Bacteriophages of Marine Roseobacter

Jacqueline Chan

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Warwick, Coventry, United Kingdom

Dept of Biological Sciences

July 2010

Table of Contents

List of Figures vii

List of Tables x

Acknowledgments xi

Declaration xii

Summary xiii

Abbreviations xiv

1 Introduction 1

1.1 The Roseobacter lineage 2
1.1.1 Distribution, abundance and diversity 2
1.1.2 Genomic diversity 3
1.1.3 Taxonomy of the Roseobacter lineage 4
1.1.4 Physiology 5
1.1.5 Sessile lifestyle and biofilms 6
1.1.6 Role of Roseobacters in cycle of sulfur 8
1.1.7 Ecological impact of DMS 9
1.1.8 Oxidation of atmospheric trace gases by Roseobacters 11
1.2 Marine phages 11
1.3 Bacteriophage abundance in the ocean 12
1.3.1 Phage detection 12
1.3.2 Temporal variation of phage abundance 13
1.3.3 Depth variation 13
1.3.4 Phage production 14
1.3.5 Phage decay 15
1.3.6 Lysogeny and pseudolysogeny 16
1.4 Impact of phage on the biosphere 18
1.4.1 Viral shunt 18
1.4.2 Killing the winner effect on community structure 19
1.5 Impact of phages on bacterial evolution 20
1.5.1 Phages as mediators of lateral gene transfer 20
1.5.2 Phage resistance mechanisms 21
1.5.2.1 Blocking phage adsorption 21
1.5.2.2 Preventing DNA entry 22
1.5.2.3 Targeting phage nucleic acids 23
1.5.2.4 Abortive infection systems 23
1.5.3 Effect of prophages on host fitness 24
1.6 Bacteriophage taxonomy: orders typically found in the ocean 25
1.7 Marine phage genomics 28
1.7.1 General genome architecture 28
1.7.2 Comparative genomics – phage mosaic theory 29
1.7.3 Prophage genomics 31
1.7.4 Viral host genes 31
1.7.5 Metagenomics 33

1.7.6 ORFans 35
1.8 Phage and biofilms 37
1.8.1 Effect of biofilms on phage attachment 37
1.8.2 Phages in natural biofilms 38
1.8.3 Lysogeny within biofilms 39
1.9 Bacteriophages of Roseobacter species 40
1.9.1 SIO1 41
1.9.2 RDJLΦ1 42
1.9.3 DSS3Φ2 and EE36Φ1 42
1.9.4 Inducible prophages in Roseobacter strains 43
1.10 Summary and aims 45
2 Material and Methods 46
2.1 Bacterial strains and culture conditions 47
2.1.1 Roseobacter cultures 47
2.1.2 Determination of cell counts by flow cytometry 48
2.1.3 Growth of *Escherichia coli* strain DH5α 49
2.2 Molecular biology kits 49
2.3 Chemicals 49
2.4 Equipment 49
2.5 Centrifuges and rotors 50
2.6 Media 50
2.6.1 Media for the growth of Roseobacter 50
2.6.1.1 Liquid Medium – MAMS-PY 50
2.6.1.2 Solid media 51
2.6.1.3 Purification of agar 52
2.6.2 Medium for growth of *Escherichia coli* 52
2.6.3 Medium for the storage of phage stocks 52
2.6.4 Media for the supplementation of seawater 53
2.7 Phage techniques 53
2.7.1 Plaque assay 53
2.7.2 Spot test 53
2.7.3 Co-culturing 54
2.7.4 Concentration of phages from environmental samples 54
2.7.5 Phage enrichment 54
2.7.6 Phage purification 55
2.7.7 Production of *Roseovarius* phage stocks 77
2.7.8 Host range 77
2.7.9 Preparation of lytic Roseovarius phage samples for pulsed field gel electrophoresis (PFGE) 77
2.7.10 Caesium chloride purification of Roseovarius phage 78
2.7.11 Transmission electron microscopy (TEM) 78
2.7.12 Preparation and negative staining of grids for electron microscopy 78
2.7.13 Imaging 78
2.7.14 Absorption assays 79
2.7.14.1 Liquid 79
2.7.14.2 Solid 79
2.7.15 Phage infected Roseovarius growth curve 80
2.7.16 Modified Roseovarius phage one-step growth curve 80
2.7.17 Prophage techniques 80
2.7.17.1 Mitomycin C exposure 81

2.7.17.2 Epifluorescent microscopy....81
2.7.17.3 Purification of temperate phage....81
2.8 Molecular biology techniques....82
2.8.1 Phage DNA extraction....82
2.8.2 Bacterial genomic DNA extraction....82
2.8.3 Pulsed field gel electrophoresis (PFGE)....83
2.8.4 Restriction enzyme digestion....83
2.8.5 *Bal31* digestion....84
2.8.6 Polymerase chain reaction....84
2.8.7 Primer design....85
2.8.8 PCR product clean-up....86
2.8.9 Agarose gel electrophoresis....86
2.8.10 Genome sequencing....86
2.8.11 Contig assembly....87
2.8.12 TA cloning....87
2.8.13 DNA sequencing....88
2.8.14 DNase digestion....88
2.9 Protein techniques....88
2.9.1 Preparation of phage structural proteins....89
2.9.1.1 Whole phage extraction....89
2.9.1.2 Phage ghosts....89
2.9.1.3 Trichloroacetic acid extraction of phage proteins....89
2.9.2 Protein concentration determination....89
2.9.3 Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)....90
2.9.4 Coomasie staining of polyacrylamide gels....90
2.9.5 Silver staining of polyacrylamide gels....90
2.9.6 Roseobacter outer membrane protein enrichment....91
2.9.6.1 OMP enrichment of liquid grown *Roseovarius*....91
2.9.6.2 Harvesting plate grown *Roseovarius*....92
2.9.7 Lipopolysaccharide (LPS) extraction....92
2.9.8 SDS-PAGE analysis of LPS....92
2.9.9 Silver staining of modified polyacrylamide gels....93
2.9.10 Mass spectrometry (MS)....93
2.10 Bioinformatic analyses....94
2.10.1 Open reading frame (ORF) identification....94
2.10.2 Database searches....94
2.10.3 ORF/genome comparisons....94
2.10.4 Motif and regulatory element identification....95
2.10.5 Phylogenetics....95
2.10.6 Prophage finder....96
2.10.7 Protein domain prediction....96
3 Isolation and characterization of two *Roseovarius* bacteriophages....77
3.1 Introduction....78
3.2 Results and Discussion....80
3.2.1 Isolation attempt 1: Co-culturing....80
3.2.2 Isolation attempt 2: Concentration of Environmental Samples....81
3.2.3 Isolation attempt 3: Viral enrichment/amplification....81
3.2.4 Plaque morphology....82

3.2.5 *Roseovarius* phage morphology and classification 82
3.2.6 Host range 84
3.2.7 Pulsed field gel electrophoresis 86
3.2.8 DNA restriction pattern 88
3.2.9 One-step growth assay 89
3.3 Concluding comments 93
4 Binding properties of *Roseovarius* phages RLP1 and RPP1 99
4.1 Introduction 100
4.2 Results and Discussion 103
4.2.1 Infection 103
4.2.2 Hypothesis 1 105
4.2.2.1 Liquid adsorption assay – method 1 105
4.2.2.2 Liquid adsorption assay – method 2 105
4.2.3 Hypothesis 2 106
4.2.3.1 Shaken batch cultures 106
4.2.4 Hypothesis 3 107
4.2.4.1 Age of culture adsorption assay 108
4.2.5 Hypothesis 4 109
4.2.5.1 Types of Marine agar plates 109
4.2.6 Hypothesis 5 100
4.2.6.1 Liquid adsorption assay of plate grown cells 100
4.2.6.2 Plate adsorption assay 101
4.2.7 Surface receptor comparison 103
4.2.7.1 Outer membrane protein (OMP) enrichment 103
4.2.7.2 LPS 103
4.3 Concluding comments 106
5 Genome characterisation and analysis of *Roseovarius* phage RLP1 and RPP1 108
5.1 Introduction 109
5.2 Results and Discussion 111
5.2.1 Genome assembly 111
5.2.1.1 RPP1 111
5.2.1.2 RLP1 112
5.2.2 General genome properties 114
5.2.3 Restriction digests 114
5.2.4 *Bal31* digestion 115
5.2.5 Identification of open reading frames 116
5.2.6 vRNAP 123
5.2.7 Single-stranded DNA-binding protein 124
5.2.8 Host-like genes 125
5.2.9 Major capsid protein 128
5.2.10 Regulatory elements 129
5.2.11 Gene module order 130
5.2.12 Comparative genomics 133
5.2.12.1 RLP1 vs RPP1 133
5.2.12.2 N4-like Roseobacter phages (RN4-like) 134
5.2.12.3 N4-like phages 134
5.2.13 Bipartite genome 139
5.3 Conclusions 141

6 Proteomic analysis of *Roseovarius* phages RLP1 and RPP1 141
6.1 Introduction.....142
6.2 Results and Discussion 143
6.2.1 Optimisation of protein extraction 143
6.2.2 Virion structural proteins 144
6.2.3 Conserved N4-like virion proteins 149
6.2.4 Structural proteins gp13/14, gp15 and gp16 151
6.2.5 Structural protein gp33 152
6.2.6 Structural proteins gp64 – 71 153
6.3 Concluding comments 154
7 Induction and characterisation of temperate phages from Roseobacter hosts.....158
7.1 Introduction.....159
7.2 Results and Discussion 161
7.2.1 *In silico* analysis of Roseobacter species 161
7.2.2 Induction of phage-like particles by exposure to Mitomycin C 164
7.2.3 Confirmation of induction.....166
7.2.3.1 Epifluorescence microscopy 166
7.2.3.2 TEM 167
7.2.4 Pulsed field gel electrophoresis 169
7.2.5 Host range 174
7.2.6 Restriction digest pattern 174
7.3 Concluding comments 175
8 General discussion 177
8.1 General discussion 178
8.1.1 Genome sequencing of two new N4-like phages: implications for the N4-like genus 178
8.1.2 Shared peripheral genes found in Roseobacter phages.....180
8.1.3 Definitions of phage genera and the evolution of tailed phages.....181
8.1.4 What is a phage species?.....181
8.1.5 Control of planktonic and sessile phenotypes.....183
8.1.6 Comparison of agar plates and biofilms 184
8.2 Future work and prospects.....184
8.2.1 Phage-based studies 184
8.2.2 *Roseovarius*-based studies 185
9 Bibliography.....187
10 Appendix.....206

List of Figures

Figure 1.1 A consensus maximum likelihood tree based on the alignment of a concatemer of 70 universal single-copy genes present in the 32 sequenced Roseobacter genomes.....5
Figure 1.2 A model of the molecular mechanisms involved in the symbiosis between Roseobacter and phytoplankton.....8
Figure 1.3 The feedback system linking DMS-producing oceanic plankton and climate through the production of atmospheric sulfur and cloud albedo as proposed by Charlson *et al.* (1987). 10
Figure 1.4 Schematic representation of the four possible phage lifecycles. 15
Figure 1.5 The viral shunt in the marine food web. 18
Figure 1.6 Lytic phage replication cycle. 21
Figure 1.7 Blocking the entry phage of DNA using *E. coli* proteins Imm and Sp. 23
Figure 1.8 Basic morphology of the three phage families of the *Caudovirales* order. 26
Figure 1.9 Genomic re-organizations and transfers between four T7-like phages. 30
Figure 1.10 Illustration of the growing number of phage ORFans. 36
Figure 1.11 Plate showing the "halo" caused by a typical polysaccharase-inducing phage..... 38
Figure 1.12 Hypothesized lifestyle for *B. anthracis* in the environment. 40
Figure 3.1 Plaques formed on *Roseovarius* bacterial lawns. 82
Figure 3.2 TEM micrograph of RLP1 stained with uranyl acetate. 83
Figure 3.3 TEM micrograph of RPP1 viral particles stained with uranyl acetate. 83
Figure 3.4 Phylogram (made using MrBayes) based on average branch length of 16S rRNAgenes from 18 *Roseovarius* species.....86
Figure 3.5 Pulsed field gel electrophoresis of purified RLP1 and RPP1 genomic DNA. 87
Figure 3.6 Restriction pattern of digested phage DNA. 88
Figure 3.7 Number of free phage in the agar/media during a one-step growth curve experiment.....91
Figure 3.8 One-step growth curves for RLP1 and RPP1. 92
Figure 3.9 Image of the isolation locations for RPP1 and *Rsv. nubinhibens*. 93
Figure 3.10 The North Atlantic Gyre and currents in the North Atlantic Ocean. 94
Figure 3.11 Sampling locations of BLAST search hits using the 16S rRNA sequence of *Rsv. nubinhibens* as query against the Global Ocean Survey (GOS) database (7/1/10).....94
Figure 3.12 SYBR Green I stained samples of 12 hr ACR04 cultures. 95
Satellite view of the South coast of the UK indicating the isolation location of RLP1, in Langstone Harbour, and *Rsv.* 217, in L4. 95
Figure 3.13 A Landsat image of the South West coast of UK of a coccolithophore bloom. 97
Figure 4.1 Adsorption profile of live (○) and dead (●) bacteria with phage illustrating the velocity of the reaction, taken from Krueger, 1931.....100
Figure 4.2 Liquid batch cultures of a) RLP1 with *Rsv.* 217 and b) RPP1 with *Rsv. nubinhibens* infected at different MOIs. 103
Figure 4.3 Spot test plates of a) RLP1 on *Rsv.* 217 and b) RPP1 on *Rsv. nubinhibens*. 104
Figure 4.4 Adsorption of RLP1 to *Rsv.* 218 and RPP1 to *Rsv. nubinhibens* in liquid media over five hours. 105

Figure 4.5 Adsorption assay of RLP1 on *Rsv.* 217 and RPP1 on *Rsv. nubinhibens* at an MOI of 0.1, in 2.8 ml of marine broth. 106
Figure 4.6 Change in a) *Rsv.* 217 and b) *Rsv. nubinhibens* cell number over 6 days when infected with RLP1 and RPP1, respectively, at a number of MOIs 107
Figure 4.7 Adsorption assays using host cells of different age. a) RLP1 on *Rsv.* 217 and b) RPP1 on *Rsv. nubinhibens*. 108
Figure 4.8 Growth of host bacterial cells over the period of 5 days. 109
Figure 4.9 Spot test plates of RLP1 on *Rsv.* 217 with a) Bacto agar, b) purified agar and c) agarose as a setting agent. 110
Figure 4.10 Spot test plates of RPP1. 99
Figure 4.11 Adsorption assay of plate grown cells. 100
Figure 4.12 Adsorption of a) RLP1 with *Rsv.* 217, b) RPP1 with *Rsv. nubinhibens*. 101
Figure 4.13 Adsorption assay of) RLP1 with *Rsv.* 217, b) RPP1 with *Rsv. nubinhibens* carried out both in liquid (filled squares) and on solid (open squares) media. 102
Figure 4.14 SDS-PAGE analysis of a) *Rsv.* 217 and b) *Rsv. nubinhibens* outer membrane proteins grown on plate (P) and in liquid (L). 104
Figure 4.15 Modified SDS-PAGE analysis of a) *Rsv.* 217 and b) *Rsv. nubinhibens* LPS grown on plate (P) and in liquid (L). 105
Figure 4.16 Comparison of AHL's dectected from shaken and static grown cultures of Roseobacter, taken from Bruhn *et al.*, 2007. 107
Figure 5.1 Assembly of RPP1 contig.. 112
Figure 5.2 Contigs of RLP1 mapped onto a RPP1 scaffold. 113
Figure 5.3 PCR to confirm the predicted order of contigs. 113
Figure 5.4 Restriction digest of purified RLP1 and RPP1 genomic DNA with rare cutters.. 115
Figure 5.5 *Nde*1 digested a) RLP1 and b) RPP1 genomic DNA after treatment with *Bal31* for the indicated time intervals.. 115
Figure 5.6 *Psi*I digest of phage genomic DNA.. 116
Figure 5.7 *In silico* analysis and comparison of the two *Roseovarius* phages RLP1 and RPP1. 122
Figure 5.8 Amino acid sequence alignment of motifs T/DxxGR, A, B and C from RLP1, RPP1, N4 and other T7 superfamily RNA polymerases. 124
Figure 5.9 Analysis of N4-like gp45, the single-stranded binding protein involved in late gene transcription.. 125
Figure 5.10 Phylogram (made using MrBayes) of aligned amino acid sequences of thioredoxin.. 127
Figure 5.11 Phylogram (made using MrBayes) of aligned amino acid sequences of ribonucleotide reductase.. 128
Figure 5.12 Sample origin of best BLAST hits of MCP against the GOS database. 129
Figure 5.13 Module order found in RLP1, RPP1, DSS3Φ2 and N4. 132
Figure 5.14 Dotplot analysis of RLP1 and RPP1.. 133
Figure 6.1 Comparison of protein extraction protocols using purified RPP1 on a 12 % SDS-polyacrylamide separating gel. 144
Figure 6.2 Polypeptides of purified RLP1 and RPP1 phage particles.. 145
Figure 6.3 N4 portal assembly showing the crown, wing and stalk domains.. 149
Figure 6.5 Predicted secondary structure of RLP1/RPP1 gene product 16. 152
Figure 6.6 Location of structural proteins on RPP1. 156

Figure 7.1 Comparison of treated (■) and untreated (□) *Roseovarius* 217 cultures measured by change in optical density at 600 nm. 164
Figure 7.2 Comparison of Mitomycin C-treated (■) and untreated (□) Roseobacter cultures assessed by measuring change in optical density at 600 nm over time.....165
Figure 7.3 SYBR Green I stained samples of Mitomycin C-treated and untreated/control ACR04 cultures. 166
Figure 7.4 SYBR Green I stained samples of 12 hr ACR04 cultures. 167
Figure 7.5 Electron micrographs of induced prophage from a) *Marinovum algicola*, b) ACR04 and c) "*Ruegeria*" sp. 198. 168
Figure 7.6 Caesium Chloride gradients of induced prophages from *Marinovum algicola* and ACR04. Two whitish phage bands are present in both tubes. 169
Figure 7.7 PFGE of lysogens and purified induced prophage. 170

List of Tables

Table 1.1 Genera in the order of *Caudovirales.* 27
Table 1.2 The five most common viral-encoded proteins found in four oceanic viriomes.....34
Table 1.3 Table of isolated Roseobacter phages (as of May 2010). 41
Table 2.1 Warwick Roseobacter collection.....47
Table 2.2 Chemical composition of MAMS-PY.....50
Table 2.3 Chemical composition of MS Stock solution. 51
Table 2.4 Chemical composition of Na_2WO_4 solution. 51
Table 2.5 Chemical composition of PO_4 Stock.....51
Table 2.6 Chemical composition of Trace elements, SL10. (Widdel *et al.*, 1983) 51
Table 2.7 Chemical composition of Vitamin solution. (Kanagawa *et al.*, 1982).....51
Table 2.8 Chemical composition of LB. 52
Table 2.9 Chemical composition of modified ASW (Wilson *et al.*, 1996).....53
Table 2.10 Inoculum groups for phage enrichment. 56
Table 2.11 Commonly used buffers and reagents. 82
Table 2.12 Restriction enzymes. 84
Table 2.13 Components of a typical PCR reactions.....84
Table 2.14 Primers utilized. 85
Table 2.15 Table of common reagents and buffers.....88
Table 2.16 Composition of polyacrylamide separating gels. 90
Table 2.17 Programs used in bioinformatic analyses.....95
Table 3.1 Host range of RLP1 and RPP1.....84
Table 3.2 Genera of the Podoviridae family. 87
Table 3.3 Morphological types of free viruses in the marine environment. 99
Table 4.1 Surface receptors of various bacteriophages.....102
Table 5.1 RLP1 contigs assembled from pyrosequencing. 113
Table 5.2 *in silico* predicted restriction enzyme sites. 114
Table 5.3 *In silico* analysis of the RLP1 and RPP1 genomes. 118
Table 5.4 Best BLAST hits of MCP from the GOS database. 128
Table 5.5 Codon usage in the Roseovarius phages and their hosts.....130
Table 5.6 Comparison of RLP1 and RPP1 to various N4-like phages. 136
Table 5.7 Core genes for N4-like phage genus. 140
Table 6.1 Structural proteins in Enterobacteria phage N4. 143
Table 6.2 Identification of proteins in bands analysed by mass spectrometry.....146
Table 6.3 Proteins identified by mass spectrometry. 148
Table 7.1 Summary of Prophage Finder results.. 161
Table 7.2 Summary of Prophage Finder results for *Rsv.* 217 predicted prophages 4 and 6.....162
Table 7.3 Virion sizes of induced prophages. 168
Table 7.4 List of temperate phages and the lysogen from which they were induced. 174

Firstly, I would like to thank my supervisors Nick Mann and Hendrik Schäfer for all their help and support over the last four years. I would also like to thank the members of my committee, Prof Andrew Easton, Prof Colin Murrell and Prof Liz Wellington for their suggestions and guidance.

Katrin and Andy, thank you for all your help and for acting as both metaphorical (on Tues and Thurs) and actual (on Mon, Weds and Fridays) punch-bags when needed. Edd, thanks (I think) for all your pearls of wisdom and for stealing my things. Just so you know, it was always a two way thing, I'm just a little more subtle. John, thank you for helping me bring up the many plates from stores and Shivani for letting me steal you P20 all the time. Fran, things would definitely not have gone half as well if you hadn't been there to offer advice or just to talk about our collective pets. I would also like to thank the rest of Micro II, past and present, for putting up with my ~~random~~ philosophical questions and their forbearance of cake Wednesday (long may it survive).

Thank you to Jess, Fay and Nayana who have all, at some point, have been the unwitting participants of long phone calls mainly consisting of me venting over failed experiments. You have another three years of it to come!

Nick, thank you for reminding me there is a world outside of the lab even if this new world is mostly filled with computer games and sky tv.

Finally I'd like to thank my family for all their love and support. Yes, I will get a "proper" job now, well sort of.

Declaration

I declare that the work presented in this thesis was conducted by me under the direct supervision of Prof Nicholas Mann and Dr Hendrik Schäfer, with the exception of those instances where the contribution of others has been specifically acknowledged.

None of the work presented here has been previously submitted for any other degree.

Jacqueline Chan

Summary

The oceans cover *ca.* 70% of the Earth's surface and due to their depth encompass around 300 times the habitable volume of the terrestrial environment. The exact proportion of life on Earth that exists in the oceans is unknown as many ocean species remain undiscovered; in particular this holds true for the viruses that infect marine bacterioplankton.

It is currently thought that viruses that infect bacteria, bacteriophages or phages, can numerically exceed their hosts by a factor of ten, however, this abundant and diverse group of organisms is still poorly understood. This is especially true of phages that infect members of the Roseobacter clade. Globally, members of the Roseobacter lineage can comprise up to a quarter of the marine microbial community and often dominate the alga-associated bacterial community. In this study phages capable of infecting species of Roseobacter were isolated and characterised.

Two *Roseovarius*-specific phages, RLP1 and RPP1, were isolated from UK coastal waters; morphological and sequence data identified them as belonging to the N4-like genus of *Podoviridae*. Comparative genomic analysis of both *Roseovarius* phages to other N4-like phages such as *Escherichia coli* phage N4 and *Sulfitobacter* sp. EE-36 phage EE36Φ1, revealed a number of conserved core genes involved in DNA metabolism, transcription control and virion structure. Comparison of N4-like Roseobacter phages (RLP1, RPP1, EE36Φ1 and *Ruegeria pomeroyi* DSS-3 phage DSS3Φ2) also revealed a number of peripheral genes which are likely to interact directly with host proteins/machinery specific to the Roseobacter group.

Unusually, both RLP1 and RPP1 appeared to only infect host cells when in semi-solid agar matrix, but not in liquid culture. Comparison of the outer surface of agar-embedded and planktonic cells revealed different outer-membrane protein and lipopolysaccharide expression profiles. This suggests that some Roseobacter species (spp.) change components of their bacterial cell surface according to their physiological state: agar-embedded/sessile or planktonic and RLP1 and RPP1 exploit this by binding to (a) receptor(s) only expressed during sessile conditions.

A number of prophage-like elements were also induced from three Roseobacter spp. by exposure of growing cultures to the DNA-damaging chemical Mitomycin C. These were identified by electron microscopy as belonging to the *Siphoviridae* family.

The results of this project suggest that within the marine environment there remain many uncharacterised phages with peculiar biochemical properties and a wealth of genomic information.

Abbreviations

°C	degree(s) Celsius
Å	Angstrom
AAnP	Aerobic anoxygenic photosynthesis
Abis	Abortive infection systems
AHL	Acyl homoserine lactone
Amp	ampicillin
ASW	Artificial seawater
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	base pairs
CAMERA	Community Cyberinfrastructure for Advanced marine Microbial Ecology Research and Analysis
Cas	CRISPR associated
CCN	Cloud-condensation nuclei
CFU	colony forming unit
CRISPR	Clustered Regularly Interspaces Short Palindromic Repeats
CsCl	Caesium chloride
Cys	cysteine
Da	dalton(s)
DMS	Dimethylsulfide
DMSP	Dimethysulfoniopropionate
DNA	Deoxyribose nucleic acid
DNAP	DNA polymerase
DOM	Dissolved organic matter
ds	double-stranded
E. coli	*Escherichia coli*
EcoSSB	*E. coli* single stranded DNA binding protein
e.g.	exempli gratia
EB	elution buffer
EDTA	ethylenediamine tetraacetic acid
EOP	efficiency of plating
EPS	extracellular polysaccharide
Fig.	figure
g	standard gravity
g	gram(s)
Gln	glutamine
GOS	Global Ocean Survey
gp	gene product
GTA	gene transfer agent
H	Hour(s)

i.e.	id est
ICTV	International Committee on Taxonomy of Viruses
Ile	isoleucine
ISM	*Roseovarius nubinhibens* sp. ISM
k(x)	kilo(unit): 10^3
Kan	kanamycin
kb	kilobases
KtW	Killing the winner
L	Litre
LB	Lysogeny Broth
LGT	Lateral gene transfer
LPS	lipopolysaccharide
M	molar (mol L-1)
m(x)	milli(unit): 10^{-3}
M. algicola	*Marinovum algicola*
MAMS	Marine Ammonium Mineral Salts
MAMS-PY	Marine Ammonium Mineral Salts - peptone yeast
MB	Marine broth
min	minutes
MOI	Multiplicity of infection
MRSA	methicillin-resistant *Staphylococccus aureus*
MS	Mass spectrometry
MS/MS	tandem mass spectrometery
n(x)	nano(unit): 10^{-6}
n.b	nota bene
NCBI	Nation Centre for Biotechnology Information
nt	nucleotide
O_2	Oxygen
OD	optical density
OMP	outer membrane protein
ORF	Open reading frame
P	Phosphate
p(x)	pico(unit): 10^{-9}
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFGE	Pulsed field gel electrophoresis
PFU	Plaque forming unit
Pro	proline
psi	pound per square inch

PSII	Photosystem *II*
QS	Quorum sensing
R/M	Restriction modification
RLP1	Roseovarius Langstone Podovirus 1
RN4-like	Roseobacter N4-like
RNAP	RNA polymerase
rnr	ribonucleotide reductase
rpm	revolutions per minute
RPP1	*Rose*ovarius *Plymouth Podovirus 1*
rRNA	Ribosomal-RNA
RNA	ribonucleic acid
Rsv.	*Roseovarius*
Rsv. 217	"*Roseovarius*" 217
Rsv. nubinhibens	*Roseovarius nubinhibens*
SDS	sodium dodecyl sulfate
sec	second(s)
SET	Sucrose EDTA TE
SOG	Strait of Georgia
spp.	species
ss	single-stranded
SSB	ssDNA-binding protein
T4SS	Type 4 secretion system
TBE	Tris borate EDTA
TCA	trichloracetic acid
TDA	Tropoditheietic acid
TE	Tris EDTA
TEM	Transmission electron microscopy
TEMg	Tris EDTA Mg
TM1040	*Silicibacter* sp. strain TM1040
tRNA	transfer-RNA
trx	thioredoxin
UA	uranyl acetate
UV	Ultra-violet
v	volume
V	volt(s)
VLP	virus-like particle
vRNAP	virion RNA polymerase
w	weight
μ(x)	micro(unit): 10^{-6}

Chapter 1

Introduction

Introduction

The oceans cover *ca.* 70% of the Earth's surface and due to their depth, encompass around 300 times the habitable volume of the terrestrial environment. The exact proportion of life on earth that exists in the oceans is unknown as many ocean species remain undiscovered; in particular this holds true for the viruses that infect marine bacterioplankton.

It is currently thought that viruses that infect bacteria, bacteriophages or phages, can numerically exceed their hosts by a factor of ten (Wommack and Colwell, 2000) however, this abundant and diverse group of organisms is still poorly understood. This is especially true of phages that infect members of the Roseobacter clade. The isolation and characterisation of phages capable of infecting species of *Roseovarius* will be discussed here.

1.1 The Roseobacter lineage

In spite of the wealth of bacterial diversity present in the world's oceans, the majority of marine bacteria fall into as few as nine major clades (Buchan *et al.*, 2005) of which the Roseobacter clade is one. Species of this group are Gram-negative, usually ovoid or rod-shaped cells that grow at mesophilic temperature ranges and, apart from the genus *Ketogulonicigenium* (Urbance *et al.*, 2001) and clones from a South African gold mine, members are found in marine or hypersaline habitats (Buchan *et al.*, 2005). All members cluster closely together within the *Rhodobacteraceae* family of α-Proteobacteria, but there is a comparatively large, up to 11%, sequence variation amongst the 16S rRNA genes (Newton *et al.* 2010). Descriptions of Roseobacter species, *Roseobacter denitrificans* and *Roseobacter litoralis* first appeared in 1991 (Shiba, 1991), since then the group has expanded to at least 45 described genera (Newton *et al.* 2010) with wide-ranging physiologies exploiting various ecological niches.

1.1.1 Distribution, abundance and diversity

Isolates and Roseobacter-specific 16S rRNA clones have been obtained from a variety of marine habitats ranging from coastal seawater to open oceans, marine snow,

a number of micro- and macro-algae, microbial mats, sediments, polar sea ice, hydrothermal vents and marine invertebrates (Brinkhoff *et al.*, 2008). In addition, culture-independent studies have shown Roseobacter species to be abundant in phytoplankton blooms, on the surfaces of algae and dinoflagellates or in association with surfaces or particles (Zubkov *et al.*, 2001; Mayali *et al.*, 2008; Miller and Belas, 2004; Dang and Lovell, 2000; 2002). Consequently, surface colonization and a propensity for a sessile lifestyle have been suggested as traits of Roseobacters (Slightom and Buchan, 2009, see Section 1.1.5).

One of the first studies to show the abundance of Roseobacter spp. was carried out by González and Moran (1997); they found a group belonging to the α-subclass of the class *Proteobacteria*, later recognised as the Roseobacter clade, was numerically dominant in coastal seawater. However, due to differences in methodology, habitat and seasonal variability, subsequent studies of Roseobacter abundance have produced differing results. For example, screening of a bacterial artificial chromosome (BAC) library from different depths in Monterey Bay indicated that *Silicibacter*-like sequences represented 21.1% at 0 m and 23.6% at 80 m depths (Suzuki *et al.*, 2004). In contrast, a shotgun clone library from the Sargasso Sea suggested that they only represented 3% (Venter *et al.*, 2004). Nevertheless, it is generally believed that the Roseobacter clade comprises 20% of the coastal and 15% of the mixed-layer ocean bacterioplankton community (Buchan *et al.*, 2005) and as such is considered a highly abundant group of bacteria found throughout the marine biosphere.

1.1.2 Genomic diversity

To date, the genomes of 32 Roseobacter species have been sequenced (5 closed and 27 draft) and the size of their genomes range from 3.5 to 5.4 Mbp in size. *Loktanella vestfoldensis* SKA53 has the smallest genome of 3.06 Mbp, whilst *Roseovarius* sp. HTCC2601 has the largest, 5.4 Mbp (for a list of sequenced species see Appendix Table A1). Unlike the SAR11 clade which exhibits "genome streamlining" (Giovannoni *et al.*, 2005), Roseobacter members have been found to have multiple mechanisms for sensing and reacting to their environments, thus allowing them to acquire a diverse range of substrates and nutrients for growth (Brinkhoff *et al.*, 2008). Many members also contain extrachromosomal elements such as circular and linear plasmids ranging from 4.3 to 82.7 kb; for example,

Dinoroseobacter shibae harbours seven linear plasmids which comprise 20% of its genome content (Wagner-Döbler *et al.*, 2010). As in other genome projects, a relatively high proportion of the predicted genes are unknown; in the five completed Roseobacter genomes, 28% are without "known" function (Brinkhoff *et al.*, 2008). Despite this, it is clear that the Roseobacter lineage contains a considerable degree of genomic variability and diversity which is perhaps a reflection of their physiological and ecological diversity as well as an explanation for their successful dominance in the marine habitat.

1.1.3 Taxonomy of the Roseobacter lineage

The taxonomy of the Roseobacter group is fraught with problems primarily due to the rapid increase in the number of described species and genera in the last twenty years. Classification of new isolates based solely on 16S rRNA gene sequences is become increasingly problematic as differences on a genus level often fall below 4%. For example, the difference between the 16S rRNA encoding genes of *Ruegeria atlantica* and *Silicibacter lacuscaerulensis* is only 1.7% which suggests they should belong to one genus. However, they possess differential phenotypes such as presence of inclusion bodies and flagella, which argues against such a classification. Nevertheless, phylogenetic analyses of the Roseobacter lineage have been carried out, the most recent of which was based on a concatenation of 70 universal single-copy genes present in each of the 32 sequenced Roseobacter genomes (Newton *et al.*, 2010). This tree suggests the lineage is comprised of five deeply branching clades, see Figure 1.1. Supporting this phylogenetic inference is the ability to map twelve of the 13 major 16S rRNA sequence clusters described by Buchan *et al.*, in 2005 onto the tree.

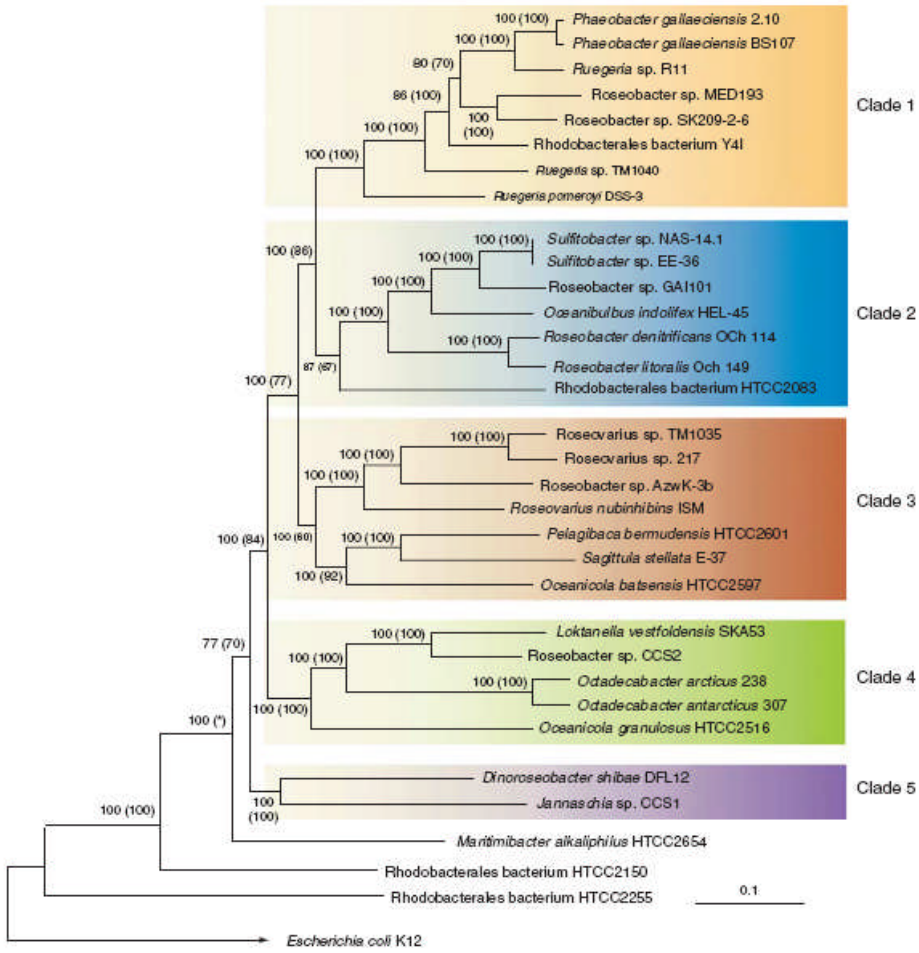


Figure 1.1 A consensus maximum likelihood tree based on the alignment of a concatemer of 70 universal single-copy genes present in the 32 sequenced Roseobacter genomes. Clades 1-5 are shown on the right, taken from Newton *et al.* 2010

1.1.4 Physiology

Members of the Roseobacter lineage have a wide range of physiologies, a fact which has led to their intensive study for the last twenty years. *Roseobacter denitrificans*, the first Roseobacter representative, was found to carry out aerobic anoxygenic photosynthesis (AAnP), photosynthesis performed in the presence of oxygen, but without the generation of oxygen, a trait subsequently found to be shared by other clade members. Since then other species have been identified as being capable of oxidation of carbon monoxide, methyl halides, degradation of multiple sulfur compounds (see Section 1.1.6) and aromatic compounds, reduction of trace metals, and production of a variety of bioactive secondary metabolites such as acyl homoserine lactone (AHL) signalling molecules involved in quorum sensing and

tropoditheietic acid (TDA), a sulfur-containing antibiotic (Wagner-Döbler and Biebl, 2006, Geng *et al.*, 2008, Schaefer *et al.*, 2002; Schäfer *et al.*, 2005). Several members have also been found in symbiotic (e.g. probiotics for dinoflagellate, *Pfiesteria piscicida*) and pathogenic (causative agent of juvenile oyster disease in Eastern oysters and black band disease in scleractine coral) relationships with both vertebrates and invertebrates (Geng and Belas, 2010; Boettcher *et al.*, 2005; Frias-Lopez *et al.*, 2004; Wagner-Döbler and Biebl, 2006). Due to their broad niches and tolerance for environmental changes, most species in the Roseobacter lineage can be characterized as ecological generalists.

1.1.5 Sessile lifestyle and biofilms

A trap fallen into by many microbiologist is the belief that a suspension culture is the normal state of growth for organisms. Instead, the reality is that most prokaryotes spend part or all of their life attached to surfaces (Marshall, 2006). In 1943, ZoBell suggested that in oligotrophic regions of constant flux, like the oceans, there is an advantage to an attached, sessile lifestyle because when a clean surface is introduced into a habitat, simple soluble macromolecules e.g. glucose and amino acids, and smaller hydrophobic molecules rapidly bind, forming a nutrient rich, molecular film. By colonising the surface, the probability of access to the nutrients accumulated on it increases, and so the bacterium thrives (Zobell, 1943); this was later proved by Jannasch in 1958.

When a community of microorganisms aggregate and adhere to a surface, the biological layer is termed a biofilm. As well as cells, a matrix of extracellular polysaccharides (EPS) is produced (by the organisms present) and acts as a mechanically stable, protective layer which helps to facilitates cell-to-cell communication through the diffusion of biochemical signalling molecules. The microorganisms living within biofilms are sessile, attached to a surface, and so display different phenotypes to compared their planktonic, unicellular counterparts (Marshall, 2006).

The propensity of many Roseobacter isolates to form associations such as biofilms on living or non-living surfaces is becoming increasingly apparent. As mentioned previously, studies have shown that Roseobacter clade members are the most common and dominant primary surface colonizers (Dang and Lovell, 2000) and

new species continue to be isolated from surfaces. Two recent examples are *Nautella italica* and *Ruegeria scottomollicae* which were isolated from marine electroactive biofilms (Vandecandelaere *et al.* 2008; 2009). Furthermore, examination of cultured Roseobacter species have confirmed the presence of many characteristics expected in sessile/surface-associated bacteria, such as the possession of holdfast structures, motility, chemotaxis and the production of quorum-sensing molecules and antimicrobial metabolites (for a review on the surface colonization features found in Roseobacters see Slightom and Buchan, 2009). Indeed the phenotypic differences caused by a sessile lifestyle were highlighted in a paper by Bruhn *et al*. (2007) which found the presence of antibacterial compounds in the filter-sterile supernatants obtained from cultures of *Silicibacter* sp. strain TM1040 (hereafter referred to as TM1040) and *Phaeobacter* strain 27-4, increased when the bacteria were grown in static rather than shaken conditions. In addition, these two species appeared to be predisposed towards attachment and biofilm formation when pre-cultured under static conditions; static growth also resulted in formation of rosettes. Collectively, these results suggest that many Roseobacter species have a biphasic "swim-or-stick" lifestyle i.e. they can either take the form of motile cells which respond to molecules via chemotaxis or sessile cells which readily attach to surfaces (Geng and Belas, 2010). An example of this can be found in the establishment of symbiosis between the dinoflagellate *Pfiesteria piscicida* and TM1040. This species possesses three flagella and is attracted to dinoflagellate homogenates, specifically to the amino acids and to DMSP or its metabolites (Miller *et al*., 2004). Once in close proximity, TM1040 suspends motility and forms rosettes, which is thought to aid attachment and development of a biofilm on the surface of the dinoflagellates. Concurrent with this process is the production of TDA, a potential probiotic for the dinoflagellates (Geng *et al*., 2008), see Figure 1.2. Though it is possible to observe the physiological and morphological changes when a bacterium "switches" between the planktonic/motile and biofilm/sessile state, the molecular mechanism(s) involved remain unknown, for a review of Roseobacter/phytoplankton symbioses and the processes involved see Geng and Belas, (2010).

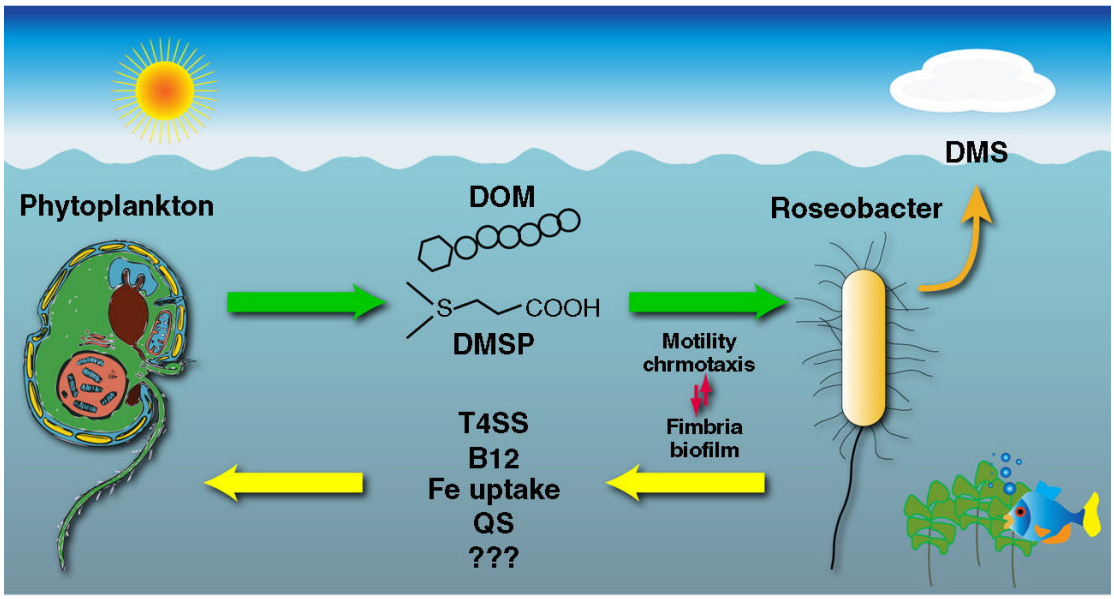


Figure 1.2 A model of the molecular mechanisms involved in the symbiosis between Roseobacter and phytoplankton. Taken from Geng and Belas (2010). Whilst in the planktonic form the Roseobacter is motile; if it encounters an attractant such as (Dimethysulfoniopropionate) DMSP it "swims" up the chemical gradient (by chemotaxis). Near the surface of the algal cell, an unknown molecular "switch" is flipped and the cell transforms to its sessile morphology where it has access to dissolved organic matter (DOM). The transformation involves the loss of flagella, formation of fimbrial adhesins and/or holdfasts which collectively allow it to "stick" to the phytoplankton and form a biofilm. Symbiotic exchanges between phytoplankton and bacteria, which may include the transfer of Vitamin B12 and iron-binding siderophores, are thought to be mediated through quorum sensing (QS) and/or a *vir*-gene-mediated Type 4 Secretion System (T4SS).

1.1.6 Role of Roseobacters in cycle of sulfur

Perhaps the most intensively studied trait of Roseobacter species is its ability to metabolise organic sulfur. Dimethysulfoniopropionate (DMSP) is an organic sulfur compound which acts as a biological protectant, for example it is involved in osmotic regulation in halophytic plants and marine phytoplankton; it is released into seawater by leakage, death or grazing (Vairavamurthy *et al.*, 1985; Dacey and Wakeham, 1986; Nguyen *et al.*, 1998; Wagner-Döbler and Biebl, 2006). It is thought to be the most important sulfur and carbon source for marine bacteria and can metabolised either by single or double demethylation/demethiolation, which results in the uptake of both the carbon- and sulfur-containing moieties of DMSP, or by cleavage, which ultimately produces dimethylsulfide (DMS) and acrylate (Wagner-Döbler and Biebl, 2006; Johnston *et al.*, 2008).

Both laboratory-based and field experiments have implicated members of the Roseobacter group in DMSP transformation (Moran *et al.*, 2003). Some isolates, e.g. *Ruegeria pomeroyi* DSS-3, can carry out both the cleavage and the demethylation step of DMSP; however, this dual metabolic route appears to be the exception rather than the rule (Howard *et al.*, 2006). Around 50% – 85% of the released DSMP is thought to be demethylated, whilst only ~30% is metabolised to DMS (Kiene *et al.*, 2000; Yoch, 2002; Zubkov *et al.*, 2001). Due to their ability to transform DMSP it is perhaps not surprising to find that Roseobacter spp. are often found in association with dinoflagellates, one of the major producers of DMSP, (Miller and Belas, 2004) or that they often dominate the microbial communities associated with DMSP-producing algal blooms (Gonzalez *et al.*, 2000).

To date three mechanisms for the cleavage of DMSP have been identified in marine bacteria, these pathways are mediated by the enzymes DMSP-dependent DMS L (DddL) - a protein of unknown function (Curson *et al.*, 2008), DddD – a predicted acyl Coenzyme A transferase (Todd *et al.*, 2007) and most recently, DddP – a novel lyase (Todd *et al.*, 2009; Kirkwood *et al.*, 2010). Homologues of the *dddD* and *dddL* genes are not found in all DMS-emitting strains (Roseobacter or otherwise) nor do they occur frequently in the (2007) GOS metagenomes (Howard *et al.*, 2008). However, *dddP* is around 10 times more abundant than *dddL* or *dddD* in almost all GOS sites which suggests it is the most prevalent pathway (Todd *et al.*, 2009).

Some Roseobacter isolates also have the ability to transform inorganic sulfur compounds such as elemental sulfur, sulfide, sulfite and thiosulfate (Buchan *et al.*, 2005). Such lithoheterotrophic strains are important for inorganic sulfur oxidation in coastal and benthic marine environments (Buchan *et al.*, 2005; Teske *et al.*, 2000).

1.1.7 Ecological impact of DMS

DMS is a ubiquitous, volatile compound in seawater and has been found to be emitted at a significant rate to the atmosphere, and as such is a key step in the global sulfur cycle (Lovelock, *et al.*, 1972). The total annual release of DMS from the world's oceans ranges from 26 to 45 Tg S $year^{-1}$ (Yoch, 2002) which far exceeds the flux from all other sources, such as soils – 0.29 Tg S $year^{-1}$, plants (excluding tropical forests) – 1.58 Tg S $year^{-1}$, tropical forests – 1.6 Tg S $year^{-1}$, salt marches – 0.07 Tg S $year^{-1}$ and freshwater wetlands – 0.12 Tg S $year^{-1}$ (Watts, 2000). It is a marker for high

marine productivity and has been found to act as a signalling molecule to animals such as birds and seals (Nevitt and Bonadonna, 2005; Kowaleswky *et al.*, 2006). DMS also plays a role in the control of Earth's climate as atmospheric DMS is rapidly oxidised to an acidic aerosol which, as well as having direct heat-reflecting properties, is also a major source of cloud-condensation nuclei, CCN (Charlson *et al.*, 1987). As the albedo or reflectance of clouds is sensitive to CCN density, the Earth's radiation budget and so climate is affected by the amount of DMS emitted (Charlson *et al.*, 1987). It is thought that DMS emission in turn is controlled by the climate for, as the temperature decreases, it has been hypothesised that so too does the DMS output by DMS-producing phytoplankton, thus a negative feedback loop may operate, see Figure 1.3. This feedback system is known as the CLAW hypothesis (originally proposed by Charlson, Lovelock, Andreae and Warren); for a review of the relationship between atmospheric sulfur and oceanic plankton see Simó (2001).

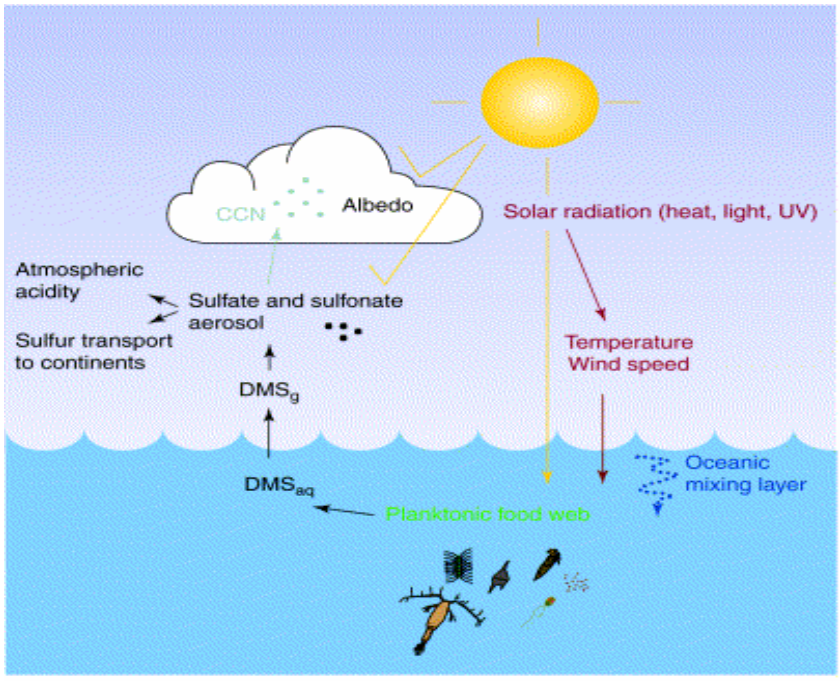


Figure 1.3 The feedback system linking DMS-producing oceanic plankton and climate through the production of atmospheric sulfur and cloud albedo as proposed by Charlson *et al.* (1987). Taken from Simó (2001). The CLAW hypothesis proposes that a negative feedback loop operates between marine ecosystems and the Earth's climate.

1.1.8 Oxidation of atmospheric trace gases by Roseobacters

Methyl halides are found in low atmospheric concentrations, but are significant sources of ozone-depleting halide ions in the troposphere (Butler, 2000). The second largest source of methyl halides are the oceans and some isolates of *Roseobacter* are able to degrade methyl halides, for example *Leisingera methylohalidivorans* and *Roseovarius* sp. 217, strain 198 and strain 179 (Schaefer *et al.*, 2002; Schäfer *et al.*, 2005). As well as this, some *Roseobacter* strains are capable of CO oxidation for example *S. pomeroyi* (Moran *et al.*, 2004) and sp. 217 (Schäfer, University of Warwick, unpublished). Consequently, there is considerable interest in *Roseobacter* with respect to their role in atmospheric chemistry and impact on Earth's climate.

1.2 Marine phages

Phages are believed to be the most abundant group of organisms on Earth and can be found in all habitats. As viral predators of bacteria, they affect many aspects of their prey's life thus influencing global biogeochemical cycles. In the oceans, marine microorganisms are the major driver of biogeochemical cycles and so their viruses can greatly influence of the cycle of various elements.

A native marine phage was defined by Børshein (1993) as "one which parasitizes a bacterial host actually growing in the marine environment". Based on this definition, the first marine phage was isolated in 1955 from seawater samples taken from the North Sea, (10 miles off the coast of Aberdeen, Scotland); it lysed isolates of the luminescent bacteria *Photobacterium phosphoreum* (Spencer, 1955). However, at the time it was thought that the levels of phages in unpolluted seawater were low and therefore ecologically irrelevant. It was not until a study by Bergh *et al.*, in 1989 showed that high numbers of viral particles (~ 2.5×10^8 per ml) could be found in seawater that phages were considered important players in the marine food web.

1.3 Bacteriophage abundance in the ocean

1.3.1 Phage detection

For the past twenty years it has been widely acknowledged that bacteriophages are highly abundant in the sea. In his frequently referenced study, Bergh *et al.* (1989) concentrated the viral fraction of seawater by ultracentrifugation and used transmission electron microscopy (TEM) to directly count total numbers of viral particles. However, this method is both costly and time consuming; consequently, more rapid and cheaper techniques have been developed to enumerate phage abundance in seawater samples.

Culture-based methods e.g. plaque counts and most-probable number assays are sometimes used, but rely on a cultivable host and therefore only a relatively small number of specific phage can be enumerated. Consequently, culture-independent techniques such as direct enumeration of virus-like particles (VLP), which includes phages, through epifluorescence microscopy are preferred. Epifluorescence microscopy is a relatively quick and simple technique for counting VLPs compared to TEM and has been shown to yield similar results (Wommack and Colwell, 2000). As phages have no intrinsic fluorescence, dsDNA-binding fluorochromes (UV-excitable dyes) such as 4', 6-diamidino-2-phenylindole (DAPI), YO-PRO, SYBR Green I and SYBR Gold are used to stain VLPs which can be captured on small-pore-size filters. However, the overlap between the size of small bacterioplankton cells and large VLPs can be a source of error.

More recently, flow cytometric analysis has been utilized in the enumeration of phages (Brussaard *et al.*, 2000). This method has been extensively used by marine microbiologists for the detection and sorting of different populations of cells within mixed samples. During analysis, a laser beam is directed through a sample and each particle within the sample scatters the light. The amount and direction of the scatter is measured by various detectors and the information used to derive the size, shape and chemical nature of the particles present. For example, due to the different measurements and natural fluorescence of *Prochlorococcus*, *Synechococcus*, and picoeukaryotes, it is possible to differentiate and count their numbers in a mixed marine sample. However, as mentioned previously, phages require staining before enumeration due to their lack of natural florescence and, due to their size, are often "lost" in the background noise of stained natural samples. Recent improvements in the

sensitivity of flow cytometers and optimization of protocols, have allowed flow cytometry to become a rapid, high-throughput method for the enumeration of bacteriophages of different morphology and size (Brussaard, 2004).

1.3.2 Temporal variation of phage abundance

Phage abundance in the oceans has been shown to fluctuate according to the seasons, though daily and even hourly changes have been observed (Jacquet *et al.*, 2002; Bratbak *et al.*, 1996). Seasonal variation has been measured in a number of marine environments such as Chesapeake Bay and the Adriatic Sea (Winget and Wommack, 2009; Weinbauer *et al.*, 2004) and these numbers often mirror those of bacterioplankton with higher levels during spring/summer compared to that of autumn and winter. For example in Norwegian coastal waters, levels fell from *ca.* 7 x 10^6 ml^{-1} during spring through to autumn to below 10^4 ml^{-1} in winter (Bergh *et al.*, 1989).

