	● (blue)	0.487	● (blue)	0.580	1.606	Late
S-PM2p133	*Hsp20*	85772	86182	+	● (blue)	0.672	● (blue)	0.662	● (blue)	0.694	2.028	Late
S-PM2p134	*S-PM2p134*	86252	86719	+	● (red)	0.468	● (red)	0.584	● (red)	0.462	1.514	Early
S-PM2p135	*S-PM2p135*	86716	86967	+	● (red)	0.465	● (red)	0.505	● (red)	0.535	1.504	Early
S-PM2p136	*gp43*	86967	89459	+	● (red)	0.783	● (red)	0.777	● (red)	0.817	2.377	Early
S-PM2p137	*UvsX RecA-like*	89470	90501	+	● (red)	0.725	● (red)	0.737	● (red)	0.772	2.234	Early
S-PM2p138	*gp41*	90461	91873	+	● (red)	0.627	● (red)	0.646	● (red)	0.671	1.944	Early
S-PM2p139	*MazG*	91875	92282	+	● (red)	0.490	● (red)	0.432	● (red)	0.361	1.283	UC
S-PM2p140	*S-PM2p140*	92242	92418	+	● (red)	0.729	● (red)	0.520	● (red)	0.629	1.878	Early
S-PM2p141	*S-PM2p141*	92603	93181	+	● (red)	0.510	● (red)	0.473	● (red)	0.591	1.574	Early
S-PM2p142	*S-PM2p142*	93210	94382	+	● (red)	0.005	● (red)	0.104	● (red)	0.033	0.142	UC
S-PM2p143	*S-PM2p143*	94379	96139	+	● (red)	0.020	● (red)	0.189	● (red)	0.194	0.403	UC
S-PM2p144	*NrdB*	96136	96411	+	● (red)	0.081	● (blue)	0.218	● (blue)	0.295	0.594	UC

Gene	Name	Start	End	Strand		Value		Value		Value	Value	Class
S-PM2p145	*S-PM2p145*	96386	97111	+	● (blue)	0.874	● (blue)	0.878	● (blue)	0.913	2.665	Late
S-PM2p146	*S-PM2p146*	97120	97395	+	● (blue)	0.848	● (blue)	0.862	● (blue)	0.889	2.599	Late
S-PM2p147	*S-PM2p147*	97424	97618	+	● (blue)	0.831	● (blue)	0.863	● (blue)	0.854	2.548	Late
S-PM2p148	*S-PM2p148*	97712	98662	+	● (blue)	0.888	● (blue)	0.894	● (blue)	0.917	2.699	Late
S-PM2p149	*S-PM2p149*	98669	99025	+	● (blue)	0.869	● (blue)	0.898	● (blue)	0.918	2.685	Late
S-PM2p150	*S-PM2p150*	98998	99711	+	● (blue)	0.872	● (blue)	0.871	● (blue)	0.894	2.637	Late
S-PM2p151	*S-PM2p151*	99662	100156	+	● (blue)	0.885	● (blue)	0.891	● (blue)	0.919	2.694	Late
S-PM2p152	*gp61*	100153	101142	+	● (red)	0.242	● (red)	0.202	● (red)	0.180	0.624	UC
S-PM2p153	*NrdA*	101127	103457	+	● (blue)	0.187	● (red)	-0.101	● (blue)	0.197	0.282	UC
S-PM2p154	*NrdB*	103438	104622	+	● (red)	0.131	● (red)	0.080	● (red)	0.151	0.362	UC
S-PM2p155	*S-PM2p155*	104778	105125	+	● (blue)	0.901	● (blue)	0.899	● (blue)	0.925	2.725	Late
S-PM2p156	*S-PM2p156*	105112	105807	+	● (blue)	0.845	● (blue)	0.850	● (blue)	0.886	2.581	Late
S-PM2p157	*S-PM2p157*	106077	106295	+	● (blue)	0.902	● (blue)	0.892	● (blue)	0.924	2.719	Late
S-PM2p158	*S-PM2p158*	106518	106688	-	● (blue)	0.881	● (blue)	0.645	● (blue)	0.918	2.444	Late
S-PM2p159	*S-PM2p159*	106693	106848	-	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p160	*S-PM2p160*	107016	107378	+	● (blue)	0.405	● (red)	0.361	● (blue)	0.409	1.175	UC
S-PM2p161	*S-PM2p161*	107409	107807	+	● (blue)	0.846	● (blue)	0.886	● (blue)	0.861	2.594	Late
S-PM2p162	*S-PM2p162*	107769	108125	-	● (red)	0.690	n.d	n.d	● (red)	0.738	1.428	UC
S-PM2p163	*S-PM2p163*	108826	108975	+	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p164	*S-PM2p164*	109088	109387	+	● (red)	0.835	● (red)	0.600	● (red)	0.887	2.322	Early
S-PM2p165	*S-PM2p165*	109393	109677	+	● (red)	0.055	● (blue)	0.457	● (red)	0.074	0.586	UC
S-PM2p166	*NrdC*	109926	110171	+	● (red)	0.542	● (red)	0.667	● (red)	0.614	1.823	Early
S-PM2p167	*S-PM2p167*	110156	110371	+	● (red)	0.829	● (red)	0.617	● (red)	0.598	2.044	Early
S-PM2p168	*S-PM2p168*	110368	110607	+	● (red)	0.838	● (red)	0.826	● (red)	0.890	2.554	Early
S-PM2p169	*S-PM2p169*	110613	110897	+	● (red)	0.707	● (red)	0.779	● (red)	0.753	2.239	Early
S-PM2p170	*S-PM2p170*	112694	113533	+	● (red)	0.356	● (red)	0.265	● (red)	0.345	0.966	UC
S-PM2p171	*S-PM2p171*	113540	113890	+	● (blue)	0.264	● (blue)	0.157	● (blue)	0.303	0.724	UC
S-PM2p172	*Hli03*	113940	114134	+	● (blue)	0.263	● (blue)	0.495	● (blue)	0.279	1.038	UC
S-PM2p173	*S-PM2p173*	114220	114420	+	● (red)	0.717	● (red)	0.687	● (red)	0.757	2.161	Early
S-PM2p174	*Hli03*	114417	114536	+	● (red)	0.715	n.d	n.d	● (red)	0.752	1.468	UC