A study by Winget and Wommack (2009) of Chesapeake Bay and coastal Californian surface waters showed that during a period of 24 hours, fairly constant, but significant diel variations in viral production could be measured, and over longer periods of time these patterns also displayed seasonality. However, no significant correlation could be found between viral production and time of day (though the authors noted that this was likely due to seasonal changes). These observations combined with other investigations suggest that there are clear seasonal variations in phage numbers which corresponds to the fluctuations in the host community. In studies based around phytoplankton blooms, these changes are magnified due to the extreme increases in bacterial abundance. Diel changes can also be observed, though other physio-chemical factors such as UV radiation probably play a more significant role in these short-term fluctuations (see Sections 1.4.5 and 1.4.6).

1.3.3 Depth variation

It is not surprising that viral abundance appears to be determined by the factors which also affect the density and productivity of the bacterioplankton community. This is well documented in surface waters, but very little data describing the vertical distribution of bacterioplankton and their viruses (in the ocean) exist. In a

study of the euphotic zone in the western Mediterranean Sea viral abundance was shown to fall from 1 x 10^7 ml^{-1} at seventy meters to 6 x 10^6 ml^{-1} at two hundred meters indicating little variability (Guixa-Boixereu *et al.*, 1999). In contrast an open ocean investigation below 100 m by Boehme *et al.* (1993) in the Gulf of Mexico found an average of 2.4 x 10^4 viral particles per ml of seawater. Deeper still in the bathypelagic zone (below 1000m), Hara *et al.*, (2006), reported a viral concentration of 4 x 10^5 ml^{-1}. However, the authors also concluded the picoplankton abundance at this depth was too low to sustain the reported viral population and instead suggested this community may have originated from suspended and sinking particles (Hara *et al.*, 1996). In contrast to the mixed open ocean, in the permanently stratified Lake Saelenvannet (western Norway) viral numbers increased 2-3 times at the chemocline (layer separating the anoxic, sufidic bottom water from the oxic, top water) compared to that of the surface. However, this also corresponded to the increase in bacteria found at the boundary layer (Tuomi *et al.*, 1997).

1.3.4 Phage production

Production of phage particles is dependent on successful infection of a host bacterium followed by lysis of the host to release the nascent phage. The infection cycle can either be lytic, lysogenic, pseudolysogenic or chronic. During the lytic cycle, the host bacterium's metabolism is hijacked and redirected towards to production of new phages before they are all released in one burst event. In the lysogenic cycle, the phage genome is repressed (i.e. phage genes are not expressed) and either integrates into the host DNA or becomes a self-replicating plasmid (Weinbauer, 2004). It remains in this dormant state replicating alongside the host genome until the lytic cycle is induced. The decision of whether or not to enter the lysogenic state is made after injection of the phage genome and is termed "the lysogenic decision" (Ackermann and DuBow, 1987). During pseudolysogeny, there is a delay between initial infection and lysis of the host similar to that found in the lysogenic cycle. However, during pseudolysogeny, the phage genome is not replicated into all the host progeny. During chronic infection, phage progeny are constantly released from the host by budding or extrusion without lysis. It should be noted that currently there are no examples of tailed phages that undergo chronic infection.

In lytic phage infection, host abundance is a key factor as there is a threshold below which the cycle cannot be maintained though this varies between phage/host systems. Terrestrial phages require a concentration of 10^4 cells ml^{-1} (Wiggins and Alexander, 1985) whereas for marine cyanophages, Suttle and Chan (1994) found that 10^3 cells ml^{-1} were required. In a study by Weinbauer and Peduzzi (1994) of bacteria in the North Adriatic Sea, it was suggested that the host's morphotype also affected the threshold value. They found that a density of ca. 2 x 10^5 rods ml^{-1} was required for infection. However, no threshold value could be observed for cocci and spirillae cells indicating that in these cell morphotypes phage production is not dependent on host density. The authors suggested that this was due to a high percentage of lysogeny in cocci and spirillae cells which would also account for the high infection frequency (79 and 100%, respectively) observed (Weinbauer and Peduzzi, 1994).

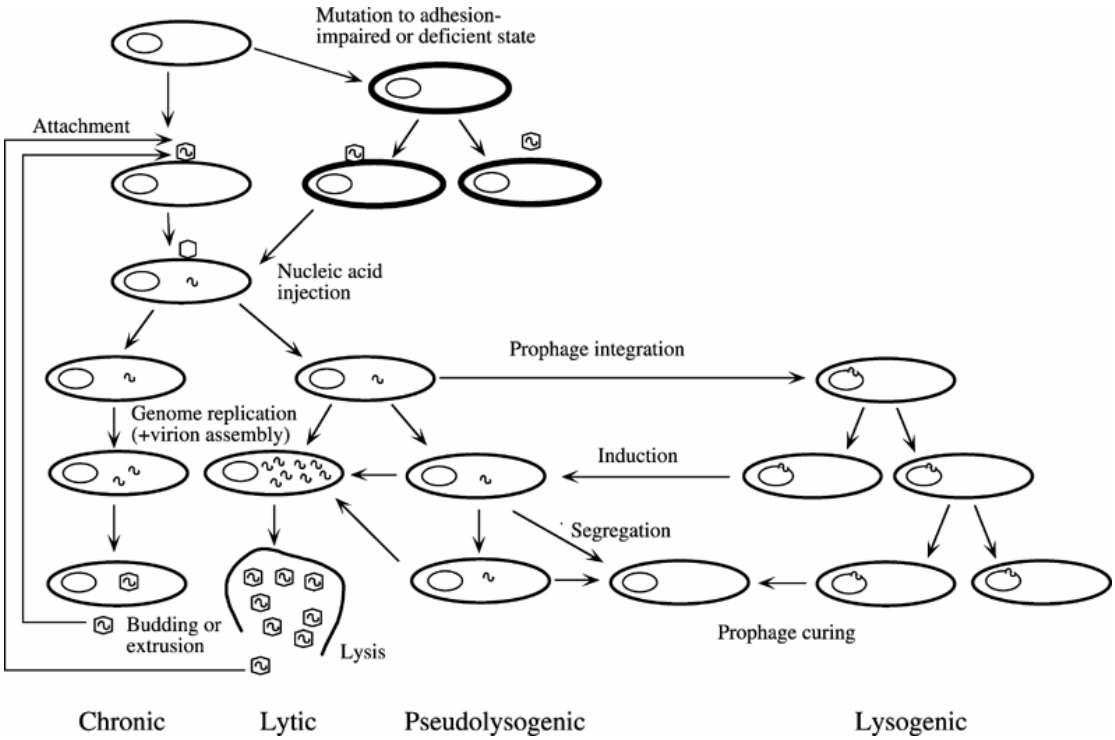


Figure 1.4 Schematic representation of the four possible phage lifecycles. Taken from Weinbauer (2004).

1.3.5 Phage decay

The decay of phages in the environment is dependent on a number of biological, physical and chemical factors; for marine phages, sunlight, specifically the

ultra-violet (UV) fraction, is the major cause of decay (Wommack and Colwell, 2000). Wavelengths of <320 nm (UV-B) generally appear to have the greatest virucidal effect on marine phages, in one study it was responsible for 1/3 to 2/3 of total decay (Noble and Fuhrman, 1997). However, in another study in the Gulf of Mexico, UV-A radiation (320 to 400 nm) was found have to the greatest affect on natural cyanophage populations (Garza and Suttle, 1998). Other factors include grazing, temperature, adsorption on particulate material and salinity though the susceptibility to each of these varies as native phages are resistant or can develop resistance to the challenges of their natural habitat (Wommack and Colwell, 2000).

The hyper-sensitivity of marine phages to sunlight appears counter-intuitive to the high concentrations found in surface water; this paradox can be resolved by the phenomenon of photoreactivation. During this process, sunlight-inactivated bacteriophages are reactivated by host- or phage-mediated DNA repair. In the Gulf of Mexico, it was estimated in the presence of the natural bacterial community, infectivity was restored in 21-26% and 51-52% of damaged phages in ocean and coastal/estuarine waters respectively (Weinbauer *et al.*, 1997).

1.3.6 Lysogeny and pseudolysogeny

The planktonic community, particularly in oligotrophic regions, can be characterised by its relatively low concentrations of slow-growing bacteria. Consequently and in spite of photoreactivation, lytic phages are often left exposed and rapidly decay before they can infect. As such, a lysogenic lifestyle would be advantageous and indeed that does seem to be the case as it is believed that between 21-60% of environmental marine bacteria are lysogens (Miller, 2005) though higher values have been suggested. Stopar *et al.* (2004) found that 71% of their isolates from the Gulf of Trieste were lysogens. A study by Williamson *et al.*, (2008) showed that lysogeny was mainly detected during the winter months (in Tampa Bay, Florida) during periods of low primary and bacterial production, nutrient input and water temperature. Similar findings have led to the suggestion that lysogeny is a survival strategy for phages when host cell abundance is low.

Levels of lysogeny can also vary between marine environments, for example, Jiang and Paul (1996) found that 11 out of 15 coastal/estuarine water samples showed evidence of phage induction compared to only 3 of 11 open water/oligotrophic

samples. This appears to contradict the previous statement that lysogeny is prevalent when bacterial abundance is low, however, it has been suggested that: a) the increased metabolic activity of the coastal/estuarine bacterioplankton also increases their susceptibility to inducing agents, b) the toxicity of the agents is greater for oligotrophic bacteria and c) counting errors and the naturally low viral levels in oligotrophic waters masked the changes in direct viral counts and so lowered the detection of lysogeny (Wommack and Colwell, 2000).

Another caveat to the marine lysogeny story is the lack of a single, reliable method for induction of all bacterial lysogens; researchers have tended to rely on UV irradiation or exposure to the antibiotic mitomycin C, though other chemical agents such as the pollutant naphthalene can also induce prophages (Jiang and Paul, 1996). Due to this, the current estimation of the incidence of lysogeny calculated by direct experimental means maybe still an underestimation. Nevertheless, it is generally believed that in eutrophic, productive areas, lytic phages dominate, but in oligotrophic or seasonally stressed environments, lysogeny is favoured (Paul, 2008).

Pseudolysogeny has been observed in various marine phage/host systems and one of the earliest was reported in the phage Hs1, which infects the halophilic archaebacterium, *Halobacterium salinarium*. It was found that the infection dynamics could be altered significantly by changing salt concentrations. At low concentrations, 17.5% (w/v), the phage appeared to be highly virulent causing a majority of the host cells to lyse, however, at 25% salinity, the majority of cells were phage carriers. Indeed, if cells were infected at 20% salinity, more than 77% of the original inoculum were able to form colonies on agar plates (of 30% (w/v) NaCl) and the majority of these were lysogens (Torsvik and Dundas, 1980). Phage S-PM2 also displays pseudolysogeny during infection of phosphate-starved *Synechococcus* sp. WH7803. It was observed that in P-deplete conditions only 9.3% of cells were lysed whilst in P-replete conditions, 100% lysis occurred (Wilson *et al.*, 1996). However, as pseudolysogeny is poorly characterised, it is difficult to define the differences between this lifestyle, and the effect on environmental conditions on the lysogenyic decision of a temperature phage.

1.4 Impact of phage on the biosphere

1.4.1 Viral shunt

It is well documented that bacteriophage concentrations are high and on average exceed those of their bacterioplankton hosts by a factor of 3-10 (Wommack and Colwell, 2000). As such it is not surprising to find that phages are major players in microbial mortality. It has been estimated that between 10-50% of total bacterial mortality is due to phages (Fuhrman, 1999; Fuhrman and Schwalbach, 2003) though each method of determining this number has flaws, for a review see Suttle (2005). Temperate phages can also account for as much as 5% of the total bacterial mortality through spontaneous induction, the effect of environmentally important pollutants such as pesticides, and the action of UV-light (Miller, 2005). Consequently, the phage-mediated diversion of carbon and nutrients away from higher topic levels and back into the pool of dissolved organic matter in the oceans is significant. This "short-circuit" or viral shunt was proposed by Wilhelm and Suttle in 1999, see Figure 1.6.

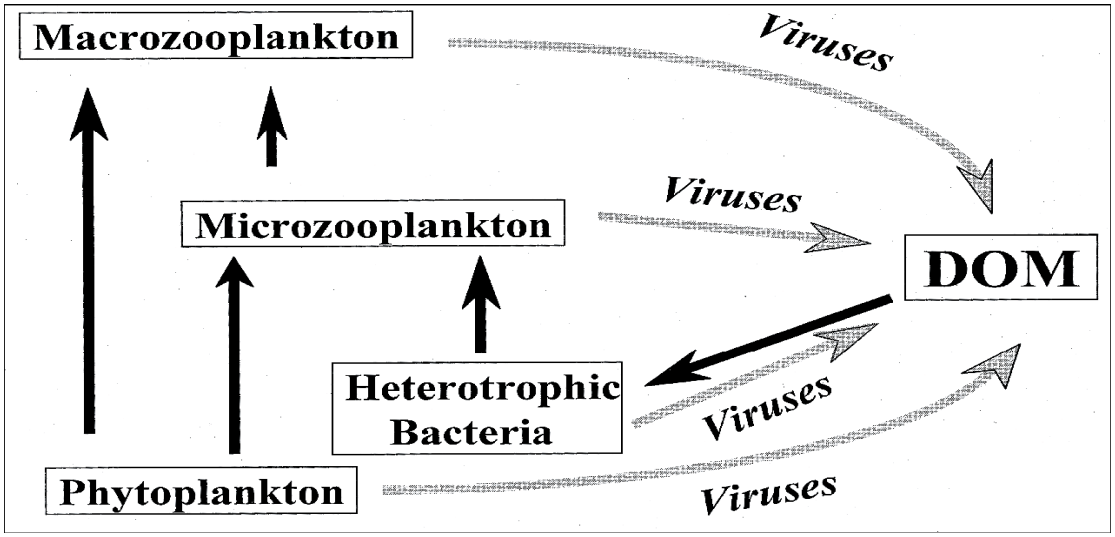


Figure 1.5 The viral shunt in the marine food web. By destroying host cells, viruses divert the flow of carbon and nutrients back to the pool of dissolved organic matter (DOM) in the ecosystem (grey arrows). Taken from Wilhelm and Suttle, 1999.

In their model, they proposed that viruses, of both bacterioplankton and grazers, could be responsible for the redirection of up to a quarter of organic carbon back into the pool of dissolved organic carbon (Wilhelm and Suttle, 1999). As organisms are composed of more that just carbon, cycling of other nutrients such as nitrogen, phosphorus, sulfur and trace elements are also affected by viral lysis. The

viral shunt also helps to retain nutrients as dissolved organic matter (DOM) in the top layer of the ocean promoting bacterial productivity, instead of sinking out as organic particulates and being exported to deeper layers (Fuhrman, 1999). Through this shunt, phage activity has a significant impact on the marine food web and in turn global biogeochemical cycles.

1.4.2 Killing the winner effect on community structure

In his paper, "The paradox of the plankton" (1961) Hutchinson examined the paradoxical situation found in bodies of water where diverse, abundant numbers of phytoplankton coexist and compete for the same nutrients, when according to the "competitive exclusion" principle, there should only be one or relatively few species that have out-competed all others (Hutchinson, 1961). He proposed that the effects of the changing environmental conditions caused by both biotic (predation) and abiotic (turbulence, light etc) factors, resolved the paradox. In the last decade, another highly idealized mathematical solution dubbed the "killing the winner" (KtW) hypothesis, was proposed and has been widely accepted (Thingstad and Lignell, 1997; Thingstad, 2000).

In the KtW theory, maintenance of bacterial community diversity is mediated by the action of phages that "kill the winner". It assumes that in a given habitat, the total biomass is finite due a limited resource (such as phosphate) and the bacterial community can be divided into two competing populations: competition specialists, who rely on their high reproductive rate to out-compete their rivals and rare, defence specialists, who have invested in defence mechanisms resulting in decreased fitness and slow growth. Phages prevent the former group from dominating for if they did, their high abundance and lack of defences would result in their death by phage lysis. This leaves open a niche with sufficient nutrients to enable the otherwise out-competed defence specialists to survive. Consequently, bacterial diversity is maintained and Hutchinson's paradox is resolved (Thingstad and Lignell, 1997; Thingstad, 2000; Winters *et al.*, 2010).

However, this is only one aspect of the KtW concept as it is much more complex than the relatively intuitive explanation above. KtW also factors in non-specific protozoan predation and makes predictions on the effect of bacterial species on viral abundance. It should also be noted that due to its idealized phage/host (and grazer/bacteria) relationships and the assumption of a steady-state habitat, the theory

has many shortcomings (for a review of KtW see Winter *et al.*, 2010). Nevertheless, the majority of experimental studies (carried out in both marine and estuarine environments) that have examined the various parameters in the hypothesis, have supported its predictions (Winter *et al.*, 2010) and as such it can concluded that phages are instrumental in regulating prokaryotic abundance, community composition and population dynamics.

1.5 Impact of phages on bacterial evolution

As well as regulators of community structures, bacteriophages have long been accepted as drivers of prokaryotic evolution. Some of the mechanisms by which they achieve this are outlined below.

1.5.1 Phages as mediators of lateral gene transfer

The phenomenon of phage-mediated lateral gene transfer (LGT) or transduction has been known about for almost 60 years (Zinder and Lederberg, 1952) and is described in many marine phage-host systems. The genetic material transferred can have repercussions on the individual host e.g. conferring toxicity and in turn altering microbial genetic diversity. A recent example can be found in a study of vibriophages in southern California coastal waters. Here it was shown that environmental phage isolates were able to infect toxigenic *V. cholerae* strains and transfer a genetic marker, CTXΦ, to another environmental non-toxic strain (Choi *et al.*, 2010).

Another study carried out by Jiang and Paul (1998) quantified the transduction frequency of a plasmid in two concentrated samples of mixed bacteria as being between 1.58×10^{-8} to 3.7×10^{-8} transductants PFU^{-1}. Using known bacterial and viral concentrations, as well as the water volume of the Tampa Bay Estuary, the authors calculated the number of transduction events per year to be up to 1.3×10^{14}. Such high rates would make transduction an important mechanism for gene evolution in the marine environment and highlights the role of phages in exchanging DNA between bacterial populations.

1.5.2 Phage resistance mechanisms

When a phage encounters a bacterium, successful infection is not always guaranteed due to a number of host defence mechanisms. Figure 1.7 shows the basic infection cycle of a lytic phage and each step of this cycle can fall foul of an antiphage system. Brief explanations and examples of some of these strategies will outlined below, for a more comprehensive review of phage resistance mechanisms see Labrie *et al.* (2010).

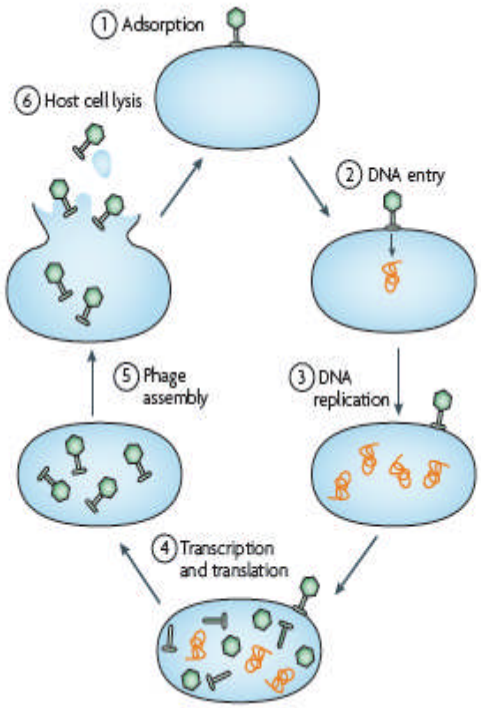


Figure 1.6 Lytic phage replication cycle.
Each step of the cycle can be targeted by antiphage mechanisms. Hosts can contain multiple mechanisms, though the effect of such combinations has rarely been assessed. Taken from Labrie *et al.* (2010).

1.5.2.1 Blocking phage adsorption

By altering or completely removing the cell surface structures to which phage attach, bacteria can halt infection at the first step. A mutant strain of *Roseobacter denitrificans* OCh114, M1 that is resistant to infection by phage RDJLΦ1 was isolated by Huang *et al.*, (2010). Comparative proteomics of wild type and the M1 strain revealed that five membrane proteins were down-regulated in the resistant strain which suggests that one of more of these proteins were the phage receptors.

Alternatively, other molecules may be produced by the potential host to mask the receptor and reduce (but not completely prevent) phage binding e.g. the immunoglobulin G-binding protein A produced by *Staphylococcus aureus* (Nordström and Forsgren, 1974) which is covalently bound to mucopeptide, thus masking the *O*-acetyl groups of the mucopeptide that are required for phage adsorption. Other bacteria restrict access to potential receptors by the production of structured extracellular polymers. For example, phage infection of the soil-dwelling *Azotobacter chroococcum* was reduced when immobilised in sodium alginate, an exopolysaccharide produced by several *Azotobacter* spp., compared to that of liquid cultures (Hammad, 1998)

Finally, molecules naturally present in the environment, such as microcins (small bacteriocins, comprised of a few peptides), can be competitive inhibitors of phages. Microcin J25 binds to the *E. coli* iron transporter, FhuA which is also the receptor for phages T1, T5 and Φ80. This was demonstrated *in vitro*, by the pre-incubation of purified FhuA with J25 at varying concentrations, and YO-PRO-1 (a fluorescent DNA dye) followed by the addition of phage T5. Fluorescence, which is proportional to the release of phage DNA through binding, was found to decrease concomitantly with increasing concentration of J25 to FhuA. The authors calculated phage adsorption decreased from 100 to 45% when concentrations of J25 increased from 0.1 to 3.2 μM (Destoumieus-Garzón *et al.*, 2005)

1.5.2.2 Preventing DNA entry

Blocking the entry of phage DNA into a host cell, prevents it from sequestering the host cellular machinery; such systems are known as superinfection exclusion (Sie) systems and involve membrane-anchored or membrane-associated proteins. Interestingly, the genes responsible for such proteins are often found in prophages (Labrie, *et al.*, 2010). One such example can be found in T4 phage-resistant *E. coli* already infected with T4, which has two Sie systems encoded by the *imm* and *sp* genes. Imm acts in conjunction with another membrane protein, to change the conformation of the phage DNA injection site (Lu *et al.*, 1993), whereas Sp inhibits the T4 lysozyme preventing the creation of new holes in the host cell wall. These two systems not only prevents superinfection by T4, but all T-even-like phages (Lu and Henning, 1994).

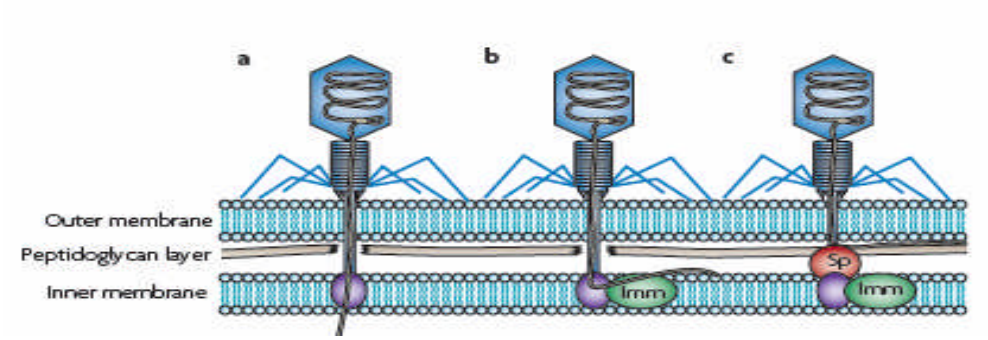


Figure 1.7 Blocking the entry phage of DNA using *E. coli* proteins Imm and Sp. Taken from Labrie *et al.*, 2010. The Imm protein binds to the phage DNA injection protein and blocks DNA entry whilst the Sp protein inhibits the phage lysozyme so the peptidoglycan layer cannot be breached.

1.5.2.3 Targeting phage nucleic acids

Two systems are known to target phage nucleic acids; the restriction modification system (of which there are at least four types) recognises foreign DNA and degrades it and the CRISPR-Cas system (Clustered Regularly Interspaces Short Palindromic Repeats), whose mechanisms are not yet fully elucidated, for a recent review on CRISPRs see Vale and Little (2010). Though the latter system was first described in 1987, its role in phage-resistance was not fully appreciated until recently. In the last three years it has been shown that bacterial cells are able to incorporate new spacers that are 100% identical to phage genetic material after infection (Barrangou *et al.*, 2007). Acquisition of the new spacer promotes host resistance to further phage infection, whilst removal of the spacer renders it susceptible again. It was also shown that mutation of the phage genome is sufficient to counter the newly acquired resistance (Deveau *et al.*, 2008). The antiphage response is mediated by CRISPR-associated (Cas) proteins; a helicase, Cas3, and the mature CRISPR RNAs work in tandem to interfere with phage replication (Brouns *et al.*, 2008). Regardless of the system used, R/M or CRISPR-Cas, both allow bacterial cells to destroy foreign genetic material thus halting any phage infection.

1.5.2.4 Abortive infection systems

Unlike the examples described above, abortive infection systems (Abis) result in the "altruistic" death of the infected cell and so work on the community, not the individual level. Most Abis have been found in *Lactococcus lactis* (23 to date, Labrie *et al.*, 2010), but all appear to target stages in phage multiplication. Perhaps the best characterised system is the two component Rex system found in λ-lysogenic *E. coli*

strains. During phage infection, the RexA protein is activated which in turn switches on an ion channel RexB. Loss of membrane potential due to RexB activation, results in a drop in ATP level which causes abortion of phage infection followed by cell death (Molineux, 1991). In contrast, the Abis in *Lactococcus* spp. as well as the *E. coli* Lit and Prr systems appear to work on the transcriptional level. In these systems, during phage infection, otherwise dormant enzymes are activated and cleave highly conserved, essential components of the translational machinery. Consequently, protein synthesis is halted, phage infection is aborted and the infected cell dies (Chopin *et al.*, 2005)

Despite the variety of the antiphage mechanisms found in prokaryotes, of which only a small number have been described above, phages have found strategies by which they can overcome all of these barriers. Consequently, in turn, bacteria must alter their strategies in order to survive resulting in a cyclic arms-race or the "Red queen effect" where there is a continuous cycle of co-evolution maintaining the genetic diversity of both bacteria and phages (Van Valen, 1973).

1.5.3 Effect of prophages on host fitness

For over 30 years, the ability of prophages to enhance host fitness has been observed as many contain genes that are not essential for the phage lifecycle. These lysogen conversion factors include virulence proteins, metabolic enzymes and transcriptional repressors which can down-regulate essential genes. Though this last point may seem at first counter-intuitive a recent study by Chen *et al.* (2005) found that during phage λ infection, expression of the host *pckA* gene (critical in gluconeogenesis) was suppressed by the phage cI repressor. They postulated silencing of this gene resulted in slower growth in the challenging glucose-free environment, thus ensuring the survival of the lysogen over its uninfected clone. Other mechanisms by which prophages can enhance host fitness are discussed in reviews by Brussow *et al.*, (2004) and Paul (2008) and in Chapter 8.

Since high numbers of cultivable marine bacteria have been found to contain prophage-like elements (as discussed in Section 1.4.6) it has been proposed that prophages are not dangerous molecular time bombs, as traditionally viewed, but are instead advantageous to their hosts. They allow their hosts to survive in resource-replete conditions by repressing metabolic genes and provide mechanisms by which

the host (and prophage) can sense changes in the environment. As well as this, as discussed in Section 1.6.1, prophages can mediate transduction, improving host fitness thus helping to drive bacterial evolution in the sea.

1.6 Bacteriophage taxonomy: orders typically found in the ocean

Historically, viral taxonomy has been based on virion morphology and nucleic acid composition and is managed by the International Committee on Taxonomy of Viruses (ICTV). Since 1991 the ICTV has used the "polythetic concept" for viral species delineation (Ackermann, 1992). This concept, introduced by Morton Beckner in *The Biological Way of Thought* (1959), is based on the understanding that not all members of a group share identical trait combinations. Instead the group is defined by an agreed set of properties, all of which may or may not present in every member (Ackermann, 1992). The ICTV defines a viral species as "a polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche" (Van Regenmortel, 1992). It should be noted, however, that phage taxonomy is in constant flux; with the high degree of gene transfer between phages and increased amounts of information coming from proteomic and bioinformatic studies, what is believed now may not hold true in a few years time.

There are currently 14 officially accepted phage families with another five awaiting classification (Ackermann, 2009). However, over 95% of all described phages fall into the order *Caudovirales*, defined as tailed, double-stranded (ds)DNA phages with binary symmetry. This order only contains three families; the *Myoviridae*, *Podoviridae* and *Siphoviridae*. The remaining eleven families contain dsDNA, (ss)DNA, ssRNA or dsRNA genomes contained within virions are defined as polyhedral, filamentous or pleiomorphic. As the phages isolated from the marine environment typically belong to the *Caudovirales*, only these three families will be discussed in detail.

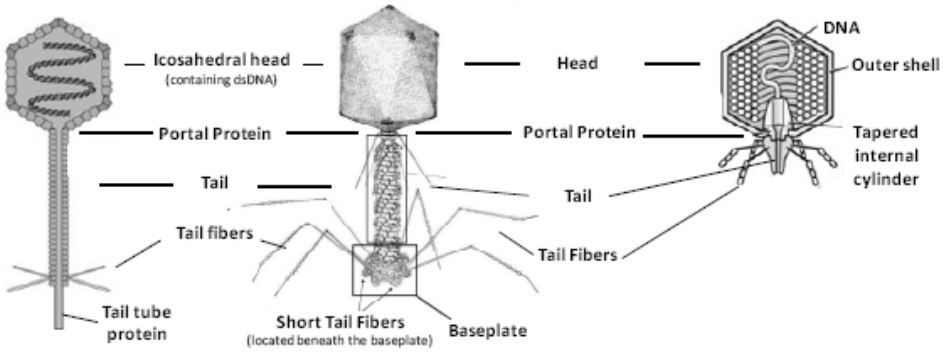


Figure 1.8 Basic morphology of the three phage families of the *Caudovirales* order. From left to right, diagrams of a Siphovirus, Myovirus and Podovirus. Taken from Ceyssens, 2009a.

The conserved icosahedral head and "helical" tail structure of the *Caudovirales* is unique in virology (Ackermann, 2005). Though other viruses, such as the polydnaviruses and tectiviruses, do have tail-like structures they are not as regular and constant. This, along with other properties has led to the suggestion that this order is of monophyletic origins (Ackermann, 2005).

The phage chromosome is highly condensed and found within the capsid head; it forms between 20-50% of the virion mass (Earnshaw and Harrison, 1977). The head itself displays cubic symmetry, has 20 sides/12 vertices with triangulation numbers of T = 4, 7, 13, 16 and 52 and is assembled separately from the tail/tail fibres from one or two major head/capsid proteins (Ceyssens, 2009). Proteins connecting the head and tail, termed "portal proteins" are homo-oligomeric structures (though the oligomerization state of these connects has been a matter of debate, see Valpuesta *et al*., 2000) whose conformational change during infection allows the DNA to exit the virion and pass into the host cell. As previously mentioned, phage tails are helical and are comprised of stacked disks; in most phages, accessory structures such as base plates, spikes or terminal fibres can also be found at the distal end (Ackermann, 2005). Myoviruses are identified by their contractile tail consisting of a sheath and central tube. The portal protein is connected to the tail by a neck region, which in *Escherichia coli* (hereafter referred to as *E. coli*) phage T4, is comprised of gene product (gp)3, gp13, gp14, gp15 and gp *wac* (Leiman *et al.*, 2004). Siphoviruses have long, non-contractile tails whilst podoviruses have short, non-contractile tails. These families can be further divided into subfamilies and genera and shown in Table 1.1.

Three recent papers have noted the presence of ssDNA phage genomes similar to chlamydiphages and microphages, which are dominant in many marine habitats (Breitbart *et al.*, 2004; Angly *et al.*, 2006; Desnues *et al.*, 2008). As yet no bacterial host have been determined and so no isolates of such phages exist.

Table 1.1 Genera in the order of *Caudovirales.* Based on the ICTV viral taxonomy list, as of May 2010

Family	Genus	
Myoviridae	I1-like viruses	
	Mu-like viruses	
	P1-like viruses	
	P2-live viruses	
	Φ H-like viruses	
	Φ KZ-like viruses	
	SPO1-like viruses	
	T4-like viruses	
Podoviridae	Φ KMV-like viruses	Autographivirinae
	SP6-like viruses	
	T7-like viruses	
	AHJD-like viruses	Picovirinae
	Φ29-like viruses	
	BPP1-like viruses	
	ε 15-like viruses	
	LUZ24-like viruses	
	N4-like viruses	
	P22-like viruses	
	Φ Eco32-like viruses	
Siphoviridae	c2-like viruses	
	L5-like viruses	
	λ -like viruses	
	N15-like viruses	
	Φ C31-like viruses	
	ψ M1-like viruses	
	SP β -like viruses	
	T1-like viruses	
	T5-like viruses	

1.7 Marine phage genomics

Pseudoaltermonas espejiana BAL-31 ΦPM2 was the first marine phage genome to be completely sequenced (Mäannistö *et al.*, 1999); now advances in sequencing technology have made obtaining a complete genome sequence *de rigeur* in the characterisation of novel phages. However, of the ca. 550 sequenced phages, less than 5% are marine. Despite this, what little is known has led to them being described as the largest untapped reservoir of genomic information (Paul and Sullivan, 2005). As before, as the majority of marine phages appear to belong to the *Caudovirales*, this discussion will mainly focus on the genetics of this order of viruses.

1.7.1 General genome architecture

All genomes of tailed phages contains genes for transcriptional regulation, DNA replication, DNA packaging, structural genes and finally, lysis. This can be found in genomes as small as 10 kb in ΦPM2, or as large as the 253 kb of *Prochlorococcus* NATL1a phage P-SSM2. It is presumed that as genome sizes increase, the likelihood of phage interference with host cellular activities increases concurrently.

Phage gene expression during the lytic cycle, is largely time-ordered as groups of genes are sequentially expressed by RNA polymerases; genes are defined as being early, middle or late. Early proteins are generally directed toward the take-over of host metabolism by defending against anti-phage mechanisms or through the establishment of optimal conditions for the synthesis of new virions. In a study by Roucourt and Lavigne (2009) it was revealed that the majority (64%) of phage-host protein interactions involved phage early proteins. (Delayed early, middle and later protein take part in 7%, 25% and 4% of phage-host interactions respectively.) Early proteins can be relatively phage-specific, poorly conserved and many may be ORFans (see Section 1.7.6). During middle gene expression, replication of the phage DNA begins. Replication machinery maybe of host cell origin as commonly found in temperate phages or can be encoded by the phage itself, the situation often found in virulent phages. The variety of replication systems utilized by phages has been reviewed extensively by Weigel & Seitz (2006). Late genes are primarily concerned with the creation of procapsids and the packaging of genomes into them. Interestingly,

the two protein most widely conserved amongst tailed phages are the large terminase subunit responsible for providing energy (from the hydrolysis of ATP) to translocate the DNA molecule into the procapsid and the portal protein (Casjens, 2008). Finally, the host cell is lysed, again by late genes, and new virions are released into the environment.

1.7.2 Comparative genomics – phage mosaic theory

Perhaps the most striking observation that has been drawn from the comparative genomics of tailed phages is the degree of genetic moasicism. This evidently arose from recombination events between ancestral phages and likely has been ongoing for the last three billion years (Hendrix *et al.*, 1999). When viewed at the DNA sequence level, it is possible to observe the precise junction, on either side of which, the genetic material has distinct evolutionary origins, these are known as mosaic boundaries.

The non-random distribution of mosaic boundaries has been recognised for many years and led Susskind and Botstein (1978) to propose their "modular theory" of phage evolution. They hypothesized that modules, defined as genes or group of genes, were exchanged by recombination due to "linker" sequences between them. Ownership of such linkers would afford phages a selective advantage as it would provide a simple mechanism by which progeny of increased fitness could be produced. Clark *et al.*, (2001) reported examples of short conserved sequences in lambdoid phage genomes, which could serve as points for exchanges using either host- or phage-encoded recombinases. However, these sequences are not widespread enough to account for the vast majority of exchange events that must have occurred (Hendrix, 2002). An alternative model proposes that genetic recombination is instead a random event with little or no sequence preference. This model would result in many invalid progeny as the majority of exchanges would result in non-functional phage progeny. However, due to the vast number of phages, such low probability events can occur and be selectively amplified in the viriosphere. Thus, novel advantageous genetic combinations can be formed and maintained (Hendrix, 2002).

Recombination events can result in the exchange of not only individual genes, but large blocks of genetic information (as seen in Figure 1.9).

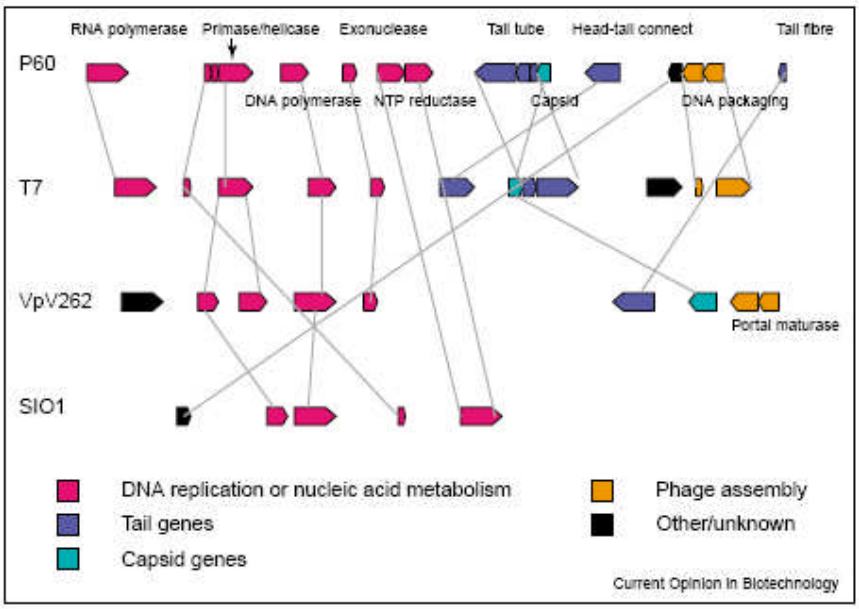


Figure 1.9 Genomic re-organizations and transfers between four T7-like phages. Taken from Paul and Sullivan (2005). The grey lines indicate homologous genes that appear to have been exchanged during gene transfer event between three marine T7-like podophages; Cyanophage P60, *Vibro* phage VpV262, Rosophages SI01, and coliphage T7.

An extreme example of recombination can be found in *Xanthomonas oryzae* bacteriophage Xp10 (Yuzenkova *et al.*, 2003). Morphologically it is a Siphovirus, but sequencing revealed that it encodes its own single-subunit RNA polymerase characteristic of T7-like Podoviruses. The left half the Xp10 genome contains genes for structural and host lysis proteins which are similar to those found in temperate λ-like phages (Siphoviruses), whilst the right half codes for DNA replication and transcription and includes the T-odd like RNAP. Both sets are transcribed divergently from a regulatory region which separates the two and so it appears Xp10 arose through recombination between genomes of widely different phages.

Results such as this led to Hendrix *et al.*, (1999) proposing that within the tailed phages, all members share a common ancestry and their genetic structure is due to large amounts of horizontal gene exchange. This ability of genes to "walk" from phage to phage can be viewed as all phages having access to a common genetic pool and as a result of many small steps, the phages we see today are mosaics. Hendrix *et al.*, described this as a sequence's ability to take a "random walk through phylogenetic space". In some cases, such as that of Xp10, a marathon rather than a light stroll, seems to have occurred.

1.7.3 Prophage genomics

In his review of marine prophages, Paul (2008) probed 113 marine bacterial genomes for the presence of prophages based on their existing annotations. He found 64 prophage-like elements in 49 bacterial strains (43%). However, he classified 33% of these as gene transfer agent (GTA)-like elements or defective prophages.

As with non-marine prophages, it is possible to classify marine prophages into types A, B and C. Paul found that the type A prophages contained a lysogeny module followed by replicative genes such as terminases, capsid and tail genes. (The lysogeny model was defined as containing a coliphage λ-like integrase and at least one phage repressor.) The genome sizes of type A prophages range from 37 to 41 kb. Type B prophages also contained an integrase gene, but they did not contain a recognizable repressor protein; they were similar to type A in length, 31-49.5 kb. Paul also noted that there was no discernable genome organization in this group. The type C prophages formed the largest group and had genomes sizes ranging from 12 to 29 kb in length though, as this is shorter than most temperate phages previously reported, it was hypothesized they maybe remnants and not fully functional (Paul, 2008).

The remaining prophages named in the review did not fit the criteria above, but did bear resemblance to other known terrestrial prophages such as coliphage Mu. Transposable phage Mu replicates in a copy-and-paste fashion though it lacks an integrase gene (Toussaint *et al.*, 1994). The lysogeny switch is controlled by the Rep repressor encoded by the c-gene; in addition Mu phages cannot be induced by Mitomycin C (Paul, 2008).

As prophages, by their very nature, can exist in tandem with their bacterial host's genome for an indeterminate length of time, it is logical to assume that they display a greater degree of moasicism due to the increased probability of horizontal gene transfer events occurring between host and phage. However, due to the lack of sequencing data it is not yet possible to examine this facet of phage evolution until more genomes of both marine and terrestrial phages are available.

1.7.4 Viral host genes

As agents of gene transfer it is not surprising to find phages themselves sometimes contain genes that are of bacterial origin. The best studied example of phage-encoded host proteins can found be amongst the cyanophages. Core

photosynthetic genes of photosystem II (PSII) such as *psbA* and *psbD* are frequently found in phages of *Synechococcus* and *Prochlorococcus* (Millard, 2004); recently core genes from photosystem I were also identified in viral metagenomic datasets (Sharon *et al.*, 2009). When first observed in cyanophage SPM-2 (Mann *et al.*, 2003) it was postulated that expression of the phage-encoded D1 protein (which in bacteria has a high turnover rate) helps to maintain energy production through photosynthesis during infection. This was confirmed experimentally as *psbA* transcripts as well as phage D1 polypeptide were detectable during infection (Clokie *et al.*, 2006; Lindell *et al.*, 2005; 2007). Modelling studies have also shown that possession of *psbA* provides a fitness advantages over phages that do not, especially in high-light conditions (Bragg and Chisholm, 2008; Hellweger, 2009). Interestingly, phylogenetic analysis suggests that *psbA* has been inherited from cyanobacteria on a number of occasions and since then there has been significant intragenic recombination between host and phage copies (Lindell *et al.*, 2004; Millard *et al.*, 2004; Zeidner *et al.*, 2005; Sullivan *et al.*, 2006).

In 2006, a survey of marine metagenomic sequences suggested that up to 11% of the genes found in cyanophages were host-like and included photosynthetic genes (*psbA, psbD*, *hli*), as well as the phosphate-scavenging genes (*phoH, pstS*) and the cobalamin biosynthesis gene, *cobS* (DeLong *et al.*, 2006). Results from another metagenomic study showed that there is a relationship between type of host-like genes and biogeography (Williamson *et al.*, 2008). Their data suggested that the frequency of host-derived viral genes increased from temperate, mesotrophic waters to tropical, oligotrophic waters. For example, there was a positive correlation between abundance of *pstS* sequences (a phosphate-binding protein), salinity and water depth. The authors noted that nutrient concentrations often decrease with water depth, due to less input from land-based sources, which would explain the need for a phosphate-binding protein. These results imply environmental pressures specific to a habitat influence the types of genes acquired by phages.

Recently it has been suggested that viral host genes have a site-specific acquisition mechanism. A comparative genomics study of five cyanomyoviruses by Millard *et al.* (2009) identified a hyperplastic region within a highly conserved structural gene module, which often contained host-like genes e.g. *petE* and *ptoZ*, (encode plastocyanin and plastoquinol terminal oxidase respectively). Despite their localization, phylogenetic analysis suggested that these genes were acquired

independently of each other. The region contained no conserved boundary sequences or any features that might suggest a method for genetic exchange and the G + C content, G + C skew and di-nucleotide frequency of the genes were in keeping with the corresponding genomes. The authors suggested that fitness-increasing host-like genes can be transferred into hyperplastic regions by an as yet unknown mechanism, and over time either become stable features of the genome or can be relocated to other areas (Millard *et al.*, 2009).

This and other findings suggest that bacterial metabolic genes in phages, whether they are from a closely related or divergent host, can be functionally active during infection and in doing so, confer a fitness advantage.

1.7.5 Metagenomics

Investigations of the marine phage metagenome offer glimpses into the community genomics of various marine environs without the inherent bias that culture-based experimentation creates. So far the Global Ocean Survey (GOS) has sampled various locations including the coasts of North America, the Artic and Antarctic, deep sea ocean vents and even whale falls (Angly *et al.*, 2006; López-Bueno *et al.*, 2009; Hallam *et al.*, 2004; Tringe *et al.*, 2005). Information also exists for marine sediments at various locations as well as from stromatolites and thrombolites (Kim *et al.*, 2009; Breitbart *et al.*, 2004; Desnues *et al.*, 2008). These studies have revealed that that the majority of sequences from the viral fraction (which includes phages) showed no similiarity to any other sequences previously deposited. For example 75% of the DNA sequenced was unknown in the viral community of a near-shore marine-sediment (Breitbart *et al.* 2004). (It is worth noting that this is similar to the number of unknown genes found in newly sequenced marine phages, around 60% (Paul and Sullivan, 2005).) In the same paper, the authors used the distribution of overlapping sequence fragments to estimate that in one kilogram of sediment there would be approximately 10,000 different viral genotypes (Breitbart *et al.*, 2004). Such un-expectantly high numbers reveal that the marine virome is the most genetically diverse biological entity on Earth.

Sequences that do have hits to known genes are usually phage-related such as DNA and RNA polymerases, helicases, terminases, exonucleases and structural proteins. Though as discussed in Section 1.7.4, host-like genes are also relatively

prevalent. The five most abundant virus-encoded host genes, as found by Angly *et al.*, (2006), are shown in Table 1.2.

Table 1.2 The five most common viral-encoded proteins found in four oceanic viriomes.
Data taken from Angly *et al.* (2006).

Marine Region	Enzyme Name	EC number	Gene occurrence
Sargasso Sea	Ribonucleotide reductase of class Ia (aerobic), alpha subunit	1.17.4.1	89
	Ribonucleotide-disphosphate reductase	1.17.4.1	75
	Ribonucleotide reductase of class II (coenzyme B12-dependent)	1.17.4.1	50
	GTP cyclohydrolyase I, type 2	3.5.4.16	37
	Adenine-specific, methyltransferase	2.1.1.72	22
Gulf of Mexico	Formate dehydrogenase-O, major subunit	1.2.1.2	27
	Carbamoyl-phosphate synthase large chain	6.3.5.5	25
	Cytochrome c oxidase polypeptide I	1.9.3.1	24
	Ribonucleotide reductase of class II (coenzyme B12-dependent)	1.17.4.1	23
	DNA polymerase III alpha subunit	2.7.7.7	23
British Columbia coast	Ribonucleotide reductase of class II (coenzyme B12-dependent)	1.17.4.1	34
	DNA polymerase III alpha subunit	2.7.7.7	22
	3-polyprenyl-4-hydroxybenzoate carboxylase	4.1.1.-	18
	Cytochrome c oxidase polypeptide I	1.8.3.1	18
	Ribonucleotide reductase of class Ia (aerobic), alpha subunit	1.17.4.1	18
Arctic Ocean	3-polyprenyl-4-hydroxybenzoate carboxylase	4.1.1.-	205
	DNA polymerase III alpha subunit	2.7.7.7	185
	Cytochrome c oxidase polypeptide I	1.9.3.1	175
	Isoleycyl-tRNA synthase	6.1.1.5	157
	Methylcrotonyl-CoA carboxylase carboxyl transferase subunit	6.4.1.4	155

Interestingly, a PCR-based metagenomics investigation by Breitbart *et al.*, (2004) of a number of habitats including the marine biome, found that the T7-like DNA polymerase could be found nearly everywhere. Furthermore, they found most sequences clustered into two groups dubbed, HECTOR and PARIS. The former group was >99% identical at the nucleotide level and occurred in 49 out of 66 samples. In contrast, a purely marine study (also carried out using T7-like degenerate PCR primers) by Labonté *et al.*, (2009), found that none of their environmental sequences clustered with HECTOR and PARIS, but instead formed three new groups, ENV1, ENV2 and ENV3 with the former two containing the majority of sequences. However, currently there are no cultured representatives of any of the new marine groups (Labonté *et al.*, 2009).

Another intriguing observation gained from mining the viral metagenome was the prevalence of motility and chemotaxis proteins. A study by Dinsdale *et al.* (2008)

found a total of 130 such proteins out of a possible 157 when examining the sequences from nine biomes. They found the flagella biosynthesis protein FlhA, the chemotaxis response regulator proteins CheA and CheB and deacetylases were particularly abundant in the viriomes, whilst the twitching motility protein PilT, type II secretary pathways and GldJ (gliding motility lipoprotein) were over-represented in the micro biomes. It is not yet clear what role these proteins have in phage infection.

Metagenomic data have also been used to show the geographic distribution of phage types. It appears that myovirus-like sequences are more abundant in oligotrophic, tropic regions whilst Podovirus-like sequences are largely found in temperate regions such as the east coast of the North American continent (Williamson *et al.*, 2008).

To date, the majority of marine viral metagenomic studies have been carried out using high-throughput sequencing and have concentrated on DNA phages. Very little information exists about the RNA virus metagenome due to the inherent technical challenges of dealing with RNA which is unstable and requires reverse transcription prior to sequencing. However, from the studies that have been carried out it appears that no RNA phages exist thus far (Cullen *et al.*, 2006).

1.7.6 ORFans

ORFans are defined as genes without known function and/or database homologues (Fischer and Eisenberg, 1999). However, their existence and prevalence is an ongoing puzzle to biologists especially in today's metagenomic age, for if we believe that proteins in different organisms have descended from a common ancestor, why do many still show no similarity to each other? It is logical to assume that as more sequencing data are accumulated, the number of ORFans will decrease, however, a study by Yin and Fischer (2008) revealed the opposite to be true, see Figure 1.10. When they calculated the average number of ORFans in a random selection of phage genomes, and the percentage of ORFans, they found though the percentage of ORFans is gradually decreasing (with increasing number of phage sequences), it is not likely to drop significantly even if hundreds more genomes are sequenced. Instead, the total number of ORFans is likely to increase, see Fig. 1.10.

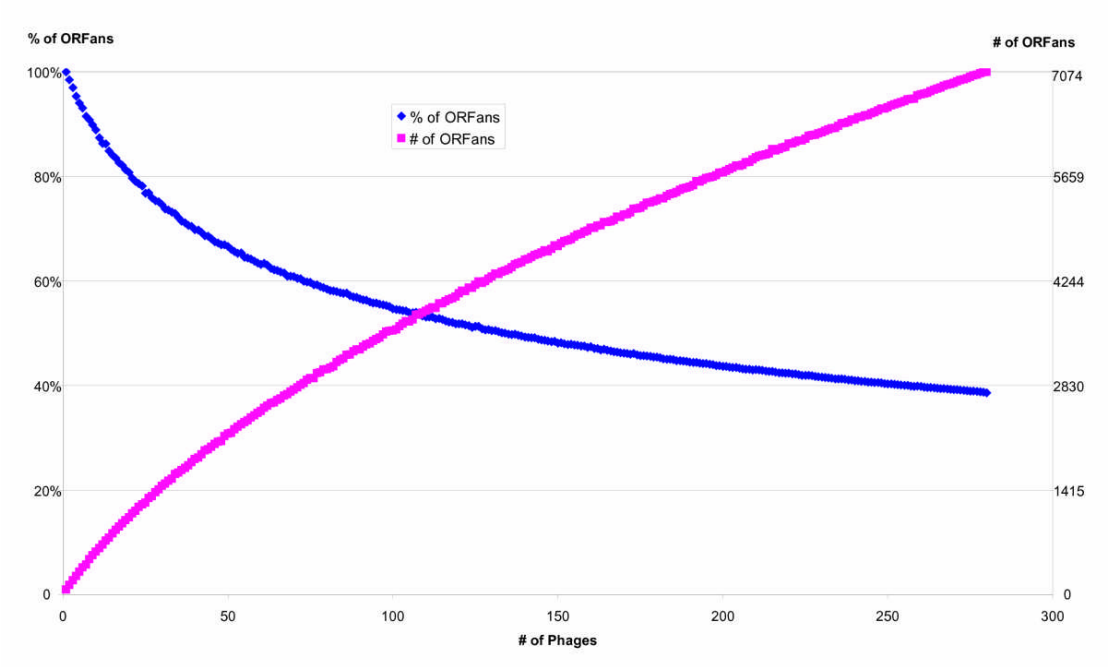


Figure 1.10 Illustration of the growing number of phage ORFans. Despite the increase in number of phages, the % of ORFans is only gradually decreasing and is unlikely to drop significantly even after hundreds of more phage genomes are sequenced. Taken from Yin and Fischer (2008).

ORFans make up *ca.* 30% of a phage genome compared to around 15% of a cellular genome (Frost *et al.*, 2005). Though this may in part, be due to the bias against phages in the sequence databases, their abundance and heterogeneity in the phage ORFome becomes understandable when the diversity of phage genes and genomes are taken into consideration.

Frost *et al.*, (2005) also concluded that phage ORFans, on average, are shorter than non-ORFans. This and some key experimental results have led to speculation that ORFans function as "molecular splints"; inhibiting or modifying host proteins by binding to them. For example the 90 amino acid protein, AsiA, in T4 binds to the host σ^{70} transcription initiation factor thereby preventing its interaction with host RNA polymerase. Instead, a cascade of T4-encoded transcription factors dominates the infected cell's transcription profile (Hinton *et al.*, 2005). Such a role would explain the extreme sequence diversity and high level of expression especially amongst early genes and the unique nature of ORFans.

The bidirectional trafficking of ORFans should not be overlooked, as a recent study by Cortez *et al.*, (2009) showed that archaeal and bacterial genomes contained a significant proportion of recently acquired foreign genes, including ORFans.

Furthermore, 56% of these likely originated from integrative elements (such as viruses, plasmids and transposable elements) compared to only 7% from distant cellular sources via LGT. This is another example of how phages can drive the gene diversity and so the evolution of their bacterial hosts.

1.8 Phage and biofilms

As biofilms and phages are prevalent throughout the marine environment (see Sections 1.1.5 and 1.2) it is logical to assume that phages must interact with marine biofilms. However, the inherent properties of biofilms such as the EPS matrix are likely to alter the known phage/bacterium interactions seen with planktonic hosts (Sutherland *et al.*, 2004). Unfortunately, due to the increased interest in phage therapy as a control against biofilms much of what is known comes from model, nuisance or pathogenic bacteria. Very little information exists about natural marine biofilms, but what is known about the interactions of phage and biofilms, both natural and experimental, will be discussed below.