S-PM2p175	*S-PM2p175*	114607	117894	-	🔵	0.890	🔵	0.892	🔵	0.915	2.696	Late
S-PM2p176	*S-PM2p176*	117952	121128	-	🔵	0.876	🔵	0.881	🔵	0.913	2.670	Late
S-PM2p177	*S-PM2p177*	121139	124672	-	🔵	0.888	🔵	0.892	🔵	0.921	2.701	Late
S-PM2p178	*PsbA*	124904	126195	+	🔴	0.830	🔴	0.815	🔴	0.871	2.516	Early
S-PM2p180	*F-CphI*	126262	126690	+	🔵	0.850	🔵	0.834	🔵	0.887	2.572	Late
S-PM2p181	*S-PM2p181*	126687	127316	+	🔵	0.812	🔵	0.834	🔵	0.865	2.511	Late
S-PM2p182	*PsbD*	127398	128459	+	🔴	0.827	🔴	0.799	🔴	0.873	2.499	Early
S-PM2p183	*S-PM2p183*	128492	129004	+	🔴	0.795	🔴	0.811	🔴	0.817	2.423	Early
S-PM2p184	*NrdC*	129008	129256	+	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p185	*S-PM2p185*	129253	129432	+	🔴	0.831	n.d	n.d	🔴	0.891	1.722	Early
S-PM2p186	*S-PM2p186*	129448	129684	+	🔴	0.495	🔴	0.525	🔴	0.540	1.559	Early
S-PM2p187	*S-PM2p187*	129671	130534	+	🔴	0.813	🔴	0.799	🔴	0.857	2.469	Early
S-PM2p188	*S-PM2p188*	130531	130965	+	🔵	0.599	🔵	0.556	🔵	0.626	1.780	Late
S-PM2p189	*S-PM2p189*	130947	131234	+	🔴	0.295	🔴	0.048	🔴	-0.015	0.328	UC
S-PM2p190	*S-PM2p190*	131231	131494	+	🔵	0.863	🔵	0.850	🔵	0.879	2.592	Late
S-PM2p191	*SpeD*	131501	131833	+	🔵	0.171	🔴	0.489	🔵	0.175	0.835	UC
S-PM2p192	*S-PM2p192*	132024	132194	+	🔴	0.360	n.d	n.d	🔴	0.463	0.823	UC
S-PM2p193	*S-PM2p193*	132187	132300	+	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p194	*S-PM2p194*	132336	132491	+	🔵	0.902	n.d	n.d	🔵	0.924	1.826	Late
S-PM2p195	*Td*	132573	133208	+	🔴	0.169	🔴	0.365	🔴	0.199	0.734	UC
S-PM2p196	*S-PM2p196*	133211	133450	+	🔵	0.819	🔵	0.807	🔵	0.829	2.455	Late
S-PM2p197	*PhoH*	133457	134209	+	🔵	0.766	🔵	0.728	🔵	0.797	2.291	Late
S-PM2p198	*S-PM2p198*	134209	134898	+	🔵	0.868	🔵	0.848	🔵	0.878	2.594	Late
S-PM2p199	*gp33*	134876	135124	+	🔵	0.762	🔵	0.752	🔵	0.794	2.309	Late
S-PM2p200	*DUF1825*	135121	135468	+	🔵	0.327	🔵	0.217	🔵	0.339	0.883	UC
S-PM2p201	*gp32*	135547	136434	+	🔴	0.813	🔴	0.808	🔴	0.838	2.459	Early
S-PM2p202	*gp53*	136459	137121	-	🔵	0.893	🔵	0.893	🔵	0.919	2.705	Late
S-PM2p203	*gp48*	137121	138119	-	🔵	0.899	🔵	0.900	🔵	0.922	2.721	Late
S-PM2p204	*gp2*	138119	138781	-	🔵	0.848	🔵	0.843	🔵	0.867	2.558	Late
S-PM2p205	*gp4*	138784	139221	-	🔵	0.866	🔵	0.882	🔵	0.899	2.647	Late