1.8.1 Effect of biofilms on phage attachment

An obvious difference between planktonic and biofilm bacteria is that the latter exist in agglomerations with excreted products such as EPS (see Section 1.1.5). These may impede viral access to the bacterial cell surface and thus protect the potential hosts from infection. However, many phages produce polysaccharases or polysaccharide lysases and the action of these enzymes can be observed as halos around phage plaques where the polysaccharide has been removed from viable bacteria, but the bacteria have not yet been lysed, see Figure 1.11. The ability of phages to reach the cell surface using their associated enzymes was demonstrated by Bayer *et al.*, (1978) with phage K29 which was found to penetrate its host's capsule by binding to, then destroying its polysaccharide receptor. For an overview on polysaccharide-degrading phages see Scholl and Merill (2005). Confocal scanning of biofilms has revealed their heterogeneity; numerous voids and channels through which water can flow are present, allowing the supply of nutrients, the removal of wastes and an access point for phages (Wood *et al.*, 2000). In addition, the continual sloughing of cells as they age and die can reveal potential receptor sites.

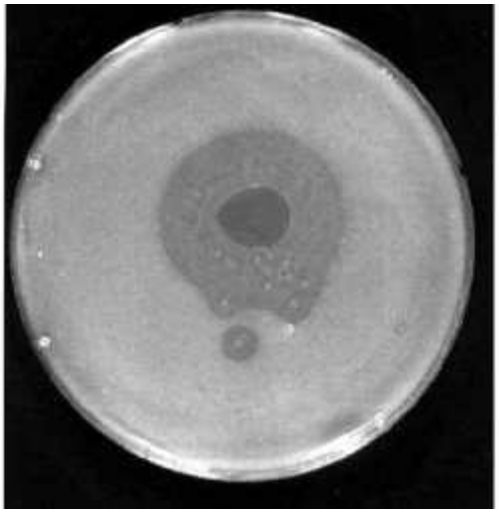


Figure 1.11 Plate showing the "halo" caused by a typical polysaccharase-inducing phage. Taken from Sutherland *et al.* (2004). The "halo" consists of viable bacteria from which the polysaccharide has been degraded by phage lyases.

In an investigation by Doolittle *et al.*, (1996) *E. coli* phage T4 and *P. aeruginosa* phage E79 were labelled with fluorescent and chromogenic probes then used to infect mature biofilms. They found that T4 was able to attach throughout the entire biofilm whereas E79 only absorbed to cells on the outer surface. The study illustrated that phages are able to bind to bacteria in biofilms despite the EPS matrix, though the fact that E79 was somehow prevented from infecting the interior of the biofilm demonstrates that accessibility can vary between biofilms and phages due to inherent structural differences.

1.8.2 Phages in natural biofilms

Biofilms have been shown to act as "sponges" in natural environments, trapping phages and viruses non-specifically. In wetland biofilms, Flood and Ashbolt (1999) found that viral-sized particles were 100-fold more concentrated compared to the surrounding water column. Unfortunately, the study did not examine if any infections could be observed in the biofilm. This was done by Filippini *et al.* (2006) in sediments, decomposing plant litter and biofilms on aquatic vegetation; using TEM they observed only four visibly infected bacteria in the 1.5×10^4 benthic bacteria examined in the various locations. Despite the apparent scarcity of infected cells, it is still possible to imagine a biofilm where phage infection is present. Hypothetically, if

the concentration of bacteria in a biofilm is 10^5 per cm^2 and roughly one in every 10^4 bacteria are infected ($4 / 1.5 \times 10^4$) then within this area there would be on average, 10 infected bacteria (Abedon, 2010). So it can be concluded that, though seemingly rare, phage infection in natural biofilms may still occur. Furthermore, it has been hypothesized that a steady state/equilibrium between phage and biofilm bacteria exists which allows the persistence of biofilms (Abedon, 2010). Support for this theory can be found in a study by Corbin *et al.* (2001) who exposed a biofilm of *E. coli* (in a glucose-limited chemostat) to T4 phage at various MOIs. After 90 minutes, the biofilm was disrupted and phage titre increased, but 6 hours post-infection, phage concentration decreased and a stable equilibrium in phage titre was seen. Interestingly, the levels at which the phage stabilized appeared dependent on the initial MOI. The phage titre was approximately one order of magnitude higher at a MOI of 100 than that seen at 10 (Corbin *et al.*, 2001).

1.8.3 Lysogeny within biofilms

The presence of prophages in bacteria is a well documented fact and so the potential role that lysogens may play in biofilms cannot be ignored. The advantages from a phage's perspective of lysogeny when in a biofilm are obvious; reproduction of its genome by the host whilst in its repressed state means that it becomes widely disseminated and may form new biofilm or planktonic colonies. However, recent studies have shown that presence of a prophage may also have advantages for the host bacterium.

Resch *et al.* (2005) showed that biofilms of lysogenic *Staphylococcus aureus* spontaneously released phages into their surrounding at a rate comparable to planktonic cultures. They postulated that the lysis of some bacteria promoted the persistence and survival of the remaining cells through the release of the sequestered nutrients within the induced lysogens (Resch *et al.*, 2005).

In another study, the presence of a number of prophages in *Bacillus anthracis* were shown to have an impact on sporulation, EPS expression and biofilm formation in both laboratory and environmental strains (Schuch and Fischetti, 2009). Screening of the genomes of two of the temperate phages found that phage-encoded RNA polymerase sigma factors were responsible for the phenotype alterations. The authors speculated that through infection by temperate phages of recently shed vegetative

cells (after death of the animal host), the new lysogens with their altered phenotypes could avoid dormancy and instead form stable biofilms or become earthworm gut endosymbiotes thus promoting survival in an active form.

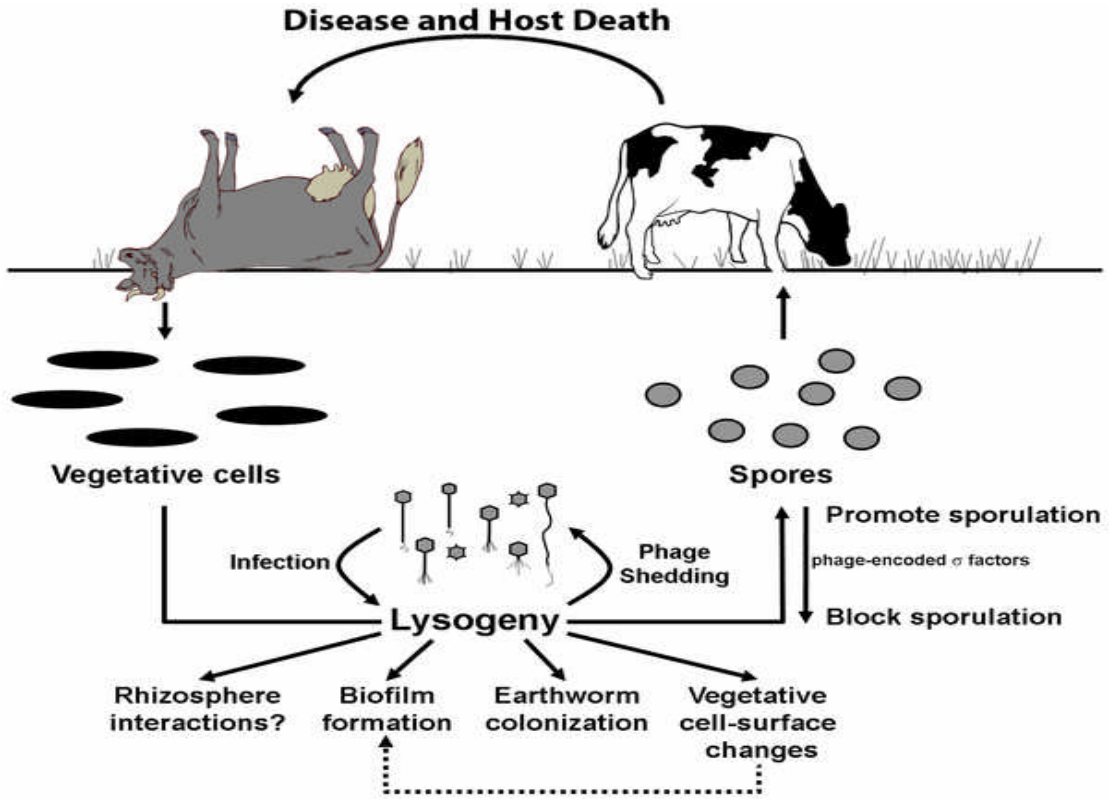


Figure 1.12 Hypothesized lifestyle for *B. anthracis* in the environment. Taken from Schuch and Fischetti, (2009). Lysogenic *B. anthracis* display different phenotypes such as EPS and biofilm formation, soil survival and earthworm colonization; these changes favour saprophytic or endosymbiotic lifestyles over dormancy.

1.9 Bacteriophages of Roseobacter species

Marine heterotrophic bacteria can greatly influence the biogeochemical cycles in the world's ocean; this is particularly true of species belonging to the Roseobacter lineage. As discussed in Section 1.1.1, Roseobacter species have been found in diverse marine habitats and can comprise up to a quarter of the marine bacterial community (Wagner-Döbler and Biebl, 2006). They have displayed a considerable degree of genomic diversity which reflects their wide-ranging physiologies (Brinkhoff *et al.*, 2008). As such, interest in the phages that infect this group of bacteria is high. To date four lytic phages and four prophage elements have been isolated from various

Roseobacter species and the current state of characterisation of these phages can be found in Table 1.3 and will be reviewed below.

Table 1.3 Table of isolated Roseobacter phages (as of May 2010).

Name	Original Host	Type	Genome size (kb)	Reference
SIO1	*Roseobacter* SIO67	*Podoviridae*	39.9	Rohwer *et al.*, (2000)
RDJLΦ1	*Roseobacter denitrificans* OCh114	*Siphoviridae*	-	Zhang and Jiao (2009)
DSS3Φ2	*Ruegeria pomeroyi* DSS-3	*Podoviridae*	74.6	Zhao *et al.*, (2009)
EE36Φ1	*Sulfitobacter* sp. EE-36	*Podoviridae*	73.3	Zhao *et al.*, (2009)
Prophage 1	*Silicibacter sp.* strain TM1040	*Siphoviridae*	73.6	Chen *et al.*, (2006)
Prophage 3	*Silicibacter sp.* strain TM1041	*Siphoviridae*	39.2	Chen *et al.*, (2006)
Prophage 4	*Silicibacter sp.* strain TM1042	*Siphoviridae*	36.0	Chen *et al.*, (2006)
ISM-pro1	*Roseovarius nubinhibens* ISM	*Siphoviridae*	26.9	Zhao *et al*., (2010)

1.9.1 SIO1

SIO1 was the first lytic Roseobacter phage isolated from a sample of Californian coastal seawater spotted onto a lawn of *Roseobacter* sp. SIO67 (Rohwer *et al.*, 2000). Both host and phage were isolated from seawater collected from coastal waters around the Scripps Institute of Oceanography, California. Though 13 closely related Roseobacter species were tested for susceptibility, SIO1 was found to exclusively lyse its original host. In a follow-up paper, a further four strains of SIO1 were isolated and all but one showed the same host range; isolate SBRSIO67-2001 was found to also infect *Roseobacter* GAI-101 (Angly *et al.*, 2009).

The complete genome of SIO1 was one of the first marine phages to be determined and (in 2000) was found to contain 30 open reading frames (ORFs). Recent re-sequencing and re-annotation found four new ORFS and assigned functions to seven additional genes. Angly *et al.*, (2009) also found that the genome could be divided into three modules: nucleotide synthesis and DNA replication (ORFs 7-13), phosphate metabolism (ORFs 14-17) and capsid structural proteins (ORFs 24-27).

In the Rohwer *et al.* (2000) paper, the SIO1 primase, DNA polymerase and endodeoxyribonuclease I were found to have greatest similiarity to those of coliphages T3 and T7. With the increased data from various sequencing projects, the DNA polymerase now shows most significant similarity to one from the Strait of Georgia (SOG) found using PCR with degenerate T7-like primers (Angly *et al.*, 2009).

Phylogenetic analysis by Labonté *et al.*, (2009), found that SIO1 now clustered closest to their SOG and other environmental sequences, but was distinct from the T7 enterophages.

1.9.2 RDJLΦ1

Roseophage RDJLΦ1 was isolated from the South China Sea on the host *Roseobacter denitrificans* OCh114, an aerobic anoxygenic phototrophic bacterium (Zhang and Jiao, 2009). The phage DNA was found to be resistant to three restriction enzymes; *Hah* I, *Hae* III and *Xba* I despite there being cleavage sites for *Hah* I and *Hae* III in a 1.65 kb fragment cloned from the phage. Interestingly, the host *Roseobacter denitrificans* does contain the genes for a type 1 restriction modification system, (Swingley *et al.*, 2007) and the authors concluded RDJLΦ1 must modify its DNA in order to escape restriction.

The viral proteome of RDJLΦ1was elucidated and revealed the presence of four host proteins, an outer membrane porin, a hypothetic protein and most intriguingly two 50S ribosomal proteins L14 and L22 (Zhang and Jiao, 2009). It is possible, however, that these latter two proteins are passenger proteins accidentally packaged into the virion due to their location in the cytoplasm and high copy number in bacterial cells.

In recent paper by Huang *et al.*, (2010), a phage-resistant mutant of *R. denitrificans* was isolated. Comparative proteomics found that five membrane proteins were down-regulated in the mutant, whilst several outer membrane porins and an OmpA family domain protein were up-regulated. Consequently, it was concluded that resistance was likely due to blocking of phage binding through alterations on the cell surface which were compensated for by expression of several outer membrane proteins (Huang *et al.*, 2010).

1.9.3 DSS3Φ2 and EE36Φ1

Until the isolation of DSS3Φ2 and EE36Φ1, the enterobacteria phage N4 was a genetic orphan which had been studied for 40 years without a comparable system. Sequencing of the two Roseobacter phages revealed a high degree of similarity to each other and to N4; 70 ORFs were shared by DSS3Φ2 and EE36Φ1 and both phages share 26 ORFs with N4. Also identified were four host-like genes and several

other phage-like genes (Zhao *et al.*, 2009). Perhaps the most intriguing aspect of N4 is its RNA polymerase which is encapsulated alongside the phage genome in the virion (Kazmierczak and Rothman-Denes, 2005). Both DSS3Φ2 and EE36Φ1 were found to contain homologues of the virion RNA polymerase (vRNAP) gene with the same four conserved T7-like RNAP motifs, as well as the two constituents of the N4 RNAP polymerase II (RNAP1 and RNAP2). The authors interpreted this as meaning the two Roseobacter phages used early and middle transcription machinery similar to that of N4.

Of the four host-like genes found in DSS3Φ2 and EE36Φ1 only two had assigned function: ribonucleotide reductase (rnr) and thioredoxin (trx). The former was found by metagenomic analysis to be among the most abundant genes found in the Sargasso Sea (Angly, *et al.*, 2006), see Section 1.7.5 whilst the latter has been found in other T7-like marine phage genomes (Hardies *et al.*, 2003). The absence of *trx* in N4 and T7 was interpreted by the authors as evidence of its importance for phage survival in marine environments.

The gene encoding the enzyme for host cell wall lysis posed another interesting difference between N4 and the two new Roseobacter phages. The murein hydrolase found in N4 was absent in DSS3Φ2 and EE36Φ1; instead both Roseobacter phages contained a late gene with similarities to a lytic enzyme of *Sagittula stellata* E-36. A likely explanation is that the cell walls of *E. coli* and the two Roseobacters are different enough to require a change in lytic enzymes, but those of *Ruegeria pomeroyi* DSS3 and *Sulfitobacter* sp. EE-36 (the hosts of DSS3Φ2 and EE36Φ1 respectively) are similar enough to use the same lyase.

1.9.4 Inducible prophages in Roseobacter strains

Silicibacter sp. strain TM1040 and *Roseovarius nubinhibens* have both been shown to contain Mitomycin C-inducible prophages (Chen *et al.*, 2006; Zhao *et al.*, 2010). Of the five prophage-like elements identified in TM1040, only three were detected by PCR in the lysate post induction. These inducible prophages shared very limited homology though they all contained genes encoding a terminase, major capsid protein and an integrase. Interestingly, prophage 2 contained three integrase and several DNA metabolism genes whilst prophage 5 contained genes responsible for

termination, lysis and structural genes, but neither were inducible suggesting they are prophage relics (Chen *et al.*, 2006).

Prophage 1, the largest (73.6 kb) found in TM1040 contains a GTA-like element. There are several possibilities for this phenomenon; it may be the result of an illegitimate recombination event between a GTA and a temperate phage either during co-infection or, as TM1040 contains a GTA in addition to the one found in prophage 1, after infection. Another possibility is that the prophage is the progenitor of the external GTA (Paul, 2008).

Unlike the *Silicibacter* prophages, the one induced from *Rsv. nubinhibens* (ISM-pro1) was not identified during initial annotation of the host genome (Zhao *et al.*, 2010). Instead, experimental laboratory results first indicated there might be a prophage-like element. Only upon subsequent examination of the ISM genome was a "hidden" prophage identified. ISM-pro1 contains 40 predicted ORFs, 25 of which showed high sequence similarity to *Rhodobacterales* bacterial genes and 11 to known phage genes.

It was observed that after prophage induction several small DNA fragments (ca. 3, 4 and 12 kb) were also present among the viral-like particles, but when not induced by Mitomycin C, these bands were missing. These were hypothesized to be GTAs however, prior studies have shown that they are not inducible (Solioz and Marrs, 1977). The authors suggested that viral lysis due to induction served to increase the release of more GTA-like elements to the extent of which they became visible bands on PFGE gels. It should also be noted that it is possible that these were additional cellular DNA segments accidentally packages into prophage capsids.

1.10 Summary and aims

It is apparent that members of the Roseobacter lineage are group of organisms with many unusual physiologies, varying lifestyles and a high genetic diversity. It is also well documented that phage abundance in the world's oceans is high. The broad aim of the work presented in this thesis was to isolate and characterise novel bacteriophages which infect Roseobacter species. The specific project objectives were to:

- Isolate Roseobacter phages from seawater samples
- Analyse the life-cycle of any isolated phages
- Characterise the genomes of isolated phages
- Ascertain what host receptors are required for successful viral infection
- Determine if any of the Roseobacter species in the Warwick culture collection harbour inducible prophage

Chapter 2

Material and Methods

2.1 Bacterial strains and culture conditions

2.1.1 Roseobacter cultures

Table 2.1 Warwick Roseobacter collection.

Roseobacter culture	Isolated from	Physiology/property of interest	Similar strains in culture collection	Reference
Ruegeria pomeroyi	Costal seawater enriched with DMSP from Georgia (USA)	Heterotroph, demethylates DMSP		Gonzalez *et al.*, 2003
Ruegeria atlantica	Northeastern Atlantic Ocean bottom sediments			Ruger and Hofle, 1992; Uchino *et al.*, 1998
Marinovum algicola	Phycosphere of *Prorocentrum lima*			Lafay *et al.*, 1995
Roseovarius nubinhibens	Caribbean Sea	DMSP demethylation		Gonzalez *et al.*, 2003
Roseovarius crassostreae	Oyster farming water, Martha's vineyard (USA)	Thought to be cause of Juvenile oyster disease		Boettecher *et al.*, 2005
Roseovarius mucosus	Dinoflagellate culture of *Alexandrium ostenfeldii* KO287 (Biological Institute, Helgoland)			Biebl *et al.*, 2005
Sagittula stellata	Georgia (USA)	DMS oxidation to DMSO		Gonzalez *et al.*, 1997
Leisingera methylohalidivorans	Marine tidal pool, Pacific coast, California	Degrades methyl halides, cleaves DMSP		Schaefer *et al.*, 2002
Rhodobacteraceae bacterium 179	Coastal seawater from Achmelvich Bay, Scotland, enriched with methyl bromide	Degrade methyl halides	*Rhodobacteraceae* 183, 176 and 181	Schäfer *et al.*, 2005
"Ruegeria" sp. 198	Coastal seawater from Achmelvich Bay, Scotland, enriched with methyl bromide	Degrades methyl halides	*Ruegeria* 193, 197 and 257	Schäfer *et al.*, 2005
"*Roseovarius*" sp. 217	Seawater from L4 sampling station, Plymouth (UK), enriched with methyl bromide	Degrades methyl halides, cleaves DMSP, oxidises CO	*Roseovarius* 218, 216 and 210	Schäfer *et al.*, 2005
ACR04	Seawater from L4, Plymouth (UK), enriched with acrylate	Grows on acrylate		Schäfer, unpublished
"Antarctobacter" sp. ACR05	Seawater from L4, Plymouth (UK), enriched with acrylate	CO oxidation		Schäfer, unpublished

Table 2.1 cont.

Roseobacter culture	Isolated from	Physiology/property of interest	Other strains in culture collection	Reference
"Sulfitobacter" sp. ACR07	Seawater from L4, Plymouth (UK), enriched with acrylate	Grows on acrylate		Schäfer, unpublished
"Sulfitobacter" sp. ACR09	Seawater from L4, Plymouth (UK), enriched with acrylate	Grows on acrylate		Schäfer, unpublished

All liquid cultures were grown in Marine Ammonium Mineral Salts (MAMS) amended with peptone and yeast extract (MAMS-PY) see Section 2.6.1.1 at 20 °C either static or shaken at 150 rpm in an orbital shaker (New Brunswick Scientific). Solid cultures were grown on Marine Broth supplemented with Bacto agar final concentration of 1.5 % (w/v).

2.1.2 Determination of cell counts by flow cytometry

Flow cytometry was utilised to determine the cell numbers present in bacterial cultures to derive the relationship between absorbance at 600 nm and cell number. A Becton Dickinson FACScan flow cytometer with an air-cooled argon 488 nm laser was used.

Samples of known optical density (OD_{600} between 0.01 – 0.200) were fixed using glutaraldehyde at a final concentration of 2.5 % (v/v) before storage at 4 °C. The cells were stained with SYBR Gold for at least 30 min and diluted with autoclaved and filtered ASW prior to analysis. Cell numbers were determined by green fluorescence measured by channel FL1; each cell sample was examined in triplicate. The flow rate was determined by CaliBRITE beads (BD Biosciences) of which a known number was added to each sample. Each measurement was taken over a period of 2 min. All data files obtained were analysed using Cytowin v4.31 software which was developed by D. Vaulot, Roscoff, France; http://www.sb-roscoff.fr/phyto/ (Vaulot, 1989) and the data extracted for further analysis. Cell number ml^{-1} was determined using the following equation taking into account the dilution factor and the flow rate:

$$Cell\,ml^{-1} = \left(\frac{cell\,count}{2}\right) \times \left(\frac{1}{flowrate}\right) \times dilution\ factor \times 1000$$

2.1.3 Growth of *Escherichia coli* strain DH5α

E. coli strain DH5α (Hanahan, 1983) cultures, were grown at 37 ºC in lysogeny broth (LB) under agitation at 150 rpm in an orbital shaker (New Brunswick Scientific). Solid LB medium contained in addition to the above, Bacto agar at a final concentration of 1.5% (w/v). Both media were supplemented with antibiotics when required; kanamycin (Kan) 50 μg ml^{-1} or ampicillin (Amp) 100 μg ml^{-1}.

2.2 Molecular biology kits

pCR2.1 – TOPO from the TOPO® TA cloning® kit (Invitrogen, UK) was used for the cloning of PCR fragments of phage DNA, see Section 2.8.12. QIAquick PCR purification kit, QIAquick Gel extraction kit and QIAprep Spin Miniprep kit were obtained from Qiagen Ltd, UK. All kits were used according to manufacturer's instructions.

2.3 Chemicals

All chemicals were of analytical grade from Sigma Chemicals, unless otherwise stated. Restriction enzymes were supplied by New England Biolabs or Fermentas. Bacto-agar was supplied by Difco Laboratories Ltd.

2.4 Equipment

Gel tanks for running DNA gels were supplied by Pharmacia, Bucks, UK. Gels were digitised using a GelDoc-IT™ system, Ultra-Violet Products Ltd, Cambridge UK. Gel tanks for running protein gels were supplied by BioRad, Herts, UK. Sterilisation was done using a Dixons autoclave; the conditions used were to maintain a temperature of 121 ºC, 15 psi for 15 min inside the autoclave unless otherwise stated.

2.5 Centrifuges and rotors

Eppendorf tubes for centrifuging small volumes (< 2 ml) of material were centrifuged in a bench top Biofuge Pico (Heraeus) between 4000 rpm (1340 x *g*) and 13000 rpm (16000 x *g*). Volumes up to 50 ml were centrifuged in Oakridge tubes in a Beckman JA-25.50 rotor, at speeds of up to 20000 rpm (75600 x *g*) over a range of temperatures or in Falcon tubes in a Hettich Zentrifugen Rotina 46R between 2000 rpm (440 x *g*) and 4000 rpm (1780 x *g*). Volumes greater than 100 ml and less than 300 ml were centrifuged in polycarbonate tubes in a Beckman JLA-10.500 rotor, with centrifugation speeds of up to 10000 rpm (18600 x *g*) over a range of temperatures. Caesium chloride gradients were carried out in 14 ml Thinwall, Ultra-Clear™, 14 x 95 mm tubes in a Sw40Ti rotor using a Beckman Coulter Optima™ L-80 XP ultracentrifuge.

2.6 Media

Water used in the preparation of media was obtained from a Milli-Q plus 185 water purification system, Millipore Gloucester, UK

2.6.1 Media for the growth of Roseobacter

2.6.1.1 Liquid Medium – MAMS-PY

The MAMS medium was adapted from that of Goodwin *et al*, (2001) which in turn was adapted from that of Thompson *et al.* (1995). In this recipe, a higher salt concentration was used than that of Goodwin *et al.* (2001).

Table 2.2 Chemical composition of MAMS-PY.

	Per litre
NaCl	20 g
MilliQ Water	959 ml
$(NH_4)_2SO_4$	1 g
CaCl	0.2 g
MS stock solution	10 ml
Trace elements	1 ml
Peptone	5 g
Yeast extract	1 g
Post-sterilisation at 121 ºC for 15 min	
PO_4 Stock	10 ml
Vitamin solution (filter sterilised)	5 ml

Table 2.3 Chemical composition of MS Stock solution.

	Per litre	Final concentration/ mM
$MgSO_4.7H_2O$	100 g	410
$FeSO_4.7H_2O$	0.2 g	0.72
Na_2WO_4 solution	10 ml	0.001
$Na_2MoO_4.2H_20$	2 g	8.12

Table 2.4 Chemical composition of Na_2WO_4 solution.

	Per litre	Final concentration/ mM
Na_2WO_4	29.4 mg	0.1
1 M NaOH	20 ml	20

Table 2.5 Chemical composition of PO_4 Stock.

	g/L	Final concentration/M
KH_2PO_4	36	0.26
$K_2HPO_4.3\ H_2O$	234	1.03

Table 2.6 Chemical composition of Trace elements, SL10. (Widdel *et al.*, 1983)

	per litre	Final concentration/mM
HCl (7.7 M)	10 ml	77
$FeCl_2.4H_2O$	1.5 g	7.54
$ZnCl_2$	70 mg	0.51
$MnCl_2.4H_2O$	100 mg	0.51
H_3BO_3	6 mg	0.097
$CoCl_2.6H_2O$	190 mg	0.69
$CuCl_2.2H_2O$	2 mg	0.012
$NiCl_2.6H_2O$	24 mg	0.1
$NaMoO_4.2H_2O$	36 mg	0.16

$FeCl_2$ was dissolved in HCl first, then Milli-Q water was added to 1 L before the remaining components.

Table 2.7 Chemical composition of Vitamin solution. (Kanagawa *et al.*, 1982)

Vitamin	mg/L	Final concentration/mM
Thiamine HCl	10	0.03
Nicotinic acid	20	0.16
Pyridoxine HCl	20	0.12
Para-aminobenzoic acid	10	0.073
Riboflavin	20	0.053
Biotin	1	0.004
Cyanocobalamin	1	0.0007

2.6.1.2 Solid media

Marine Broth (MB) from Pronadisa (Conda) was used for agar plates (85 mm diameter Petri dishes). Bacto agar (from Difco, (BD) was added to a final concentration of 1.5 % (w/v) and autoclaved at 121ºC for a holding time of 30 min. Top agar used in the phage enumeration overlay technique was prepared at x2 concentration, autoclaved at 121ºC for 15 min and mixed with sterilized 0.8 % (w/v) Bacto agar prior to use.

2.6.1.3 Purification of agar

Bacto Agar was prepared as described by Millard (2004). Briefly, Bacto agar was washed by stirring 250 g with 5 l of water in a large beaker for 30 min. After this time the agar was allowed to settle, the water poured off and the agar filtered through a Buchner funnel with 2 sheets of 3MM Whatman acting as a filter. This procedure was repeated 3 times or until the filtered water was clear. The agar was then washed with 5 l of ethanol before a final wash with 5 l of acetone. The purified agar was then allowed to dry in a fume hood for approximately 3 days. When completely dry, it was stored in a plastic airtight container. This method was first described by Waterbury and Willey (1989).

2.6.2 Medium for growth of *Escherichia coli*

Table 2.8 Chemical composition of LB.

	g/L
Bacto tryptone	10
Bacto yeast extract	5
NaCl	5

pH was adjusted to 7.5 with 10 M NaOH

2.6.3 Medium for the storage of phage stocks

Artificial Seawater (ASW), modified as described in Wilson *et al.* (1996) was routinely used as phage buffer to maintain phage stocks. HCl was used to adjust the pH of the medium to 8.0 before sterilisation by autoclaving at 121 °C for 30 min. ASW was originally made from (x100) concentrated stock solutions; later x2 stocks of complete ASW were made routinely by the University media preparation service.

Table 2.9 Chemical composition of modified ASW (Wilson *et al.*, 1996).

	g/L
NaCl	25
$MgCl_2$. $6H_20$	2
KCl	0.5
$NaNO_3$	0.75
$MgSO_4.7H_2O$	3.5
$CaCl_2.H_2O$	0.5
Tris base	1.1
$K_2HPO_4.3H_2O$	0.03

2.6.4 Media for the supplementation of seawater

A x10 solution of yeast extract 10 g L^{-1} and peptone 50 g L^{-1} was made as supplement to seawater to allow the growth of Roseobacter cultures.

2.7 Phage techniques

2.7.1 Plaque assay

A culture of exponentially growing culture under test was concentrated by centrifugation at 6000 rpm (4020 x *g*) for 10 min. The cell pellet was resuspended in MB to give a final concentration of ~ 5 x 10^8 CFU ml^{-1}. 0.3 ml of this cell solution was added to the sample to be tested (usually 50 µl) and left at room temperature for at least 30 min to allow phage adsorption. For each cell/phage suspension, 2.5 ml of 0.4 % (w/v) molten MB agar was added, briefly mixed and poured over a 1.5 % (w/v) MB agar plate. After 30 min the plate was inverted and incubated at 20°C. Faint plaques were observed after 36 hours, but plates were left for 3 days before further testing. This procedure was used to isolate phages, create and enumerate phage stock concentrations.

2.7.2 Spot test

A lawn of Roseobacter was prepared as described in Section 2.7.1, but with no added test sample. Instead, 10 µl of sample was spotted onto the lawn and left to dry. The plate was then inverted and incubated at 20 °C. Plates were left for two or three days before observations were made.

2.7.3 Co-culturing

Roseobacter species were grown to exponential phase and approximately equal amounts (i.e. cell number) of two species were mixed together and made up to a final volume of 20 ml with MAMS-PY. The culture was then grown at 25 °C in an orbital shaker at 150 rpm, until stationary phase was judged to have been reached. It was then centrifuged at 6000 rpm (4020 x *g*) for 20 min, the supernatant removed, and split into two equal volumes. One was filter-sterilised (0.22 µm, Millipore) and the second mixed with a few drops of chloroform. The two samples were then tested for the presence of lytic phages by plaque assay and spot test using the two original species.

2.7.4 Concentration of phages from environmental samples

The viral portion of 18 L of surface seawater samples from L4 sampling station (50° 15.00' N, 4° 13.02' W), approximately 10 miles off the coast of Plymouth were was pre-filtered through low-protein binding GF/F 1.6 µm filters with 142 mm diameter (Whatman). It was then was concentrated to a final volume of approximately 20 ml by spiral cartridge ultrafiltration (300,000 Da molecular weight cut-off; Quixstand™ GE Healthcare), to concentrate the virus fraction. This work was carried out by K. Weynberg, at Plymouth Marine Laboratory. Samples were stored at 4 °C in the dark before being used in plaque assays and spot tests.

2.7.5 Phage enrichment

Seawater samples were filtered through a low-protein binding GF/F 1.6 µm filters with 142 mm diameter (Whatman). They were then supplemented with Yeast/ Peptone (see Section 2.6.1.1) to provide organic nutrients for the inoculum. The species that made up the inoculum are listed in Table 2.10 and were grouped according to the proximity of their isolation locations. 2.5 % (v/v) of freshly growing potential hosts were added and incubated for seven days at 20 °C. 20 ml of the enrichment was removed and the cell/cellular debris removed by centrifugation at 4000 rpm for 20 min at 4 °C. The supernatant was removed and filter sterilised through a 0.22 µm pore filter (Millipore). The samples were then used in plaque

assays against the species in the original inoculum. Any plaques formed were noted and purified as described in Section 2.7.1.

2.7.6 Phage purification

Roseovarius phages were purified by the removal of a single plaque using a sterile Pasteur pipette. The plug was resuspended in 1 ml sterile ASW, vortexed for 30 sec and left at 4 °C in the dark, overnight to allow for uniform resuspension of phage particles. The resultant solution was serially diluted and a repeat plaque assay performed; this process was repeated three times to ensure a clonal phage stock.

Table 2.10 Inoculum groups for phage enrichment.

Inoculum	Species/strains	Isolation location
1	*Roseovarius* 217	L4
	Roseovarius 218	L4
	Roseovarius 216	L4
	Roseovarius 210	L4
2	ACR 5	L4
	ACR 4	L4
	ACR 7	L4
	ACR 9	L4
3	*Roseovarius* 179	Achmelvich Bay, Scotland
	Roseovarius 183	Achmelvich Bay, Scotland
	Roseovarius 176	Achmelvich Bay, Scotland
	Roseovarius 181	Achmelvich Bay, Scotland
4	*Ruegeria* 198	Achmelvich Bay, Scotland
	Ruegeria 193	Achmelvich Bay, Scotland
	Ruegeria 197	Achmelvich Bay, Scotland
	Ruegeria 257	Achmelvich Bay, Scotland
5	*Ruegeria pomeroyi*	Georgia, USA
	Sagittula stellata	Georgia, USA
	Roseovarius nubinhibens	Caribbean Sea
	Ruegeria atlantica	Atlantic Ocean
	Roseovarius crassostreae	Oyster farming water, Martha's vineyard, USA
6	*Leisingera methylohalidivorans*	Marine tidal pool, Pacific coast, California
	Marinovum algicola	Stationary culture of *Prorocentrum lima* PL2V (Instituto Espanol de Oceanographia, Vigo, Spain)
	Roseovarius mucosus	Dinoflagellate culture of *Alexandrium ostenfeldii* KO287 (Biological Institute, Helgoland)

2.7.7 Production of *Roseovarius* phage stocks

The clonal phage sample made using agar plugs was used in a plaque assay procedure to produce plates with confluent lysis of the *Roseovarius* lawn. The top agar layer was removed using a flame-sterilised glass microscope slide and mixed with 3 ml (per plate) of ASW. Chloroform was added to a final concentration of 25 % (v/v), vortexed thoroughly for at least 1 min and left for at least 30 min at room temperature in the dark. The top agar and chloroform was removed by centrifugation at 4000 rpm (1780 x *g*) for 10 min at 4 °C. This regularly produced stocks of 1 x 10^8 plaque-forming units (PFU) ml^{-1}.

2.7.8 Host range

To determine the host range of the *Roseovarius* phage isolated, spot tests and plaque assays were performed. Both these procedures have been previously described in Sections 2.7.1 and 2.7.2 respectively.

2.7.9 Preparation of lytic Roseovarius phage samples for pulsed field gel electrophoresis (PFGE)

A high titre (10^5 – 10^6 PFU ml^{-1}) phage solution was prepared by removal of the top agar layer from plaque assay plate where confluent lysis could be observed. This was resuspended in 1 ml of ASW, vortexed thoroughly for at least 1 min and left overnight at 4 °C in the dark. The agar was removed by centrifugation at 13000 rpm (16000 x *g*) for 5 min and the supernatant used to make agarose plugs. These were made using 50 µl of the high titre phage solution mixed with 50 µl of molten 2 % (w/v) low melt agarose (Sigma) which was allowed to set in plug moulds. Once solid, the plugs were incubated in lysis buffer containing 100 mM EDTA, 100 mM Tris-HCl pH 9.0, 1 % (w/v) SDS and 0.5 mg ml^{-1} proteinase K, at 55 °C overnight. The plugs were then dialysed three times in Tris/EDTA buffer (TE) for 1 hour and stored at 4 °C until use. Samples were run in a 1 % (w/v) PFGE grade agarose gel.

2.7.10 Caesium chloride purification of Roseovarius phage

Roseovarius phage particles were pelleted by centrifugation of phage stock solutions at 25000 rpm (75600 x *g*) at 4 °C for 20 min in a JA-25.50 (Beckman Coulter). The glassy pellet was resuspended in ASW and layered onto a step gradient of aqueous caesium chloride (CsCl) solutions with the following concentration: 1.7, 1.5, 1.45 g ml^{-1} (w/v). The gradient was made in a Beckman Ultra-Clear^{TM} centrifuge tube (14 - 95 mm) and spun at 35000 rpm (91000 x *g*) in a SW40Ti rotor (Beckman Coulter) for at least 2.5 hours at 4 °C. The phage formed a whitish band between the 1.5 and 1.45 g ml^{-1} layers and was removed using a syringe. The resultant purified sample was dialysed twice using size 3/MWCO 12–14,000 Da, dialysis tubing for at least 2 hours in ASW. Samples were stored in glass universals at 4 °C in the dark until further analysis.

2.7.11 Transmission electron microscopy (TEM)

Phage samples were examined by TEM to confirm presence of the virus and to examine phage morphology. A Joel 1200EX TEM at 80 kV was used.

2.7.12 Preparation and negative staining of grids for electron microscopy

400 mesh Cu carbon film grids (Agar Scientific) were glow-discharged, shiny side up, for 30 sec using an Emitech K100X Glow Discharger (EM Technologies Ltd.) on a glass slide. 5 µl of CsCl purified phage sample was applied to the grid and left for 1 min to adsorb. Filter paper (Whatman) was used to remove the sample followed by immediate staining with 25 µl of 1 % (w/v) uranyl acetate (UA) for 45 sec. The UA was removed using filter paper and the staining process repeated twice. The grid was then rolled on its edge against filter paper to remove excess UA and stored in a dessicator in the dark until viewing.

2.7.13 Imaging

Negatively stained grids were placed in the TEM and objects of interest were identified on the fluorescent screen before image capture using a 1K Gatan Camera

via DigitalMicrograph™ (Gatan Inc.). Images were also processed with DigitalMicrograph™.

2.7.14 Absorption assays

2.7.14.1 Liquid

A number of modified versions of liquid absorption assay were performed, but they all followed the same basic protocol. *Roseovarius* cultures were grown in MAMS-PY to the desired concentration, harvested by centrifugation then resuspended in MB. Phage was added to a known multiplicity of infection (M.O.I.) and the same amount of phage was added to a Marine broth only control. The tubes were mixed briefly and a sample removed immediately from each for the 0 min reading. At the required points, 100 µl of sample was removed and added to 900 µl of ASW and 50 µl of chloroform, vortexed for 30 sec then centrifuged at 13000 for 5 min. The samples were then stored at 4 °C in the dark until the number of PFU remaining in the media could be determined by plaque assay. Absorption of phage to bacteria was expressed as a percentage of the number of PFU in the "bacteria-free" control. Plaque assays for each sample were carried out three times and each absorption assay was repeated three times.

2.7.14.2 Solid

Roseovarius cultures were grown in MAMS-PY to the desired concentration, harvested by centrifugation then resuspended in MB to an approx concentration of 1 x 10^8 CFU per plate. Phage was added to an M.O.I. of 0.01 and aliquots of the phage/host mixture removed for pouring of top agar. The same amount of phage was added to a MB only control. During pouring, the time of plating was noted for each plate. At the required points, the top agar layer was removed using a flame-sterilised glass slide and added to 3 ml ASW and 3 ml chloroform. The tube was then vortexed briefly to halt phage adsorption by lysis of the cells and left overnight at 4 °C in the dark. The samples were then centrifuged for 10 min at 4 °C and the number of free PFU in the supernatant determined by dilution and plaque assay. As for the liquid adsorption assays, test samples were expressed as a percentage of the "host-free" control and there were three technical replicates per sample and three biological replicates per binding assay.

2.7.15 Phage infected Roseovarius growth curve

A culture of early exponential susceptible bacterial hosts in MAMS-PY was split into five equal aliquots and infected with a range of M.O.I.'s (5, 1, 0.1, 0.01 and 0/control). Growth was then monitored through the measurement of their absorbance at 600 nm until the control culture had reached stationary phase. Growth curves were determined in triplicate for each phage-infected *Roseovarius* culture.

2.7.16 Modified Roseovarius phage one-step growth curve

An early exponential culture of host bacterial cells grown in MAMS-PY of known optical density (at 600 nm) was harvested by centrifugation (4000 rpm/ 1300 x *g*, 15 °C for 10 min). The cells were then washed in MB and centrifuged again at 13000 rpm (16000 x *g*) at room temperature for 10 min. The pellet was then resuspended in fresh MB containing enough phage to have an M.O.I. of 0.001. Prior to addition of bacterial, aliquots of the MB + phage solution had been removed to act as control samples. Both "bacteria + phage" and "phage-only" samples were then plated using the top agar overlay technique (see Section 2.7.1) and the time noted for each plate. The plates were then transferred to a dark, 20 °C incubator for the duration of the experiment.

At appropriate intervals plates were removed and the top agar layer removed with a flame-sterilised glass slide. This was mixed with 3 ml ASW and 3 ml chloroform or cold 3 ml ASW. The period of time between plating and mixing with the ASW:choloroform or cold ASW only solution was taken as time of incubation.

All samples were left at 4 °C in the dark overnight then centrifuged at 4000 rpm (1300 x *g*) at 4°C for 10 min to separate the agar and chloroform. The number of free PFU in the supernatant was then analysed by appropriate dilution and plaque assays. Each time point for bacterial/phage samples were assayed in triplicate, control samples in duplicate and each growth curve was repeated three times.

2.7.17 Prophage techniques

All cultures of Roseobacter examined for the presence of a prophage were grown in MAMS-PY at 25 °C in an orbital shaker, shaking speed 150 rpm.

2.7.17.1 Mitomycin C exposure

Cultures of Roseobacter were exposed to Mitomycin C using a method described by Chen *et al*, (2006). Briefly a 200 ml mid-exponential phase culture of Roseobacter (OD_{600} ~0.4) was split into two and one part treated with Mitomycin C to a final concentration of 0.5 µg ml^{-1} (w/v) for 30 min. The remaining culture served as a control. After exposure both cultures were washed twice by centrifugation at 7500 rpm (10460 x *g*) in a Beckman JLA-10.500 centrifuge for 10 min at 25 °C and resuspended in 100 ml of fresh MAMS-PY. Both control and treated cells were incubated for at least further 24 h and their growth measured by the OD_{600}. Samples were taken at each time point and fixed immediately with paraformaldehyde to final concentration of 1 % (w/v) and kept at 4 °C in the dark.

2.7.17.2 Epifluorescent microscopy

100 µl of fixed sample was filtered onto a 0.02 µm-pore-size 25 mm Anodisc membrane filter (Whatman) using a vacuum pump. The filter was stained sample side up with 100 µl of 1 x SYBR Green I solution for at least 15 min in the dark. Excess stain was removed with a Kim wipe and 30 µl of antifade (50 % v/v PBS, 50 % v/v glycerol and 0.1 % (w/v) p-phenylenediamine) was spotted onto the filter. This was mounted on a glass slide and bacterial cells and/or viral-like particles were observed under blue light excitation (488 nm) on an Olympus bx60 microscope.

2.7.17.3 Purification of temperate phage

One litre of mid-exponential phase lysogen culture was treated with Mitomycin C as described in Section 2.7.17.1. The washed culture was allowed to grow for a further eight hours before the cells were pelleted by centrifugation at 7500 rpm (10460 x *g*) in a JLA-10.500 (Beckman Coulter). The viral lysate was then treated with polyethylene glycol (PEG) 6000 to a final concentration of 10 % (w/v) overnight at 4 °C in the dark. The phage particles were pelleted by centrifugation at 10000 rpm (18600 x *g*) in a JLA-10.500 (Beckman Coulter) for 30 min at 4 °C and resuspended in ASW. The PEG was removed by adding an equal volume of chloroform, shaking, and then centrifuging the sample in a Hettich Rotina 46R centrifuge at 4000 rpm (1780 x *g*) for 10 min at 4 °C. The aqueous layer was removed and layered onto of a CsCl gradient for CsCl purification, see Section 2.7.10.

2.8 Molecular biology techniques

Table 2.11 Commonly used buffers and reagents.

TE buffer	10 mM Tris-HCl pH 8, 1 mM EDTA
SET	20 mM EDTA, 50 mM Tris-HCl pH 9, 0.75 M Sucrose
Loading buffer II	0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 15% (w/v) ficoll
TBE buffer (x10)	0.89 M Tris-HCl, 0.89 M Boric acid, 0.02 M EDTA pH 8
Elution buffer	10 mM Tris·HCl, pH 8.5

2.8.1 Phage DNA extraction

A high titre phage stock was mixed with an equal volume of phenol, vortexed for 30 sec and left to stand for 3 min. The mixture was centrifuged at 13000 rpm (16000 x *g*) for 3 min and the aqueous layer removed. This was mixed with an equal volume of phenol:choloroform (1:1 v/v), left to stand for 3 min and centrifuged again for 13000 rpm (16000 x *g*). Once more the aqueous layer was extracted and mixed with and equal volume of choloroform:iso-amylalcohol (24:1 v/v) and left for 3 min. The mixture was again centrifuged at 13000 rpm (16000 x *g*) for 3 min and the aqueous layer removed. To this 0.4 volume of ice cold 7.5 M ammonium acetate and two volumes of isopropanol were added, vortexed briefly and the resultant mixture left on ice for at least 20 min before being centrifuged at 13000 rpm (16000 x *g*) at 4 °C for 20 min.

After centrifugation, a while pellet was visible on the side of the tube; the supernatant was carefully removed with an aspirator and the precipitated DNA washed in 500 µl of 70 % (v/v) ethanol. The pellet was again sedimented by centrifugation at 13000 rpm (16000 x *g*) for 10 min, supernatant removed with an aspirator and left to air dry for one hour. It was then resuspended in TE or in nuclease-free H_2O.

2.8.2 Bacterial genomic DNA extraction

Exponentially growing Roseobacter cells were harvested by centrifugation (25 ml) at 4000 rpm (1780 x *g*) for 10 min and resuspended in 2 ml of SET buffer. The cell wall was digested by the addition of 100 µl of lysozyme (10 mg ml^{-1}) and incubation for 1 h at 37 °C or until the cell suspension turned from turbid to

opalescent. 200 µl of 10 % (w/v) SDS and 50 µl of proteinase K (20 mg ml^{-1}) was then added, mixed and incubated at 55 °C overnight to lyse the cells.

DNA was purified by the addition of an equal volume of phenol:choloroform:iso-amylalcohol (25:24:1 v/v/v) followed by incubation for 30 min at room temperature and centrifugation at 4000 rpm (1780 x *g*) for 10 min to separate the aqueous phase. This layer was harvested and the process repeated. The final aqueous layer was transferred to a clean tube and the nucleic acids were precipitated with 0.4 volume of cold ammonium acetate and two volumes of isopropanol. The tube was mixed by gentle inversion and the nucleic acids pelleted by centrifugation at 13000 for 10 min at 4 °C. The supernatant was removed and the pellet washed with 70 % (v/v) ethanol before being air dried. The dry pellet was resuspended in TE buffer or deionised H_2O.

2.8.3 Pulsed field gel electrophoresis (PFGE)

Pulsed field agarose gels were made of 1% PFGE grade agarose (BioRad) and were run in a CHEF Mapper (BioRad). The running buffer of 0.5 x TBE was maintained at 14 ^{0}C and gels were run according to conditions suggested by the machine's inbuilt algorithm program for optimal band separation. All gels were run with either a 0.1 – 200 kb pulse marker (Sigma) or with FastRuler^{TM} middle range and FastRuler^{TM} high range DNA ladders (Fermentas). Completed gels were stained in an ethidium bromide solution for 1 hour, then destained for 30 min in H_2O. Images were taken using a GelDoc-IT ™ system.

2.8.4 Restriction enzyme digestion

Restriction digests were carried out according to the restriction enzyme manufacturers' instructions (see Table 2.12 for list). Each reaction contained ~ 500 ng of genomic phage DNA extracted from CsCl-purified phage and was carried out in buffers recommended by the corresponding manufacturer. The digests were mixed with loading buffer II prior to loading and the restriction pattern resolved using PFGE.

Table 2.12 Restriction enzymes.

Enzyme	Manufacturer	Buffer
*Cfr*10I	Fermentas	Tango™
*Aas*I	Fermentas	Tango™
*Sex*AI	New England Biolabs	NEBuffer 4
*Aan*I	Fermentas	Tango™
*Nde*I	Fermentas	FastDigest®
*Eco*RI	Fermentas	FastDigest®
*Bam*H1	Fermentas	FastDigest®

2.8.5 *Bal31* digestion

Around 40 µg of phage DNA extracted from CsCl-purified phage was digested with *Bal31*, an exonuclease, (0.5 units μg^{-1}) at 30 °C. Samples were removed at 0, 5, 10, 20, 40 and 60 min after the addition of the enzyme and the digest stopped by heat inactivation (incubation at 65 °C for 10 min). All samples were purified by phenol-chloroform extraction as described in Section 2.7.1 and precipitated with the addition of 0.1 volume of 3 M sodium acetate and 2.5 volume of 100 % ethanol. The tubes were inverted gently, incubated at -20 °C for at least 15 min then spun at 13000 rpm (16000 x *g*) for 5 min. The resulting DNA pellet was washed in 70 % (v/v) ethanol and air dried. All samples were then digested with *Nde*1 fast digest as described in Section 2.8.4 and the digest patterns elucidated be PFGE.

2.8.6 Polymerase chain reaction

Polymerase chain reaction (PCR) was used extensively throughout this project using a range of conditions. The specific conditions for each reaction are outlined in the relevant chapters. PCR reactions were routinely carried out in 50 µl volumes in a Biometra® Tgradient/T3000 machine. The primers utilised for PCR are listed in Table 2.14

The components of a typical PCR reaction were as follows:

Table 2.13 Components of a typical PCR reactions.

Component	Vol (µl)
2x PCR Master Mix (Promega)	25
Forward primer (10 µM stock)	1
Reverse primer (10 µM stock)	1
DNA template (<250 ng)	1
Nuclease – free H_2O	22
Total	**50**

Table 2.14 Primers utilized.

Primer name	Primer sequence (5' - 3')	Target
1 forward	CAA ATT CAG GAG AGC CTT GG	Contig 1 of RLP1
1 reverse	GTG GAG ATG TGG CAG GTT GG	
2 forward	AGG TCA ACT ATG AGC GAA AC	Contig 2 of RLP1
2 reverse	TCA GTC CAG GTC CAT CAT TA	
Extended 2 reverse	GAG TTG AGT TGT CAA ACA CC	
3 forward	GTG ACGT TGA AGA TGC AGA AA	Contig 3 of RLP1
3 reverse	GGT GTA TGA CGA GTT CTT TG	
4 forward	GCA TGA TCA CCG CTG AGA TG	Contig 4 of RLP1
4 reverse	ATC CAT CTC ACC TGC GTT GA	
5 forward	CGG ATT TGT TGG TAC AGA AC	Contig 5 of RLP1
Extended 5 forward	CTC AAT GGA TAC GTT GGA AC	
5 reverse	ATA CTG AGT TCG TTC CTC GT	
Extended 5 reverse	TAC GAT ACT GAA TCC AAC CA	
6 forward	CGT GTC TTT GGA CCC ATC AT	Contig 6 of RLP1
6 reverse	CTT CGT TCC TAT CAG CCA CA	
7 forward	GCG GAA GAA AGA TCC GAA GT	Contig 7 of RLP1
7 reverse	GTC AGT ACA TCA CTT CAA AG	
8 forward	GGT AGT GCG GGA GAT GTA TC	Contig 8 of RLP1
8 reverse	CTG CAT CGT GTC AAT CAT AA	
9 forward	CTG ACA CCT GAA GAA GAG GA	Contig 9 of RLP1
9 reverse	ACG GCT GCG AAT CTC TTC AC	
Extended 9 reverse	GTT GGC AAT GAT GGT TCG AC	
9A forward	GAC GAA GAG TTG GAC AAC CT	
9A reverse	AGG TTG TCC AAC TCT TCG TC	
10 forward	TTC TAC TAC AAC TCC CCA CA	Contig 10 of RLP1
10 reverse	ATG TAG GCA GAG TAT TGA TG	
27f	AGA GTT TGA TCM TGG CTC AG	16S rRNA gene of Bacteria
1492r	TAC GGY TAC CTT GTT ACG ACT T	
341f	CCT ACG GGA GGC AGC AG	
vRNAP1	AGT TCC GAA GAT CAC ACG AG	vRNAP gene of N4-like phages
vRNAP2	ATA CCG GAT GAC CCG TAG TTC T	
vRNAP3	TGC CAT TCA TGG TCA TTG GT	
vRNAP4	ACA CGC ATG TCA GTC AGC TTC T	
M13F	CGC CAG GGT TTT CCC AGT CAC GAC	Plasmid pCR2.1
M13R	TCA CAC AGG AAA CAG CTA TGA C	

2.8.7 Primer design

A number of primers were designed to aid in the completion and assembly of the genome sequences of the two *Roseovarius* phages. Primers were analysed and selected using the web based Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/pcr_primer_stats.html). The criteria for primer

selection included: length, GC content, melting temperature, secondary structures and self-priming. An optimal primer had, in theory, a length around 20 bp, a mol G+C content of 50 – 60 %, melting temperature between 55 – 75 °C and no theoretical hairpins or self-annealing. Primer sets were chosen so that each primer pair had a melting temperature within a 10 °C range.

2.8.8 PCR product clean-up

PCR reactions that resulted in a single product were purified using a QIAquick PCR purification kit (Qiagen) following the manufacturer's instructions and products were eluted in 30 µl of elution buffer. If there were multiple products, the reaction was run through a 1 % (w/v) agarose gel and the band of interest cut out with a razor blade. DNA was then extracted using a QIAquick Gel extraction kit (Qiagen) following the manufacturer's instructions and products eluted in 30 µl of elution buffer. The QIAquick columns used in both the PCR purification and Gel extraction kits contain silica-gel membranes to which nucleic acids/PCR products adsorb. This allows excess primers, nucleotides, enzymes, salts, ethidium bromide, agarose and other impurities to be removed resulting in a pure DNA sample to be eluted.

2.8.9 Agarose gel electrophoresis

As previously stated 1 % (w/v) agarose gels were used and run in 1 x TBE buffer using a Power pac 300 electrophoresis power supply (BioRad) at approximately five volts cm^{-1}. Ethidium bromide was added to a final concentration of 0.5 µg ml^{-1} to molten agarose prior to pouring, to allow for DNA band visualisation using UV light excitation. Samples were mixed with loading buffer II prior to loading. All samples were run alongside a DNA size marker: Generuler^{TM} 1 kb DNA ladder (MBI Fermentas).

2.8.10 Genome sequencing

DNA was extracted from CsCl-purified phage and dissolved in TE. The genomes were commercially sequenced by The Gene Pool (http://genepool.bio.ed.ac.uk/) using SOLEXA and Roche 454 shotgun sequencing.

Assembly of the reads from SOLEXA sequencing was done using Velvet (http://www.ebi.ac.uk/~zerbino/velvet/) (Zerbino and Birney, 2008), reads from SOLEXA and Roche 454 were assembled using Minimus (Sommer *et al.*, 2007).

2.8.11 Contig assembly

Genome sequencing of RLP1 resulted in ten contigs. As RLP1 appeared to be highly related to RPP1 (based on their gene synteny) and the latter had assembled into one contig, RPP1 was used as a scaffold for RLP1 so that the order of the contigs could be elucidated. This was confirmed by PCR with primer pairs designed to overlap two contigs; Sanger sequencing of the products allowed for the closure of the gaps. For products too large for in-house sequencing (i.e. more than 700 bp), new primers were designed and/or the primer walking technique was utilised. The sequenced contigs and the forward and reverse sequences of each sample were imported into Seqman (DNAstar, Madison WI) to assemble the sequence data. Where mismatches occurred in alignment, the sequencing reaction was repeated and the consensus sequence taken. Each gap was sequenced at least three times.

2.8.12 TA cloning

Occasionally the end sequences of PCR products produced ambiguous sequence data, so to improve the quality of the sequence reads some PCR products were cloned using the TOPO - TA Cloning® kit (Invitrogen, UK). Cloning was carried out according to manufacturer's instructions. Sanger sequencing of the cloned inserts was carried out using purified plasmids and M13 forward and reverse primer, see Table 2.14. The amplified PCR products were purified using a QIAquick PCR purification kit (Qiagen Ltd, UK) prior to cloning and transformed plasmids were purified using QIAprep Spin Miniprep kits from Qiagen Ltd, UK. In this kit, bacterial cultures and lysed and the cleared lysate applied to the QIAprep column where plasmid DNA adsorbs to the silicagel membrane. Impurities such as proteins and RNA are washed away ensuring pure plasmid DNA is eluted.

2.8.13 DNA sequencing

Sanger sequencing reactions were carried out by the in-house University of Warwick Molecular Biology Service. Template DNA was mixed with 5.5 pmol of primer and Milli-Q water to a total volume of 10 µl and submitted to the service. Sequencing reactions were performed using the BigDye® Terminator v3.1 cycle sequencing kit from Applied Biosystems and the fragments were separated by capillary electrophoresis through a liquid acrylamide-based polymer in an ABI PRISM® 3130xI Genetic Analyser (16 capillary sequencer).

2.8.14 DNase digestion

Digestion of contaminating exogenous bacterial or phage DNA (see Sections 2.9.1.2 and 7.2.3.1) was done using TURBO^{TM} DNase (Applied Biosystems) according to the manufacturer's instructions. Briefly, two units of enzyme and the appropriate buffer were added to 200 µl of sample and incubated for 30 min at 37 °C. The enzyme was heat inactivated at 90 °C for 15 min. Samples were stored at 4°C until further analysis.

2.9 Protein techniques

Table 2.15 Table of common reagents and buffers.

Laemmli loading buffer x 4	50 mM Tris-HCl pH 6.8, 2 % (w/v) SDS, 10 % (v/v) glycerol, 1 % (v/v) β – mercaptoethanol, 12.5 mM EDTA, 0.02 % (w/v) bromophenol blue
LPS loading buffer x 2	0.1 M Tris-HCl pH 6.8, 2 % (w/v) SDS, 20 % (w/v) sucrose, 1 % (v/v) β – mercaptoethanol, 0.001 % (w/v) bromophenol blue
Coomassie stain	45 % (v/v) methanol, 10 % (v/v) glacial acetic acid, 0.5 % (w/v) Coomassie brilliant blue (R-25- (Sigma)), 45 % Milli-Q
Destain	45 % (v/v) methanol, 10 % (v/v) glacial acetic acid, 45 % Milli-Q
SDS-PAGE running buffer	1 % (w/v) SDS, 3 % (w/v) Tris, 14.4 (w/v) glycine
OM buffer	20 mM Tris-HCl pH 8, 10 mM $MgCl_2$
TEMg	25 mM Tris-HCl pH 8, 50 mM EDTA, 20 mM $MgCl_2$

2.9.1 Preparation of phage structural proteins

High titre suspensions of Roseophage stocks were purified twice on a CsCl step gradient to remove host cellular protein contaminants. The structural proteins were then extracted. All samples were resuspended in Laemmli buffer and heated at 100 °C for 10 min to denature the proteins prior to resolution on sodium dodecyl sulfate (SDS) polyacrylamide gels, see Section 2.9.8.

2.9.1.1 Whole phage extraction

Whole phage particles were pelleted by ultra-centrifugation in a TLA-100.3 (Beckman Coulter) at 80000 rpm (264500 x *g*) at 4 °C for 30 min. The supernatant was removed using an aspirator and the pellet resuspended in 15 µl of 1 x Laemmli loading buffer.

2.9.1.2 Phage ghosts

Phage ghosts were prepared according to the method described by Clokie *et al.* (2008) (Clokie *et al.*, 2008). Briefly, whole phage particles were pelleted by centrifugation at 80000 rpm (264500 x *g*) at 4 °C for 30 min in a TLA-100.3 (Beckman Coulter), resuspended in 100 µl of 10 M lithium chloride and incubated at 46 °C for 20 min. 40 units of DNase (Applied Biosystems) and an appropriate amount of buffer was added and mixed gently. This was incubated for 4 hour at 37 °C. The sample was then reconcentrated and resuspended in 1 x Laemmli buffer.

2.9.1.3 Trichloroacetic acid extraction of phage proteins

0.01 volume of 2 % (w/v) sodium deoxycholate was added to the phage sample and left on ice for 30 min. Trichloracetic acid (TCA) to a final concentration of 12 % was then mixed in and the sample again left on ice for 30 min. The precipitated proteins were then harvested by centrifugation TLA-100.3 (Beckman Coulter) at 30000 rpm (37200 x *g*) at 4 °C for 20 min. The supernatant was carefully removed, the pellet washed twice in cold acetone then left to air dry. The dry pellet was then re-suspended in 1 x Laemmli buffer.

2.9.2 Protein concentration determination

To determine the protein concentration of a sample a Bicinchoninic acid protein assay kit (Sigma) was used according to the manufacturer's instructions.

Bovine serum albumin (provided in the kit) was used as a protein standard in each assay performed.

2.9.3 Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium-dodecyl-sulfate (SDS) polyacrylamide gels were prepared as shown in Table 2.16. Gradient gels (20 x 30 cm) were poured by the gradual combination of 24 ml 10 % and 24 ml 20 % resolving gel using a BioRrad model 385 gradient former.

Gradient polyacrylamide gels were electrophoresed using a dual slab get kit (C.B.S. Scientific); mini gels were electrophoresed using a mini Protean II™ kit (Biorad). All gels were run in SDS PAGE gel running buffer. Mini gels were run at 80V and gradient gels were run overnight at 100V. All gels were run with a protein size marker, PageRuler™ Prestained Protein Ladder Plus (Fermentas).

Table 2.16 Composition of polyacrylamide separating gels.

Reagent	Stacking gel	Resolving gel		
Acrylamide 29:1	5 %	10 %	12 %	20 %
Tris HCl pH 6.8	0.126 M	-		
Tris HCl pH 8.8	-	0.375		
SDS	0.1 % (w/v)	0.1 % (w/v)		
Ammonium Persulfate	1.6×10^{-3} (w/v)	6.25×10^{-4} (w/v)		
Tetramethylethylenediamine	1/300 volume	1/800 volume		

2.9.4 Coomasie staining of polyacrylamide gels

SDS polyacrylamide mini gels and gradient gels were stained in Coomassie stain for one hour and overnight respectively. To allow for visualisation of protein bands, gels were destained with destain solution first quickly to removed excess Coomassie stain then subsequently in fresh destain solution, (changed every 2-3 hour) under gentle agitation. When the desired colouration had been achieved gels were rinsed in Milli-Q water prior to storage in 1 % (v/v) acetic acid.

2.9.5 Silver staining of polyacrylamide gels

Polyacrylamide gels with low protein content (1 – 10 ng) were stained with silver nitrate as follows. Gels were fixed for 15 min in 50 % (v/v) acetone, 1.25 % (w/v) TCA and 0.015 % (v/v) formaldehyde with gentle agitation (n.b. all incubation steps were carried out under gentle agitation) followed by three quick and one 5-min

wash in Milli-Q. Gels were then washed again in 50 % (v/v) acetone for 5 min. To increase band density, gels were then placed in a 0.16 mg ml^{-1} sodium thiosulfate solution for 1 min followed by three quick washes in Milli-Q. Gels were then placed in the staining solution (2.7 mg ml^{-1} silver nitrate and 0.37 % (v/v) formaldehyde) for 8 min and rinsed briefly in Milli-Q twice. Developing solution (0.01 g ml^{-1} sodium carbonate, 0.008 % (v/v) formaldehyde 0.02 mg ml^{-1} sodium thiosulfate) was then added until the desired band intensity had been reached. The gels were then removed from the developing solution and placed in a stop solution (1 % (v/v) acetic acid) for storage.

2.9.6 Roseobacter outer membrane protein enrichment

A modified method as described in Neumann *et al.* (1997) was used to obtain *Roseovarius* outer membrane protein (OMP) enrichment from both liquid and plate grown cultures of *Roseovarius*.

2.9.6.1 OMP enrichment of liquid grown *Roseovarius*

An early exponential phase culture of *Roseovarius* (OD_{600} between 0.1 – 0.2) was harvested by centrifugation and washed twice in ice-cold OM buffer (see Table 2.15) then resuspended in 3 ml of OM buffer and 50 mM protease inhibitor (Sigma). Cells were then disrupted by repeated passage through a pre-chilled Amico® French pressure cell at 1300 psi or freeze/thawing. This involved rapidly freezing samples in dry ice and ethanol for 30 sec, then thawing the sample in a 60 °C water bath. The process was repeated at least 15 times to ensure maximal cell lysis.

Whole cells were removed by centrifugation at 4000 rpm (1340 x *g*) for 20 min at 4 °C. Triton-X 100, to a final concentration of 2 % (v/v) was then added, mixed briefly and incubated at 40 °C for 40 min. The samples were then chilled briefly and centrifuged at 50000 rpm (103300 x *g*) for 45 min in a TLA-100.3 (Beckman Coulter) at 4 °C. The supernatant was removed, the pellet resuspended in OM buffer with 2 % (v/v) Triton-X 100 and incubated at 38 °C for a further 40 min. The sample was centrifuged again at 50000 rpm (103300 x *g*) for 45 min in a TLA-100.3 (Beckman Coulter). The pellet was then resuspended in OM buffer and stored at 4 °C until further analysis.

2.9.6.2 Harvesting plate grown *Roseovarius*

Early exponential phase *Roseovarius* (OD_{600} between 0.1 – 0.2) was harvested by centrifugation and approx 1 x 10^9 colony forming units (CFU) per plate were poured onto Marine broth agar plates using the double agar layer technique (see Section 2.7.1). After one day, plate-grown *Roseovarius* cells were harvested by removal of the top agar layer with a flame sterilised glass slide. Agar and cells were separated by vigorous vortexing for at least 1 min with ice-cold OM buffer, incubation at 4 °C for 15 min and centrifugation at 2000 rpm (440 x *g*) for 15 min at 4 °C to sediment the heavier agar leaving the cells suspended in the supernatant. This was removed and the process repeated on the agar layer until sufficient cells were harvested. Final traces of agar were removed by centrifugation at 4000 rpm (1780 x *g*) for 20 min at 4 °C; and removal of the top layer of cell/agar mixture from the resulting pellet. Cells were then resuspended in 3 ml of OM buffer and 50 mM protease inhibitor (Sigma) and the OMP selected for as described in Section 2.9.6.1.

2.9.7 Lipopolysaccharide (LPS) extraction

A LPS extraction kit (iNtRON biotechnology) was used according to the manufacturer's instructions to extract LPS from both solid and plate-grown *Roseovarius* cells. Cells were harvested as described in Sections 2.9.6.2 using TEMg buffer instead of OM buffer to aid preservation of the LPS. Briefly, cells were lysed in the supplied Lysis Buffer and chloroform added to separate the phenol layer (containing the unwanted cell membrane etc.) from the aqueous layer. The aqueous layer was removed and incubated with the Purification Buffer for 10 min at -20 °C to purify LPS from other cellular debris e.g. proteins and nucleic acids. The purified LPS was harvested by centrifugation at 13000 rpm (16000 x *g*) for 15 min at 4°C and the resulting pellet washed in 70 % ethanol and allowed to dry at room temperature. The extracted LPS was then resuspended in 30 µL of 10 mM Tris-HCl (pH 8.0).

2.9.8 SDS-PAGE analysis of LPS

LPS samples were mixed with an equal volume of LPS loading buffer and boiled for 5 min. Samples were run on a 12 % separating mini gels (see Table 2.16)

incorporating 4 M urea to resolve LPS bands. The gel was run at 50 V until the bromophenol blue band migrated to the end of the glass slab.

2.9.9 Silver staining of modified polyacrylamide gels

Modified polyacrylamide gels containing LPS samples were visualised by silver staining as described by Tsai and Frasch (1981). All incubation steps apart from overnight fixation were carried out under gentle agitation. Briefly, gels were fixed overnight in a solution of 40 % (v/v) ethanol and 5 % (v/v) acetic acid. The LPS in the gel was then oxidised using fresh fixing solution containing 0.7 % (w/v) periodic acid for 5 min followed by rinsing three times in Milli-Q for 15 min each. The gel was then added to freshly prepared staining solution (0.02 M sodium hydroxide, 0.6 % (w/v) silver nitrate, approx 4 % (w/v) ammonium hydroxide) for 10 min followed by washing three times, for 10 min, in Milli-Q. The water was then replaced by developer (0.0185 % formaldehyde and 0.05 mg L^{-1} citric acid) until the desired band intensity was reached. Development was terminated by placement of the gel in 10 % (v/v) acetic acid for 1 min followed by repeated rinses in Milli-Q water.

2.9.10 Mass spectrometry (MS)

Protein bands of interest from SDS PAGE were excised and tryptically digested using the manufacturer's recommended protocol on the MassPrep robotic protein handling system (Waters). The extracted peptides from each sample were analysed by means of nanoLC-ESI-MS/MS using the NanoAcquity/Q-ToF Ultima Global instrumentation (Waters) using a 45 minute LC gradient. All MS data was corrected for mass drift using reference data collected from the [Glu^1]-Fibrinopeptide B (human - F3261 Sigma) sampled each minute of data collection.

The data were then used to interrogate a provided database made up of the predicted protein sequences from open reading frame identification (see Section 2.10.1) appended with the **c**ommon **R**epository of **A**dventitious **P**roteins sequences (http://www.thegpm.org/cRAP/index.html) using ProteinLynx Global Server v2.3. All protein identification was carried out in the in-house Biological Mass Spectrometry and Proteomics Facility, University of Warwick.

2.10 Bioinformatic analyses

2.10.1 Open reading frame (ORF) identification

Open reading frames (ORFs) were predicted using the freely available gene prediction programs GeneMark™, heuristic approach (Besemer and Borodovsky, 1999) and GLIMMER 3.01 (NCBI) (Salzberg *et al.*, 1998; Delcher *et al.*, 1999). The final set of predicted ORFs for each genome was created by amalgamation of the two sets of results from GeneMark and GLIMMER. For ORFs with near complete overlap between the two programs, the longer of the two predictions was kept.

2.10.2 Database searches

A number of Basic Local Alignment Search Tool (BLAST) comparisons were carried out on the predicted ORFs using different databases (Altschul *et al.*, 1990). Initially, a search using the BLASTp algorithm of the predicted amino acid sequences from the two *Roseovarius* phages to a database containing all bacteriophage amino acid sequences freely available in July 2008 (created by A. Millard) was performed. This was then repeated using BLASTp against the non-redundant protein sequences database at the National Center for Biotechnology Information (NCBI). The results from the two searches were compared to assign putative function to each predicted ORF.

Selected sequences such as the predicted major capsid protein were also used in blast searches against the "Global Ocean Survey" (GOS) databases and the "All metagenomic ORF peptide" database using the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) resource (http://camera.calit2.net/index.php). Another database utilised was RoseoBase (http://roseobase.org/), a genomic resource for marine Roseobacter (Dept. of Marine Sciences, Uni. of Georgia). These were carried out to further annotate the predicted ORFs and to analyse the environmental distribution of orthologous genes.

2.10.3 ORF/genome comparisons

Phage genome comparisons were carried out using a variety of techniques: WebACT (http://www.webact.org/WebACT/home) (Abbott *et al.*, 2007) and Mauve

genome alignment software (Darling *et al.*, 2004) to visually compare genomes and a custom Perl script (written by A. Millard) to elucidate the % identity value between ORFs. Briefly, a nucleotide database containing all predicted ORFs of the genomes to be compared was created and a tblastx comparison with an e-value set to 0.001 was performed with the query genome. The comparison file was then truncated so that it only contained the top hit for each ORF. Then the custom Perl script performed a sequence alignment for each ORF pair using the ClustalW program (Thompson *et al.*, 1994) to obtain a percent identity matrix from which the % identity value was calculated.

Whole genome analyses were performed with the DOTMATCHER tool from EMBOSS (Ian Longden, Sanger Institute, Cambridge) with a threshold score of 50 and a window size of 15 bp.

2.10.4 Motif and regulatory element identification

The programs used to identify motifs and regulatory elements are shown in Table 2.17. Default parameters were used in all searches.

Table 2.17 Programs used in bioinformatic analyses.

Feature	Program	Reference
tRNA	tRNAcan-SE 1.21	Lowe and Eddy, 1997
Transcriptional terminators	TransTermHP	Kingsford *et al.*, 2007
RBS	RBS finder	Suzek *et al.*, 2001

2.10.5 Phylogenetics

Amino acid alignments were created in SeaView (Gouy *et al.*, 2010) using either ClustalW (Thomson *et al.*, 1994) or MUSCLE (Edgar, 2004) then imported into CHROMA (Goodstadt and Ponting, 2001) for formatting. To create phylograms alignments were converted to a nexus file format and a Bayesian estimation of phylogeny was carried out using MrBayes (Huelsenbeck and Ronquist, 2001) and subsequent trees were formatted with TreeView (Page, 1996). Sequences for

alignment were selected from top blast hits and related sequences of interest (e.g. from related phages) as well as outgroups to aid creation of phylogenetic trees.

2.10.6 Prophage finder

A web-based PHP application, Prophage finder (Bose and Barber, 2006), http://bioinformatics.uwp.edu/~phage/ProphageFinder.php, was used to search for putative lysogens in the Warwick Roseobacter culture collection. Briefly, the prophage finder program identifies possible prophage loci by performing a BLASTx search within a database of predicted amino acid sequences derived from all sequenced phage genomes available at:
www.ncbi.nlm.nih.gov/genomes/static/phg.html.

The output is then sorted according to the number of significant matches found clustered together. Sequenced Roseobacter isolates (as of Dec 2006) and freely available on Roseobase (http://www.roseobase.org/) were analysed by this program with default settings.

2.10.7 Protein domain prediction

Putative protein domains were predicted using the DomPred server (http://bioinfadmin.cs.ucl.ac.uk/dompred/). As the majority of the structural proteins identified by mass spectrometry did not contain any domains with obvious similarities to other known domains, DomSSEA was used to identify domains from secondary structure element alignment (Marsden *et al.*, 2002). Analysis of amino acid sequences using the DomPred server also produced a PRSIPRED prediction. Default parameters were used in all predictions.

Chapter 3

Isolation and characterization of two *Roseovarius* bacteriophages

3.1 Introduction

Twenty years ago Bergh *et al.* discovered that, contrary to previously held beliefs, viruses are highly abundant in a variety of aquatic ecosystems (Bergh *et al.*, 1989). Since then, further studies have shown that they are the most abundant biological entity in the oceans (Suttle, 2005) and can exceed that of host bacteria by a factor of 3-10 (Wommack and Colwell, 2000). However, the number of marine phages currently isolated and characterised is far outnumbered by those obtained from other ecosystems e.g. soil, human body etc. In practical terms this is due to two main reasons; firstly the studied i.e. cultured marine bacterial community is merely a fraction of the smorgasbord of prokaryotes found within the water column. This fact is illustrated by the high fraction of metagenomic DNA sequences produced by the Sorcerer 2 expedition that cannot be matched to culture representative microorganisms (Rusch *et al.*, 2007). Consequently, though evidence of a plethora of "unusual" phages exists in metagenomic data (such as ssDNA phages; Section 1.3 or T7-like phages containing ENV1/2/3 DNA polymerase; Section 1.73), no isolates can be identified, as the host has not yet been identified. Secondly, many isolated marine microorganisms are difficult to maintain in culture (both in liquid and on solid media) probably due to the constant flux of biological and physical parameters experienced in the ocean that can be impossible to reproduce in a laboratory. Cultivation of host cells is a necessity for isolation and propagation of phage. Taking into account these limitations, standard techniques used to isolate novel marine bacteriophage can prove challenging.

Fortuitously many members of the Roseobacter lineage are amenable to laboratory culture. As such the classical method of phage isolation, incubation of a growing culture of host bacterium with an inoculum (seawater) which is assumed to contain phage, is possible. However, bacteriophage isolation can prove time-consuming as it is, in essence, a game of chance; chance that a viable, infectious phage is present in the water sample tested and chance that the host is in the correct growth stage for the virus to successfully infect. The probability of expeditiously finding a lytic phage can be improved by attempting a number of simple isolation techniques, the relative merits of which will be discussed in this chapter.

Though the first plaque assay was probably performed by Felix d'Herelle (who along with Twort is credited as the discoverer of phages), the double-layer agar technique, used widely used today, is attributed to Gratia (1936). It is a simple method used for primary phage isolation, identification of susceptible hosts, as well as to enumerate phage particles and to produce stocks. The exact plaque assay protocol used in this project is described in Section 2.7.1 but the fundamentals remain the same regardless of phage/host used. A top layer of molten agar containing a phage/bacterial mixture is poured onto an agar plate fixing them into a semi-solid matrix. During plate incubation, the bacteria grow to cover the surface in a lawn, but concurrently the embedded phage particles infect and kill susceptible bacteria in their vicinity. Their progeny phages in turn, then infect other bacteria in their vicinity producing a zone of clearing, or "plaque". The majority of known phages form large and well-defined plaques that are easily observed, but some give rise to small, turbid plaques that are very difficult to detect and enumerate. Turbid plaques are characteristic of temperate phages; their cloudy appearance is due to the presence of lysogens within the zone of clearing. These cells survive, due to ability of most phages to prevent superinfection by various systems such as Sie, see Section 1.6.2.3. To improve the resolution of small plaques or to enhance the contrast between clear and turbid plaques some researchers have used triphenyltetrazolium chloride, a redox indicator which can differentiate between metabolically active and inactive cells; it can be added to the soft agar either before of after plating (Pattee, 1966; McLaughlin and Balaa, 2006). Others have used antibiotics to increase the size of otherwise small plaques, in some cases up to 50 times (Santos *et al.*, 2009). However, none of these methods enable the isolation of phages that lysogenize at near 100 % efficiency. If the supposition that temperate phages predominate the marine viriome (see Section 1.4.6) and the claim by Freifelder that more than 90% of known phages are temperate (Freifelder, 1989), are both true then plaque assays bias researchers towards easily cultivable lytic phages. Despite these problems, the plaque assay remains an important technique in phage biology.

Post-isolation, characterisation still remains paramount to the understanding of novel phages. As interest in the wider impact of phages grows, research has moved away from the characterisation of individuals and has begun, instead, to focus on global distribution, diversity and evolutionary relationships of groups of marine phages. Towards that end, novel phages have been identified and characterised

through commonly found genes such as portal proteins, (see Section 1.7.1) gp20 in T4-like phages, and DNA polymerase (Section 1.7.5). These have been targeted either by directed-PCR or metagenomic studies (Labonté *et al.*, 2009; Angly *et al.*, 2006). However, as there are so few Roseobacter phage isolates and no clear relationship between them, it is difficult carry out such investigations. Instead, this project's objective was to make use of genomic sequencing and proteomics to characterise the individual Roseobacter phages isolated, see Chapters 5 and 6. Other basic phage features such as host range, morphology, binding and infection profiles were also determined and are outlined in this chapter.

3.2 Results and Discussion

3.2.1 Isolation attempt 1: Co-culturing

In an investigation of the literature Ackerman and DeBow showed that between 21 % and 60 % of all environmental bacteria are actually lysogens (Ackermann and DeBow, 1987), i.e. they harbour a hidden prophage element. Previously a method has been developed in this laboratory to isolate a temperate phage from its lysogenic host known as co-culturing (Rapson, 2002). This method was successfully used to isolate several temperate phages from *Synechococcus* spp. (Millard, 2004). It was hoped that any phage isolated by this method would be able to infect other Roseobacter spp.

The process of co-culture was carried out as described in Section 2.7.3 and is a modified version to that of Millard (2004). Modifications were required due to the different growth rates and conditions required by Roseobacter isolates compared to those of *Synechococcus.*

As shown in Table 2.1 the Warwick Roseobacter culture collection consists of 15 species, three of which have a number of strains giving a total of 24. The latter number was too large to test all strains in a pairwise fashion therefore one strain from each species was selected for co-culture experimentation. Unfortunately, the filtrate from all co-cultures tested negative for bacteriophage.

3.2.2 Isolation attempt 2: Concentration of Environmental Samples

As the co-culturing technique did not yield any bacteriophages, a second approach with an emphasis on increasing the probability of a successful host/virus encounter was employed. It was thought that by concentrating the natural viral community of an environmental sample, e.g. 20 L of seawater, through tangential flow filtration and adding aliquots of the concentrate to potential hosts, the Warwick culture collection of Roseobacter spp. could be screened expeditiously for phage.

It should be noted that the actual filtering was not carried out by the author, instead phage concentrates were kindly provided by K. Weynberg at the Plymouth Marine Laboratory, Plymouth. As described in Section 2.7.4 samples were collected from the sampling station L4, (ten miles off the coast of Plymouth) between April 2006 – Feb 2007 and used in a number of double agar overlay plaque assays and spot tests as described in Section 2.7.1 and 2.7.2.

Despite numerous attempts no lytic Roseophage was obtained using viral enrichemnt. This may have been due to pre-filtration removing viruses (Paul *et al.*, 1991) though a large (1.6 µm) filter was employed. It was also noticed that plates used in spot tests often had bacterial growth on/around the test area which may have obscured any lawn clearance.

3.2.3 Isolation attempt 3: Viral enrichment/amplification

As phage isolation via concentration of the natural viral community was not successful, an alternative approach of amplifying the number of viruses already present in a volume of seawater through viral enrichment, before screening, was utilised. This is a modification of a method described by Suttle (1993) the major change of which was the introduction of several potential hosts into the nutrient enhanced seawater rather than separate enrichments for each Roseobacter strain. A detailed description of this method can be found in Section 2.7.5. Briefly, seawater samples, supplemented with Yeast/Peptone, were inoculated with Roseobacter cultures to enrich any Roseobacter phages present. After incubation for seven days, cell/cellular debris were removed by centrifugation and the supernatant used in plaque assays against the species in the original inoculum.

Clear plaques were observed on lawns of "*Roseovarius*" strain 217 (hereafter referred to as 217) tested with viral-enriched seawater from Langstone Harbour,

Hampshire collected 17th Sept 2005 and on plates containing *Roseovarius nubinhibens* (*Rsv. nubinhibens*) and enriched L4 sample (collected 24-11-98). The bacteriophages were known to have originated from the environmental samples and not from a co-culture effect as this had been previously tested for in Section 3.2.1. The plaques were then picked and made clonal as described in Section 2.7.6. The putative phage were named using the nomenclature suggested by Kropinski *et al.* (2008); vB_Rsv217_RLP1 (RLP1, Roseovarius Langstone Podovirus) and vB_RsvN_RPP1 (RPP1, Roseovarius Plymouth Podovirus) respectively. As a number of bacteria can be predatory on other bacterioplankton and produce plaque-like artefacts, the presence of phage as the infective agent was proved using TEM, see Section 3.2.5.

3.2.4 Plaque morphology

Plaques for both RLP1 and RPP1 were clear and circular as can be seen in Figure 3.1. The sizes for both ranged from 0.5 – 2 mm after *ca*. 48 hours incubation. Upon prolonged incubation the plaques did not enlarge, but as the lawn matured (slight change in colour), they became better defined. This was likely due to the phage's inability to infect their respective hosts during stationary phase. As no plaque turbidity was observed, it was concluded that RLP1 and RPP1 were lytic phages.

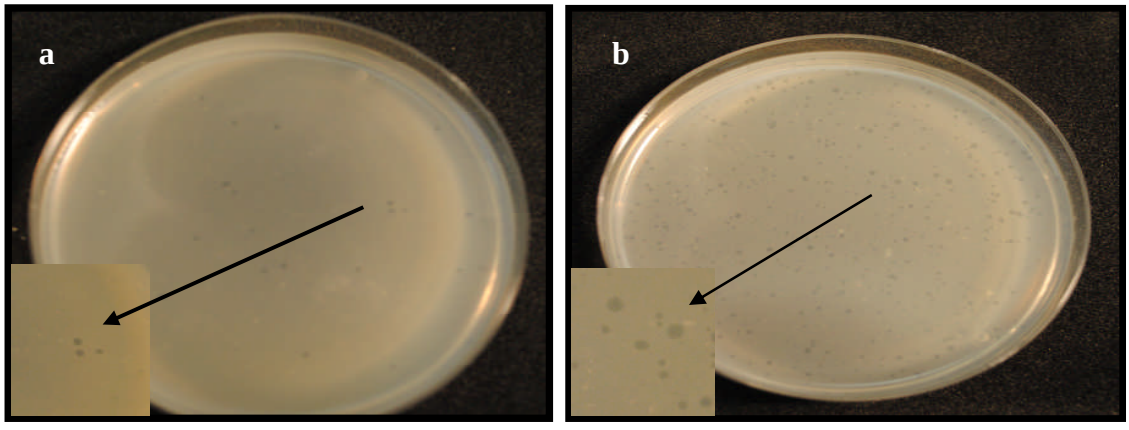


Figure 3.1 Plaques formed on *Roseovarius* bacterial lawns. a) RLP1 plaques on a lawn of *Rsv.* 217 b) RPP1 plaques on a lawn of *Rsv. nubinhibens*.

3.2.5 *Roseovarius* phage morphology and classification

To determine which family of viruses RLP1 and RPP1 belonged to, TEM was carried out. Images from CsCl-purified samples can be seen in Figures 3.2 and 3.3.

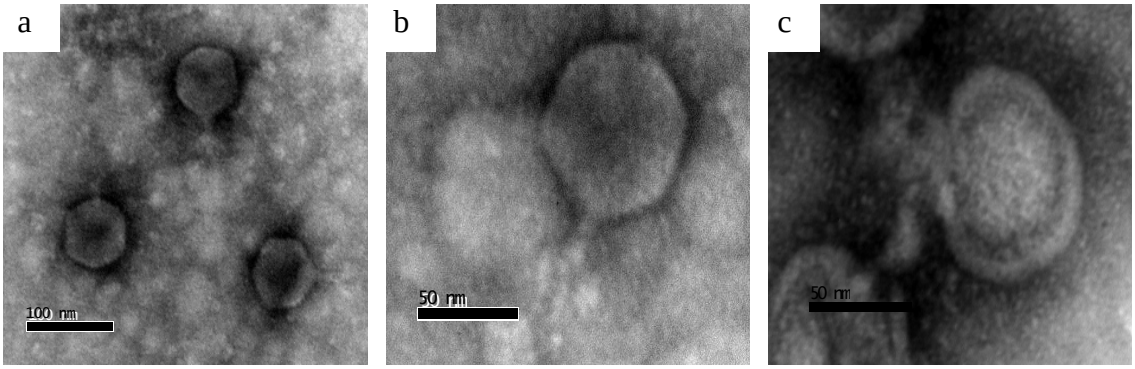


Figure 3.2 TEM micrograph of RLP1 stained with uranyl acetate . Viral particles were stained with 2 % (w/v) uranyl acetate. Magnification: (a) x 120 000, (b) x 250 000, (c) x 300 000

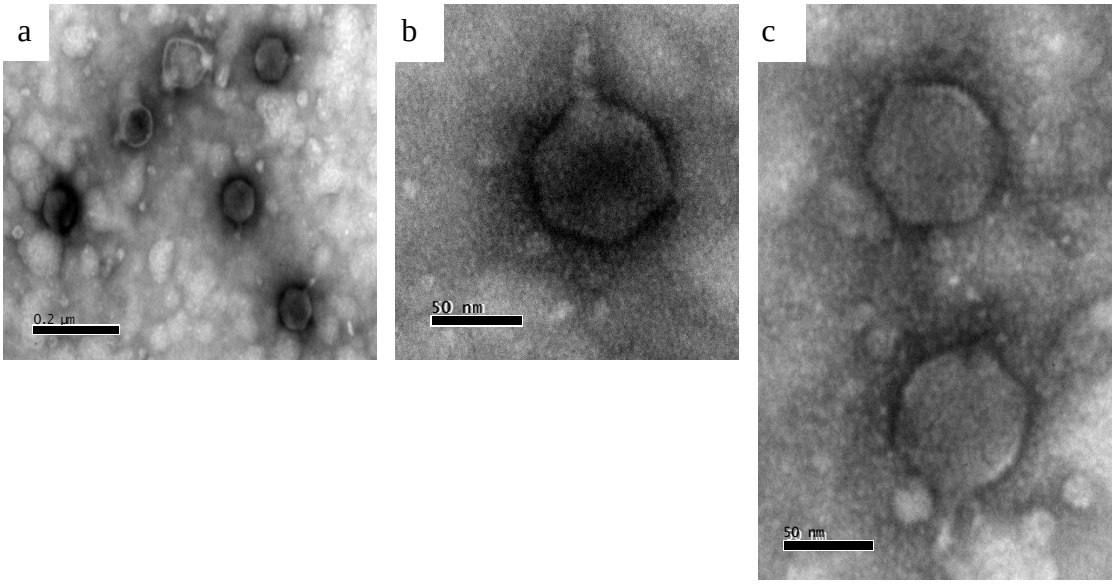


Figure 3.3 TEM micrograph of RPP1 viral particles stained with uranyl acetate. Magnification (a) x 75 000, (b) x 200 000, (c) x 250 000

TEM indicated both phages had icosahedral heads with short tails; these characteristics are typical of the family *Podoviridae*. *Podoviridae* belong to the tailed order of bacteriophages known as *Caudovirales* which consists of viruses with binary symmetry (a head with cubic symmetry and a "helical" tail) and double stranded DNA genomes, see Section 1.3 (Ackermann 2005).

Digital images of the heads of the two phages were measured in DigitalMicrograph (Gatan). The mean and standard deviation for each isolate was

calculated from a minimum of 10 particles. RLP1 and RPP1 had a capsid head size of 72.4 ± 2 nm and 77.4 ± 5 nm respectively.

3.2.6 Host range

RLP1 and RPP1 were tested against all other species and strains in the Warwick Roseobacter culture collection. The results, shown in Table 3.1, suggest that RLP1 has a narrow host range whereas RPP1 may have an increased infective potential perhaps limited to the *Roseovarius* genus. The change in RLP1 infectivity could be explained by a mutant/alternative strain of phage becoming dominant in the phage stock.

Table 3.1 Host range of RLP1 and RPP1. Ticks indicate successful infection; lawn clearance was observed at decreasing titres until single plaques were formed. Crosses indicate unsuccessful infection; no clearing and or plaques were observed either with spots test or plaque assays.

	RLP1	RPP1
Ruegeria pomeroyi	X	X
Ruegeria atlantica	X	X
Marinovum algicola	X	X
Roseovarius nubinhibens	X†	✓
Roseovarius crassostreae	X	X
Roseovarius mucosus	X	X
Sagittula stellata	X	X
Leisingera methylohalidivorans	X	X
Rhodobacteraceae bacterium 179	X	X
Rhodobacteraceae bacterium 183	X	X
Rhodobacteraceae bacterium 181	X	X
Rhodobacteraceae bacterium 176	X	X

	RLP1	RPP1
"Ruegeria" sp 198	X	X
"Ruegeria" sp 193	X	X
"Ruegeria" sp 197	X	X
"Ruegeria" sp 257	X	X
"*Roseovarius*" sp 217	✓	✓*
"Roseovarius" sp 216	✓	✓*
"*Roseovarius*" sp 210	✓	✓*
"*Roseovarius*" sp 218	✓	✓*
ACR04	X	X
*"Antarctobacter"*sp ACR05	X	X
"Sulfitobacter" sp ACR07	X	X
"Sulfitobacter" sp ACR09	X	X

† RLP1 initially did infect *Rsv. nubinhibens*, however, from Dec. '09 onwards the suspected host was no longer susceptible at any titre or to cultures re-grown from original -80 °C stocks.

* RPP1 infects the closely related strains 217, 216, 210 and 218, but at a reduced sensitivity compared to that of *Rsv. nubinhibens.*

Though RPP1 was found to infect both *Rsv. nubinhibens* and the *Rsv.* 217 strains, the infectivity of the phage against the later host was observed to be less than that against *Rsv. nubinhibens*. As such the relatively efficiency of plating (EOP), the titre of phage on a given bacterial cell line compared to the maximum titre observed, was calculated. For RPP1 plated with *Rsv.*217 this was found to be 0.01 where an EOP of 1 corresponds to a maximal number of plaques obtained from the same amount of RPP1 lysate added to culture of *Rsv. nubinhibens*. The variation observed may have been due to host factors such as O antigens masking receptor sites or the presence of restriction endonucleases.

Fig. 3.4 shows the phylogeny of various *Roseovarius* species based on the partial 16S rRNA gene. It shows that the closely related strains 217, 216, 210 and 218 are highly related to each other and to *Rsv. mucosus*. This species was amongst those tested and did not show susceptibility in plaque assays or spot tests with RPP1. In contrast, *Rsv. nubinhibens*, the initial cultivating host, does not appear to be closely related to the 217 strains. However, as many closely related species often do not share susceptibility to the same phage, such phylograms cannot be used to predict phage host ranges.

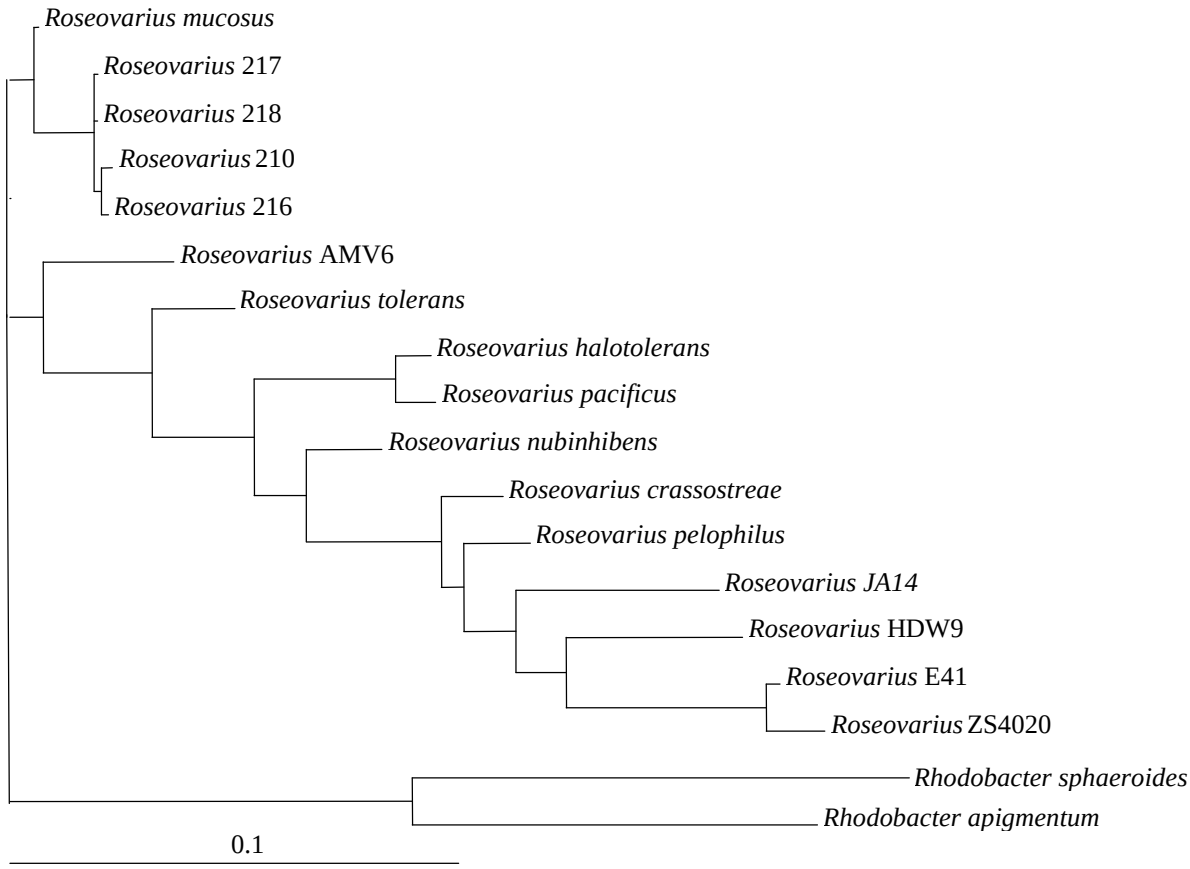


Figure 3.4 Phylogram (made using MrBayes) based on average branch length of 16S rRNAgenes from 18 *Roseovarius* species. The root was determined using *Rhodobacter sphaeroides* and *Rhodobacter apigmentum*. The scale bar indicates expected changes per site.

3.2.7 Pulsed field gel electrophoresis

The genome size of RLP1 and RPP1 was elucidated by PFGE and found to be around 70 kb (Fig. 3.5). According to the International Committee on Taxonomy of Viruses (ICTV) taxonomy list 2009, the *Podoviridae* family consists of two subfamilies and six genera as shown in Table 3.2. RLP1 and RPP1 genome sizes correspond with the N4-like and ΦEco32-like virus genera, which are larger than the average size of *Podoviridae,* but both still fall within the range providing further evidence for their classification as Podoviruses.

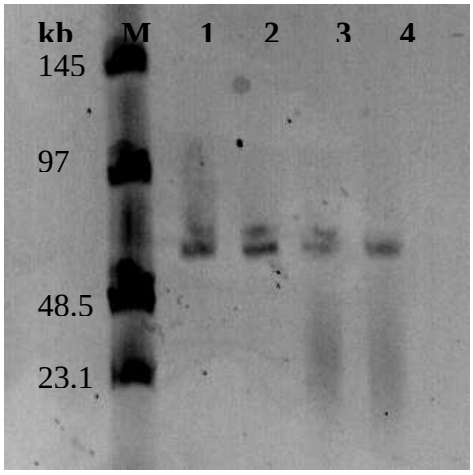


Figure 3.5 Pulsed field gel electrophoresis of purified RLP1 and RPP1 genomic DNA. Phage DNA was obtained by **i**n plug digestion of CsCl purified phage. M: Marker, sizes given in kb; lanes 1 & 2: RLP1 genomic DNA, lanes 3 & 5: RPP1 genomic DNA

Table 3.2 Genera of the Podoviridae family.

Genus	Type species	Genome size / kb
ΦKMV-like viruses†	*Enterobacteria* phage ΦKMV	42.5
SP6-like viruses†	*Enterobacteria* phage SP6	43.8
T7-like viruses†	*Enterobacteria* phage T7	39.9
AHJD-like viruses*	*Staphylococcus* phage AHJD	16.9
Φ29-like viruses*	*Bacillus* phage Φ29	19.3
BPP1-like viruses	*Bordetella* phage BPP-1	42.5
ε15-like viruses	*Salmonella* phage ε15	39.7
LUZ24-like viruses	*Pseudomonas* phage LUZ24	45.6
N4-like viruses	*Enterobacteria* phage N4	70.1
P22-like viruses	*Enterobacteria* phage P22	41.7
ΦEco32-like viruses	*Enterobacteria* phage ΦEco32	77.5

* *Picovirinae* subfamily

† *Autographivirinae* subfamily

3.2.8 DNA restriction pattern

To further characterise RLP1 and RPP1, restriction digests of the two phage genomes using various restriction endonucleases were carried out. The digest patterns for both phages resembled each other suggesting that the genomes are highly similar, but not identical. This suggests that RLP1 and RPP1 maybe isolates of the same phage, however as shown in Section 3.2.9 the infection profiles of the two phages were distinct. Their relationship, taking into account all determined characteristics, is discussed in further detail in Section 7.1.4

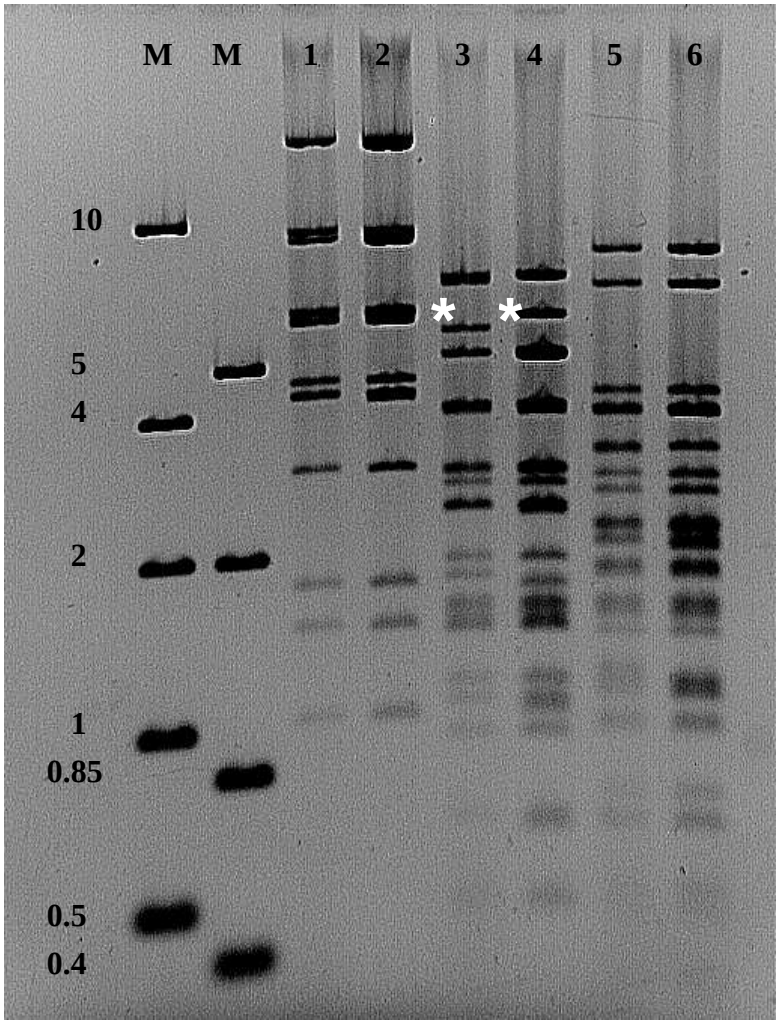


Figure 3.6 Restriction pattern of digested phage DNA. ~ 500ng of CsCl purified phage was digested using various restriction enzymes according to the manufacturer's instructions. M: Marker, sizes given in kb; lanes 1, 3 and 5 contain digested RLP1 genomic DNA, lanes 2, 4 and 6 contain digested RPP1 genomic DNA. Lanes 1 & 2: *NdeI* digests, lanes 3 & 4: *BamHI* digests and lanes 5 & 6: *EcoRI* digests. * indicates obvious band size difference between the two phages.

3.2.9 One-step growth assay

The infection of host cells by their respective phages was examined by one-step growth assays. These allow the characterisation of the infection process and produce three results: latent period, eclipse period and burst size. Burst size is the number of phage released from one infected cell during lysis. The eclipse period is defined as the time interval between infection and the appearance of the first viable infectious particle and is measured by plaque assays of chloroform-treated samples (taken during the assay). Finally, the latent period is defined as the time interval between infection and host cell lysis. This is calculated by determination of the number of free phages in the medium (after a burst event, the number of free phages increases dramatically). Samples are typically centrifuged to remove host cells so only mature, free phages present at the sampled time are counted. However, as RLP1 and RPP1 appear to only bind and infect when immobilised in agar and not in liquid (discussed in detail in Chapter 4), the standard protocol was adapted. The assay was instead carried out on plates rather than in liquid; the protocol is described in full in Section 2.7.16.

In the modified one-step assay carried out for RLP1 and RPP1, immediate processing of samples (e.g. centrifugation and subsequent plaque assay) was not possible as both infected/uninfected host and nascent/mature phage were embedded within the top agar matrix and therefore not available for plaque assays. Instead an additional overnight incubation in phage buffer, to allow diffusion of phages out of the agar, was required prior to enumeration. As such, the infection process could plausibly have continued during the incubation step unless the cycle was rapidly and completely stopped at the sampling point. In the case of eclipse period samples, treatment with chloroform would immediately quench the infection thus providing a true "snap-shot" of the process. For latent period samples, it was believed that cold ASW would rapidly slow down the infection to a rate that would render any lag between believed sampling point (time the sample was processed) and actual sampling time (point at which the infection process was halted), negligible. Unfortunately, the number of free phages measured in the first hour of infection always exceeded the original titre added which meant that the infection process continued after addition of cold ASW (see Fig. 3.7). Consequently, the concentration

of free phages in the medium during infection could not be determined and only the total PFU per sample is plotted below.

The results as shown in Fig. 3.8 suggest that the eclipse period for both phages is between 2-3 hours after which the PFU per infected cell continues to increase. PFU per infected cell then appears to plateau between 4-6 hours before increasing again. This may suggest that during this period a burst event has occurred, however, without a free phage infection profile this cannot be verified. RLP1 also appears to have a larger burst size compared to that of RPP1 shown by comparison of the rate increase in PFU per infected cell. A rough estimate would suggest that RLP1 has a burst size of ~100 PFU $cell^{-1}$, RPP1 ~10 PFU $cell^{-1}$. A precise number for burst size for either phage could not be calculated as the true latent period could not be determined. In addition, it is likely that the infected cells were not synchronised, neither could multiple infections of one cell be eliminated (since the infected cells were not diluted as occurs in a standard growth assay). As such any results from the modified version of this assay should be viewed only as a preliminary investigation into the infection profile of these two phages.

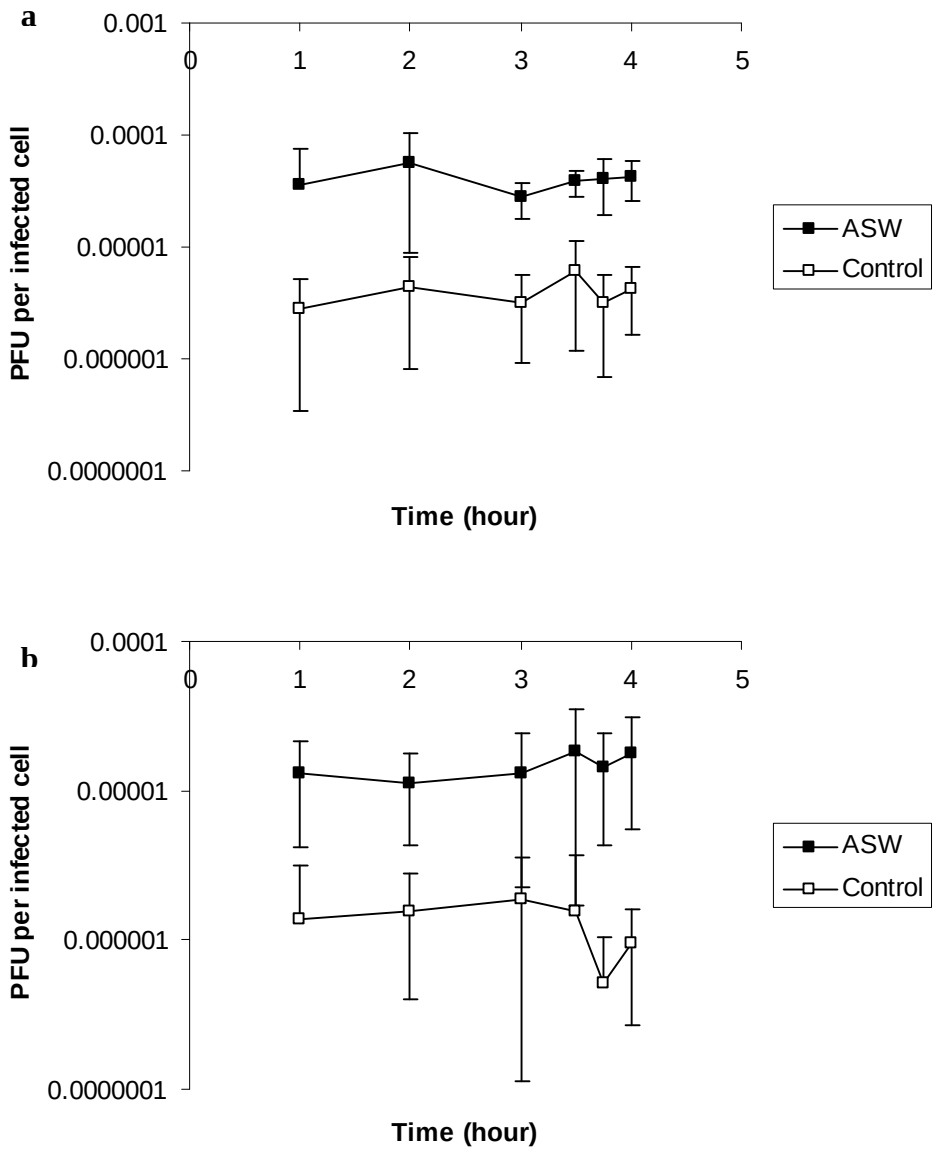


Figure 3.7 Number of free phage in the agar/media during a one-step growth curve experiment. a) RLP1 with *Rsv.* 217 b) RPP1 with *Rsv. nubinhibens.* When the sample was mixed with cold ASW the infection cycle did not stop as shown by the increased number of free phage in the agar in the ASW samples compared to the bacteria-free controls. If the sample had been "quenched" the number of free phage per infected cell would have been lower than that of the bacteria-free control as the free phage should have bound to the bacterial host cells.