245

S-PM2p206	*S-PM2p206*	139255	140043	+	●	0.844	●	0.859	●	0.893	2.595	Late
S-PM2p207	*gp26*	140049	140765	+	●	0.866	●	0.868	●	0.909	2.644	Late
S-PM2p208	*gp51*	140768	140941	+	n.d	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2p209	*S-PM2p209*	140934	144437	+	●	0.890	●	0.881	●	0.906	2.678	Late
S-PM2p210	*S-PM2p210*	144530	146779	+	●	0.894	●	0.902	●	0.919	2.715	Late
S-PM2p211	*S-PM2p211*	146779	148173	+	●	0.783	●	0.788	●	0.853	2.424	Late
S-PM2p212	*gp5*	148203	151148	+	●	0.903	●	0.902	●	0.926	2.730	Late
S-PM2p213	*S-PM2p213*	151148	152242	+	●	0.900	●	0.904	●	0.923	2.727	Late
S-PM2p214	*S-PM2p214*	152403	152747	+	●	0.847	●	0.849	●	0.862	2.557	Late
S-PM2p215	*S-PM2p215*	152769	152900	+	●	0.586	●	0.468	●	0.602	1.656	Early
S-PM2p216	*S-PM2p216*	153265	153537	+	●	0.747	●	0.803	●	0.814	2.363	Early
S-PM2p217	*CpeT*	153530	154057	+	●	0.387	●	0.450	●	0.443	1.280	UC
S-PM2p218	*S-PM2p218*	154355	154498	+	●	0.841	n.d	n.d	●	0.883	1.724	Early
S-PM2p219	*S-PM2p219*	154558	154716	+	●	0.349	n.d	n.d	●	0.451	0.800	UC
S-PM2p220	*S-PM2p220*	154885	155466	+	●	0.824	●	0.115	●	0.851	1.790	Early
S-PM2p221	*gp12*	166849	168198	+	●	0.879	●	0.885	●	0.912	2.676	Late
S-PM2p222	*S-PM2p222*	168200	169087	+	●	0.871	●	0.878	●	0.904	2.654	Late
S-PM2p223	*S-PM2p223*	169091	170794	+	●	0.865	●	0.864	●	0.894	2.623	Late
S-PM2p224	*S-PM2p224*	170791	172473	+	●	0.892	●	0.893	●	0.919	2.704	Late
S-PM2p225	*S-PM2p225*	172474	174222	+	●	0.854	●	0.855	●	0.902	2.611	Late
S-PM2p226	*gp12*	174261	177374	+	●	0.885	●	0.892	●	0.918	2.695	Late
S-PM2p227	*S-PM2p227*	177383	178087	+	●	0.898	●	0.892	●	0.920	2.710	Late
S-PM2p228	*S-PM2p228*	178084	179028	+	●	0.882	●	0.888	●	0.913	2.683	Late
S-PM2p229	*PiuC*	179028	179723	+	●	0.606	●	0.582	●	0.654	1.842	Late
S-PM2p230	*S-PM2p230*	179701	180210	+	●	0.834	●	0.834	●	0.863	2.531	Late
S-PM2p231	*S-PM2p231*	180194	180547	+	●	0.809	●	0.608	●	0.662	2.079	Late
S-PM2p232	*S-PM2p232*	180547	181434	+	●	0.848	●	0.855	●	0.901	2.604	Late
S-PM2p233	*S-PM2p233*	181422	182018	+	●	0.643	●	0.687	●	0.669	1.999	Early
S-PM2p234	*S-PM2p234*	182085	183332	+	●	0.809	●	0.778	●	0.878	2.466	Early
S-PM2p235	*S-PM2p235*	183403	183588	+	●	0.730	●	0.785	●	0.787	2.301	Early