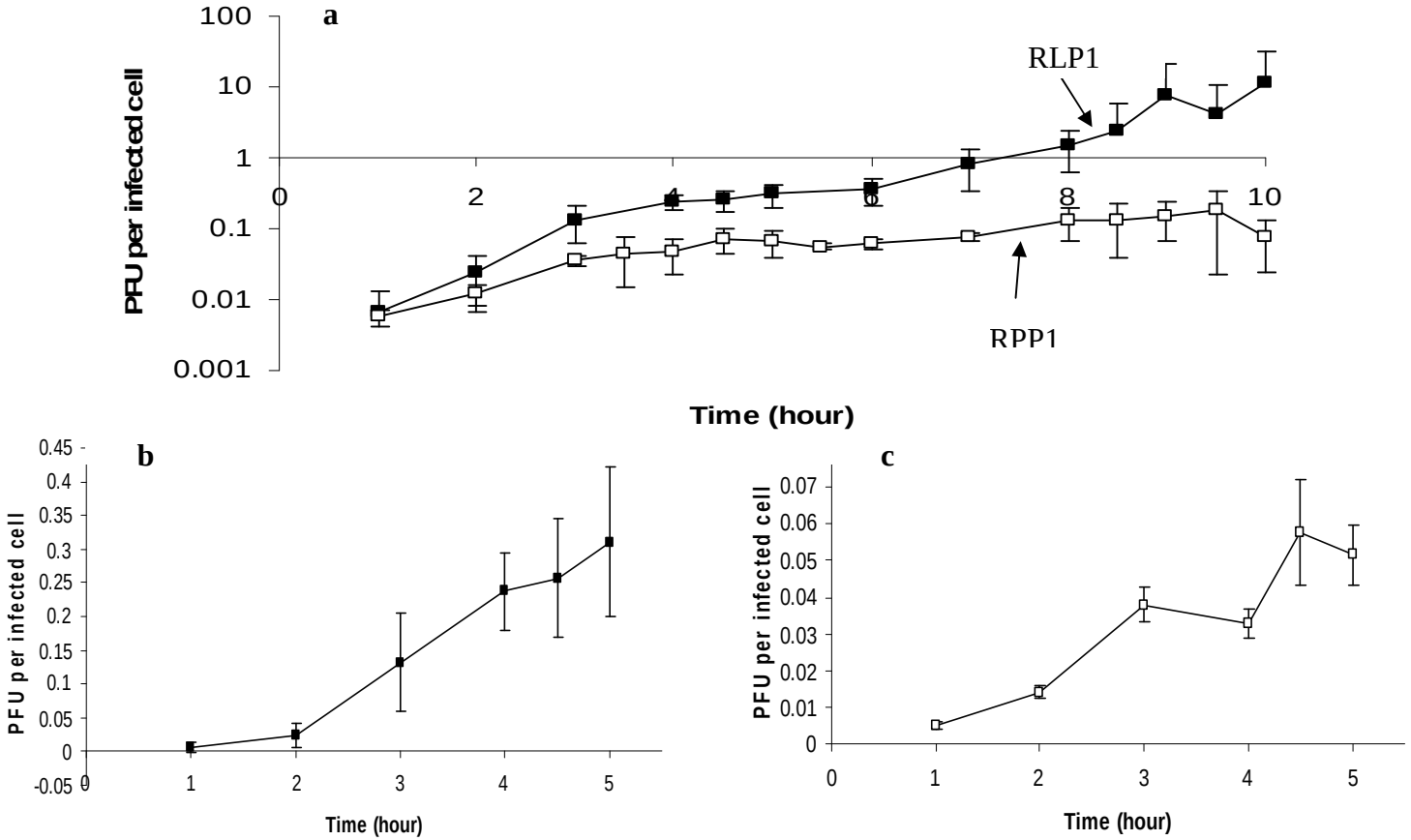


Figure 3.8 One-step growth curves for RLP1 and RPP1. a) One step growth curve of RLP1 and RPP1 on *Rsv*. 217 (■) and *Rsv. nubinhibens* (□). The number of phage increases over time indicating infection has occurred. The first 5 hours of b) RLP1 and c) RPP1; there is a marked increase in phage between 2 and 3 hours which suggests a burst event has occurred during this period.

3.3 Concluding comments

RLP1 and RPP1 are the first *Roseovarius*-specific phages to be isolated. Notably there are few Roseobacter phages in the literature (Rohwer, 2000; Chen *et al*., 2006; Zhang *et al*., 2009; Zhao *et al.*, 2009) despite the prevalence of Roseobacter in the microbial coastal community. It is of interest to note that RPP1 was isolated from L4 whereas its host, *Rsv. nubinhibens* was isolated around 6500 km away from surface waters of the Caribbean sea, 22° 3·7' N, 74° 35·2' W (Gonzales *et al*., 2003), Fig 3.8. This would suggest either that RPP1 has a broad host range and is capable of infecting other Roseobacter species present around L4 and/or that *Rsv. nubinhibens* is a geographically widely distributed species probably carried by ocean currents such as the Caribbean Current, the Gulf Stream and its extension the North Atlantic Current which collectively pass through the Caribbean to the coastal waters of the UK (Fig 3.9). A 16S rRNA BLAST search of the GOS database indicates that close relatives of *Rsv. nubinhibens* do appear to be present around the coastal areas of Americas (Fig 3.10). However, as the dataset from European waters has not yet been published, it cannot be concluded if this is the case. In contrast to this, RLP came from Langstone Harbour around 230 km away from the isolation location of 217 at L4 (Fig. 3.11).

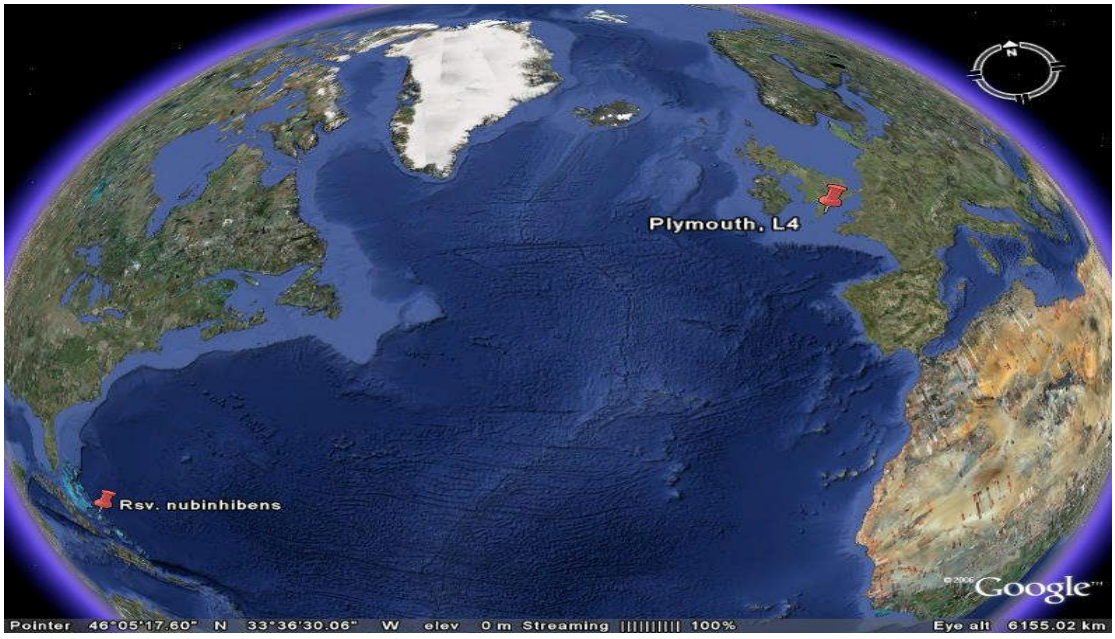


Figure 3.9 Image of the isolation locations for RPP1 and *Rsv. nubinhibens*. A Google Earth image illustrating the considerable distance between isolation locations of host and phage. http://earth.google.co.uk/

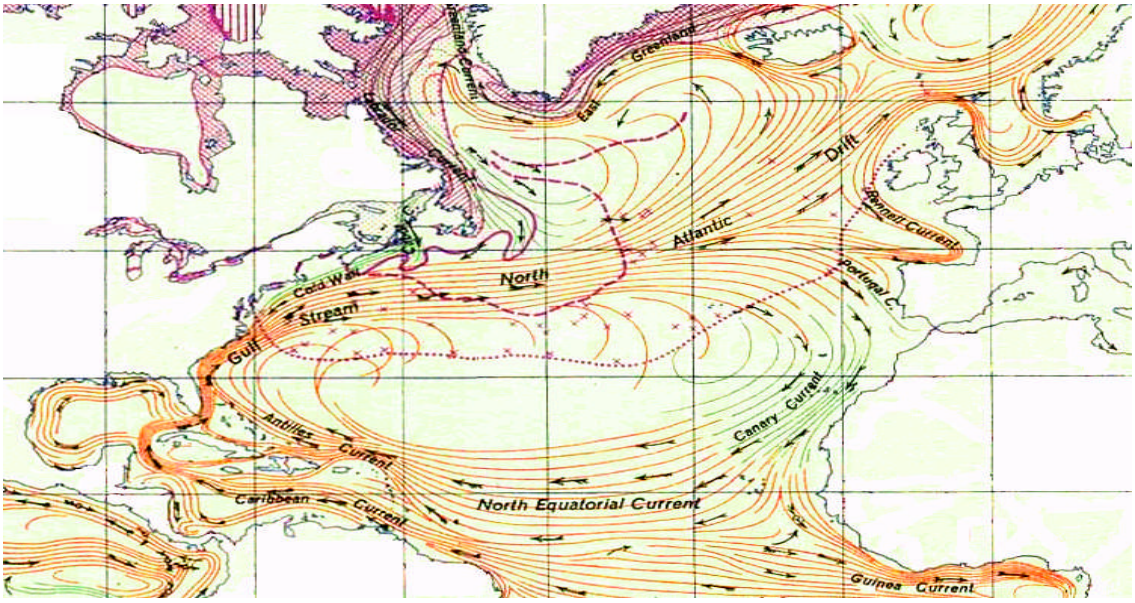


Figure 3.10 The North Atlantic Gyre and currents in the North Atlantic Ocean. Cropped image taken from Ocean Currents and Sea Ice from Atlas of World Maps, United States Army Service Forces, Army Specialized Training Division. Army Service Forces Manual M-101 (1943). http://www.lib.utexas.edu/maps/world.html

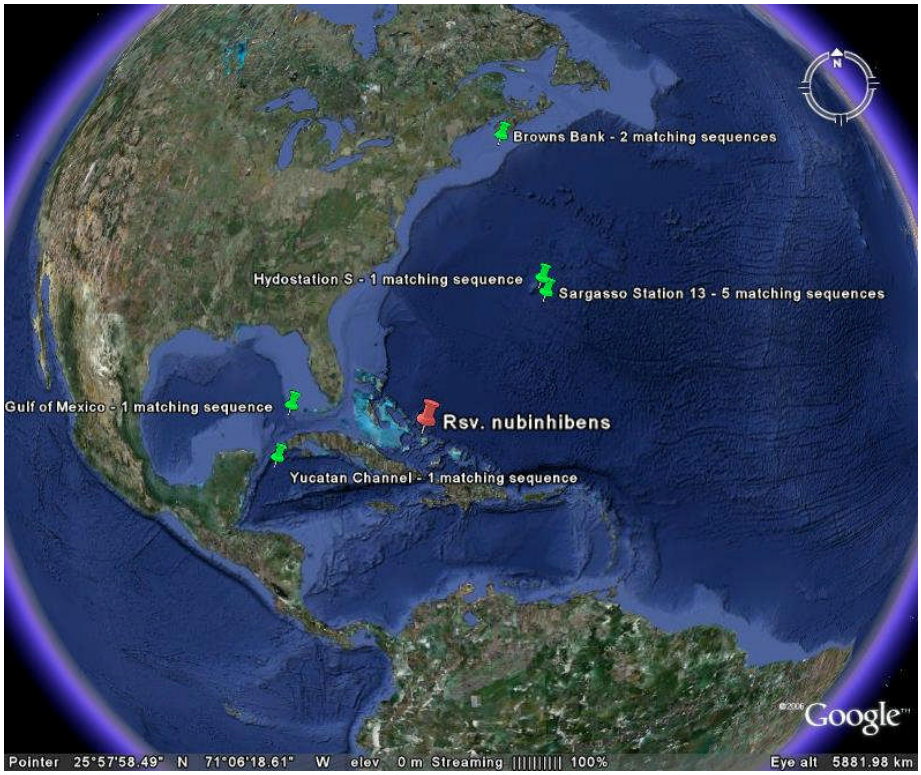


Figure 3.11 Sampling locations of BLAST search hits using the 16S rRNA sequence of *Rsv. nubinhibens* as query against the Global Ocean Survey (GOS) database (7/1/10). Taken from Google Earth http://earth.google.co.uk/

Figure 3.12 Satellite view of the South coast of the UK indicating the isolation location of RLP1, in Langstone Harbour, and *Rsv.* 217, in L4. Taken from http://earth.google.co.uk/

Isolation and cultivation of novel phages is still relevant as it allows for the elucidation of the physiology, biogeochemistry and ecosystem role of bacteriophages, but it can pose many temporal and financial limits. Consequently, protocols which hasten this process are invaluable. During the attempt to isolate a novel Roseophage three different methods; co-culture, viral concentration of environmental samples and viral enrichment were employed. Though each of these techniques have proved successful in the past (and likely will continue to do so), in this project, only the last viral enrichment, yielded a positive result. This is due to the enrichment method being most suited to the challenge: to isolate a phage that infects an easily cultured host from an environmental sample containing the presumptive phage, but in relatively low abundance. However, as discussed in Chapter 5, RLP1 and RPP1 have unusual binding properties meaning a significant degree of enrichment may not have occurred. In which case, isolation of the two phages was merely a serendipitous event.

Failure of the co-culture method to isolate lytic bacteriophage was not wholly unexpected as induction of a prophage from one species by definition means that it contains the genes for lysogeny. Consequently, it is likely to re-infect another

susceptible species and undergo lysogeny rather than enter the lytic cycle. To test for successful induction the filtrates, post co-culture, should have been screened for presence of viral-like particles (for methods see Section 1.4.1). As this was not performed, the success or failure of the co-culture method cannot be determined.

In the laboratory, temperate phage isolation through co-culturing has been used with great success with methicillin-resistant *Staphylococcus aureus* (MRSA) (Y. Jia, University of Warwick, personal communication). This species is well known as a carrier of prophages which play an important role in pathogenicity by carrying accessory virulence factors such as Panton-Valentine leukocidin, staphylokinase, enterotoxin A, and exfoliative toxin A (Goerke *et al.*, 2009). The presence of prophages in MRSA can also aid adaptation to its harsh environment by increasing genome plasticity through lateral gene transfer. A study by Goerke *et al.* (2006) showed that extensive phage dynamics was a specific trait that characterised infectious strains of *S. aureus* compared to that of nasal commensal isolates. Though the oceans can be considered a harsh environment often lacking in key nutrients, the conditions do not favour highly mobile prophage elements. Instead, the slow growth of bacterial hosts maintained at relatively low concentrations are conditions that favour a stable temperate phage only induced during blooms where many highly related and/or clonal susceptible bacteria are prevalent. As a result, such phages would have a narrow host range (Section 1.4.6; Wommack and Colwell, 2000; Miller, 2005).

It is intuitive to understand the appeal for the viral enrichment method as the probability of finding a virus increases with the volume of water screened. Unlike enrichment, concentration is relatively non-selective as it is a physical and not biological technique. Nevertheless, phage isolation after sample concentration was not fruitful. The most plausible explanation for this is that the viral concentrate samples were not taken at the times of year during which the particular Roseobacter spp. used in this study (or their close relatives) were prevalent and so neither were their phages. This is surprising as the literature indicates that members of the Roseobacter lineage appear to dominate bacterioplankton communities during the summer algal bloom, around June/July and often coincides with high concentration of dissolved DMSP (Gonzales *et al.*, 2000; Zubkov *et al.*, 2001), a compound often metabolised by Roseobacter isolates (see Section 1.1.6). Images taken by Landsat of a coccolithophore bloom (a source of DMSP) taken in the summer, support this time-

frame see Fig. 3.12. A study of a North Sea coccolithophore bloom in June 1999 found a single species related to the Roseobacter genus accounted for 24% of the bacterial population number and up to 50% of the biomass (Wilson *et al.*, 2002). As such it is logical to assume that during the summer bloom, Roseobacter spp. and their phages should be abundant in the waters surrounding L4, especially in the latter stages, i.e. the crash where phages have been shown to be the main cause of bloom termination (Martínez *et al.*, 2007). This period was represented in the viral concentrate tested therefore Roseophage isolates should have been plentiful in the sample and easily isolated. Another possible explanation for this negative result, is that the wrong hosts were used in the isolation attempts. The inability to isolate phages from this method only highlights the problem of isolation of novel bacteriophage even if a carefully planned, directed approach is used.

Figure 3.13 A Landsat image of the South West coast of UK of a coccolithophore bloom. The image includes the L4 sampling station. Taken 24th July 1999 by Andrew Wilson and Steve Groom of Plymouth Marine Laboratory (Hays *et al.*, 2005).

Studies in phage characterisation can be categorised according to five main aspects: molecular, environmental, evolutionary, ecological and applied. As the molecular/biochemical aspects of RLP1 and RPP1 will be discussed in detail in Chapter 4; mainly evolutionary and environmental/ecological facets of these novel phages will be examined here.

The key to understanding phage evolution lies in classification, however, as there are no firm criteria for genus and species delineation, no universal method for

phage classification exists. One outcome of this is the lack of an accepted phage nomenclature which has lead to differing opinions over phage description. Consequently, phage classification has been described as "much an art as a science" (Ackermann, 1999). As examined in Section 1.3, the ICTV uses the polythetic species concept mainly guided by particle morphology and nucleic acid composition. Using these criteria, RLP1 and RPP1 have been identified as putative members of the *Podoviridae* genus whose members have been widely found in the oceans. Currently tailed phages, members of the order of *Caudovirales*, have been found to dominate the marine environment. As the order is believed to be monophyletic (Ackermann, 1999) it is likely that the lineage has a dominant shared ancestor. This leads to the question of why do tailed phages prevail in the ocean? Do tails provide an advantage? However, these questions are fundamentally flawed as they are based on the assumption that the phages isolated thus far are a true representation of the virioplanktonic community. This is unlikely to be true. Many direct viral visualisation studies have shown that both tailed and non-tailed forms are present in seawater though the dominance of either group appears to differ according to sample location (Proctor, 1997) see Table 3.3. It should be noted that the studies referenced were based on direct visualisation i.e. culture-independent approaches thus differences between prokaryotic and eukaryotic viruses may not have been taken into account. In addition, virus tails can be lost in sample preparation creating tail-less artefacts.

It appears that current culture-dependent techniques favour the isolation of tailed phages. The underestimation of phage numbers by traditional plaque assays is a well documented phenomenon; Ashelford *et al.* (2003), showed that direct counts by electron and epifluorescence microscopy were around 350 times greater than the highest number estimated by plaque assay enumeration.

With the advent of metagenomics it is now possible to advance our understanding of marine viral biodiversity purely through sequencing. However, many such studies such as the 2002 study by Breitbart *et al.* have shown that the majority of DNA sequences obtained from the viral community do not have homologues in the nucleotide database, they are unique unknown sequences (Breitbar *et al.*, 2002). As it is not possible to use metagenomic data to definitively determine classification of members of the virioplanktonic community, as long as they remain unknown, the evolutionary history of marine phages will remain unclear.

Table 3.3 Morphological types of free viruses in the marine environment.

Adapted from (Procter, 1997). c: coastal, o: oceanic

Site	Tailed	Non-tailed	Reference
Yaquina Bay (c)	Majority		Torrella and Morita, 1979
Raunefjorden (c)		Majority	Borsheim *et al*., 1990
Caribbean Sea (o)	Majority		Proctor and Fuhrman, 1990
Gulf Stream (o)	Majority		Proctor and Fuhrman, 1990
Long Island Sound (o)	Majority		Proctor and Fuhrman, 1990
Chesapeake Bay (c)		57 %	Wommack *et al.*, 1992
S. California Bight (c)	Majority		Cochlan *et al.*, 1993
Gulf of Bothnia (c)	Majority		Cochlan *et al.*, 1993

The host ranges of both RLP1 and RPP1 were narrow, though the number of species tested was limited by availability of hosts and these may not have been a true representation of environmental cultures. Though laboratory-cultured Roseobacter species do appear to be closely related to cloned environmental sequences (Wagner-Döbler and Biebl, 2006), phage susceptibility can be greatly affected by strain type. In a study by Holmfeldt *et al.* (2007) isolates of *Cellulophaga baltica* showed unique phage susceptibility and sensitivity to 46 phages. Thus, although a cultured strain may be resistant, its close relative found in the ocean may be sensitive.

To add another dimension to an already complex phage-host community, infectiousness rarely remains constant as typified by the change in RLP1's host range; it is likely that in the environment neither phage nor bacteria remain static in their respective infectivity and sensitivity. Consequently, we must consider laboratory based analyses as momentary glimpses into the ever-changing, complex ecological web of phage-host interactions.

Chapter 4

Binding properties of *Roseovarius* phages RLP1 and RPP1

4.1 Introduction

For any free phage the first step in a phage's life cycle is adsorption to a host cell surface. From there, penetration of the cell wall and injection of genetic material can take place. Determination of this initial contact is indispensable to studies of phage ecology in particular the impact of the predator (phage) on the prey (host) populations. As with all binding interactions, adsorption involves two structures: one of bacterial origin (a surface receptor) and one of phage origin (a part of the phage anatomy e.g. tail fibre). As such, it is perhaps not surprising that Krueger, in 1931, established that binding of phage and host (dead or alive) broadly follows first-order kinetics defined the equation below:

$$\kappa = \frac{2.3}{\mathrm{Bt}} \log \frac{Po}{P}$$

Where κ is the adsorption rate constant, B is the concentration of bacteria cells and t is the time interval between the titre taken at *Po* (original) and P (final). A typical adsorption assay should produce results as shown in Fig. 4.1 which is taken from Krueger's original paper.

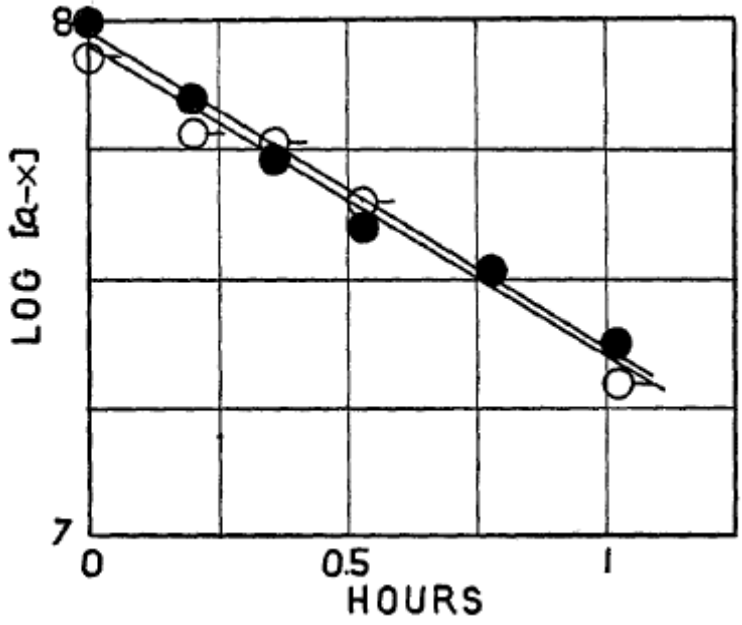


Figure 4.1 Adsorption profile of live (○) and dead (●) bacteria with phage illustrating the velocity of the reaction, taken from Krueger, 1931.

Many factors have been found to influence the adsorption rate constant; presence of salts (in particular divalent cations) and organic compounds, growth phase of the host cell, agitation and temperature. Surface receptor density was also thought to significantly affect the rate of adsorption but it was shown that this is not the case. Schwartz (1976) demonstrated that under optimal conditions the rate of adsorption of phage λ to *E. coli* K12 cells only increased tenfold when the density of receptor protein at the cell surface increased from 30 molecules per cell to 6000 molecules per cell. Furthermore, the rates calculated by Schwartz indicated that the theoretical maximum adsorption constant (based on the assumption that all collisions led to irreversible adsorption) was only 2.5 times higher than the experimentally determined adsorption constant. This and other results, illustrated the near perfect efficiency of phage binding as it seemed that positive phage capture occurred on each phage/host collision. However, the small size of receptors and their scarcity on the cell surface means that this is extremely unlikely. This paradox has led researchers to hypothesize that phage adsorption is more complex than a simple, one-step first order kinetics reaction. It now believed that phage binding consists of two distinctive steps consisting of fast, reversible adsorption and desorption, and slow, irreversible binding each of which is determined by three rate constants: k, k', and k''. Based on this, the binding process can be defined by the equation below where B is bacteria, P is phage and BP and BP* represent the transient and the stable bacterium/phage complexes (Moldovan *et al.*, 2007).

$$B + P \underset{k'}{\overset{k}{\longleftrightarrow}} BP \xrightarrow{k''} BP^{*}$$

Amongst the *Caudovirales*, which make up over 95% of all described phages (Ackermann, 2009), receptor specificity is determined by tail fibre structures found at the distal end of phage tails. It has been shown that recombination of phage adhesion genes between T-even phages can alter host specificity, and the addition of tail fibre genes can broaden the host range of a specific phage (Tétart *et al.*, 1998; Scholl *et al.*, 2001). Consequently, LGT of whole or partial gene sequences can mediate acquisition of diverse host range determinants thus allowing families of phages to cross species boundaries and infect taxonomically distant hosts.

Almost every structure exposed or extending from a bacterium's surface can be exploited as a phage receptor. Indeed, phages have been found to bind to a variety of surface receptors as shown in Table 4.1. For researchers, this has made phages useful laboratory tools for example, in the characterisation of strains or for the selection of receptor-deficient (therefore resistant) mutants; but for bacteria, phage-driven selection favours the alteration, masking or complete removal of susceptible cell surface structures (see Section 1.6.2.1). However, phage resistance often involves a trade-off with fitness; for example, mutations in the receptor molecule in phage resistant *E. coli* cells can correspond to reduced rate of resource uptake (Lenski, 1988). Furthermore, the magnitude of the cost of resistance can increase relative to number of phages the host is resistant to. A classic example of this is the much higher cost of resistance for *E. coli* strains resistant to both T4 and T7 (Lenski, 1988). Due to their relative simple requirements, bacterium/phage systems have often been used as paradigms in the study of the evolutionary arms-race and the cost for those involved (Bohannan and Lenski, 2000).

Table 4.1 Surface receptors of various bacteriophages.

Surface receptor	Example	Reference
Pili	M13	Pemberto, 1973
Flagella	SP3	Shae and Seaman, 1984
Lipopolysaccharide	T7	Kruger and Schroeder, 1981
Surface protein	λ	Randall-Hazelbauer and Schwartz, 1973
Teichoic acid	SP50	Givan *et al.*, 1982
Capsule	K29	Bayer *et al.*, 1979

Due to such wide ranging implications in areas such as evolution and biological kinetics, characterisation of phage/host binding through identification of the phage receptor and calculation of the adsorption constant can often be found in many preliminary phage studies. However, during the experiments to determine this value for RLP1 and RPP1, it was established that they exhibited an atypical phage/host relationship. The results from these investigations and its unexpected resolution are outlined below.

4.2 Results and Discussion

4.2.1 Infection

It was observed that in liquid cultures, regardless of the phage:host ratio, the culture never completely cleared. Cultures with phage added to an MOI of > 1 continued to grow albeit at a slower rate. It would appear that a steady-state was reached between cells that lysed and "resistant", growing cells. The cultures with an MOI of 5 displayed an initial decrease in growth after addition of phage possibly due to lysis from without however, it too continued to grow and did not "crash". Lysis from without is caused by the adsorption of many phage to a single living cell. Above a threshold value, the contents of the cell are liberated by a distension and destruction of the cell wall. During this process no new phages are formed.

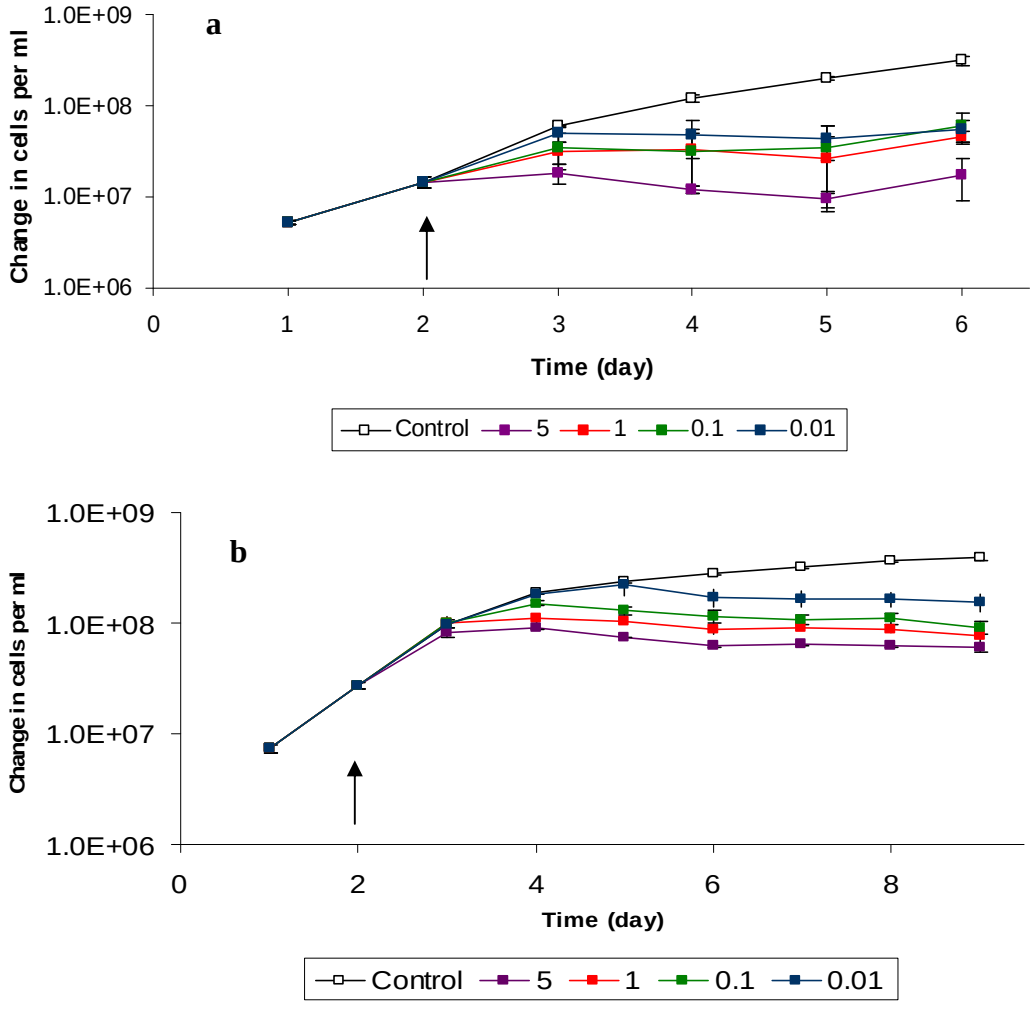


Figure 4.2 Liquid batch cultures of a) RLP1 with *Rsv.* 217 and b) RPP1 with *Rsv. nubinhibens* infected at different MOIs. Arrow indicates point at which phage was added; the key indicates the MOI.

Both RLP1 and RPP1 however, are known to be capable of true infection on agar plates as during spot assays using serially diluted phage stocks, the zone of clearance was observed to progress from confluence lysis to single plaques (see Fig. 4.3). If lysis from without had occurred, single plaques would not have been observed. To explain the lack of complete lysis in liquid cultures, five hypotheses were proposed:

1. **Low phage:host affinity** – poor adsorption in liquid cultures could be due to low affinity of the phage tail fibres to the host receptor; physical proximity by fixation in agar would increase the phage's ability to bind
2. **Oxygen content of media** – low O_2 concentration in liquid compared to solid media might affect host physiology and consequently phage binding
3. **Growth phase-dependent receptor expression** – the host receptor to which the phage binds is transiently expressed, semi-immobilisation of phage and growing host in sloppy agar ensures adsorption when the receptor is expressed
4. **Co-factor in Bacto Agar** – a co-factor present in Bacto Agar promotes irreversible phage binding
5. **Host receptor expression in top agar** – the host outer membrane protein profile changes due to its embedding in sloppy agar and the appropriate receptor is expressed

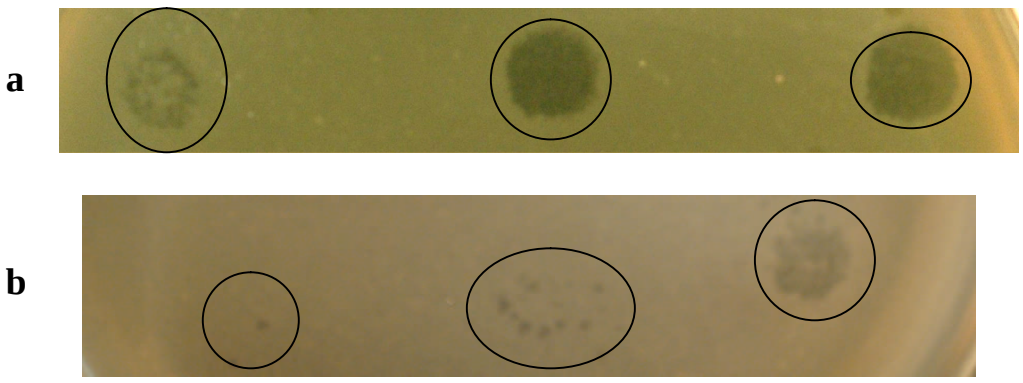


Figure 4.3 Spot test plates of a) RLP1 on *Rsv.* 217 and b) RPP1 on *Rsv. nubinhibens.* The spots were of increasing serial dilution from right to left. Circles indicate the location of the zone of clearance.

4.2.2 Hypothesis 1

RLP1 and RPP1 display a low phage:host affinity. If the phages do have a low affinity to its host, there should be a moderate decrease in free phage over time.

4.2.2.1 Liquid adsorption assay – method 1

A binding assay over 5 hours using host in a 14 ml liquid culture with a MOI of 0.1 was carried out, see Fig. 4.4. In the adsorption assays *ca.*14,500 phage were added in the *Rsv.* 217 with RLP1 binding experiment and *ca.* 8,700 phage in the *Rsv. nubinhibens* with RPP1 adsorption assay. The results showed that the percentage of free phage in the supernatant did not decrease significantly suggesting that the phages do not bind bacterial cells to an appreciable level when in liquid Marine Broth.

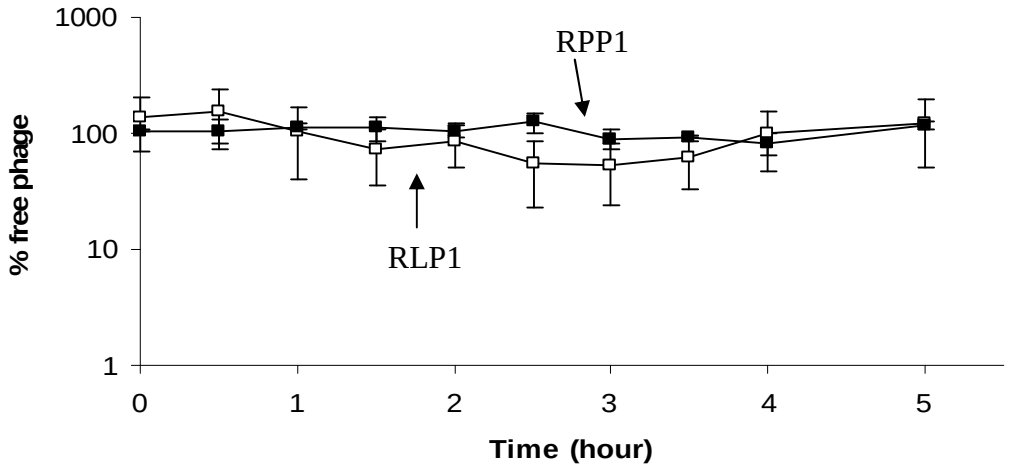


Figure 4.4 Adsorption of RLP1 to *Rsv.* 218 and RPP1 to *Rsv. nubinhibens* in liquid media over five hours. The graph shows the results of three separate experiments. Assay was carried out at an MOI of 0.1 and the supernatant was titrated every hour to determine the amount of phage unabsorbed/free phage. There was no appreciable decrease in free phage during the five hour assay.

4.2.2.2 Liquid adsorption assay – method 2

If the two phages do have a low affinity to their host, but require them to be at a minimum concentration prior to adsorption, an assay carried out in a reduced volume of broth but with the same MOI as the previous 14 ml experiment would show a noticeable decrease in free phage over 1 hour. Figure 4.5 illustrates binding in 2.8 ml did not result a demonstrable change in liquid binding though the degree of error in the assay was considerable, especially in the first twenty minutes.

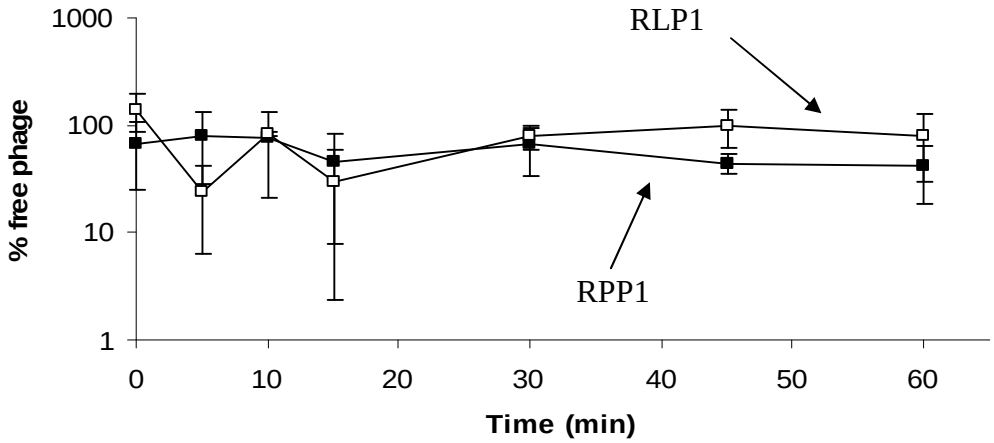


Figure 4.5 Adsorption assay of RLP1 on *Rsv.* 217 and RPP1 on *Rsv. nubinhibens* at an MOI of 0.1, in 2.8 ml of marine broth. Representative graph of three separate experiments. Though there is considerable variation in results particularly in the first 20 minutes, there results show there is no appreciable decrease in free phage after 1 hour.

From the results of the two versions of adsorption assay, hypothesis 1 does not seem likely. In addition, the lack of adsorption in 14 ml liquid cultures cannot be attributed to a volume effect as the two phages are known to infect in 2.8 ml as this is the corresponding volume of top agar on a double layer agar plate.

4.2.3 Hypothesis 2

The high oxygenation of plates compared to that of static liquid changes the physiological/proteome of the host. Consequently, highly aerated shaken cultures of host should be susceptible to phage infection.

4.2.3.1 Shaken batch cultures

Batch cultures of the relevant bacteria were grown in a shaking incubator and infected to a known MOI (5, 1, 0.1, 0.01 and 0/control). Growth was monitored through daily optical density readings. Cultures infected with RLP1 did not show a significant difference in growth whereas those infected with RPP1 did appear to grow at a reduced rate at higher MOI but no "crash" was observed. These results mirror those of a static culture infected with varying MOIs (see Fig. 4.2) which suggests oxygenation of the medium does not affect the phage's ability to absorb. It should be

noted that these experiments were done in capped tubes but with a large headspace. Consequently, shaking probably did not completely overcome low O_2 concentration.

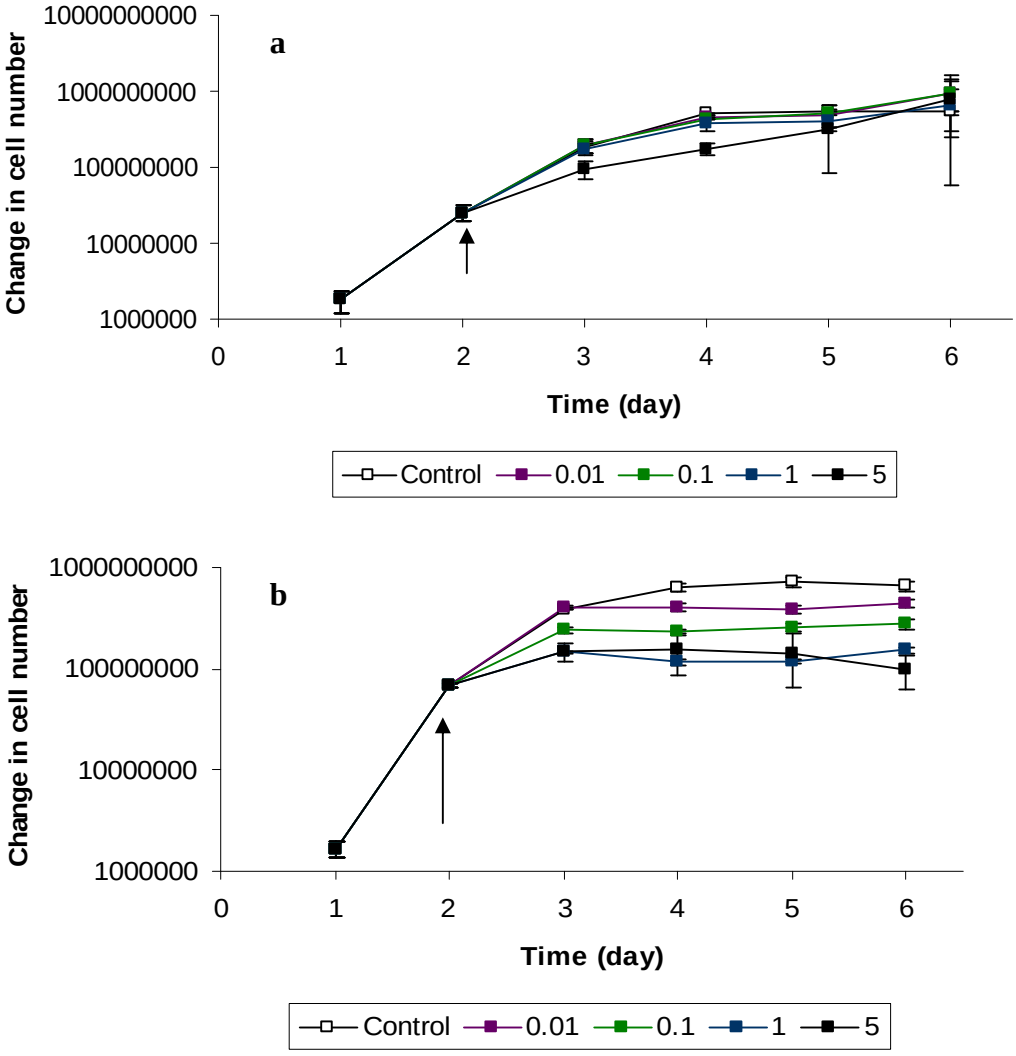


Figure 4.6 Change in a) *Rsv.* 217 and b) *Rsv. nubinhibens* cell number over 6 days when infected with RLP1 and RPP1, respectively, at a number of MOIs. Batch cultures were incubated at 25 °C in a shaking incubator. Arrow indicates point at which phage was added.

4.2.4 Hypothesis 3

The receptor to which RLP1 and RPP1 bind is differentially expressed during the growth cycle of their hosts. As such phage should be able to bind to bacteria in grown batch culture only during the appropriate stage of the growth curve.

4.2.4.1 Age of culture adsorption assay

A single batch culture of *Rsv.* 217 and *Rsv. nubinhibens* was grown in MAMS-PY; each day between 10^7 – 10^8 CFUs were removed and resuspended in 2.8 ml. Phage was added to a MOI of 0.01 and the number of free phage monitored over 10 minutes. The results (see Fig. 4.7) indicate that the age of the culture did not have an effect on adsorption of either phage. Instead, the % free phage remained fairly constant and so the phage receptor is not differentially expressed over time. (The growth of the host cells over time are shown in Fig. 4.8 .)

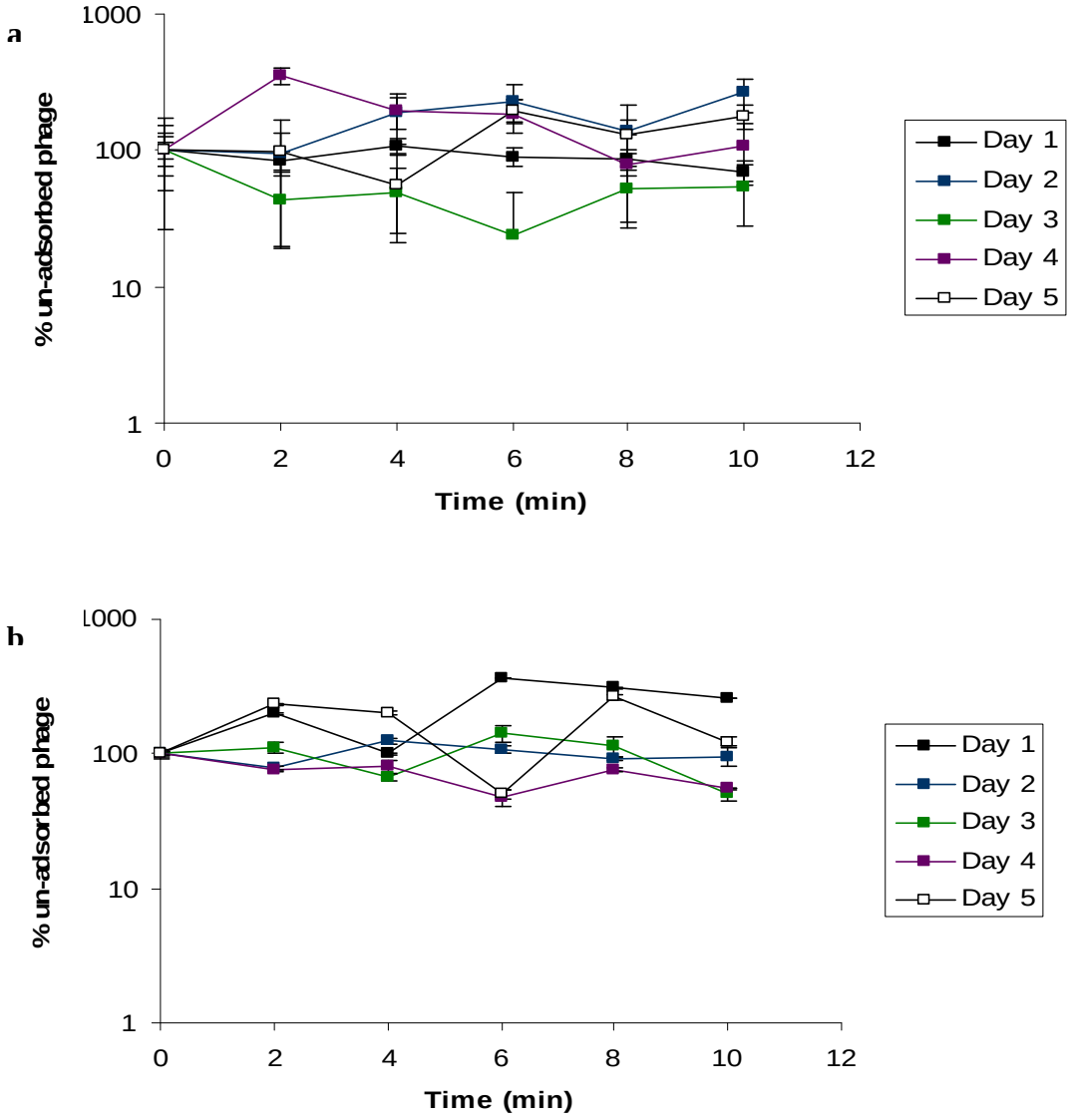


Figure 4.7 Adsorption assays using host cells of different age. a) RLP1 on *Rsv.* 217 and b) RPP1 on *Rsv. nubinhibens*. Cells from a single batch culture were removed daily and used in a 10 min adsorption assay to test if the age of the cell affects phage binding ability. Age of bacterial cells are indicated by the key.

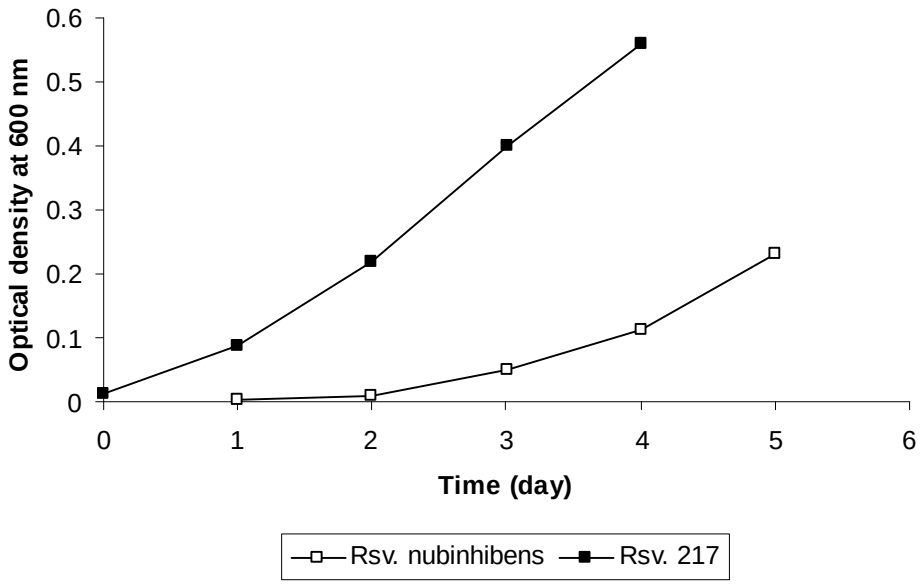


Figure 4.8 Growth of host bacterial cells over the period of 5 days. Cells were removed from the batch culture each day and used in a 10 minute adsorption assay.

4.2.5 Hypothesis 4

Bacto Agar (Difco) contains a contaminant that acts as a co-factor to allow the positive binding of RLP1 and RPP1 to their hosts. Consequently, marine broth plates made using purified agar (see Section 2.6.1.3) and agarose (a component of agar) respectively as a setting agent and used in spot assays should not display single plaques.

4.2.5.1 Types of Marine agar plates

Marine agar plates with 1.5 % (w/v) bottom agar/agarose and 0.4 % (w/v) top agar/agarose plates were made and 10 µl of phage with decreasing titres were spotted onto lawns of susceptible host. As single plaques formed on all three types of plate, it appears unlikely a co-factor only found in Bacto Agar is responsible for positive infection on plates.

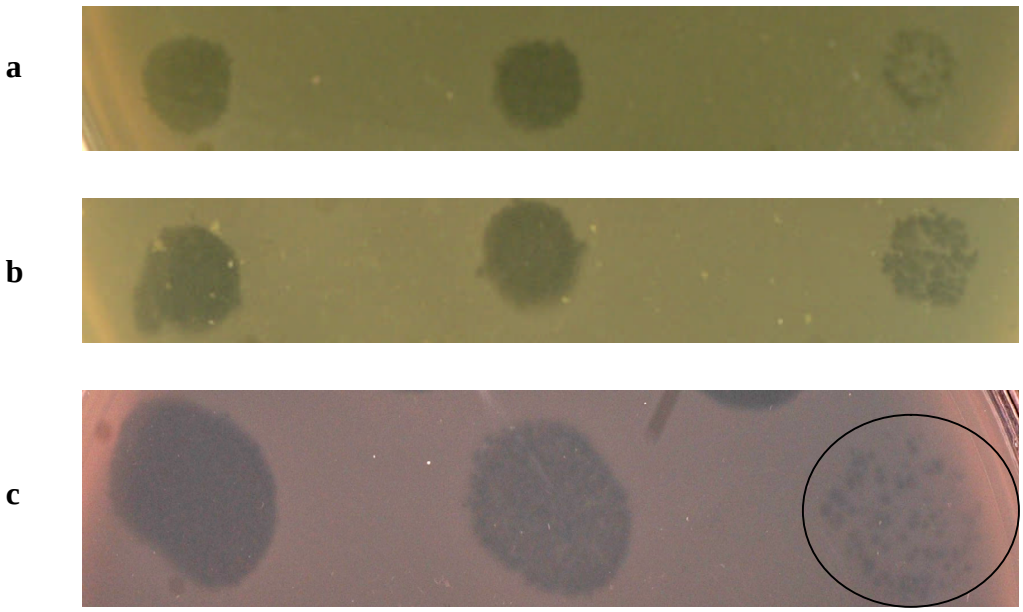


Figure 4.9 Spot test plates of RLP1 on *Rsv.* 217 with a) Bacto agar, b) purified agar and c) agarose as a setting agent. Spots of serially decreasing titre (10^8 - 10^6 PFU/ml) from left right.

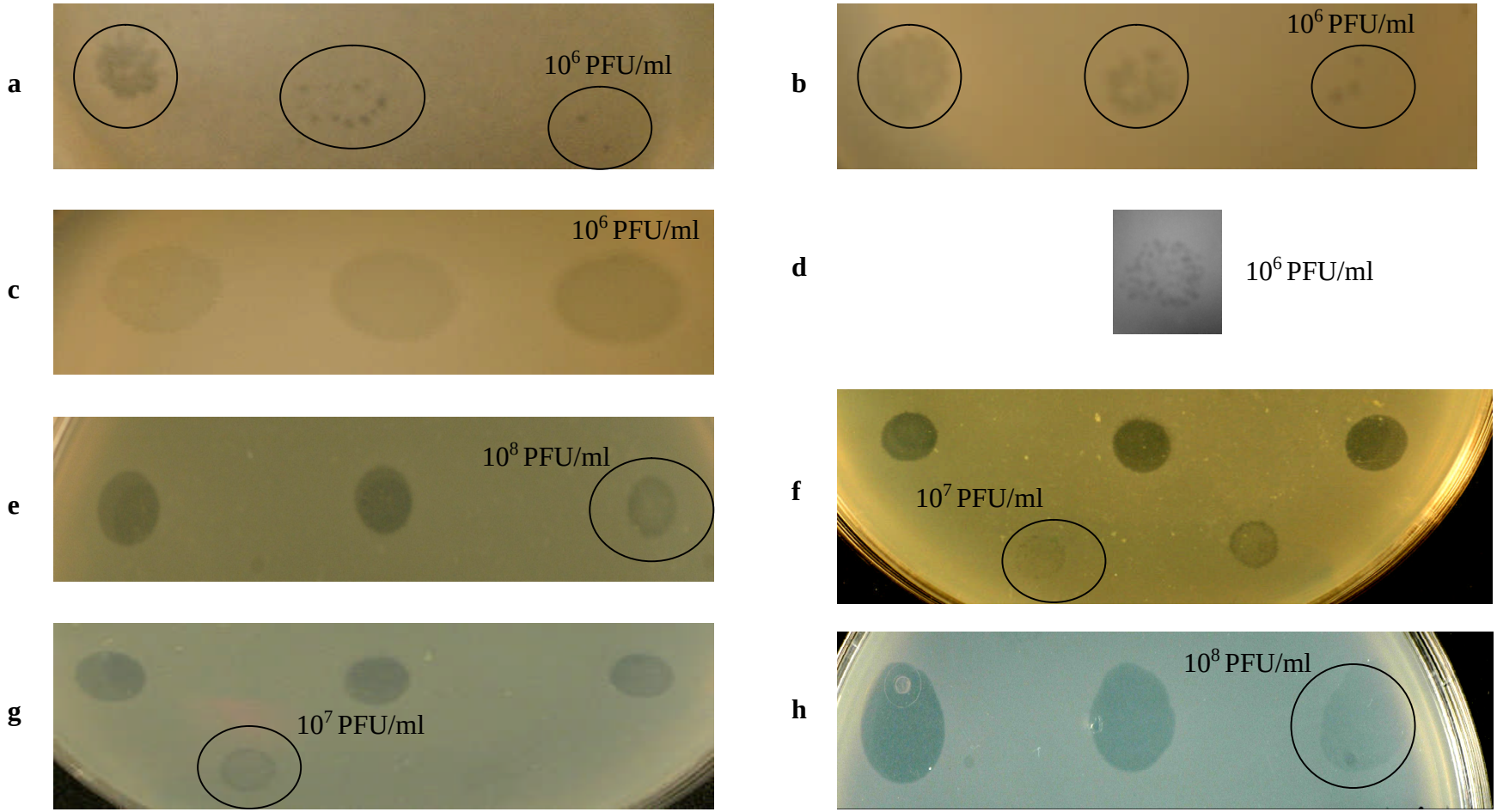


Figure 4.10 Spot test plates of RPP1. 10 µl of serially decreasing phage lysate was spotted onto a lawn of bacterial cells made with a variety of setting agents. a) – d) RPP1 on *Rsv*. 217 e) – h) RPP1 on *Rsv. nubinhibens*. Plates a) & e) Bacto agar, b) & f) purified agar and c), d), g) & h) agarose. Circles highlight spotting location.

4.2.6 Hypothesis 5

RLP1 and RPP1 are only able to bind to their respective hosts when the latter is embedded in a semi-solid matrix e.g. low percentage agar/sloppy agar.

4.2.6.1 Liquid adsorption assay of plate grown cells

Around 5 x 10^9 cells grown in liquid culture were harvested, used to create a lawn of host cells and grown for 24 hours at 20 °C. Cells were then harvested using a modified version of the protocol described in Section 2.9.6.2. The modification was to use cold ASW as a buffer and the harvested cells were subsequently used in a standard liquid adsorption assay.

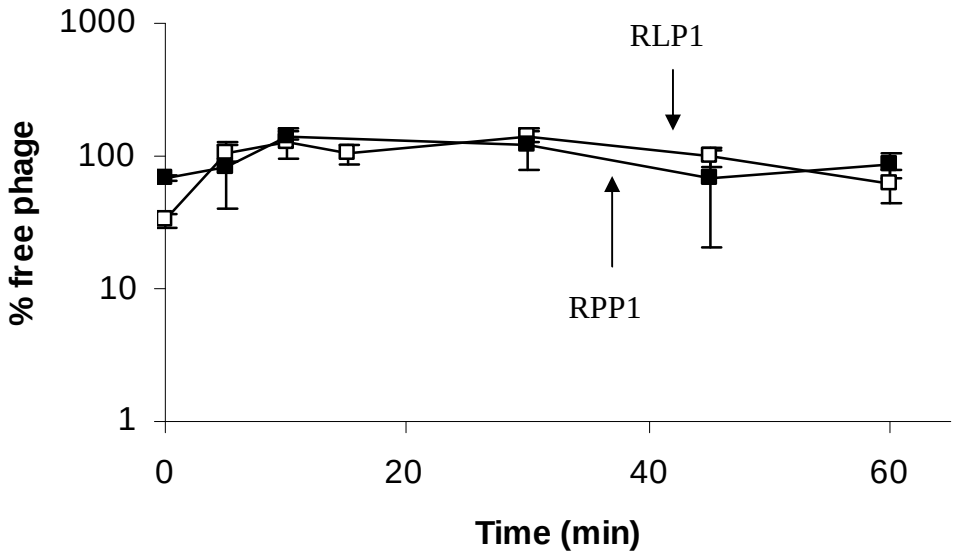


Figure 4.11 Adsorption assay of plate grown cells.

There was no appreciable decrease in free phage during the assay suggesting plate grown hosts are not different to liquid grown host cells. However, there was a slight lag (*ca.*30 minutes) between harvesting and start of the assay during which the cells could change cell expression. It also should be noted that the errors in this assay were considerable which must be taken into account. Nevertheless, its does appear that the plate grown cells did not bind to the phage.

4.2.6.2 Plate adsorption assay

In order to maintain host cells in their plate-grown physiological state during an adsorption assay, the assay was performed on plates. The method for this is described in detail in Section 2.7.14.2. These results indicate there is a sustained reduction in % free phage suggesting that there is indeed a difference caused by plating.

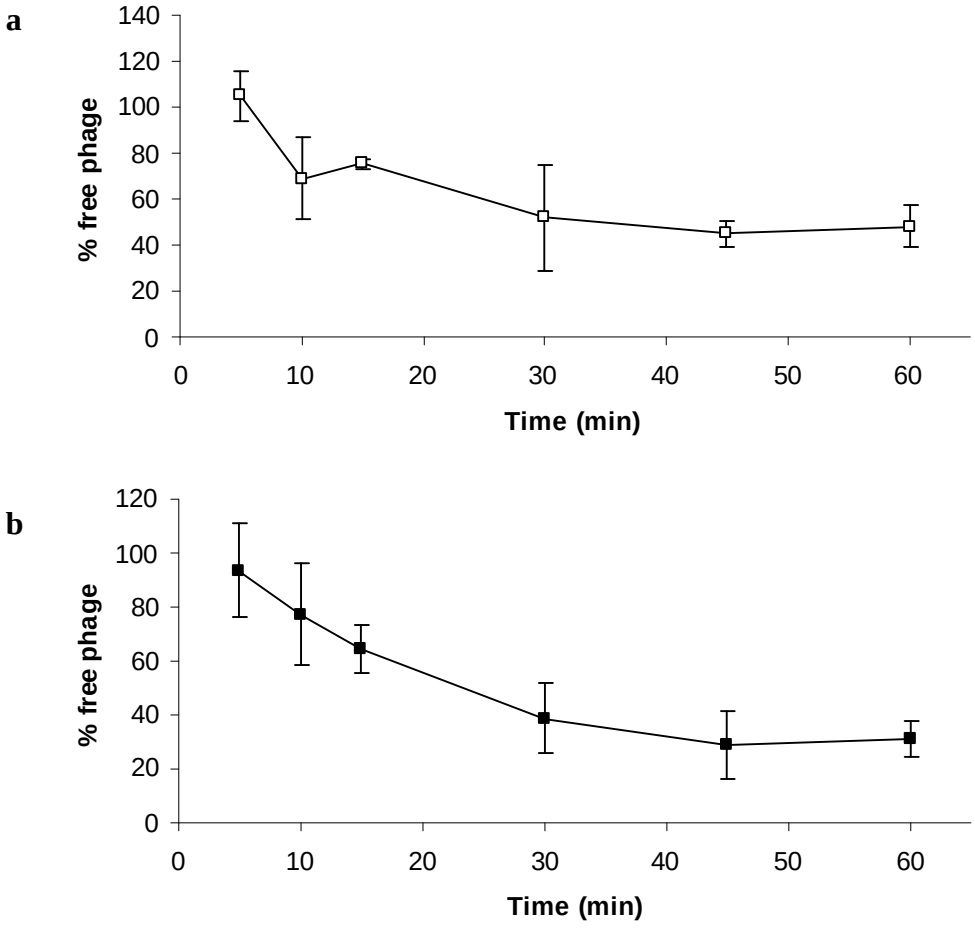


Figure 4.12 Adsorption of a) RLP1 with *Rsv.* 217, b) RPP1 with *Rsv. nubinhibens*. The assay was carried out on a plate instead of in liquid.

To further validate these results a modified version of an adsorption assay was carried out. In this modified version a single culture of host cells were grown overnight at 20°C and split into two aliquots. One was used to perform a liquid adsorption assay and the second, a plate binding assay. The results are shown in Fig. 4.13. Comparison of the two assays clearly shows a difference between assays performed in liquid and those performed on a plate.

Unfortunately, due to the large variation between biological replicates, an accurate adsorption rate cannot be calculated. It is, however, clear from the data that there is a change in adsorption when carried out on a plate which suggests a change in the host's proteome when in liquid versus when plated.

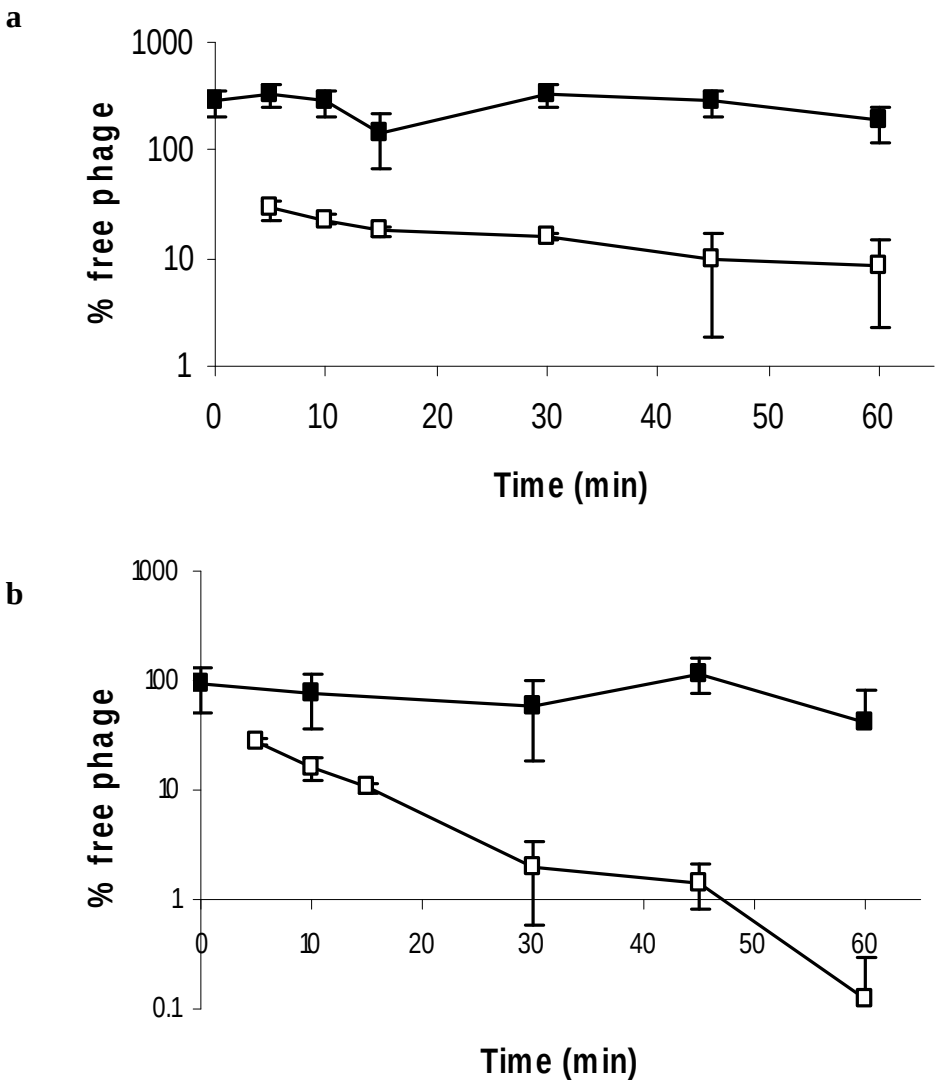


Figure 4.13 Adsorption assay of) RLP1 with *Rsv.* 217, b) RPP1 with *Rsv. nubinhibens* carried out both in liquid (filled squares) and on solid (open squares) media. A single culture of host bacterial cells was divided into two aliquots and used in a liquid and solid binding assay to compare the adsorption of phage to susceptible bacteria in different media.

4.2.7 Surface receptor comparison

In order to substantiate the preliminary observations that bacterial cells, once plated, change their surface profile, a comparison of the OM proteins and lipopolysaccharides (LPS), two common types of phage receptors, expressed in plate- and liquid-grown cells was performed.

4.2.7.1 Outer membrane protein (OMP) enrichment

The outer membrane proteins from liquid- and plate-grown cells were enriched for and analysed by SDS-PAGE (Fig. 4.14). One band consistently appeared in the samples from plate-grown *Rsv.* 217 cells but was absent from liquid extractions. As two methods for OMP enrichment were attempted and the band appeared in both, error due to sample preparation can be dismissed. The change in *Rsv.* 217 OMP profile supports the hypothesis that the proteome of the cell changes when embedded into agar. In addition the protein(s) present in the "plate-grown" band are promising candidates for phage surface receptor.

The differentially expressed bands are highlighted in blue in Fig. 4.13. These bands were excised and analysed by the in house mass spectrometry and proteomics facility. Unfortunately, the protein sample was too low in concentration and no definitive results were returned.

No extra bands were found in the samples from plate-grown *Rsv. nubinhibens* which suggests that the phage receptor is not a protein.

4.2.7.2 LPS

As some phage receptors have been identified as LPS, this was also extracted from liquid- and plate-grown cells and analysed by modified SDS-PAGE (see Section 2.9.8).

Prior to extraction, standardisation of approximate cell number was carried out to ensure roughly equal amounts of LPS were purified. From the profile seen in Fig. 4.15 there does appear to be some differences (highlighted in blue) in the *Rsv. nubinhibens* LPS profile with a few bands present in the plates lanes but absent in the liquid lanes. However, as the amounts loaded onto the gel were not known, overloading and thereby poor resolution of LPS bands was observed so they may indeed be present in both profiles. The initial evidence does suggest a LPS is a phage receptor in *Rsv. nubinhibens*, but this requires further work.

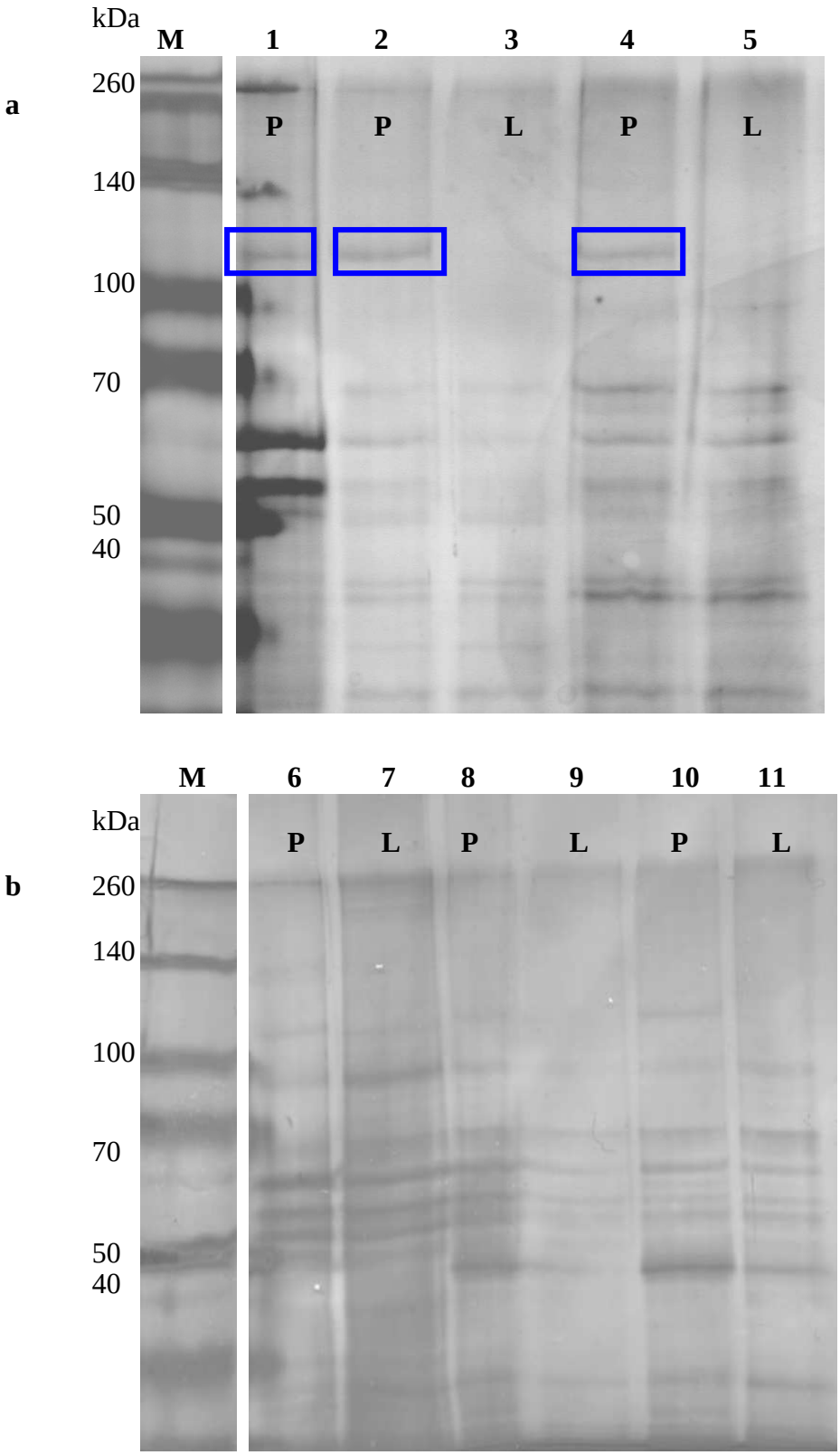


Figure 4.14 SDS-PAGE analysis of a) *Rsv.* 217 and b) *Rsv. nubinhibens* outer membrane proteins grown on plate (P) and in liquid (L). Lanes 1, 6 and 7 were processed by Freeze/thaw, lanes 2 - 5 and 8 – 11 by French cell press.

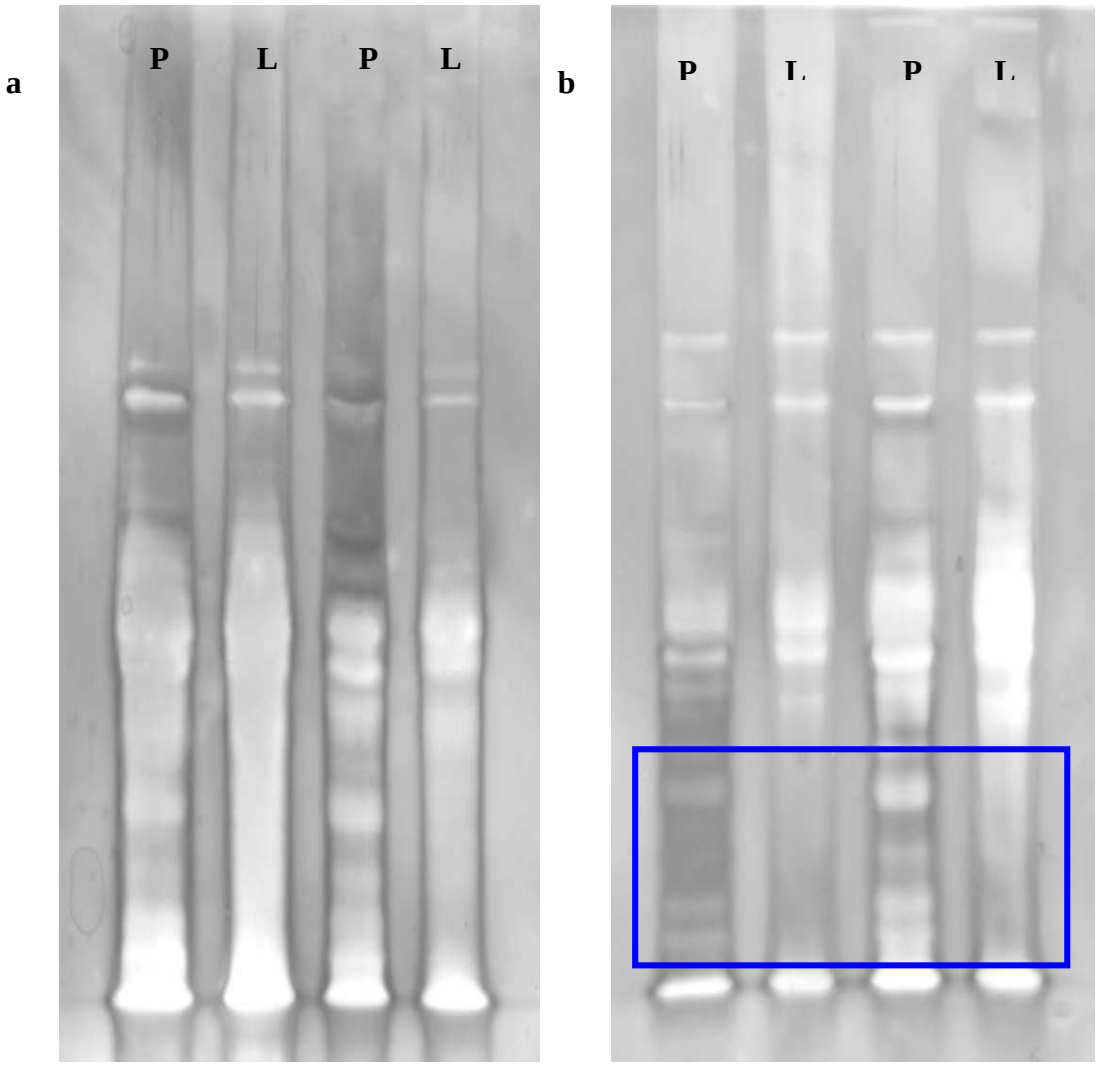


Figure 4.15 Modified SDS-PAGE analysis of a) *Rsv.* 217 and b) *Rsv. nubinhibens* LPS grown on plate (P) and in liquid (L). The area highlighted in blue shows the change in pattern between LPS found on plate and liquid grown *Rsv. nubinhibens* cells.

4.3 Concluding comments

Bacteriophages have proven to be ingenious predators able to recognise and bind to almost every exposed structure, either integral or extended, on a bacterial cell surface. Consequently, it is perhaps not surprising to find two phages that preferentially bind to host cells under certain conditions. The underlying cause responsible for the change in host surface profile however, remains unresolved as there are many plausible explanations. One appealing interpretation is that fixation in agar, a substance derived from seaweed, mimics an environmental condition that *Roseobacter nubinhibens* and *Rsv.* 217 encounter such as a biofilm on the surface of a macroalgae. Roseobacter sp. have been isolated from marine snow (Gram *et al.*, 2002) and biofilms (Rao *et al.*, 2006) and the tendency of many isolates to form associations on living or non-living surfaces is discussed in Section 1.1.5.

It has long been observed that some bacteria have a physiological response when in association with a surface, one of the best documented of these is in marine *Vibrio* spp. such as *Vibrio alginolyticus* (De Boer *et al.*, 1975; Golten and Scheffers, 1975). When grown in liquid cultures these bacterial have a single, sheathed, polar flagellum but when plated on an agar surface they produce multiple, lateral flagella. Furthermore, in a study of B*acillus subtilis* it was found that *ca.* 6% of genes were differentially expressed in biofilms and some of these genes had phage-related functions (McLean *et al.*, 2001).

A possible molecular mechanism for change in physiology is quorum sensing and production of acylated homoserine lactone (AHL); this signalling molecule of quorum sensing has been detected in many Roseobacter spp. (Wagner-Döbler and Biebl, 2006). In the investigation by Bruhn *el al.* (2007) the production of AHLs of different Roseobacter species under shaken and static conditions was examined (see Fig 4.15). Their results show that *Rsv. nubinhibens* (the host for RPP1) only produces AHLs when in a static culture. Though this does not exactly reflect the conditions of liquid vs. plate, the paper does illustrate the link between gene expression and culture conditions in some Roseobacter species. In addition to AHL production, the ethyl acetate extracts of shaken *Rsv. nubinhibens* cultures were also shown to contain antibiotics which could inhibit growth of *Vibrio anguillarum*.

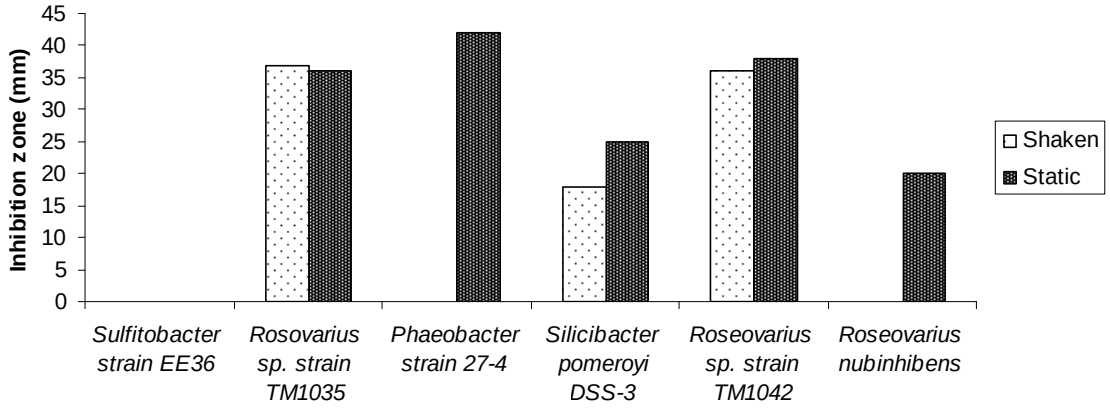


Figure 4.16 Comparison of AHL's dectected from shaken and static grown cultures of Roseobacter, taken from Bruhn *et al.*, 2007.

As discussed in Section 1.1.5, many Roseobacter species have a biphasic "swim-or-stick" lifestyle (Geng and Belas, 2010). As such, it seems likely that the change in physiology observed in *Rsv.* 217 and *Rsv. nubinhibens* is the alternation between the swimming, planktonic form to the stuck, sessile state. However, both these species were isolated as planktonic organisms and they have not been identified in any sessile/biofilm related studies. Indeed in the Bruhn *et al.*, (2007) study, apart from the change in AHL production, *Rsv. nubinhibens* did not shown any changes in attachment or cell morphology when grown under shaken conditions. As such, the functional role of such a lifestyle for these two *Roseovarius* spp. remains unclear. It is however, tempting to hypothesize that through a mechanism such as quorum sensing, a molecular switch is flipped upon plating and the sessile physiology is induced. In this state, a new protein and/or LPS is expressed on the cell surface and the phages RLP1 and RPP1 use these as an receptors from which to initiate infection. The change in gene expression appears to be rapidly implemented as there is no obvious lag in phage binding as observed in Fig. 4.11. This topic is explored further in Sections 7.15 and 7.1.6.

Regardless of the explanation and mechanism behind the difference in host expression, ultimately from a phage's perspective, it is advantageous to infect a host when other susceptible cells are in close proximity as they would be in a biofilm or a surface. This is particularly significant in the marine environment where planktonic bacteria exist in relatively low concentrations thus making a biofilm/surface infection-only strategy highly profitable.

Chapter 5

Genome characterisation and analysis of *Roseovarius* phage RLP1 and RPP1

5.1 Introduction

Since the first marine bacteriophage genome, *Pseudoaltermonas espejinana* BAL-31 ΦPM2, was completely sequenced in the 1968 it has become increasingly clear that phages in the oceans harbour vast amounts of genetic diversity. Recent studies have suggested that there are around 10^6 viral genotypes per kg of marine sediment and more than 10^{30} unique viral genotypes present in marine viriome (Breitbart and Rohwer, 2005; Kristensen *et al.*, 2010). However, as less than 5% of the phage genomes currently available are of marine origin it is, therefore, not surprising that the majority of marine phage diversity remains uncharacterised. With the rapid advances in sequencing technology, sequencing has become *de rigueur* for new phages and is a powerful and efficient method to gain insight into the workings of an organism.

The sequencing of RLP1 and RPP1 described in this chapter has identified them as putative members of the recently recognised genus of "N4–like viruses" (Hendrix and Casjens, 2006). For over 40 years Enterobacteria phage N4 was a genetic orphan and was unique in its use of three distinct RNA polymerases (RNAP) to control transcription. N4 early genes are transcribed by a virion-encapsulated RNAP (described in more detail in Section 5.2.6), middle genes by a phage-encoded heterodimeric T7-like RNAP in conjunction with the phage gp2 product (thought to be a ssDNA-binding protein) and finally late gene transcription by the host σ^{70} together with (the middle gene) activator gp45.

Early genes have an eleven nucleotide (nt) long promoter sequence, WT-P1, WT-P2, and WT-P, which forms a DNA-hairpin that is recognized by the virion RNAP (vRNAP). The hair pin consists of a 5-7 nt stem and a 3 basepair loop and two main determinants control vRNAP recognition, a purine at the centre of the hairpin loop (−11G preferred), and a specific interaction in the major groove (−8G) of the hairpin stem. Subtle differences in sequence and hairpin stem-loop length affect the strength of binding to vRNAP, for example, WT-P1 has a dissociation constant of 40 nM whilst WT-P2 K_d = 2 nM (Gleghorn *et al.*, 2008; Davydova *et al.*, 2007). X-ray crystallography of the vRNAP domain containing the polymerase functions has shown that the DNAP-hairpin is recognized by four structural motifs in the RNAP.

Interestingly, three of the four motifs can also be found in phage T4 RNAP which also recognizes dsDNA promoters (Gleghorn *et al.*, 2008).

Very recently, the N4-like genus has expanded to include other phages isolated from various environments suggesting this phage-type is more prevalent than previously thought. As well as Roseophages DSS3Φ2 and EE36Φ1 (see Section 1.9.3) Ceyssens (2009) isolated two N4-like *Pseudomonas* phages, LUZ7 and LIT1. Like N4, the early genes of the two *Pseudomonas* phages have a conserved 11 nt (5'-CAAACCATGAA-3') motif upstream of the transcription start site that may serve as an initiation site.

LUZ7 and LIT1 also share another similiarity to N4, all three phages have linear genomes with defined genomic ends. N4 has unique (non-permuted) direct terminal repeats of 390 to 440 base-pairs in length with 3' extensions. The left end has a relatively precise 5' terminus whilst the 3' extensions exhibit microheterogeneity with the predominant sequences being either 3'CATAA or 3'CATAAA. The right end is more variable with at least six discrete ends which differ by *ca.* 10 bp resulting in terminal repeats of differing length. As with the left end, the 3' extension (of the right genome end) displays microheterogeneity in length. In contrast, LUZ7 and LIT1 have terminal redundancy of 660 and 655 bp respectively, at both ends (Ceyssens, 2009).

Other phages, such as the T-even phages, have circularly permuted genomes; this means within a clonal phage sample all members have linear genomes whose sequences are circular permutations of each other. For example, if a genetic sequence is represented as ABCDEFG, then circular permutation would generate molecules ABCDEFG, BCDEFGA, CDEFGAB, DEFGABC and so on. Some circularly permutated phage genomes, such as T2 and T4, also have terminal redundancy which means the ends of the genome are duplicated i.e. ABCDEFGA. During DNA replication, a long concatemer (long, continual DNA molecule that contains multiple copies of the same DNA sequence linked in series) is generated by recombination of the end repeats; during packaging, this molecule is enzymatically cleaved into "headful" packages which are <100% of the minimum length of the phage genome (*ca.* 102% in T4) which accounts for the duplication at the ends (Hartl and Jones, 2009).

In this chapter the assembly, structure, annotation, comparison of RLP1 and RPP1 with other N4 –like phages and the implications of these results for the N4-like genus and other marine podoviruses will be discussed.

5.2 Results and Discussion

5.2.1 Genome assembly

RLP1 and RPP1 genomes were sequenced by Solexa and 454 pyrosequencing by The GenePool, Edinburgh, as described in Section 2.8.10. However, both required additional experimentation to complete their assembly.

5.2.1.1 RPP1

RPP1 was assembled into one contig of 74.7 kb with 756 fold coverage. However, upon initial annotation, the ends of the contig appeared to straddle a single predicted gene which bore striking similarity to bacteriophage N4's virion RNA polymerase (vRNAP). Consequently, it was believed that RPP1 was circularly permuted. To test this theory, primers were designed to amplify the region spanning the two ends, firstly to confirm that they were linked and secondly to elucidate any missing sequence. Additional control primers were also designed to test the suitability of the gap-spanning primers (see Fig 5.1). If the genome was circular or circularly permuted, then a PCR with primers vRNAP1 and vRNAP2 would produce a product *ca* 1.3 kb. The internal control primers would produce products of 640 and 570 bp for reactions with vRNAP1 & vRNAP3 and vRNAP2 & vRNAP4 respectively.

Presence of bands of the correct size in all three PCRs indicated that RPP1 was likely to be circular or circularly permuted. Sequencing of the 1.3 kb fragment and analysis by SeqMan indicated that there was a 9 bp overlap of the ends of the assembled contig.

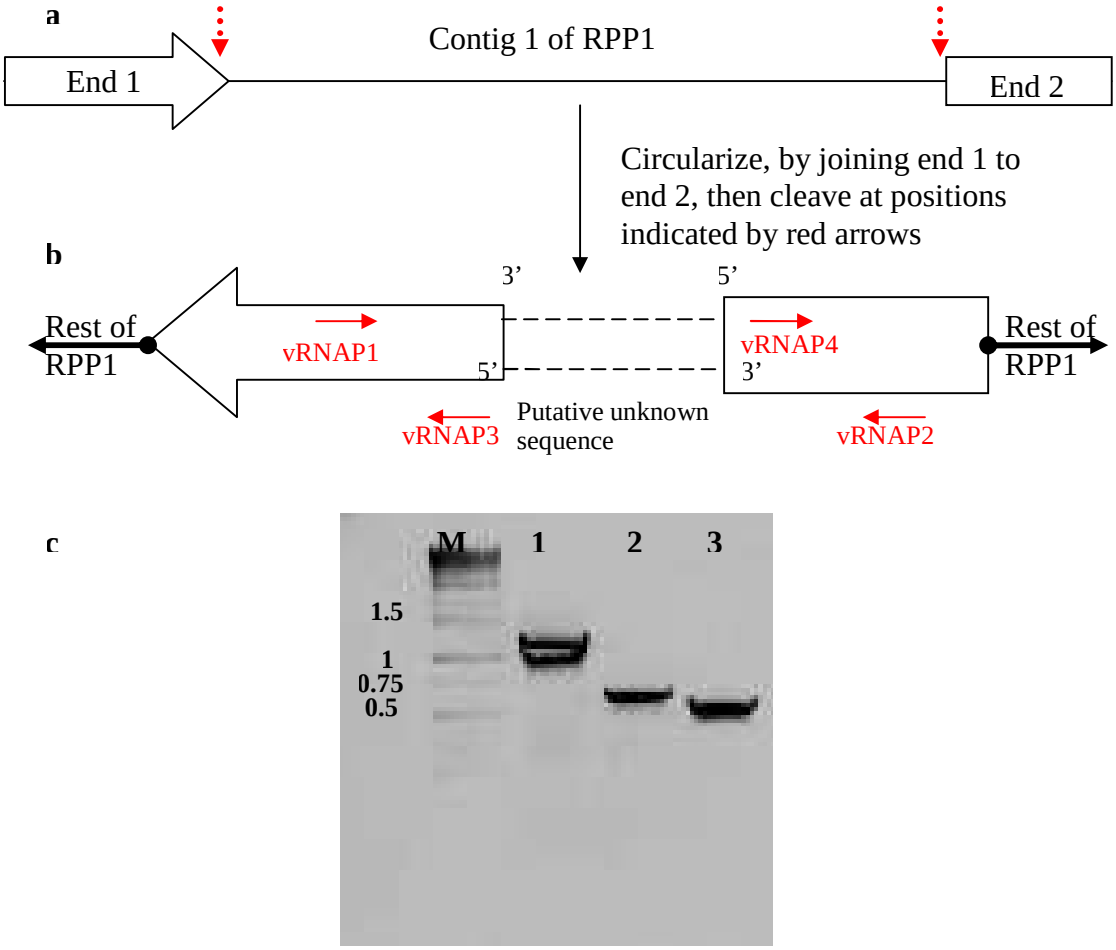


Figure 5.1 Assembly of RPP1 contig a) Contig 1 of RLP1 with two ends b) schematic showing the end sequences of RPP1 connected by possible unknown sequence with location of the primers (not to scale), c) PCR confirmation of the genome structure of phage RPP1. Lane 1 – 1.3 kb product from primers vRNAP1 and vRNAP2, 2 – 0.64 bp product from primers vRNAP1 and vRNAP3, 3 – 0.57 bp product from primers vRNAP2 and vRNAP4.

5.2.1.2 RLP1

RLP1 assembled into 10 contigs (see Table 5.1); initial annotation of the largest contig suggested a high degree of gene synteny between RLP1 and RPP1. Consequently, RPP1 was used as a scaffold for RLP1 (see Fig. 5.2) and the order of contigs was confirmed by PCR (see Fig. 5.3). Sequencing of the PCR products resulted in complete assembly of RLP1 (see Section 2.8.11).

Table 5.1 RLP1 contigs assembled from pyrosequencing.

Contig	Size (kb)
1	35.4
2	9.7
3	6.1
4	5.5
5	3.8
6	3.1
7	2.5
8	1.7
9	1.6
10	1.1
Total	70.5

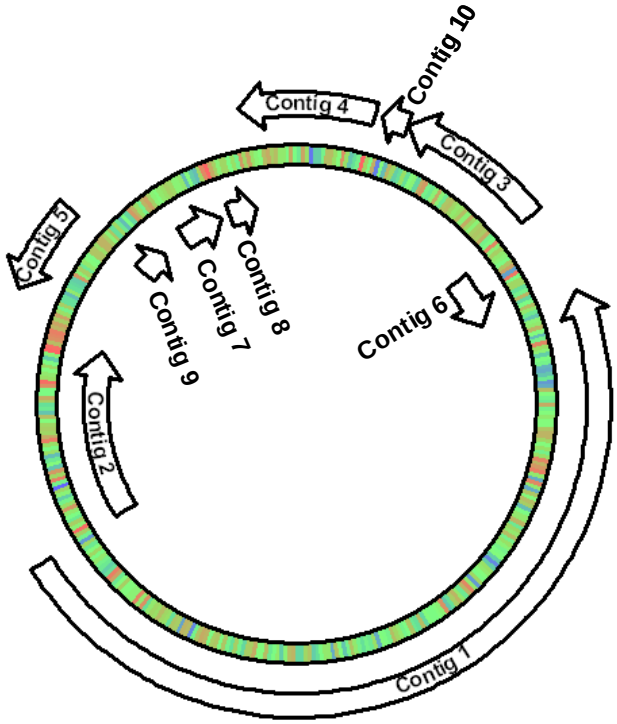


Figure 5.2 Contigs of RLP1 mapped onto a RPP1 scaffold.

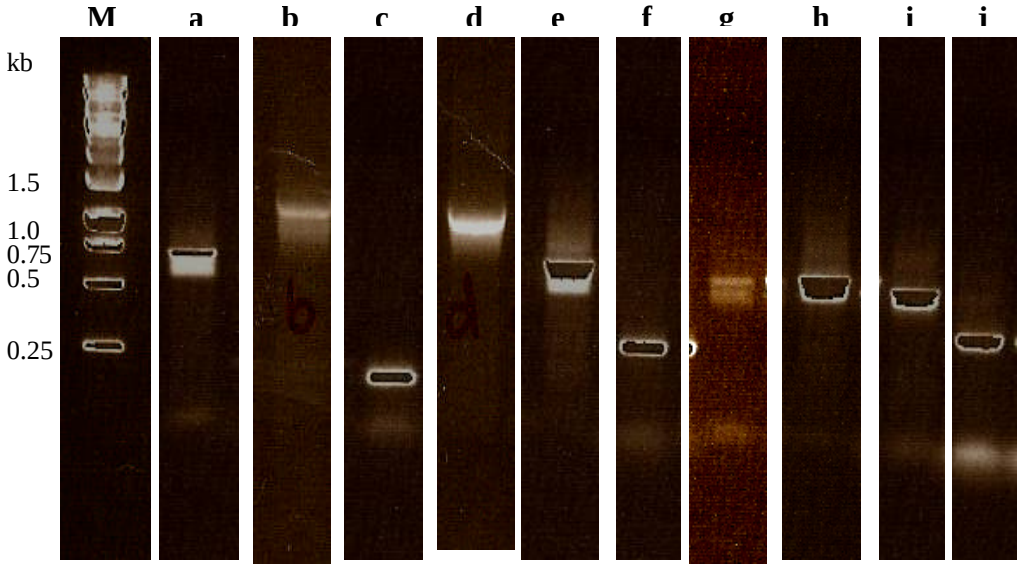


Figure 5.3 PCR to confirm the predicted order of contigs. Primers were designed to amplify the region between contigs to confirm the predicted order based on the RPP1 scaffold. Reaction a – amplification of region between contig 1 and 2, b – contigs 2 and 5, c- contigs 5 and 9, d – contigs 9 and 7, e – contigs 7 and 8, f – contigs 8 and 4, g – contigs 4 and 10, h – contigs 10 and 3, i – contigs 3 and 6, and j – contigs 6 and 1.

5.2.2 General genome properties

Genome sizes as determined by sequencing were 74704 bp and 74583 bp for RPP1 and RLP1, respectively. Both phages have a GC content of 49% in contrast to their hosts, which have a GC content of 60% and 63% for *Rsv. 217* and *Rsv. nubinhibens*, respectively. Initial sequencing assembled both phages into circular contigs which suggested they were either circular or (more likely) circularly permuted.

5.2.3 Restriction digests

Using the sequence data, restriction enzymes that would cut the phage genomes at 1-3 restriction sites (RS) were identified and used to perform digests of phage genomic DNA (Table 5.2). However, experimentally obtained digest profiles did not completely match the *in silico* predicted fragments. *Crf*101 and *Aas*I which were predicted to cut only once generated a pattern nearly identical to uncut DNA, though RLP1 digests did exhibit some smaller faint bands. In contrast, digests with *Sex*AI and *Aan*I did appear to match the predicted number of fragments for a linear or circularly permuted molecule.

Table 5.2 *in silico* predicted restriction enzyme sites.

Restriction enzyme	RLP1			RPP1		
	# RS sites	Predicted no. of fragments		# RS sites	Predicted no. of fragments	
		Circular	Linear/Circularly permuted		Circular	Linear/Circularly permuted
*Cfr*101/*Bsr*FI	1	1	2	1	1	2
*Aas*I/*Drd*I	1	1	2	1	1	2
*Sex*AI	3	3	4	2	2	3
*Aan*I/*Psi*I	3	3	4	2	2	3

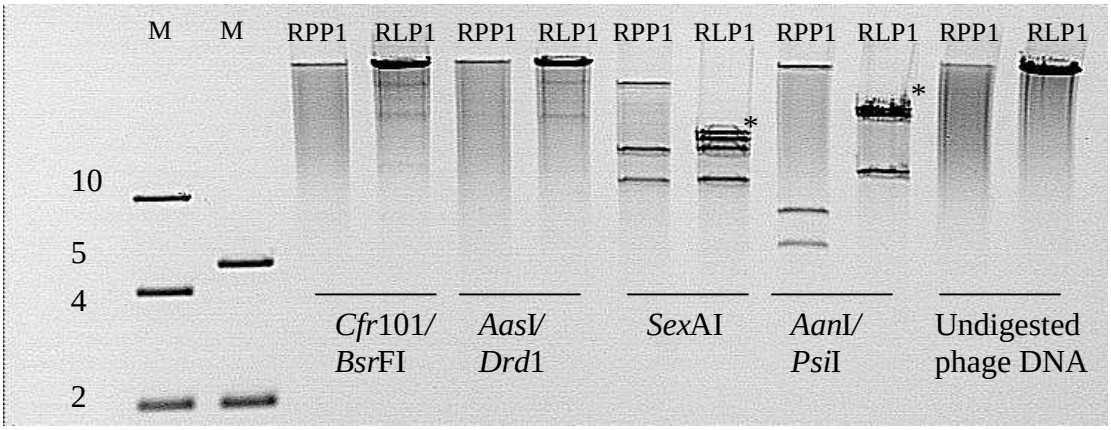


Figure 5.4 Restriction digest of purified RLP1 and RPP1 genomic DNA with rare cutters. PFGE was set to resolve between 800 – 1 kb for 4 hours at 14 °C, initial switch 0.5 sec, final switch 1.7 sec, 120°, 8 V/cm. * indicates multiple fragments. M, DNA marker (kb).

5.2.4 *Bal31* digestion

To further examine the structure of the two phages an exonuclease which only degrades the termini of dsDNA, *Bal31* was applied (see Section 2.8.5 for full protocol). The presence of two (one for each end) progressively shortening bands is indicative of a linear genome with defined ends. In contrast, circularly permuted genomes, after treatment with *Bal31* and subsequent restriction enzyme digestion, display an even, simultaneous degradation of all fragments.

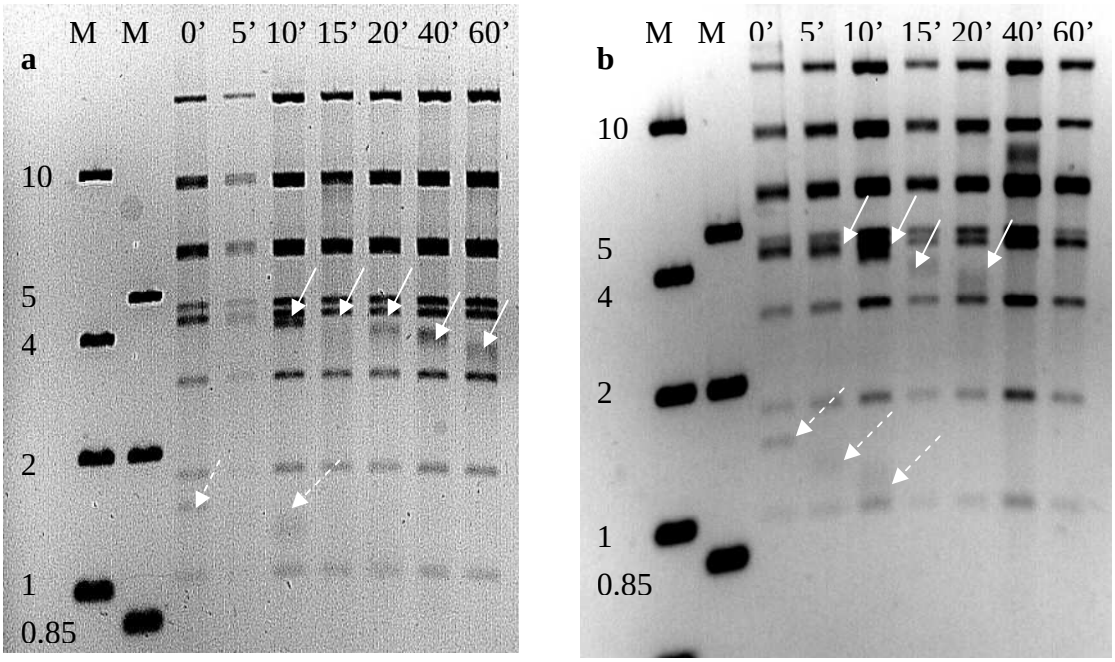


Figure 5.5 *Nde*1 digested a) RLP1 and b) RPP1 genomic DNA after treatment with *Bal31* for the indicated time intervals. Solid arrows indicate restriction fragment decreasing over time, dotted arrows indicate possible second disappearing restriction fragment. M, DNA marker (kb).

It can be concluded from the experimental data, shown in Fig. 5.5, that both RLP1 and RPP1 have linear dsDNA genomes with defined ends. These were likely missed in the assembly of the raw data from pyrosequencing due to terminally redundant ends. However, when both genomes are mapped as a linear molecule with a start point based on similarity to related phage genomes (see Fig. 5.7) the experimental digest profiles for *Psi*I still do not completely match the *in silico* predicted fragment sizes. *Psi*I digestion of RLP1 should produce four fragments: 29.7, 28.7, 13.3 and 2.9 kb whilst digestion of RPP1 should produce three fragments: 58.4, 10.6 and 5.6 kb. As shown in Fig. 5.6, the lane containing RPP1 digested DNA is missing the 10.6 kb fragment and has gained one which is >10 kb. Lane 2 containing RLP1 digested genomic DNA is missing the 5.6 kb fragment. Though the length of the unknown terminal repeats may be responsible, further examination is required before the physical structure of both phage genomes can be fully determined.

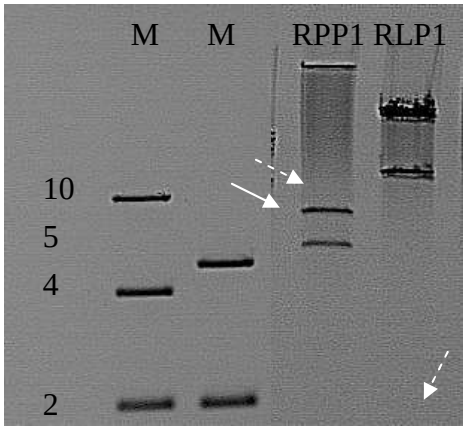


Figure 5.6 *Psi*I digest of phage genomic DNA. Solid arrow indicates additional band, dotted arrow indicates missing band. M, DNA marker (kb).

5.2.5 Identification of open reading frames

Gene prediction programs identified 92 and 91 putative open-reading-frames (ORFs) in RLP1 and RPP1 respectively. Most ORFs (in both phages) appear to initiate at an ATG codon though around 10 % use GTG or TTG as start codons. As genes with start codons of GTG and TTG are usually translated at a lower frequency, it is likely that these gene products will have a regulatory role.

All predicted ORFs were subject to BLASTp searches (see Section 2.10.2) for functional assignment. Many ORFs shared similarities with the genes from

Roseophage DSS3Φ2 and EE36Φ1 though the majority were shared between the two phages. Around 30% of the predicted ORFs were closely related to an enterobacteria phage, N4, whilst the majority of the remaining ORFs did not share any similarity with known genes in the nr database. The *in silico* analysis of RLP1 and RPP1 can be found in Table 5.3, the start point of both genomes was chosen by comparison with the related Roseobacter phages DSS3Φ2 and EE36Φ1 as well as phage N4. Comparison of the genomes of RLP1, RPP1, N4 and various N4-like phages are examined in detailed in Section 5.2.12.

Table 5.3 *In silico* analysis of the RLP1 and RPP1 genomes [a] Mutual identity expressed as % identity on the nucleotide level.

ORF RLP1	ORF RPP1	Mut. ident[a]	Mw (kDa)	Strand	Closest homologue	Comments
1	1	100.0	4.1	-	no similarity	
2	2	99.2	14.2	-	gp2 [N4/Roseophage]	
–	3		5.9	+	no similarity	
3	4	99.2	9.6	-	no similarity	
4	5	98.9	6.6	-	no similarity	
5	6	99.5	7.3	-	no similarity	
6	7	100.0	7.7	-	no similarity	
7	8	100	6.0	-	no similarity	
8	9	97.7	8.2	-	no similarity	
9	10	99.3	11.0	-	no similarity	
10	11	99.2	16.7	-	no similarity	
11	12	96.4	12.6	-	gp81 [DSS3Φ2], gp79 [EE36Φ1]	
12	–		8.1	-	no similarity	
–	13		10.9	-	no similarity	
13	14	98.9	36.0	-	gp79 [DSS3Φ2], gp78 [EE36Φ1]	Structural protein
14	–		14.7	+	no similarity	
15	15	98.1	22.3	-	gp78 [DSS3Φ2], gp76/N4 gp53-like [EE36Φ1]	Structural protein
16	16	100.0	10.4	-	gp77 [DSS3Φ2], gp75 [EE36Φ1], hyp RD1_B0001 [*Roseobacter denitrificans*]	Host-like protein/ Structural protein
17	17	99.9	25.7	+	gp69 [N4/Roseophage]	
18	18	99.4	61.6	+	gp68 [N4/Roseophage]	
19	19	99.0	25.2	+	gp67 [N4/Roseophage]	30 kDa structural protein
20	20	99.5	13.7	+	gp73 [DSS3Φ2], gp71 [EE36Φ1]	
21	21	97.5	21.5	+	gp72 [DSS3Φ2], gp70 [EE36Φ1]	
22	22	99.4	24.3	+	gp71 [DSS3Φ2], gp69 [EE36Φ1], hyp RCCS2_17771 [*Roseobacter* sp. CCS2]	Host-like protein
23	23	100.0	12.1	+	gp70 [DSS3Φ2], gp68 [EE36Φ1]	
24	24	100.0	6.1/5.1	+	gp67 [EE36Φ1]	
25	25	99.4	89.0	+	gp59 [N4/Roseophage]	94 kDa structural portal protein
26	26	99.2	14.2	+	gp68 [DSS3Φ2], gp65 [EE36Φ1], gp81 [LUZ7]	
27	27	99.2	48.6	+	gp57 [N4/Roseophage]	
28	28	99.9	51.4	+	gp56 [N4/Roseophage]	Major coat protein
29	29	99.3	27.3	+	gp55 [N4/Roseophage]	

Table 5.3 cont.

ORF RLP1	ORF RPP1	Mut. ident[a]	Mw (kDa)	Strand	Closest homologue	Comments
30	30	99.9	43.8	+	gp54 [N4/Roseophage]	Structural protein
31	31	99.7	98.5	+	gp53 [N4/Roesophage]	
32	32	99.8	16.1	+	gp52 [N4/Roseophage]	16.5 kDa structural protein
33	33	99.8	75.1	+	gp61 [DSS3Φ2], gp58 [EE36Φ1]	Structural protein, cell wall hydrolase domain
34	34	99.3	413.7	+	vRNAP [N4/Roseophage]	
35	–		6.4	-	no similarity	
36	35	100.0	7.2/5.2	-	gp59 [DSS3Φ2], gp56 [EE36Φ1]	
37	36	99.1	23.8/32.1	-	gp58 [DSS3Φ2], gp55 [EE36Φ1]	
38	37	99.6	29.8	-	gp45 [N4/Roseophage]	Single-stranded binding protein (SSB)
39	38	99.5	28.3	-	gp44 [N4/Roseophage]	
40	39	99.6	83.3	-	gp43 [N4/Roseophage]	
41	40	99.5	38.0	-	gp42 [N4/Roseophage]	
42	41	99.4	98.8	-	DNAP [N4/Roseophage]	
43	42	99.6	10.5	-	no similarity	
44	43	98.9	48.6	-	DNA helicase [N4/Roseophage]	
45	44	98.9	13.9	-	no similarity	
46	45	99.3	87.2	-	gp48 [DSS3Φ2], gp45 [EE36Φ1], Ribonucleoside diphosphate reductase [*Rsv*. sp HTCC2601]	Host-like protein
47	-		13.2	-	no similarity	
48	46	99.6	31.1	-	gp47 [DSS3Φ2], gp44 [EE36Φ1]	
–	47		10.4	-	no similarity	
49	48	99.7	21.3	-	gp46 [DSS3Φ2], gp43 [EE36Φ1]	
50	49	100.0	6.0	-	no similarity	
51	50	100.0	5.3	-	no similarity	
52	51	100.0	5.85/10.7	-	no similarity	
53	52	100.0	14.4	-	gp14 [N4/Rosoephage]	
54	53	99.6	18.7/16.9	-	gp44 [DSS3Φ2], gp41 [EE36Φ1]	
55	54	100.0	11.4	-	gp22 [N4/Roseophage]	HNH endonuclease
56	55	99.9	52.9	-	rIIB-like protein [N4/Roseophage]	
57	56	99.9	100.4	-	rIIA-like protein [N4/Roseophage]	
58	57	100.0	9.2/7.1	-	no similarity	
59	58	100.0	15.4	-	gp79 [*Rhizobium* phage 16-3]	

Table 5.3 cont.

ORF RLP1	ORF RPP1	Mut. ident[a]	Mw (kDa)	Strand	Closest homologue	Comments
60	59	100.0	12.1	-	Thioredoxin [DSS3Φ2, EE36Φ1, *Oceanicola batsensis, Silicibacter* sp. TM1040]	Host-like protein
61	60	100.0	21.2	-	gp36 [N4/Roseophage]	
62	61	100	7.7	-	gp38 [DSS3Φ2], gp35 [EE36Φ1]	
63	62	100.0	34.8	-	gp37 [DSS3Φ2]	
64	63	100.0	21.6	-	no similarity	
65	64	96.8	68.1	-	gp33 [DSS3Φ2], gp33 [EE36Φ1]	Similar to virion structural protein in *Pseudomonas* phage 201Φ2-1/Structural protein
66	65	96.7	12.4	-	gp32 [DSS3Φ2], gp30 [EE36Φ1]	Host-like/Structural protein
67	66	100.0	5.4/4.3	-	no similarity	
68	67	99.2	10.0	-	gp31 [DSS3Φ2], gp29 [EE36Φ1]	
69	68	99.1	44.4	-	gp30 [DSS3Φ2], gp28 [EE36Φ1]	Structural protein
70	69	96.5	35.6	-	gp30 [N4/Roseophage]	Thymidylate synthase complementing protein
71	70	99.1	15.7	-	gp28 [DSS3Φ2], gp26 [EE36Φ1]	
72	71	98.9	17.2	-	gp25 [DSS3Φ2], gp23 [EE36Φ1]	Structural protein
73	72	95.2	16.3	-	gp27 [DSS3Φ2], gp25 EE36Φ1]	Putative deoxycytidylate deaminase
74	73	96.9	10.3	-	gp23 [DSS3Φ2], gp21 [EE36Φ1]	
75	74	98.6	17.5	-	gp24 [DSS3Φ2], gp22 [EE36Φ1]	
76	75	100.0	47.0	-	gp25 [N4/Roseophage]	Von Willebrand factor type A (vWFA) domain
77	76	99.2	9.2	-	no similarity	
78	77	99.5	43.1	-	gp24 [N4/Roseophage]	ATPase superfamily
79	78	95.3	7.2	-	gp18 [DSS3Φ2], gp16 [EE36Φ1]	
80	79	98.5	7.1	-	no similarity	
81	80	100.0	45.7	-	gp15 [N4/Roseophage]	RNAP1
82	81	100.0	7.6	-	gp15 [DSS3Φ2], gp14 [EE36Φ1]	
83	82	100.0	11.9/9.3	-	no similarity	
84	83	98.8	5.9	-	no similarity	
85	84	100.0	9.4	-	gp12 [DSS3Φ2], gp11 [EE36Φ1]	

Table 5.3 cont.