S-PM2p236	*S-PM2p236*	183606	183893	+	🔴	0.783	🔴	0.782	🔴	0.809	2.374	Early
S-PM2p237	*S-PM2p237*	183895	184503	+	🔴	0.818	🔴	0.791	🔴	0.876	2.485	Early
S-PM2p238	*S-PM2p238*	184475	184729	+	🔴	0.810	🔴	0.795	🔴	0.875	2.481	Early
S-PM2p239	*S-PM2p239*	184741	184914	+	🔴	0.633	🔴	0.540	🔴	0.877	2.051	Early
S-PM2p240	*S-PM2p240*	184883	185173	+	🔴	0.776	🔴	0.783	🔴	0.803	2.362	Early
S-PM2p241	*S-PM2p241*	185296	185858	+	🔴	0.771	🔴	0.739	🔴	0.845	2.355	Early
S-PM2s001	*CfrI*	125091	126319	+	🔴	n.d	n.d	n.d	n.d	n.d	n.d	N.d.
S-PM2t003	*S-PM2t003*	7104	7373	+	🔴	0.814	🔴	0.820	🔴	0.880	2.513	Early
S-PM2t004	*S-PM2t004*	7794	8396	+	🔴	0.787	🔴	0.795	🔴	0.883	2.465	Early
S-PM2t006	*S-PM2t006*	8369	8560	+	🔵	0.301	🔴	0.832	🔴	0.873	2.005	Early
S-PM2t009	*S-PM2t009*	155506	166845	+	🔵	0.890	🔵	0.886	🔵	0.911	2.688	Late

Appendix 6: Prediction of transcript boundaries by Cufflinks software (http://cufflinks.cbcb.umd.edu/). Time points are shown on the left hand side. Purple arrows indicate transcripts on the Watson strand and green indicate those from the Crick strand at each time point. The below panel shows the locations of ORFs in the SPM$^{2p017:050}$ genome. Above rectangles represent ORFs on the Watson strand whilst below indicate ORFs on the Crick strand. ORFs are coloured according to the expression profile where red is early ORFs, blue are late ORFs and grey are unclustered ORFs.

Appendix 7: Thorough description of RSS bias created by the ScriptSeq^TM v2.0 library preparation kit.

Consider 3' ends of two cDNA fragments (i.e. 5' ends of the original transcript) from stage 3 of the ScriptSeq procedure (Fig. 4.2.3):

1. `3'-NNNNNNNNNNNNNNNNNNN`

2. `3'-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTAGANNNNNNNNN`

Now consider the terminal tagging oligo (TTO) which has the sequence:

`5'ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCTNNNNNN-3'`

1. The TTO will anneals to the first example as follows:

```
5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCTNNNNNN
                                        ||||||
                               3'-NNNNNNNNNNNNNNNNNNNN
```

2. However, this situation would be the preferential annealing for the TTO:

```
  5'-5'ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNN
                                      |||||||||||
   3'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTAGANNNNNNNNN
```

Thus, after extension in the 5'-3' direction by DNA polymerase, (bearing in mind the TTO 3' end is blocked) we would end up with:

```
1.5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCTNNNNNN
     |||||||||||||||||||||||||||||||||||||||||||||
     TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGACTAGANNNNNNNNNNNNNNNNNNNNN
2.5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNN
     |||||||||||||||||||||||||||||||||||||
     TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNN
```

Therefore removal of the TTO one would have

1. `TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNN`

2. `TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNNNNNNNNNNNNNNNNNNNN`

Thus after sequencing and removal of the tag and TTO sequence, one cannot differentiate between *bona fide* 5' ends and the result of annealing to a complimentary sequence within the cDNA fragment. The first example refers to the bona fide 5' end of the fragment (whether it is the 5'end of the transcript or not). In the second example the TTO annealed nowhere near the 3' end of the cDNA (5' end of the RNA fragment). Yet DNA polymerase will extend the product complementary to the TTO replacing the original sequence giving a false 5' end of the fragment. Note that the sequence on the cDNA that requires this to happen does not have to be 5'-CTAGA-3'. Rather it can be:

5'-GA-3'
5'-AGA-3'
5'-TAGA-3'
5'-CTAGA-3'
5'-ACTAGA-3'
5'-GACTAGA-3'
5'-AGACTAGA-3'
5'-TAGACTAGA-3'
5'-CTAGACTAGA-3'
5'-GCTAGACTAGA-3'
5'-GGCTAGACTAGA-3'

Increasing thermodynamic stability of annealing

Increasing chance of occurrence across the genome

Thus this explains the sequence logos in Fig. 4.3.10. The sequence conservation decreases as you go from -1 to -6 to the read start site because these sequences do not occur often in the genome. Whereas a CT motif is obviously very frequent.