ORF RLP1	ORF RPP1	Mut. ident[a]	Mw (kDa)	Strand	Closest homologue	Comments
86	85	96.8	6.8	-	no similarity	
87	86	96.7	4.7	-	no similarity	
88	87	100.0	11.9	-	gp10 [DSS3Φ2], gp9 [EE36Φ1]	
89	88	99.4	7.0	-	no similarity	
90	89	99.5	7.7	-	no similarity	
91	90	99.3	30.2	-	gp16[N4/Roseophage]	RNAP2
92	91	99.5	18.3	-	hypothetical protein DORFOR_01894 [*Dorea formicigenerans ATCC27755*]	

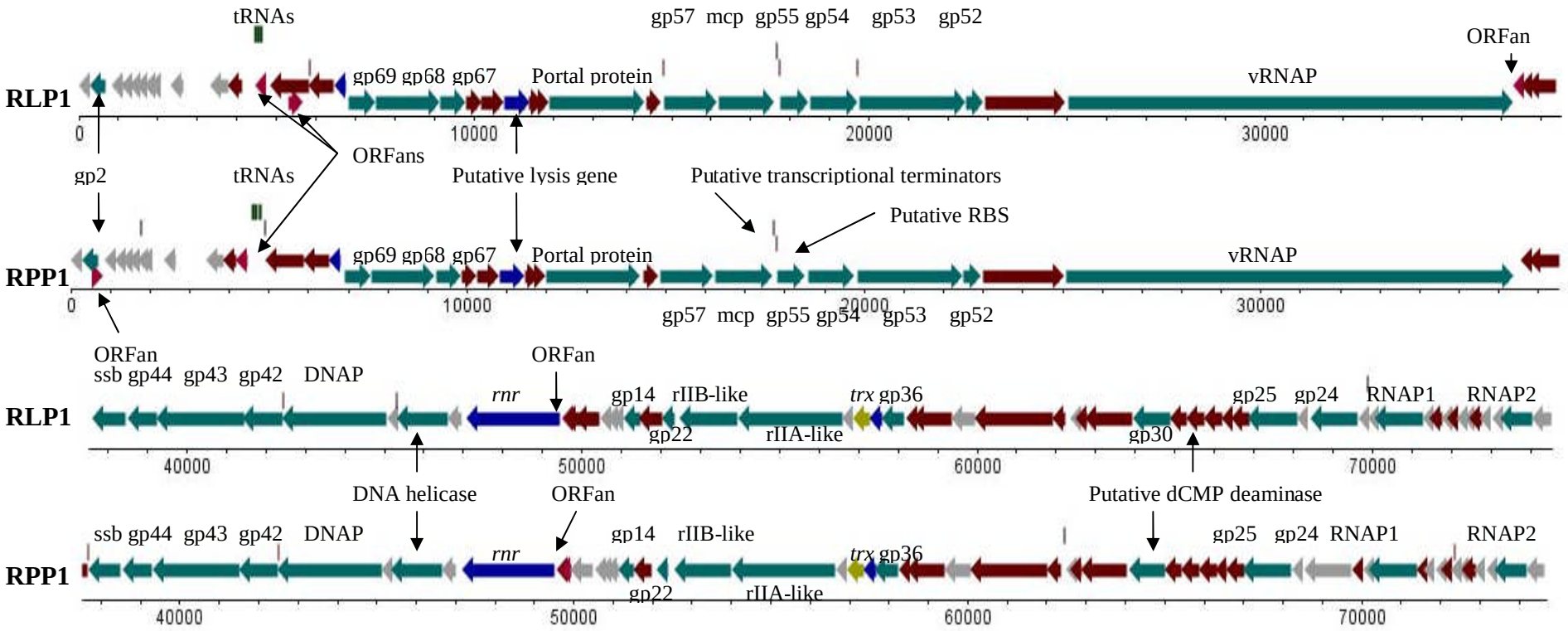


Figure 5.7 *In silico* analysis and comparison of the two *Roseovarius* phages RLP1 and RPP1. N4-like genes are highlighted turquoise, Roseophage: red, host-like: blue, ORFans: pink, other phage-like genes yellow; unknown: grey. Regulatory elements and tRNAs indicated by lines are found above the predicted ORFs.

5.2.6 vRNAP

A noteworthy gene that encompasses 15 % of the total phage genome in both phages is that of the virion RNA polymerase. In enterobacteria phage N4, the multifunctional vRNAP protein is encapsulated within the virion and is involved in DNA injection, early transcription and phage DNA replication (Kazmierczak and Rothman-Denes. 2005; Choi *et al.*, 2008). The N4 protein shares 47 % amino acid identity with the RLP1 and RPP1 homologues; all three do not contain cysteine residues a feature conserved amongst the N4-like phages. This characteristic suggests that at some point during infection the protein passes through its bacterial host's periplasm which is rich in enzymes that oxidise Cys residues into disulfide bonds (Ritz and Beckwith, 2001).

The N4 vRNAP middle domain contains four short motifs: TxxGR, A, B and C (these recognise the early promoter sequences) which are characteristic of the family of T7–like single-subunit RNA polymerases (Kazmierczak *et al.*, 2002). RLP1 and RPP1 contain all these motifs and the majority of the conserved catalytic residues apart from tyrosine in motif B which has been replaced by a valine residue (see Fig. 5.8).

Intriguingly, as well as the middle RNAP domain, both RLP1 and RPP1 vRNAP homologues contain the C-terminal domain which catalysis the enzymes' encapsulation (Gleghorn *et al.*, 2008). However, there is only weak similarity to the N-terminal domain responsible for the injection of the first 500 bp of phage genome into the host cell (Choi *et al.*, 2008). This suggests that the *Roseovarius* phage vRNAPs have a broadly similar function to their N4 vRNAP homologue in terms of polymerase activity and enzyme processing, but their involvement in DNA injection has been altered.

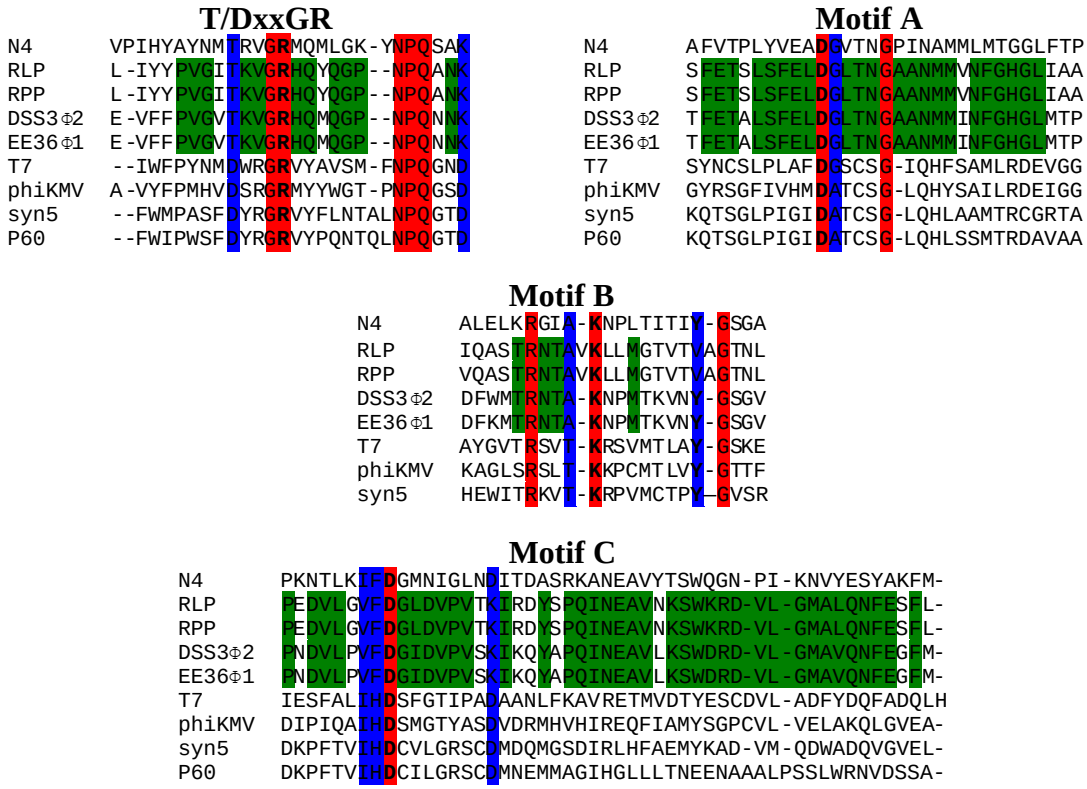


Figure 5.8 Amino acid sequence alignment of motifs T/DxxGR, A, B and C from RLP1, RPP1, N4 and other T7 superfamily RNA polymerases. Red residues are conserved, blue are 50% identical and green are shared between the Roseobacter phages.

5.2.7 Single-stranded DNA-binding protein

In phage N4 gp45 encodes a ssDNA-binding protein (SSB) that interacts with and activates the host σ^{70} transcription of late phage genes (Kazmierczak and Rothman-Denes, 2005). Site-directed mutagenesis studies of N4 SSB have identified five catalytic residues responsible for phage DNA replication & recombination as well as transcription of late genes. As with the vRNAP gene, alignments of the SSB gene suggest some alterations between N4 and the Roseophages, but significant conservation within the latter group, see Fig. 5.9.

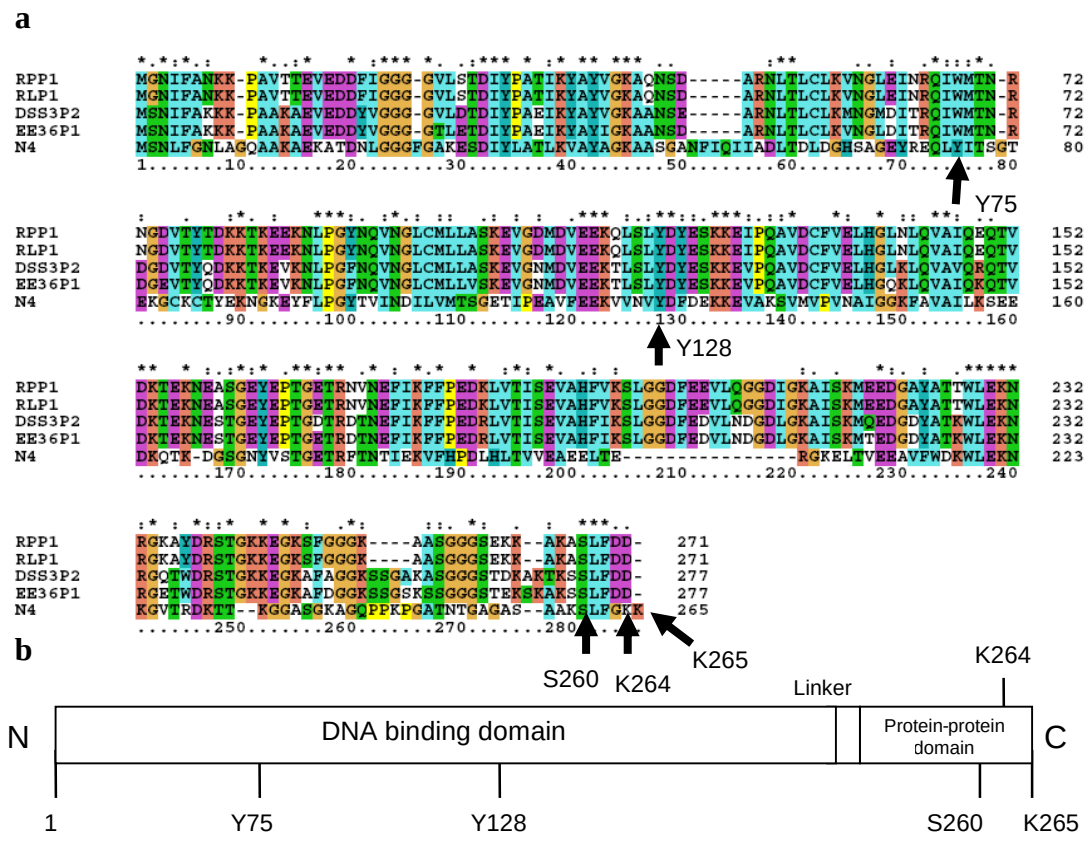


Figure 5.9 Analysis of N4-like gp45, the single-stranded binding protein involved in late gene transcription. a) Amino acid alignment of gp45–like proteins from Roseobacter phages and their homologue from N4, the arrows indicate the catalytic residues in N4. b) Basic domain structure of N4 gp45 showing the position and function of the catalytic residues.

5.2.8 Host-like genes

Annotation of the genomes of RLP1 and RPP1 revealed that both contain five genes whose closest homologues were genes occurring in Roseobacter species and that consequently may be of host origin i.e. host-like: gene products 16, 22, 46/45, 59/60 and 66/65 (RLP1/RPP1 respectively). gp16, which was found to be a homologue of the hypothetical gene RD1_B0001 in *Roseobacter denitrificans*, does not contain any known motifs (as determined by functional analysis with InterProScan and PPsearch), however, gp22 (homologue hyp. RCCS2_17771 in *Roseobacter* sp. CCS2), has two recognised domains; a putative peptidoglycan-binding domain and a C-terminal mandelate racemase/muconate-lactonizing enzyme domain. The homologue in *Sagitulla stellata* E-37 (ZP_01745311.1) is a spore cortex-lytic enzyme

precursor; it contains a putative peptidoglycan-binding domain and peptidoglycan recognition protein domain which are pattern recognition molecules that are conserved from insects to mammals, they recognise and sometimes hydrolyse the peptidoglycan in bacterial cell walls (Dziarski, 2004; Yoshida et al., 1996). These findings point to a possible lysis role for gp22 during phage infection. However, contrary to most lysis genes, gp22 is found relatively early or upstream in the genome (based on the assumption that the genes products are expressed in sequential order) which points to an alternative/multifunctional role or the assumed genome arrangement requires revision (see Section 5.3). gp66/65, was found to be the homologue of the hypothetical protein RAZWK3B_16620 in *Roseobacter* sp.AzqK-3b (ZP_0.1903543) which is believed to be the periplasmic component of an ABC-type oligopeptide transport system.

Gene products 46/45 and 59/60 have best BLAST hits to ribonucleoside diphosphate reductase (*rnr*) and thioredoxin (*trx*) genes respectively. Both these genes are commonly found in marine phage genomes, indeed *rnr* was identified in viral metagenomic studies to be the amongst the most common genes found in the Sargasso sea where it may provide vital nucleotides for DNA synthesis in a phosphate limited environment (Angly *et al.*, 2006). Thioredoxin, when bound to the enzyme, can increase the processing speed of T7-polymerase by promoting DNA/polymerase binding (Huber *et al.* 1987; Etson *et al.*, 2010). However, the 76 amino acid thioredoxin-binding domain on the T7-DNAP is absent from the RLP1 and RPP1 homologues. Nevertheless, it is of interest to note that the *trx*-like gene is found in relatively close proximity to the N4-like DNA polymerase and helicase genes which would suggest together they form a DNA replication module. Such modules have been identified in other marine phage belonging to the T7 superfamily (Hardies *et al.*, 2003).

Initial BLAST analysis of the putative *rnr* and *trx* genes found them to be similar to those found in Roseobacter species, Table 5.3. Interestingly, in phylogenetic comparisons the *trx* genes (Fig. 5.10) from Roseobacter phages cluster closer to the other marine podoviruses than to the bacterial homologues whilst analysis of *rnr* homologues (Fig. 5.11) places the RLP1/RPP1 genes in the Roseobacter species/Roseobacter phage cluster and away from other marine podovirus homologues. Conspicuous by its absence is a N4 phage homologue; intriguingly N4 does not encode a ribonucleoside diphosphate reductase nor a thioredoxin enzyme which

suggests that they are essential for phages in the marine environment. An attractive hypothesis is that a marine podovirus acquired the *trx* gene from an unknown marine host and it was then shared by lateral gene transfer to a number of marine phages. In contrast, the *rnr* gene was obtained directly from a Roseobacter host. Regardless of the method of acquisition, together they allow marine podoviruses to contend with the challenges posed by their new nutrient-limited conditions.

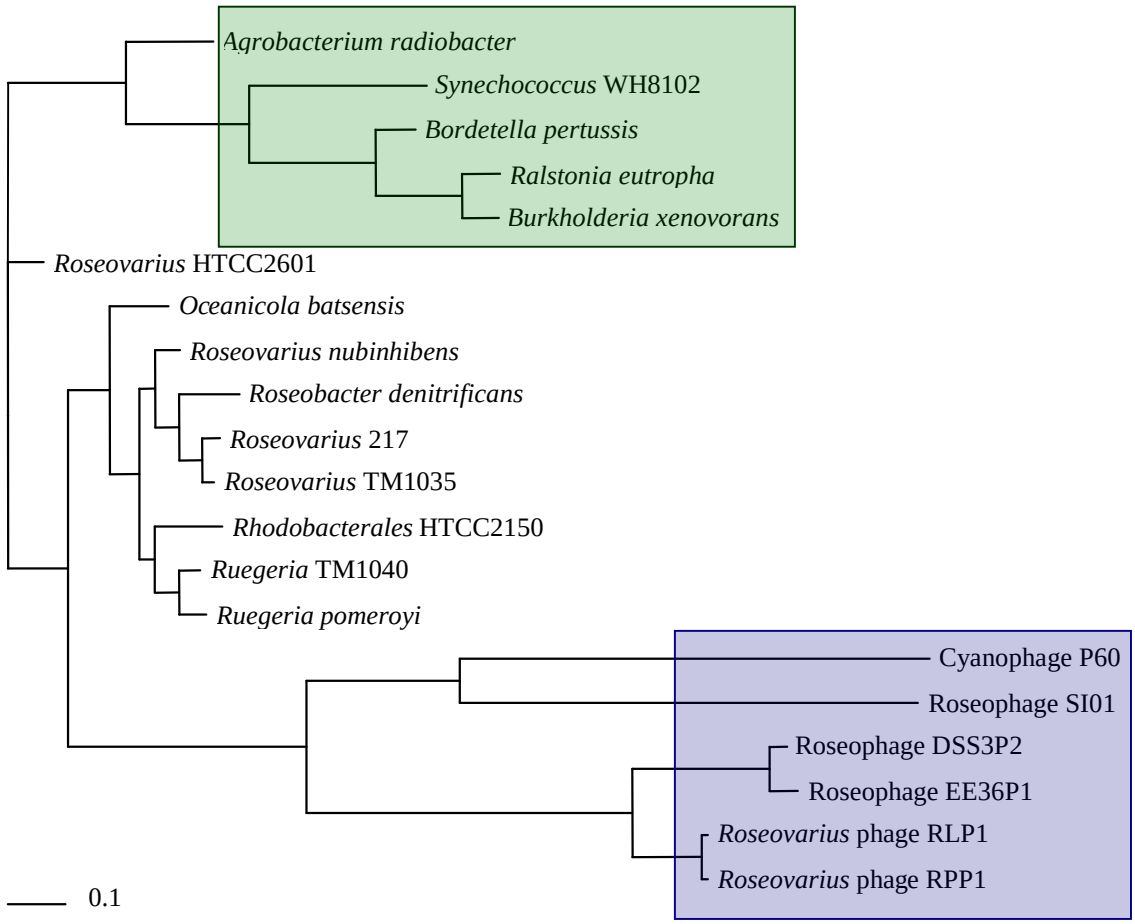


Figure 5.10 Phylogram (made using MrBayes) of aligned amino acid sequences of thioredoxin. Blue box highlights the marine phage cluster, root determined by outgroups in green box. Scale bar indicates expected changes per site.

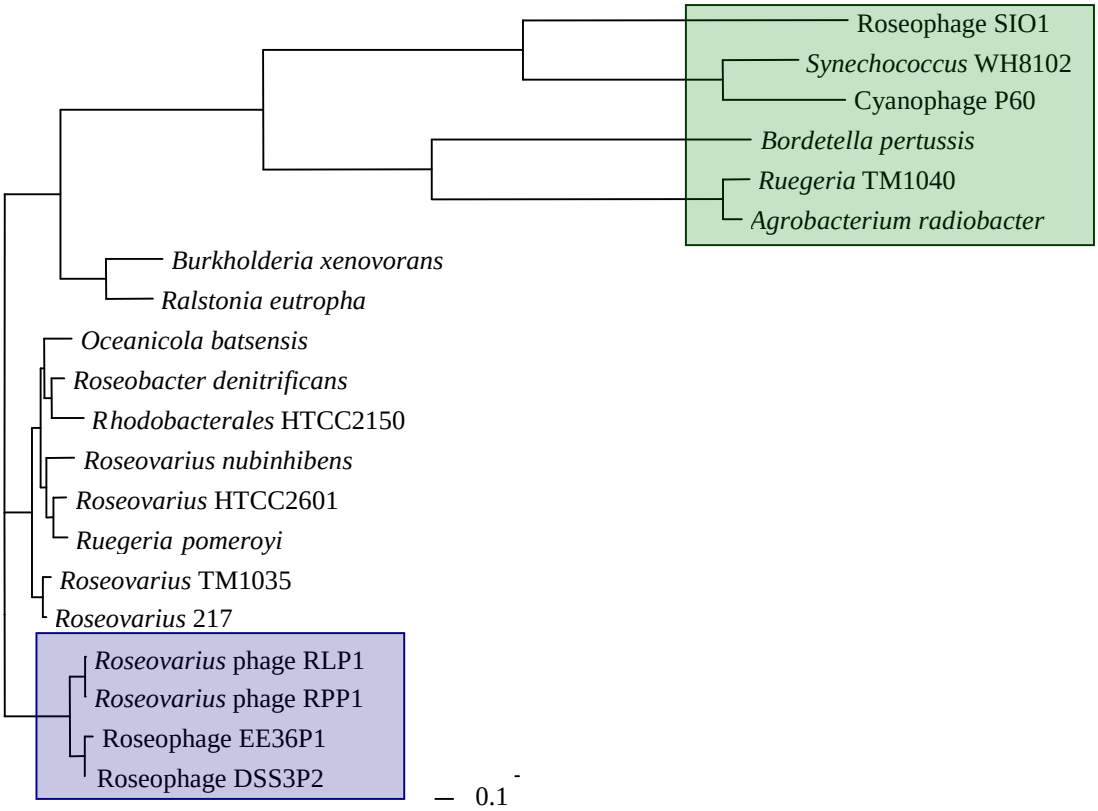


Figure 5.11 Phylogram (made using MrBayes) of aligned amino acid sequences of ribonucleotide reductase. Blue box highlights Roseobacter phage cluster which is adjacent to the Roseobacter cluster root determined by outgroups in green box. Scale bar indicates expected changes per site.

5.2.9 Major capsid protein

The putative major capsid protein sequence was used in a search against the GOS database and 12 environmental hits were obtained from 7 locations (Fig 5.12). The majority of sequences came from coastal areas which mirror the results from the study by Zhao *et al.*, (2009) where DNA polymerase sequences from two N4-like Roseobacter phages were used to interrogate the GOS database. The hits did not form any discernable geographic pattern though the majority were coastal.

Table 5.4 Best BLAST hits of MCP from the GOS database.

Location	Coastal/ Oceanic	CAMERA accession number	*E*-value
Gulf of Panama	Coastal	JCVI_PEP_1105158677525	5.56E-83
Sargasso station 3	Oceanic	JCVI_PEP_1105096277453	5.26E-41
Gulf of Panama	Coastal	JCVI_PEP_1105158677525	3.33E-11
Punta Cormorant, Hypersaline Lagoon, Floreana Island	Coastal	JCVI_PEP_1105141713275	9.06E-09

Table 5.4 cont.

Location	Coastal/ Oceanic	CAMERA accession number	*E*-value
Chesapeake bay	Coastal	JCVI_PEP_1105145032695	2.92E-07
Punta Cormorant, Hypersaline Lagoon, Floreana Island	Coastal	JCVI_PEP_1105137099813	1.00E-07
Sargasso station 11	Oceanic	JCVI_PEP_110510398569	1.45E-06
Sargasso station 11	Oceanic	JCVI_PEP_1105103995903	1.45E-06
Sargasso station 11	Oceanic	JCVI_PEP_1105145465073	1.45E-06
Yucatan channel	Coastal	JCVI_PEP_1105085698360	6.08E-05
Chesapeake bay	Coastal	JCVI_PEP_1105109986267	1.35E-04
Northern gulf of Maine	Coastal	JCVI_PEP_1105081745177	7.43E-03

Figure 5.12 Sample origin of best BLAST hits of MCP against the GOS database.

5.2.10 Regulatory elements

Three transfer RNA genes were identified in RLP1 and RPP1: proline (CCA), isoleucine (ATC) and glutamine (CAA). Analysis of the codon usage of host and phage identified the proline codon CCA in *Rsv. nubinhibens* to be rarely used, in contrast it is relatively highly monopolised in RPP1. Interestingly, the same phenomenon occurs for the Pro-tRNA (CCA) with Roseophages DSS3Φ2, EE36Φ1 and their respective hosts (Zhao *et al.*, 2009). The reason behind the presence of the

remaining two tRNAs remains unclear as their usage frequencies do not suggest a need for an additional ribozyme, see Table 5.5.

Table 5.5 Codon usage in the Roseovarius phages and their hosts. Codons in bold are those encoded by phage tRNAs.

Amino acid	Codon	Usage frequency: per thousand			
		Rsv. 217	**RLP1**	***Rsv. nubinhibens***	**RPP1**
Pro	CCT	14.8	13.6	1.6	13.7
	CCC	18.3	8.5	33.1	10.1
	CCA	**18.5**	**18.7**	**1.6**	**22.1**
	CCG	28.2	8	22.6	11.2
Ile	ATT	9.9	14.2	4	13.6
	ATC	**21.1**	**25.2**	**52.4**	**18.1**
	ATA	6.8	9.2	0	7.8
Gln	**CAA**	**16.1**	**17.8**	**10.5**	**22.1**
	CAG	18.3	25.5	16.1	21.9

Ribosomal-binding sites (43 in RLP1 46 in RPP1) and transcriptional terminators (four in RLP1 and eight in RPP1) were also identified using RBS-finder and TransTerm respectively; the majority of the predicted elements occur at the ends of ORFs. Those that do not are probably false positives, but may also indicate the presence of hidden ORFs and so cannot be discounted.

5.2.11 Gene module order

Annotation of RLP1 and RPP1 identified the presence of all of the major components required for early (vRNAP), middle (RNAP1, RNAP2, gp2) and late (gp45/SSB with host σ^{70}) gene expression as defined by N4 transcriptional control; as well as putative structural and lysis related genes. However, they do not appear to be in the expected order for sequential expression as is usually found in phage genomes. In Roseophages DSS3Φ2 and EE36Φ1 gene order appears to have been conserved with the RNAPII genes occurring amongst the early genes, followed by the DNA replication module, structural and finally lysis genes. (The only exception can be found upstream of the DNA replication module where gp33/rIIA, gp3r4/rIIB, gp22 and gp14 cluster together.) In contrast the module order appears to have been reversed in RLP1 and RPP1 and the structural genes divided into two clusters see Fig 6.11. Two possible explanations can be inferred from the first issue (i.e. the reversal of genes): either the start point of the linear genome is incorrect (n.b. this was chosen by

comparisons with DSS3Φ2 and EE36Φ1 so that the N4 gp2 was the beginning) or these two phages circularise upon infection so that module order is conserved, albeit in reverse. The apparent division of the structural module remains unclear and will be discussed further in Chapter 6.

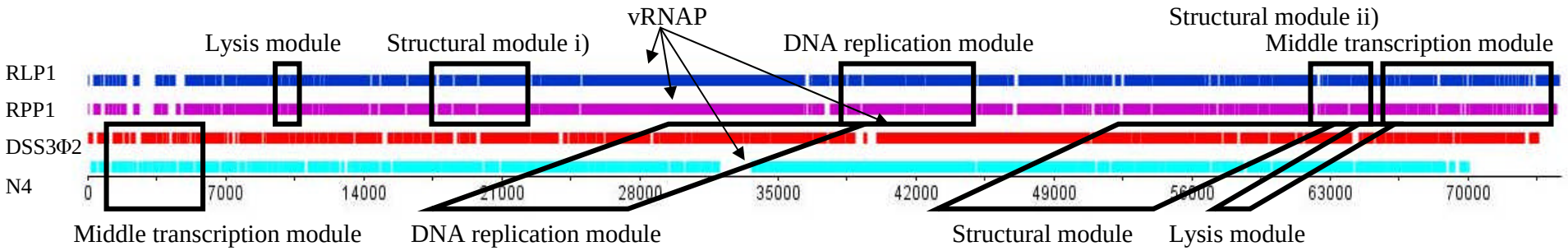


Figure 5.13 Module order found in RLP1, RPP1, DSS3Φ2 and N4 (top to bottom).

5.2.12 Comparative genomics

5.2.12.1 RLP1 vs RPP1

The two *Roseovarius* phage genomes are highly related in almost all putative ORFs; RLP1 has only four unique ORFs and RPP1 has three. Furthermore, at the nucleotide level the genes are highly conserved with a 95 – 100% similarity (see Table 5.3). There do not appear to be any large-scale genomic rearrangements such as duplications, deletions or insertions; consequently, it is possible that the two are strains of the same phage. However, the difference in host range and restriction patterns suggest otherwise. Other highly related phages with above 90% identity have been reported e.g. *Pseudomonas aeruginosa* phages SD1-M/ΦKZ – 99% identity, *Staphylococcus aureus* phages K/GI – 90% and Mycobacteriophages Bxz1/Catera and Cjw1/224 - < 90% (Kwan *et al.*, 2005; 2006; Hatfull *et al.*, 2006).

An additional pertinent fact to consider is the passage of time between the collection of the seawater samples. RLP1 came from a sample collected in 2005 and RPP1 from one harvested in 1998. With the highly dynamic nature of phage populations, turnover rates estimated to be a week or less (Wommack and Colwell, 2000) and problems with delineation of species (due to the mosaic nature of phages and their high propensity for horizontal gene transfer and recombination) it is unlikely that the two phages are the same (see Section 7.1.4 for further discussion).

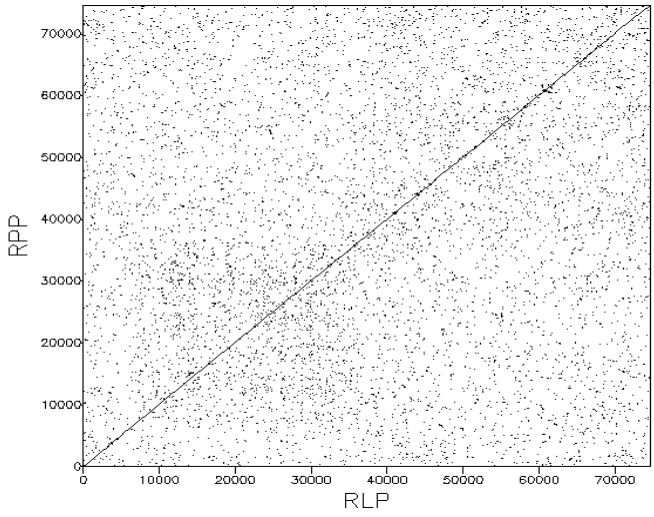


Figure 5.14 Dotplot analysis of RLP1 and RPP1. Threshold score of 50, window size of 15 bp.

5.2.12.2 N4-like Roseobacter phages (RN4-like)

Comparative analysis of Roseobacter phage ORFs identified 65 potential Roseophage-specific genes, see Table 6.6. Previously DSS3Φ2 and EE36Φ1 were thought to have 15 and 11 ORFans respectively however, with comparison with RLP1 and RPP1, gp67 from EE36Φ1 and gp37 from DSS3Φ2 can be reclassified as Roseophage-specific. This is in keeping with the belief that as more phage genomes are sequenced, a minority of ORFans will be reclassified but the majority will still remain unknown (see Section 1.7.6). Interestingly, gp67 from EE36Φ1 is relatively small, 96 amino acids long, which again agrees with the hypothesis that ORFans are molecular splints. In this case, it could bind to a protein common the Roseobacter spp. infected by EE36Φ1, RLP1 and RPP1.

On a nucleotide level, % identity ranges from 45 – 85% and apart from the apparent inversion event (see Section 6.2.11), there appears to be a high degree of synteny between the four phages. It is important to note that the four Roseobacter phages were isolated in different locations (both geographically and environmental conditions e.g. temperature and climate) and in the case of DSS3Φ2 and EE36Φ1, predatory against relatively unrelated hosts. Therefore their high degree of relatedness would not necessarily be expected especially considering there have only been six Roseobacter phages isolated thus far. It is however, far too premature to draw any firm conclusions as to the nature of all Roseobacter phages from these four N4-like examples especially considering SIO1, RDJLΦ1 and the inducible prophages are not N4-like.

5.2.12.3 N4-like phages

The percent identity between shared homologous genes in N4, LUZ7 and LIT1 ranges from 40 – 60% and as with the Roseobacter phages, gene order is relatively conserved. The main deviation from this can be found in the "DNA replication module" which nominally runs from the N4-like gp45 (single-stranded-binding protein) to *trx* (gp59/60 in RLP1 and RPP1 respectively). Intriguingly, in all the Roseobacter phages the N4 gp14 homologue has been translocated into the replication module whereas in LIT1, the gp14-like gene remains amongst the early genes. The N4 gp22-like (a putative HNH endonuclease) has also been rearranged into the DNA replication module, but in this case also is present in the Roseobacter phages and LIT1. Additionally, it is noteworthy that neither the gp14 or gp22

homologues are present in LUZ7. However, with so few N4-like phages isolated and the lack of function assigned to either gp14, speculation into the cause of these rearrangements would be premature.

Table 5.6 Comparison of RLP1 and RPP1 to various N4-like phages. ORFs in bold with - % identity values are homologous but have below threshold e-value (> 0.001).

RLP1 ORF	RPP1 ORF	DSS3Φ2			EE36Φ1			N4			LUZ7			LIT1			Comments
		ORF	% identity		ORF	% identity		ORF	% identity		ORF	% identity		ORF	% identity		
			RLP	RPP		RLP	RPP		RLP	RPP		RLP	RPP		RLP	RPP	
2	2	3	73.5	73.2	4	73.7	73.5	**2**	**-**	**-**							
11	12	81	65.2	64	79	66.6	66.3										
13	14	79	50.5	50.2	78	51.7	51.5										Structural protein
15	15	78	49.2	54.4	76	47.7	48.8										Structural protein
16	16	77	60.4	60.4	75	60.8	60.8										Host-like/Structural protein
17	17	76	78	77.8	74	76.5	76.4	69	51.2	51.3	86	55.2	55.1	85	54.1	54.3	
18	18	75	83.9	83.9	73	82.9	83.1	68	58.9	59	85	57.5	57.5	84	58.3	58.4	Terminase, large subunit
19	19	74	72.6	72.6	72	72.3	72.3	**67**	**-**	**-**	**84**	-	-	**83**	-	-	30 kDa, structural protein
20	20	73	81.4	81.9	71	77.4	77.7										
21	21	72	67.7	67.4	70	65.2	64.7										
22	22	71	79.4	80.1	69	79.9	80.5										Host-like protein
23	23	70	79.5	79.5	68	76.8	76.8										
24	24				67	80.7	83										
25	25	69	78.2	78.3	66	78.2	78.2	59	55.4	55.3	82	53.4	53	80	55.7	55.7	94 kDa portal protein/Structural protein
26	26	68	75.8	75	65	71.9	72.1				81	54.3	-				
27	27	67	74.7	74.5	64	74.5	74.5	57	51.6	51.8	80	51	51.1	78	51	50.7	
28	28	66	84.1	84	63	82.4	82.3	56	55.7	55.7	79	59	59	77	59	59.1	Major coat protein
29	29	65	70.9	71.2	62	70.1	70.2	55	48.2	48.6	78	48.5	48.7	76	50	50.8	
30	30	64	74.7	74.7	61	76.2	76.1	54	52	52	77	47.7	47.7	75	52.8	52.8	Structural protein
31	31	63	65.7	65.7	60	71.6	71.7	53	48.7	48.8							
32	32	62	56.9	56.9	59	65.8	65.8	52	49.9	49.9	75	42	42	73	46	46.2	16.5 kDa, structural protein
33	33	61	51.4	51.3	58	50.1	50.1										Structural protein, cell wall hydrolase domain
34	34	60	59.9	59.9	57	59.2	59.2	50	48.5	47.3	73	47.6	47.7	71	47.7	47.5	vRNAP

Table 5.6 cont.

RLP1 ORF	RPP1 ORF	DSS3Φ2			EE36Φ1			N4			LUZ7			LIT1			Comments
		ORF	% identity		ORF	% identity		ORF	% identity		ORF	% identity		ORF	% identity		
			RLP	RPP		RLP	RPP		RLP	RPP		RLP	RPP		RLP	RPP	
36	35	59	52.1	-	56	51.6	-										
37	36	58	43.7	81.5	55	76.4	76.7										
38	37	57	79	79.2	54	78.7	78.3	45	51	51	**65**	-	-	62	53.4	53.4	SSB
39	38	56	73.9	74	53	73.6	73.7	44	57.2	57.2	64	58.2	58.3	61	56.3	56.3	
40	39	55	78.6	78.5	52	77.9	77.8	43	56.2	56.1	63	56.6	56.5	60	54.5	54.6	
41	40	54	78	78.3	50	76.2	76.5	42	53.4	53.8	62	53.4	53.4	59	55.2	55.4	
42	41	52	75.7	75.6	48	75	75	39	58.6	58.5	42	54.1	54.2	38	54.7	54.8	DNA polymerase
44	43	50	77.5	77.9	47	76.5	76.5	37	50.8	51	40	49.7	50.9	36	51.3	51.9	DNA helicase
46	45	48	79.5	79.3	45	78.7	78.6										*rnr*
48	46	47	62.2	60.9	44	59.7	58										
49	48	46	65.8	65.8	43	65.1	65										
53	52	45	76.7	76.7	42	75.1	75.1	14	57.1	57.1				15	57.4	57.4	
54	53	44	60.7	61.4	41	62.4	50.7										
55	54	43	80.7	80.7	40	81.7	81.7	22	61.3	61.3				41	60	60	
56	55	42	67	67.1	39	66.9	66.9	34	49.4	49.4	47	51.6	51.6	43	49	49	rIIB-like
57	56	41	56.9	56.6	38	55.3	55.3	33	45.6	45.3	46	47.8	47.6	42	49.1	49.5	rIIA-like
60	59	40	72.4	72.4	37	72.7	72.7										*trx*
61	60	39	68.4	68.4	36	66.7	66.7	36	44.2	44.2							
62	61	38	47.5	47.5	35	45.7	45.7										
63	62	37	48.3	48.3													
65	64	33	52	49.8	33	46.9	45.9										Structural protein
66	65	32	64.8	66.3	30	67	68.7										Host-like/Structural protein
68	67	31	65.7	65.7	29	66.2	66.2										
69	68	30	77.5	77.8	28	78	78.5										Structural protein

Table 5.6 cont.

RLP1 ORF	RPP1 ORF	DSS3Φ2			EE36Φ1			N4			LUZ7			LIT1			Comments
		ORF	% identity		ORF	% identity		ORF	% identity		ORF	% identity		ORF	% identity		
			RLP	RPP		RLP	RPP		RLP	RPP		RLP	RPP		RLP	RPP	
70	69	29	73.1	73.1	27	70.2	69.3	30	54.3	54.6							Thy synthase complementing protein
71	70	28	73	73	26	73.7	73.7										
72	71	25	54.5	55	23	53.7	54.4										Structural protein
73	72	27	75.5	76.2	25	76.2	76										dCMP deaminase
74	73	23	67	67	21	71.1	70										
75	74	24	66.7	66.1	22	69.4	69.5										
76	75	22	70.1	69.8	20	69.9	69.8	25	47.3	46.4	39	45.7	45.5	35	47	47.5	vWFA domain
78	77	20	78	78.2	18	77.7	78	24	49.8	50	37	52.6	53.5	33	51.9	51.6	
79	78	18	72	71.5	16	72	71.5										
81	80	16	74.3	74.3	15	74.7	74.5	16	52.7	53.1	22	50.8	52.6	23	51.8	51.8	RNAP2
82	81	15	75.1	75.1	14	72	72										
85	84	12	54.6	55	11	55.4	56.3										
88	87	10	57.1	57.6	9	56.3	56.7										
91	90	6	77.4	77.3	7	74.7	74.8	15	54	53.6	20	47.3	47.1	19	50.6	50.3	RNAP1

5.2.13 Bipartite genome

It is clear from genome comparisons of the seven N4-like phages that they share a number of conserved genes. These core genes appear to broadly fall into three categories; transcriptional control, DNA metabolism/replication and structural proteins, see Table 6.7. In addition, these genes are largely syntenous (apart from gp14 and gp22 as mentioned in Section 6.2.12.3) which suggests a stable association within each core module has been formed. Interrupting these core genes are hyperplastic regions of which the majority of putative ORFs are unknown, though they are often shared between the four Roseobacter N4-like phages. For example, between the homologues of N4 gp26 and gp33, the RN4-like phages share 12 genes. This is reminiscent of the T4 superfamily where the genomes have been defined as bipartite (Krisch and Comeau, 2008); a conserved core comprised of the minimal essential genes required for viral multiplication and a larger, highly variable set of facultative genes which collectively create an optimal environment, particular to that host, to enable successful infection.

Another intriguing observation is the number of early genes that fall into the hyperplastic region before and after the N4 gp15 & gp16 homologues. Though two genes are shared between the Roseobacter N4-like phages (RLP1 gp2 and gp79) the remainder are unique to the two Roseovarius phages. As early proteins are generally directed towards take-over of the host metabolism (see Section 1.7.1) this result is understandable.

It would appear that in clade, only genes with general functions such as DNA metabolism are shared between phages with phylogenetically distant hosts, whereas those with specific functions, such as a structural protein with a cell wall hydrolase domain (RLP1 gp33) are restricted to phages with closely related hosts.

Table 5.7 Core genes for N4-like phage genus.

gp67 in bold and - indicates there is a homologue but it falls below the threshold e-value of 0.001.

Gene in N4	Function	% identity						
		RLP1	RPP1	DSS3Φ2	EE36Φ1	LIT1	LUZ7	
15	RNAP1	54.0	53.6	54.9	53.5	51.4	49.9	Transcriptional control
16	RNAP2	52.7	53.1	53.2	54.9	49.7	49.4	
24	Unknown	49.8	50.0	52.3	53.5	49.8	51.8	
25	vWFA domain	47.3	46.4	47.8	48.4	47.8	48.4	DNA metabolism/ replication
33	rIIB-like	45.6	45.3	45.7	47.3	46.5	45.6	
34	rIIA-like	49.4	49.4	49.6	49.0	52.0	51.1	
37	DNA helicase	50.8	51.0	49.5	49.6	46.9	46.9	
39	DNA polymerase	58.6	58.5	59.0	59.5	51.5	51.4	
42	Unknown	53.4	53.8	54.5	54.3	54.6	53.7	
43	Unknown	56.2	56.1	56.8	57.1	53.0	54.0	
44	Unknown	57.2	57.2	55.8	56.5	59.4	57.0	
45	SSB	51.0	51.0	51.5	50.6	48.7	49.0	
50	vRNAP	48.5	47.3	48.1	47.9	45.4	45.2	
52	Structural protein	49.9	49.9	53.5	-	46.8	45.5	Structural proteins
53	Unknown	48.7	48.8	51.4	48.1	46.6	-	
54	Structural protein	52.0	52.0	52.0	48.3	45.4	45.5	
55	Unknown	48.2	48.6	53.1	47.7	50.7	52.4	
56	Major coat protein	55.7	55.7	56.4	45.4	59.6	61.1	
57	Unknown	51.6	51.8	48.5	45.3	50.6	48.7	
59	94 kDa portal protein/Structural protein	55.4	55.3	55.4	49.9	54.1	55.6	
67	**Structural protein**	**-**	**-**	**-**	**-**	**47.3**	**-**	
68	Terminase, large subunit	58.9	59.0	59.9	47.9	58.4	57.4	
69	Unknown	51.2	51.3	53.4	43.5	51.6	52.7	

5.3 Conclusions

Through analysis of module order and comparative genomics it is apparent that during the annotation of the RLP1 and RPP1 genomes, gene order was reversed. Consequently the chosen start point of the two linear genomes is incorrect and instead is likely found between ORFs 3 - 9. However, as this point has not yet been defined experimentally, correction of this would be precipitous and likely wrong. Instead, runoff Sanger sequencing of the ends is required to determine both the extent of terminal redundancy and the true gene order.

The end structures of enterobacteria phage N4 was determined by Ohmori *et al.* (1988); in their study the left genomic end was found to be relatively conserved whereas the right end was of variable length. As the terminal structures of double stranded phage DNA are pertinent to the mechanism by which DNA is replicated then packaged into nascent virion particles and this is probably conserved in N4-like phages, it is highly probable that genomic ends of RLP1 and RPP1 will be similar to that of N4.

The characterisation of the two Roseovarius bacteriophage RLP1 and RPP1 has many implications for the growing N4 –like phage genus. It would appear that the conserved genes identified in the comparison of the seven N4-like phages broadly fall into three categories: transcriptional control, DNA metabolism/replication and structural proteins. The remaining plastic genes are probably responsible for host interactions and so appear conserved only amongst the Roseobacter or *Roseovarius* specific phages. It is likely that future N4-like phages, when sequenced, will probably be found to contain homologues of the general host hijacking related genes identified in this study as highly conserved core genes, but those responsible for specialized host/phage interactions (e.g. host lysis) are likely to be ORFans i.e. novel and/or individual to that bacteriophage and its host(s).

A puzzling observation is the presence of four structural proteins (identified by mass spectrometry, see Chapter 7) in the DNA replication module; RPP1 gps64, 65, 68 & 71. None of these genes have homologues in N4, though they do in DSS3Φ2 and EE6Φ1. In the paper by Zhao *et al.* (2009), they identified the first three proteins as being homologues of a hypothetical protein from *Acidovorax avenae* ssp. citrulli AAC00-1, *Roseobacter* sp. AzeK-3b, hypothetical protein and *Erwinia amylovora*

phage Era 103 hypothetical protein g26, respectively. Their location would suggest expression during middle transcription but most structural proteins are found amongst late genes and so their function remains unclear. This apparent division of the structural proteins between two transcription phases will be discussed in more detail in Section 7.2.6.

In addition these two novel phages appear to be distantly related to other marine podoviruses (see Section 6.2.8) and there is growing evidence to suggest a DNA replication/metabolism module conserved amongst such phage. However, this will only be confirmed or revised upon characterisation of more marine podoviruses and general N4 –like phages.

Chapter 6

Proteomic analysis of *Roseovarius* phages RLP1 and RPP1

6.1 Introduction

Once a phage genome has been sequenced, identification of phage structural proteins is often the next step in characterisation of a novel phage and has been increasing in popularity over recent years (Ceyssens *et al.*, 2006; 2009b; Lavigne *et al.*, 2006; 2009; Lingohr *et al.*, 2008; Lecoutere *et al.*, 2009). Apart from the obvious advantage of confirming *in silico* gene annotations, identification of virion proteins can also reveal many facets of a phage's infection cycle. For example, identification of tail proteins can reveal the method by which a phage injects its genetic material into a host. In the myovirus T4, a baseplate at the tip of the phage tail attaches to cell surface fibres; binding triggers a conformational change which drives the tail tube (which contains three lysozyme domains) through the cell envelope (Rossmann *et al.*, 2004). After the cell wall is breached, one of the most efficient DNA transport processes known occurs and the phage genome is injected into the cell in only 30 seconds (Boulanger and Letellier, 1988). In contrast Podoviridae have short, non-contractile tails which are too short to form a conduit from the cell surface to the cytoplasm. Consequently in T7, phage-encoded structural proteins are injected with the DNA during infection; they are thought to form a tunnel through which the genetic material can pass and reach the cell cytoplasm (Molineux, 2001).

Enterobacteria phage N4 has an additional handicap as it must also transport approximately four 3500 amino acid proteins (the vRNAP) from the capsid into the host cell. It is presumed that that the transport of vRNAPs occurs prior to the first ~500 bp of genomic DNA. The polymerases then begin transcription from a promoter present in this region pulling the next 10 – 40 kbp of the genome out of the virion into the host cell (Choi *et al*., 2008). However, transport of such a large protein requires it to be in an unfolded or semi-unfolded form as the narrowest section of the tail tube is 25 Å in diameter (Choi *et al.*, 2008). As such, some of the structural proteins must contain domains with protein-chaperone properties in order to ensure safe passage of this key enzyme (Choi *et al.*, 2008).

To date ten gene products have been positively identified as structural proteins in N4 (see Table 6.1), however, little is known about seven of these proteins. Homologues of some of these proteins have also been identified in N4-like phage LIT1 (Ceyseens, 2009a) which points to a number of key conserved proteins in the

N4-like genus, see Section 5.2.13. Putative structure proteins have also been suggested in the Roseobacter phages DSS3Φ2 and EE36Φ1 (Zhao *et al.*, 2009). The analysis of the structural proteins identified in RLP1 and RPP1 and the impact of this new information in terms of both Roseobacter phages and the growing N4-like genus are examined in this chapter.

Table 6.1 Structural proteins in Enterobacteria phage N4. Adapted from Choi *et al.* 2008

Gene product	Number of Amino acids	Predicted molecular weight (kDa)	Proposed role
17	279	32	Decorating protein
50	3,500	382.5	vRNAP
51	644	66	
52	150	16.5	
54	299	32.4	
56	401	44	Major capsid protein
59	764	94	Portal protein
65	1,382	160	Non-contractile tail sheath
66	556	60	Appendage
67	236	30	

6.2 Results and Discussion

6.2.1 Optimisation of protein extraction

Three methods of phage structural protein extraction were compared to determine the optimal protocol. These were: whole phage extract, phage ghosts and trichloroacetic acid (TCA) precipitation (see Section 2.8.1). Analysis of SDS-PAGE comparing the products of each method showed the whole phage extract and TCA precipitation to be relatively similar; however, the protein bands from the latter were better defined allowing for a greater degree of band resolution. As a result it was utilised as the extraction method of choice.

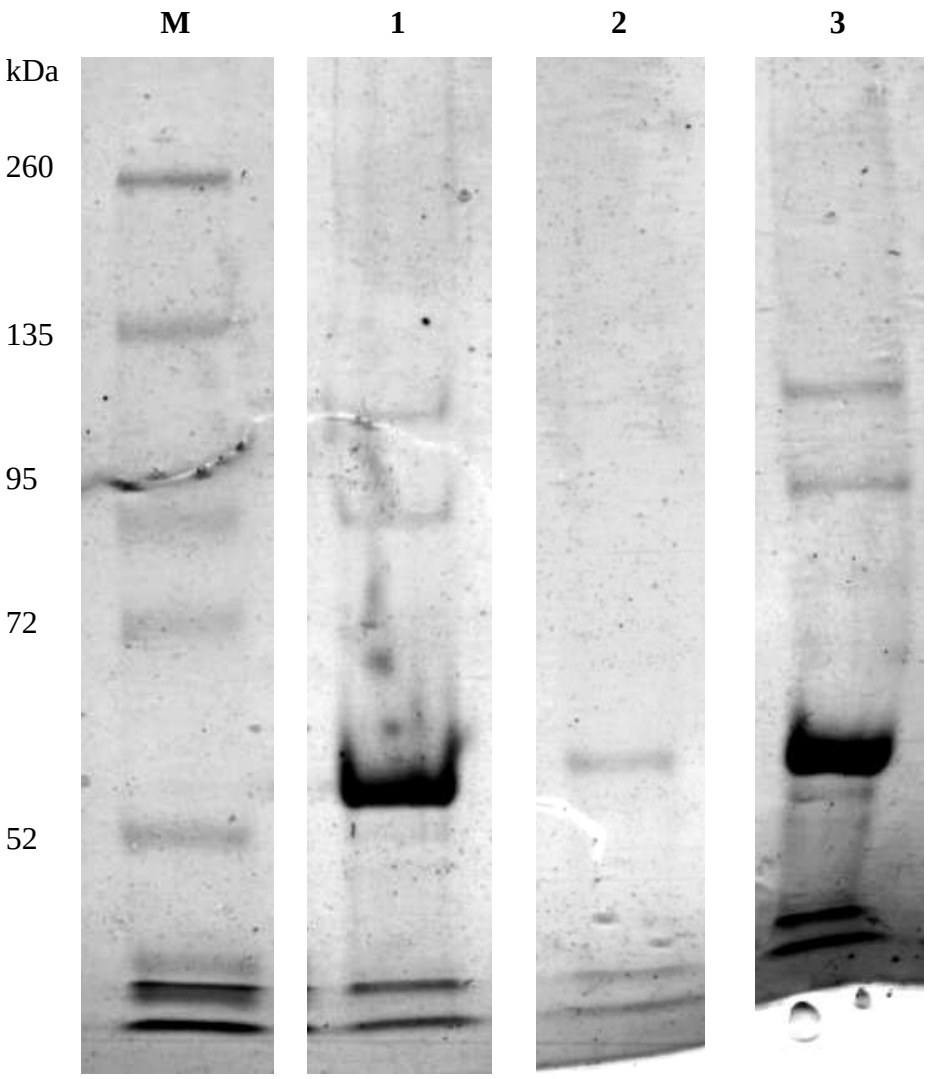


Figure 6.1 Comparison of protein extraction protocols using purified RPP1 on a 12 % SDS-polyacrylamide separating gel. Lane 1 – Whole phage extract, 2 – Phage ghosts, 3 – TCA precipitation. Approximately 10^{12} ultra-pure virions (generated by two consecutive CsCl purifications) were loaded per lane; the subsequent gel was stained with Coomassie Brilliant Blue.

6.2.2 Virion structural proteins

TCA precipitation of virion structural proteins followed by SDS-PAGE revealed the protein profile of phages RLP1 and RPP1. Chosen bands (see Fig. 6.2) were excised and identified by mass spectrometry a list of which can be found in Table 6.2.

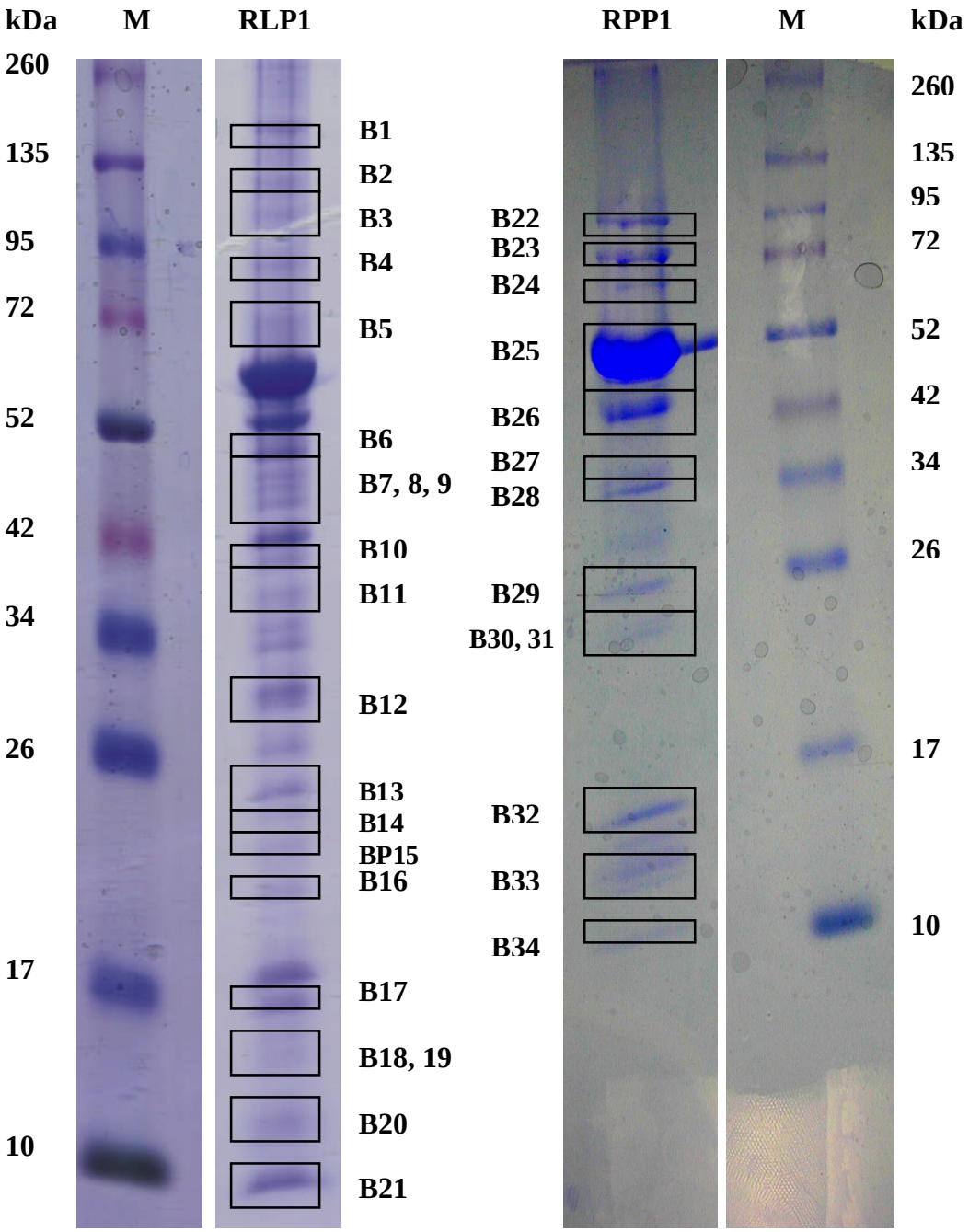


Figure 6.2 Polypeptides of purified RLP1 and RPP1 phage particles. Proteins, extracted by TCA precipitation from approximately 10^{12} double CsCl purified virions, were separated on a 10 – 20 % gradient SDS-polyacrylamide gel set at 100 V for 18 hours. Boxes indicate the bands excised for mass spectrometry. As the genomic information indicated that both phages were highly similar it was assumed that the phage proteins would be highly similar. As such duplicate bands present in both gels were only analysed once e.g. B25, the predicted major capsid protein was only analysed in the RPP1 protein gel. This was done to ensure the maximum coverage of bands present in both phages.

Table 6.2 Identification of proteins in bands analysed by mass spectrometry.

Band	Protein	No. of peptides	Mass/ kDa	*in silico* predicted mass/kDa	Observed MW/kDa
RLP1					
B1	gp33	5	7.5	58.9	109
	gp28	4	51.4		
B2	gp28	6	5.1	90.5	92.3
	gp16	4	10.4		
	gp33	2	75		
B3	gp33	8	7.5	117.3	82.2
	gp28	5	51.4		
	gp30	2	43.7		
	host contamination, putative lipoprotein	2	14.7		
B4	gp28	8	51.4	58.9	70.3
	gp33	7	7.5		
B5	gp28	12	51.4	73.8	63.3
	gp33	8	7.5		
	host contamination, putative lipoprotein	2	14.9		
B6	gp28	10	51.4	58.9	48.7
	gp33	4	7.5		
B7	gp33	10	7.5	147.8	46.6
	gp28	7	51.4		
	gp25	7	88.9		
B8	gp28	8	51.4	176.2	44.7
	gp25	5	88.9		
	gp13	3	35.9		
B9	gp28	11	51.4	120.8	44
	gp19	3	25.1		
B9 cont.	host contamination, putative outer membrane porin	2	36.8		
	gp33	2	7.5		
B10	gp28	12	51.4	113.3	39.3
	gp19	9	25.1		
	host contamination, putative outer membrane porin	2	36.8		
B11	gp28	11	51.4	183.7	37.2
	gp33	4	7.5		
	gp25	4	88.9		
	gp13	2	35.9		
B12	gp28	10	51.4	220.5	31
	gp25	6	88.9		
	gp13	6	35.9		
	host contamination, putative outer membrane porin	2	36.8		
	gp33	2	7.5		
B13	gp28	13	51.4	177.1	25.6
	gp25	6	88.9		
	host contamination, putative outer membrane porin	4	36.8		
B14	gp25	13	88.9	228.3	24.9
	gp30	9	43.7		
	gp28	7	51.4		
	host contamination, putative outer membrane porin	4	36.8		

Table 6.2 cont.

Band	Protein	No. of peptides	Mass/ kDa	*in silico* predicted mass/kDa	Observed MW/kDa
B14	gp33	3	7.5	228.3	24.9
B15	gp30	13	43.7	228.3	23.1
	gp28	12	51.4		
	gp25	6	88.9		
	host contamination, putative outer membrane porin	2	36.8		
	gp33	2	7.5		
B16	gp28	14	51.4	191.5	20.7
	gp30	10	43.7		
	gp33	9	7.5		
	gp25	4	88.9		
B17	gp25	17	88.9	147.8	16.3
	gp33	10	7.5		
	gp28	10	51.4		
B18	gp28	11	51.4	147.8	15.1
	gp33	7	7.5		
	gp25	5	88.9		
B19	gp33	21	7.5	191.5	14
	gp28	11	51.4		
	gp25	8	88.9		
	gp30	2	43.7		
B20	gp28	15	51.4	147.8	12.1
	gp25	11	88.9		
	gp33	3	7.5		
B21	gp28	14	51.4	147.8	10
	gp25	6	88.9		
B21	gp33	3	7.5	14.7	10
RPP1					
B22	gp25	25	88.9	177.8	99.8
	gp28	3	51.4		
B23	gp33	22	75.1	126.5	86.1
	gp28	4	51.4		
B24	gp33	7	75.1	218.4	78
	gp64	5	68.2		
	gp33	1	75.1		
B25	gp28	17	51.4	51.4	62.5
B26	gp28	9	51.4	51.4	51.4
B27	gp14	5	36	87.4	41.2
	gp28	2	51.4		
B28	gp14	6	36	87.4	39.2
	gp28	4	51.4		
B29	gp19	19	25.2	76.6	27.1
	gp28	3	51.4		
B30	gp68	1	44.4	95.8	23.4
	gp28	1	51.4		
B31	gp28	2	51.4	73.7	12.3
	gp15	1	22.3		
B32	gp32	2	16.1	33.3	10.4
	gp71	1	17.2		
B33	gp65	3	12.5	12.5	9.6
B34	gp16	2	10.4	10.4	8.1

An extended version of this table containing further information on each protein can be found in the Appendix, Table A.2.

Many of the bands, in particular in the RLP1 protein gel, were found to contain peptide fragments from more than one gene product. This is unexpected as the sample was heated with ß-mercaptoethanol (which breaks di-sulfide bonds) prior to loading, and the gel used was a denaturing gel which should have disrupted both the tertiary and quaternary structure of the protein complexes. In addition, the observed protein mass from the SDS-PAGE gels do not correspond well to the *in silico* predictions, however this could be due to degradation of the proteins. In the RLP1 gel, gps 25, 28, 30 and 33 were often found together which suggests they are closely associated in the phage virion and bands 14-21 represent the degraded isomers of this complex which arose due to incomplete denaturation. Notable exceptions to this were bands 25, 26, 33 and 34 which were found to contain trypic digest fragments from only one gene product. Comparison of the two phage gels, suggests that the RPP1 gel was of better quality and trypic digest fragments from gps 64, 65, 68 and 71 were also identified, though in the case of the latter two, only one peptide was found. The possible functions of these proteins are discussed in Section 6.2.6.

Many structural gene products were found to be present in many of the bands analysed. For example, gp28 (the predicted major capsid protein) can be found in all analysed bands from the RLP1 protein gel. As such, the data from Table 6.2 were collated to simplify the mass spectrometry results and is presented in Table 6.3.

Table 6.3 Proteins identified by mass spectrometry. Proteins in bold were only identified in the RPP1 protein gel.

Protein	Homologous genes	Comments
gp13/14 [a]	gp79 [DSS3Φ2], gp78 [EE36Φ1]	2 putative cell adhesion domains
gp15	gp77 [DSS3Φ2], gp76 [EE36Φ1]	2 putative glycoprotein domains
gp16	gp77 [DSS3Φ2], gp75 [EE36Φ1]	Host-like protein, 10 predicted β-strands
gp19	gp67 [N4]	30 kDa protein, approx 10 copies/virion [b]
gp25	gp59 [N4]	94 kDa portal protein, approx 14 copies/virion [b]
gp28	gp56 [N4]	Major capsid protein, approx 534 copies/virion [b]
gp30	gp54 [N4]	Approx. 30 copies/virion [b]
gp32	gp52 [N4]	16.5 kDa protein. Approx 41 copies/virion [b]
gp33	gp61 [DSS3Φ2], gp58 [EE36Φ1]	Possible similarity to C-terminal sequence of Roseophage SI01 gp24, hydrolase domain (residues 215-310)
gp64	gp33 [DSS3Φ2], gp33 [EE36Φ1], gp230 [*Pseudomonas* phage 201Φ2-1]	Abundant phage virion protein in phage 201Φ2-1, 10 putative domain of extracellular low-density lipoprotein receptor, 3 putative hydrolase, tail associated lysozyme in T4 domains
gp65	gp32 [DSS3Φ2], gp30 [EE36Φ1]	Host-like protein

Table 6.3 cont.

Protein	Homologous genes	Comments
gp68	gp30 [DSS3Φ2], gp28 [EE36Φ1], p16 [T1]	Many phage hypothetical protein homologues, 2 putative protein transport domains
gp71	gp25 [DSS3Φ2], gp23 [EE36Φ1]	1 putative chaperone domain

[a] RLP1/RPP1 [b] Based on average count from S-Met, SDS and CryoEM studies (Choi *et al.*, 2008)

6.2.3 Conserved N4-like virion proteins

Due to the high degree of similarity of both *Roseovarius* phages to N4 on both the genomic and morphological level, it is not surprising to find that five of the proteins identified by mass spectrometry were homologues of the enterobacteria phage. Unfortunately, of the ten structural proteins identified so far in phage N4, only six are of known function: gp56, the major capsid protein, gp17, a decorating protein, gp50, the vRNAP, gp65, the tail sheath , gp66, the appendages and gp59, the portal protein (Choi *et al.*, 2008). Of these, only the homologues of the major capsid protein and the portal protein were found in RLP1 and RPP1.

The structure of the N4 portal protein was determined in 2008 (Choi *et al.*, 2008) and was found to be similar in structure to the Podovirus Φ29 portal connector assembly protein (which is also found in other tailed phages). These proteins consist of three domains: the crown, wing – mainly composed of α-helices and stalk – made up of β-strands, see Fig 6.3. Due to the relatively high degree of similiarity between N4 gp59 and RLP1/RPP1 gp25 (around 55% at the nucleotide level), it seems likely that the Roseovarius phage portal protein will also contain these three domains. Indeed there is a degree of similarity in the order and length of α-helices and β-strands of the two portal proteins as seen from the predicted protein structure made using PSIPRED, see Fig. 6.4. Results from domain prediction search using DomSSEA also indicated that the RLP1/RPP1 putative portal protein contains nucleotidyltransferase, transferase and lyase domains.

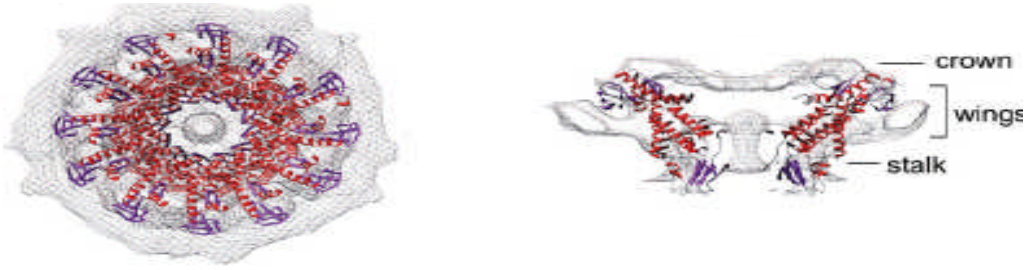


Figure 6.3 N4 portal assembly showing the crown, wing and stalk domains. Taken from Choi *et al.* 2008.

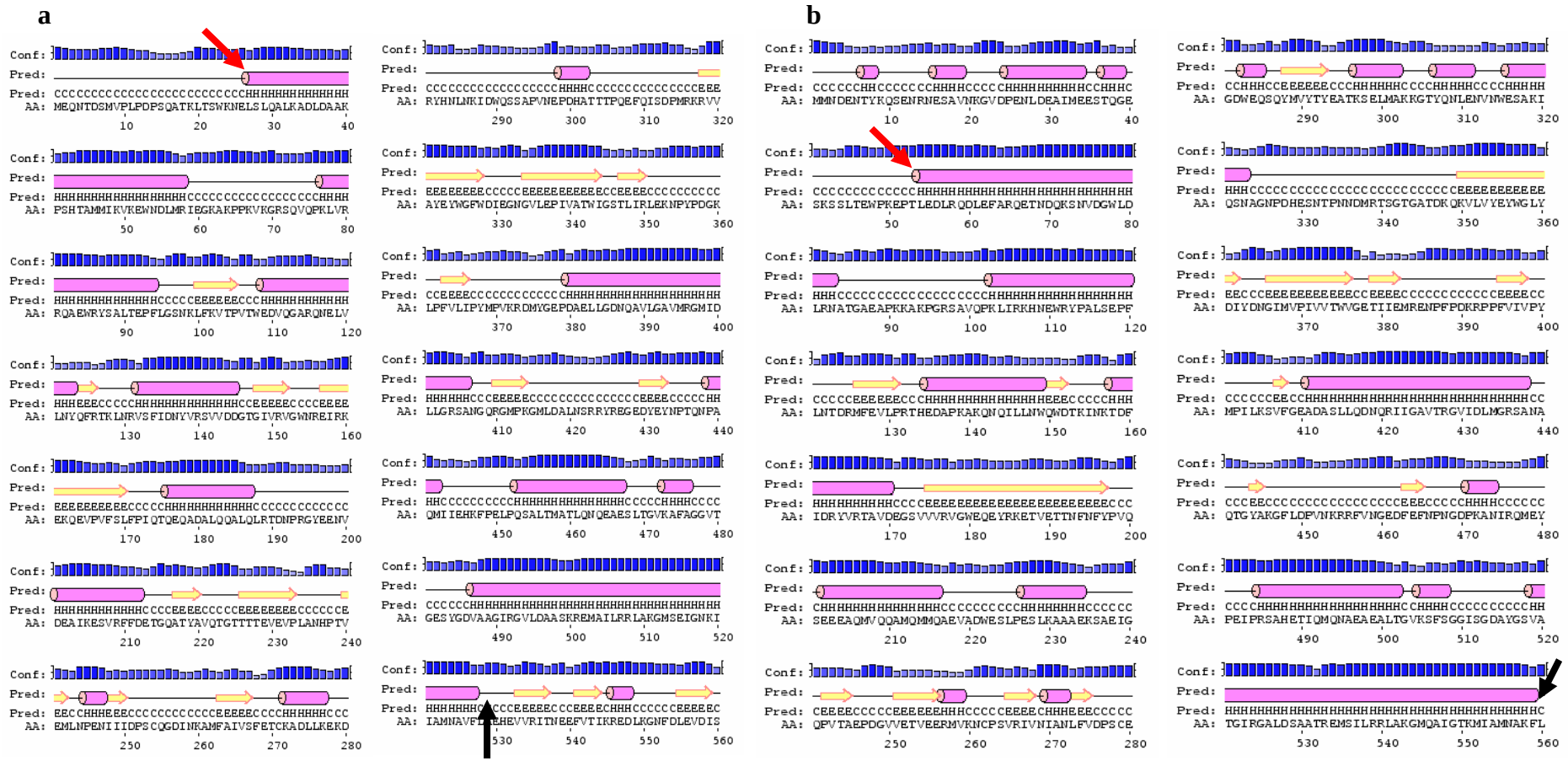


Figure 6.4 Predicted secondary structure of a) N4 and b) RPP1 gp25 portal protein. Red arrows indicate start of similarity; black, the end. The RLP1 copy of gp25 is 99.4% similar on the nucleotide level (see Table 5.3)

In phage N4, the major capsid protein is 44 kDa which is similar in size to the capsid proteins in other N4-like phages for example, LIT1 - 44.2 kDa and some Podoviruses e.g. Φ29 (48 kDa) and P22 (47 kDa) (Choi *et al.*, 2008; Ceyessens, 2009a). The MCPs of RN4-like phages appear to larger: RLP1 – 51.4 kDa, RPP1 – 51.4 kDa, DSS3Φ2 – 51.1 kDa and EE36Φ1 – 51.2 kDa. Based on their amino acid sequence similarity to the MCP of N4, it is likely that the N4-like phages belong to the pseudo-hexameric class of icosahedral viruses. Viruses belonging to this class contain six monomers each consisting of HK97-like folds (named after the structure of the major capsid protein of enterobacteria phage HK97 (Wikoff *et al.*, 2000)), the centres of which are separated by *ca.* 135 Å (Choi *et al.*, 2008). In phage HK97, the capsid protein is 31 kDa and so it has been suggested that the extra residues in similar proteins, such as in N4, Φ29, P22 and T4, form insertion domains on the virion surface (Choi *et al.*, 2008; Morais *et al.*, 2005; Jiang *et al.*, 2003; Fokine *et al.*, 2005). It is likely this is also the case in the RN4-like phages and the increased size of their capsid proteins corresponds to larger insertion domains. It is also interesting to note that TEM micrographs have shown RLP1 and RPP1 to have a capsid head size of 72.4 ± 2 nm and 77.4 ± 5 nm respectively, see Section 3.2.5. The N4 virion particle, in contrast, is slightly smaller at 70 nm (Kazmierczak and Rothman-Denes, 2005).

6.2.4 Structural proteins gp13/14, gp15 and gp16

Intriguingly, the host-like protein gp16 was identified in both RLP1 and RPP1 phage protein gels; the PSIPRED predicted secondary structure of the protein indicates it is likely made of ten β-strands, see Fig. 6.5. Unfortunately, gp16 does not contain any known protein domains or motifs so a putative function in the virion cannot be assigned. However, it is tempting to speculate that by mimicking a Roseobacter protein, gp16 is a tail accessory involved in the binding between the host and phage.

Similarly, gps 13/14 and 15 maybe involved in the process of absorption to a Roseobacter host as analysis with DomSSEA revealed similar secondary structure to cell adhesion and glycoprotein domains, respectively. As this is an interaction specific to Roseobacter phages it is not surprising that homologues to gps13/14, 15 and 16 are not found in N4, but are present in phages DSS3Φ2 and EE36Φ1, making gp13/14 and gp16 RN4-like specific genes. Further evidence to support the hypothesis that

these two proteins are Roseobacter-specific tail or tail fibre proteins, is their position along the phage genomes. In the N4 structural protein module the tail-associated proteins are found after the MCP and portal protein, which in RLP1 and RPP1 is the position occupied by these three genes.

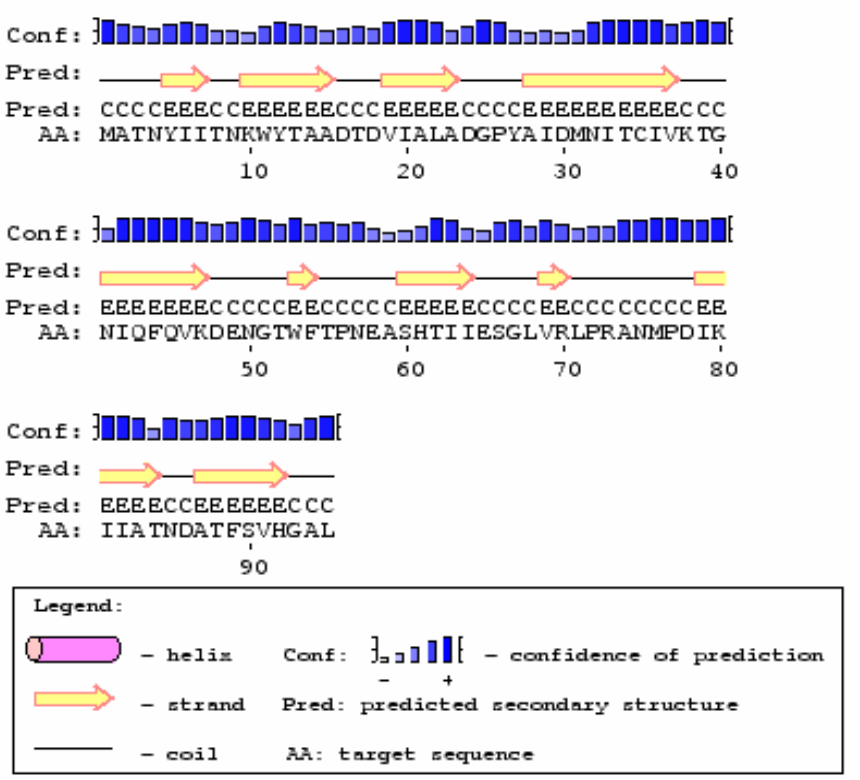


Figure 6.5 Predicted secondary structure of RLP1/RPP1 gene product 16. gp16 contains no recognized protein domains or motifs, but it is a host-like protein which suggests it may be a tail accessory protein which interacts, during binding, with Roseobacter-specific cell surface proteins

6.2.5 Structural protein gp33

gp33 was present in many of the SDS-PAGE bands along with three N4-like structural proteins gp25, the portal protein, gp28, the major capsid protein and gp30, however, gp33 appears to be specific to RN4-like phages as it has homologues in phages DSS3Φ2 and EE36Φ1. Domain prediction search programs carried out on gp33 revealed possible similarities to protein/DNA structural proteins often found in nucleosome core particles and to cell wall hydrolases, a common feature on phage structural proteins. Interestingly, the homologues of gp33 in DSS3Φ2 and EE36Φ1, gp61 and gp58 respectively, share a partial C-terminal sequence with another (non N4-like) Roseobacter phage SIO1. Consequently, the Roseobacter protein with which

these five phages interact may be the same and also suggests gp33 in RLP1/RPP1, gp61 in DSS3Φ2, gp58 in EE36Φ1 and gp24 in SIO1 are all involved in host recognition and/or phage binding.

6.2.6 Structural proteins gp64 – 71

Proteins gp64, 65, 68 and 71 all fall outside the expected N4-like structural module (see Section 5.2.11) and are likely expressed either late amongst the early genes or during the middle transcription phase of the infection cycle based on their genomic location after the RNA polymerase-containing early gene module (see Fig. 5.12). A known structural protein that occupies a similar position in N4 is gp17 a decorating protein, which is also found downstream of the RNAP2 gene. gp17 is thought to contain three IgG domains which though not essential to infectivity, act to stabilise the capsid (Choi *et al.*, 2008). Many Ig-like domain-containing proteins have been found in dsDNA phages and have been postulated to act as aids in the initial bacterial cell surface interaction probably by binding to carbohydrates (Fraser *et al.*, 2007). Unfortunately, none of the structural proteins identified by mass spectrometry appear to contain any recognisable Ig-like domain.

BLAST analysis shows gp64 shares some similarity to an abundant virion protein, gp230 in *Pseudomonas* myovrius 201phi2-1, which in turn is a fusion of homologues of phiKZ gp145 and gp146 both tail proteins. This suggests that gps 64 – 71 represent a putative tail module. Interestingly, gps 68 and 71 both contain protein chaperone-like domains which, if these gene products are component of the tail structure, would be essential in the transport of the 3500 aa vRNAP out of the virion head into the host cell (see Section 6.1). In addition, analysis of gp64 by DomSSEA also identified three hydrolase/tail associated lysozyme in T4 domains. Such domains would probably play a key role in phage binding in particular in the penetration of the host cell wall possibly similar to that of phage T4 as described in Section 6.1 and Rossmann *et al.* (2004).

Another interesting feature of gp64 is that it much larger than its homologues in DSS3Φ2 and EE36Φ1, gp33 and it should also be noted that these two genes are not homologues of each other nor of gp230 in *Pseudomonas* myovirus 201phi2-1. Comparison of the amino acid sequence of these three genes shows that the DSS3Φ2 gp33 aligns with the N-terminal portion of gp64 with fairly large gaps, whilst the

EE36Φ1 gp33 aligns with the C-terminus with no gaps, see Fig 6.6. As such it may be that gp64 is a fusion of the two genes in DSS3Φ2 and EE36Φ1 or the similarities seen between gp64 and gp33 from DSS3Φ2 are not significant and the RPP1 protein only has a homologue in EE36Φ1.

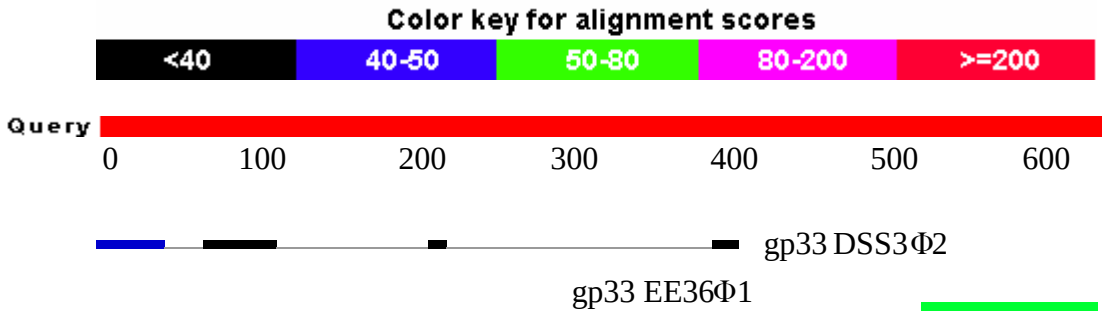


Figure 6.6 Alignment of gps 33 from DSS3Φ2 and EE36Φ1 against gp64 from RPP1. Made using bl2seq.

gp65, (like gp16) is a host-like protein whose homologue can be found in *Roseobacter* sp. AzwK-3b; this protein is believed to be the periplasmic component of an ABC-type oligopeptide transport system. As with gp16, this phage protein may facilitate adsorption to the host cell wall through interaction with the other components of the bacterial transport system.

It should also be noted that though gps 64 -71 are found in both the RLP1 and RPP1 genomes, they were only identified in the RPP1 protein gel. However, as RLP1 and RPP1 are highly similar it is likely that their homologues in RLP1 also encode structural proteins and were missed in the protein gel due to bad sample preparation.

6.3 Concluding comments

Conspicuous by its absence in the phage proteins identified by mass spectrometry was vRNAP; this protein unique to the N4-like genus, is thought to exist at around four copies per virion (Choi *et al.*, 2008). In contrast, a qualitative study of the proteome of a N4-like phage, LIT1, identified 76 peptides of the vRNAP homologue (Ceyssens, 2009a). Another concern was that many of the bands identified by mass spectrometry only had one peptide hit reducing the confidence in the results. Furthermore, in these bands, the predicted molecular weights of the proteins identified did not match the observed mass. In addition, the bands from the RLP1 phage gel

included many contaminating host proteins despite double CsCl purification prior to protein extraction. These results clearly demonstrate the need for a revised protocol for phage preparation, protein extraction and perhaps also mass spectrometric identification (an alternative method such as MS/MS could be utilised).

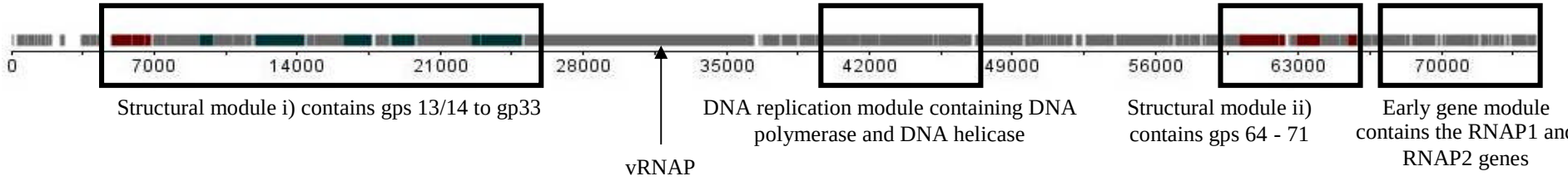
Despite these problems it can be concluded that phages RLP1 and RPP1 share five of their major structural proteins (e.g. the portal protein gp59 and the MCP gp56) with the enterobacteria phage N4. The remaining structural proteins identified are shared with the Roseobacter phages DSS3P1 and EE36Φ1 which previous to this project were not identified as structural proteins, in particular those found after the early gene module (gps 64 - 71).

As shown in Fig. 6.6 it appears that the structural proteins have been split into two clusters though the reason for this is not yet clear. Their positions in the two phage genomes indicate that the two modules are probably expressed with the late early/middle and late cohort of transcripts respectively. This may point to a gene regulation requirement and a possibility that the gp64 -71 module proteins require maturation prior to assembly on the virion, and/or they are involved in the initial steps of pro-capsid formation. In general, the constituent parts of phage virions particles i.e. the heads, tails and tail fibres, are made separately via subassembly pathways rather than a single linear pathway. Upon completion of the virion segment, the heads and tails combine first, forming complexes that are visible by electron microscopy, then the distal tail fibres are added (Campbell, 2007). It is conceivable that structurally complex tail portion of the virion involves many steps and perhaps the assistance of helper proteins whilst the head is relatively simple to construct. Consequently, the tail genes are expressed earlier than the MCP, the portal protein and other tail fibre proteins.

When comparing the relative positions (in the genome) of the structural proteins with the other members of the N4-like phage genus is it interesting to note that in the *Pseudomonas* phage LIT1 a similar extra putative tail module has been identified (Ceyssens, 2009a). However, in this phage the module is in reverse orientation to the surrounding genes and appears in a different location, downstream of the DNA replication genes DNA helicase and DNA polymerase, and so is likely expressed with the middle genes compared to the late early/middle expression of RLP1/RPP1. Despite these differences, this demonstrates a major gene rearrangement

Figure 6.6 Location of structural proteins on RPP1.

The location of structural proteins on RLP1 are similar to that of RPP1 though it should be noted that gp 64- 71 were not identified by mass spectrometry in the RLP1 protein gel. N4-like genes are highlighted in green, RN4-like genes in red.



that has occurred between the *Pseudomonas* N4-like, the RN4-like phages and N4, a change which likely determines the host range of the phage.

Through BLAST and domain analysis, putative functions of the MS/MS identified structural proteins may now be assigned. Gps 19, 25, 28, 30, 32 and 33 are core genes with homologues in all N4-like phages sequenced to date; they likely form the foundations upon which other host-specific structural proteins may attach. gp33, is likely another foundation-type protein as it is expressed alongside the N4 homologues. The known N4 tail genes gp 65 and 66 (tail sheath and appendage protein respectively) are missing in the RN4-like phages and are replaced by gps 64, 65, 68 and 71 which collectively form a tail gene module expressed early in the infection cycle. Finally, gps 13/14, 15 and 16 are putative tail fibre genes responsible for binding of the phage virion to the host cell.

Though the function of many of the shared Roseobacter phage proteins cannot yet be fully elucidated, many of their domains (as identified by DomSSEA) are often found conserved in the *Caudovirales* class (Fraser *et al.*, 2007) which begs the questions: what are the mechanisms by which they have acquired them, are they conserved between host:phage groupings and are these domains limited to the tailed phages? In light of these results it would appear that in addition to improving the genomic characterisation of RLP1 and RPP1, the identification of their structural proteins by mass spectrometry has opened up a new area of research into the elements required for the binding of both Roseobacter-specific and general tailed phages. As such, these results provide a starting point for more detailed investigations in the future.

Chapter 7

Induction and characterisation of temperate phages from Roseobacter hosts

7.1 Introduction

In 1921, Bordet and Ciuca found that certain strains of *Escherichia coli* often produced phages and in the same year Gildemeister also found a similar phenomenon in filtrates of faeces (Bordet and Ciuca, 1921; Gildemeister, 1921). Initially researchers believed this was due to viral contamination, but attempts to obtain pure, phage-free cultures failed and so these phage-producing bacteria were termed lysogenic: able to cause lysis (Bordet and Ciuca, 1921). Though this finding was initially contested (most notability by d'Herelle (d'Herelle, 1922)), a few years later in 1925, both Bordet and Bail separately proved that the bacteria were indeed the source of the phages as every bacterium in a lysogenic strain could produce a lysogenic colony whilst clones from phage contaminated cultures did not (Bordet, 1925; Bail, 1925). Bordet went further and concluded that the phage was lodged in "the hereditary weft of the bacterium" (Bordet, 1925).

Though lysogeny is usually stably maintained, a proportion of lysogenic bacteria (within a clonal population) can spontaneously lyse and produce phages. Consequently, after the discovery of lysogeny, much effort was directed towards finding external factors that could induce a lysogenic bacterium into producing phages. Unfortunately the answers were not forthcoming and it took several decades until such a trigger was found. In 1950, Lwoff, Siminovitch and Kjeldgaard found that DNA damage caused by UV irradiation could induce phages from lysogenic cultures of *Bacillus megaterium* (Lwoff *et al.*, 1950). In their subsequent work in Lwoff's laboratory, Jacob and Monod hypothesized the regulatory mechanism governing the prophage switch was similar to the control of activity of nuclear genes (Jacob and Monod, 1961). Since then, lysogeny has been identified as a widely distributed phenomenon throughout the bacterial kingdom but, most of what is known about the biology and molecular basis of the interaction between host and temperate phage, in particular the prophage switch, is derived from coliphage λ and its relatives. In particular, little is known about lysogeny and the lysogenic decision in marine bacteria.

As discussed in Section 1.4.6, due to the high rates of phage decay and low concentrations of slow-growing host bacteria, lysogeny is thought to be prevalent in the marine habitat. Indeed, high proportions, 43%, of cultivable, heterotrophic marine

bacterial isolates in Tampa Bay, Florida were found to contain inducible prophage-like particles and induction appeared to occur more frequently in coastal/estuarine environments. (Jiang and Paul 1994; 1996). In the same study, environmentally relevant pollutants such as fuel oil and naphthalene were all found to induce prophages. (For further information see Section 1.4.6.)

The advantages for the host to maintain a metabolically expensive phage element are not immediately obvious; however, the ability of prophages to enhance host fitness has been a widely accepted fact for 35 years (Edlin *et al.*, 1975). So far seven mechanisms by which prophage genes can improve host fitness have been suggested (Barnodess and Beckwith, 1995; Brussow *et al.*, 2004; Paul, 1998):

1. prophage elements can contain genes which encode fitness-enhancing functions e.g. *bor* gene of phage λ (Barnodess and Beckwith, 1995)
2. elements can serve as anchor points for gene rearrangements
3. prophage insertion into the genome can disrupt non-essential gene functions resulting in a net decrease in the host's metabolic load
4. phage infection can infer homoimmunity (the ability to resist re-infection by other phages)
5. induction and release of temperate phage can affect closely related host strains; if these are lysed, competition for nutrients is reduced (kill the relatives)
6. transduction and/or conversion can occur through the introduction of new genes and
7. essential genes may be down-regulated by phage repressors (see Section 1.5.3)

As nutrient limitation in the marine environment is a constant challenge, it is perhaps not surprising that many marine bacteria appear to be lysogens when such advantages can be gained from phage insertion.

In contrast, from a phage's perspective the advantages of lysogeny are clear; protection from several mechanisms of inactivation that a phage is subject to in the marine environment. These include DNA-damaging UV radiation, proteolytic digestion and grazing as well as the ability to survive in times of low host abundance. The latter is of particular import for marine phages as although hosts are often present, they are found in relatively low abundance for most of the year until conditions become favourable and they bloom (Brown *et al.*, 2005). Indeed a study by McDaniel *et al.* (2002) found that primary productivity of natural *Synechococcus* populations in

Tampa Bay appeared to have an inverse relationship with the incidence of lysogeny. They surmised that during conditions unfavourable to autotrophic growth, lysogeny was the preferred option for *Synechococcus* phages (McDaniel *et al.*, 2002).

To date, two members of the Roseobacter lineages have been identified as lysogens: *Silicibacter* sp. strain TM1040 and *Roseovarius nubinhibens* ISM (Chen *et al.*, 2006; Zhao *et al.*, 2010). The search for lysogens amongst the Warwick Roseobacter culture collection is described below.

7.2 Results and Discussion

7.2.1 *In silico* analysis of Roseobacter species

The web based PHP application, Prophage Finder (Bose and Barber, 2006) was used to predict if *Roseovarius* 217, *Roseovarius nubinhibens* ISM and *Roseobacter denitrificans* (all of which were sequenced as of Dec 2006) harboured prophages, the results are shown in Table 7.1 (for full description of the BLAST hits see Appendix Table A.3)

Table 7.1 Summary of Prophage Finder results. The predicted prophages in *Rsv. nubinhibens* do not match the prophage induced by Zhao *et al.*2010.

Roseobacter spp.	Predicted prophage	Number of BLAST hits
Roseovarius 217	1	6
	2	5
	3	5
	4	10
	5	8
	6	20
Roseovarius nubinhibens	1	6
	2	6
	3	9
Roseobacter denitrificans	1	6
	2	6

In order to determine if these predicted prophages were real prophages, the results from Prophage Finder were analysed according to 3 criteria:

1. Number of BLAST hits: the greater the number of hits, the more likely the predicted prophage is an actual prophage. Bose and Barber, the authors of Prophage Finder (Bose and Barber, 2006), suggest predicted prophages with greater than ten hits are generally proved to be actual prophages.
2. Hit descriptions: tail, integrase, portal, protease, capsid, terminase, tape measure, methylase, methyltransferase, packing and helicase proteins mediate key processes in the lysogenic phage replication. Presence of two or more of these genes provides further support for a predicted prophage being an actual prophage.
3. GC content: Deviations from the GC content of the host sequence may suggest that a predicted prophage is real.

Those that did not fit the criteria are likely to be prophage relics, i.e. genes left behind during past infections by temperate phages.

The most promising candidates were predicted prophages 4 and 6 of *Rsv.* 217 shown in detail in Table 7.2 as based on the criteria above, these predictions appear likely to be real prophages. Prophage 4 has hits to genes from seven different phages, but never more than two from each phage, whilst prophage 6 has hits to eleven different phages. Intriguingly, it has four hits to both *Bacillus* phage Bcepμ and *Pseudomonas* phage B3. However, the validity of both these predicted prophages could not be determined conclusively based purely on an *in silico* analysis. Consequently, *Rsv.* 217 was one of the first Roseobacter spp. tested for induction of phage-like particles after exposure to Mitomycin C.

Table 7.2 Summary of Prophage Finder results for *Rsv.* 217 predicted prophages 4 and 6

Prophage	Host GC content	Predicted prophage GC content	Best BLAST hit	Phage origin of hit	Accession number	E value
Rsv. 217 predicted prophage 4	60.9	57.6	integrase	Enterobacteria phage KH97	NP_037720	5.00E-05
			tail length tape measure protein	*Xanthomonas* phage XP10	NP_858965	5.00E-07
			tail length tape measure protein	Enterobacteria phage HK022	NP_037676	5.00E-05
			terminase large subunit	Enterobacteria phage HK97	NP_037698	2.00E-44
			putative terminase (small subunit)	*Burkholderia* phage ΦE125	NP_536357	3.00E-09
			gp9	Enterobacteria phage HK022	NP_037670	0.15

Table 7.2 cont.

Prophage	Host GC content	Predicted prophage GC content	Best BLAST hit	Phage origin of hit	Accession number	E value
Rsv. 217 predicted prophage 4	60.9	57.6	ORF19	*Bacillus* phage Φ105	NP_690803	5.00E-07
			putative major capsid protein	Enterobacteria phage ΦP27	NP_543092	3.00E-38
			putative prohead protease	Enterobacteria phage ΦP27	NP_543091	3.00E-13
			gp3	*Burkholderia* phage Φ1026b	NP_945033	7.00E-34
Rsv. 217 predicted prophage 6	60.9	64.5	cI repressor protein	*Pseudomonas* phage D3	NP_061565	2.00E-08
			DNA transposition protein	*Pseudomonas* phage D3112	NP_938214	2.00E-43
			gp5	*Burkholderia* phage Bcepμ	YP_024678	3.00E-36
			hypothetical protein, p16	Enterobacteria phage μ	NP_050620	6.00E-05
			ORF19	*Vibrio* phage VHML	NP_758912	2.00E-06
			hypothetical protein, p26	Enterobacteria phage μ	NP_050630	8.00E-05
			gp27	*Burkholderia* phage Bcepμ	YP_024700	1.00E-05
			gp28	*Burkholderia* phage Bcepμ	YP_024701	2.00E-95
			portal protein	*Pseudomonas* phage B3	YP_164068	4.00E-98
			hypothetical protein, ORF33	*Pseudomonas* phage B3	YP_164069	2.00E-34
			hypothetical protein, ORF35	*Pseudomonas* phage B3	YP_164071	4.00E-13
			gp32	*Burkholderia* phage Bcepμ	YP_024705	1.00E-51
			capsid protein	*Pseudomonas* phage B3	YP_164075	4.00E-60
			unknown, p9	Enterobacteria phage SfV	NP_599041	7.00E-04
			ORF53	*Pseudomonas* phage D3	NP_061549	0.19
			putative tape measure protein	*Mycobacterium* phage TM4	NP_569753	0.15
			hypothetical protein, p19	*Yersinia* phage PY54	NP_892065	0.004
			tail protein	*Yersinia* phage PY54	NP_892067	0.03
			gp20	*Burkholderia* phage Φ1026b	NP_945050	1.00E-04
			gp21	*Klebsiella* phage ΦKO2	YP_006601	8.00E-06

7.2.2 Induction of phage-like particles by exposure to Mitomycin C

As outlined in Section 2.6.17.1 the growth of Mitomycin C-treated and an untreated culture were compared over at least 24 hours through measurement of optical density at 600 nm. It was found that all treated cultures had a reduction in growth, as expected after exposure to a toxic chemical, but putative lysogens had a marked decrease, see Fig. 7.1.

Both *in silico* analysis and comparison of the growth of treated and untreated cultures through absorbance readings (an indicator of cell density), strongly suggested that *Rsv.* 217 could be a lysogen, see Fig. 7.2. However, no phage/viral-like particles were observed during confirmation of induction when the bacterial lysate, post-exposure, was stained with SYBR-Green and viewed under an epifluorescent microscope.

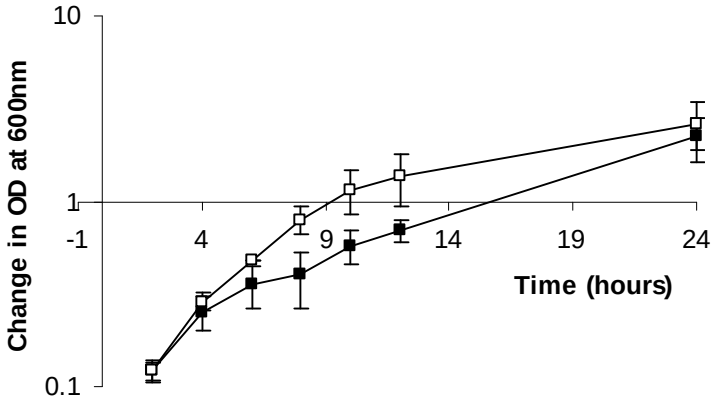


Figure 7.1 Comparison of treated (■) and untreated (□) *Roseovarius* 217 cultures measured by change in optical density at 600 nm.

Though no phages were induced after exposure, *Rsv.* 217 may still be a lysogen as not all prophages are induced by treatment with Mitomycin C. However, the temperate phage(s) must be highly stable as none of the *Rsv.* 217 lysates showed evidence of spontaneous induction.

Mitomycin C exposure did however, identify three other lysogens in the Warwick Roseobacter culture collection: *Marinovum algicola*, "*Ruegeria*" sp. 198 and ACR04, see Fig. 7.2.

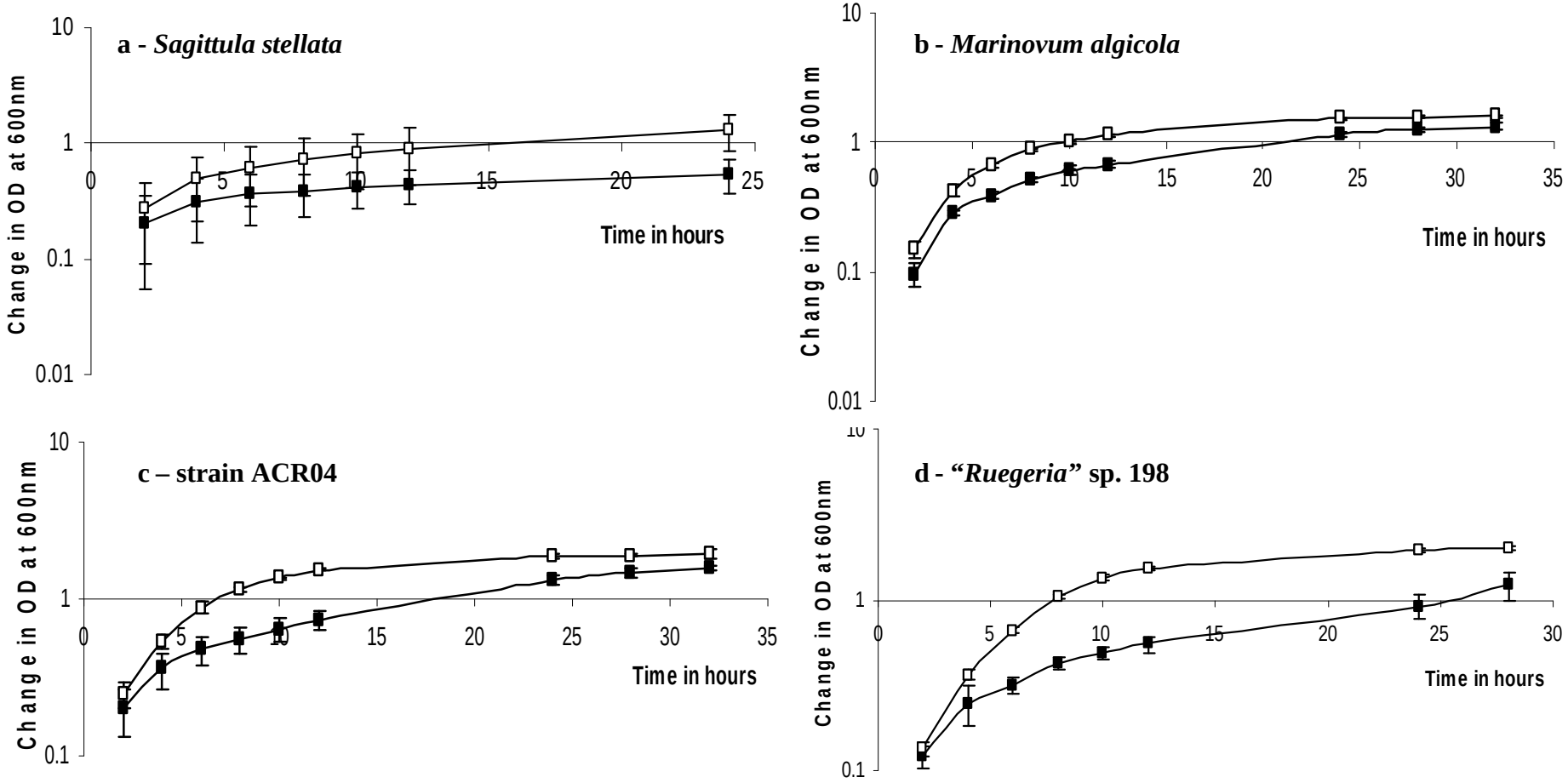


Figure 7.2 Comparison of Mitomycin C-treated (■) and untreated (□) Roseobacter cultures assessed by measuring change in optical density at 600 nm over time.

a) *Sagittula stellata* (not a lysogen), b) *Marinovum algicola,* c) ACR04 and d) "*Ruegeria*" sp. 198. Unlike panels b, d and c, *Sagittula stellata* did not show a marked reduction in growth after exposure to Mitomycin C unlike *M. algicola*, "*Rugeria*" sp. 198 and ACR04.

7.2.3 Confirmation of induction

7.2.3.1 Epifluorescence microscopy

Confirmation of prophage induction from the putative lysogens, *Marinovum algicola*, "*Ruegeria*" sp. 198 and ACR04, was carried out by filtering samples from time points 0, 6 and 12 hours onto a 0.02 µm-pore-size filter then staining the immobilized bacterial and potential induced phages with SYBR green (as described in Section 2.6.17.2). The filters were subsequently examined in an epifluorescence microscope. Virus-like particles (VLPs) appeared as dots or pin-pricks on the filter compared to the larger bacterial cells. Lysogens were confirmed by the presence of VLPs which were observed to increased in number over time, see Fig 7.3. This was confirmed by examination of 5-10 random fields of view and counting the number of VLPs present.

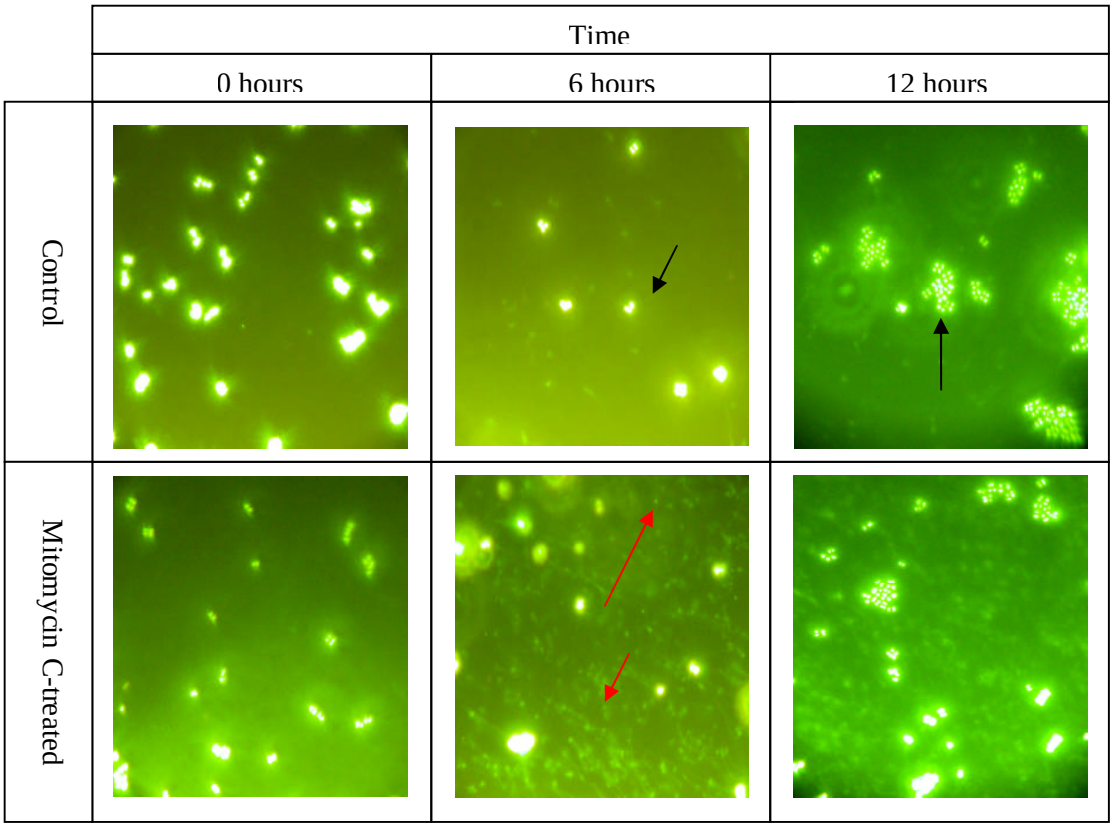


Figure 7.3 SYBR Green I stained samples of Mitomycin C-treated and untreated/control ACR04 cultures. Black arrows indicate bacterial cells, red arrows indicate suspected induced phages. x 5000 Magnification

It was observed that in all the 0 hour samples, no VLPs could be found in both the Mitomycin C–treated culture and the control culture. Six hours post-induction, there was *ca.* 5-10 fold increase in VLP in comparison to the control, for all cultures (*M. algicola,* ACR05 and "*Rugeria*" 198) treated with Mitomycin C. In the 12 hour sample, there was *ca.* 10-15 fold increase in VLP in the treated compared to the un-treated control.

Threads of stained DNA were also observed during epifluorescence microscopy which was thought to be DNA released during cell lysis and stained by the non-specific SYBR Green I. To test this theory, samples were treated with DNase (see Section 2.8.14) and compared to undigested aliquots, Fig 7.4.

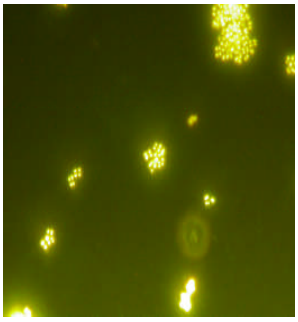
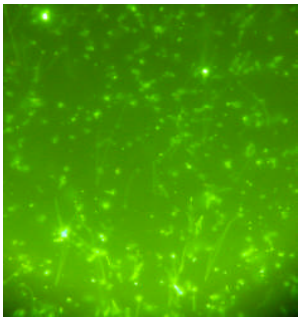
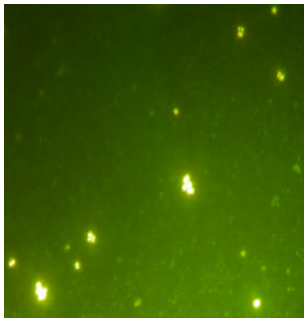
	a	b	c
DNase	+	-	+
Mitomycin C	-	+	+
			

Figure 7.4 SYBR Green I stained samples of 12 hr ACR04 cultures. a) DNase digested, control sample, b) undigested and exposed to Mitomycin C, c) DNase digested and exposed to Mitomycin C.

The disappearance of the threads illustrates the amount of DNA released into the bacterial media during lysis. In the environment, this may be advantageous to the bacterial community as lateral gene transfer through transformation, uptake of exogenous DNA, may provide a fitness increase to transformants.

7.2.3.2 TEM

Further confirmation of prophage induction was carried out by electron microscopy (see Section 2.6.11) and the sizes of the virion particles determined, see Table 7.3. TEM also allowed the morphology of the phage to be elucidated, Fig 7.5.

Micrographs indicated that all the prophages induced belonged to the *Siphoviridae* family characterised by their long, flexible tail structures.

Table 7.3 Virion sizes of induced prophages.

Prophage induced from	Head (nm)	Tail (nm)	Micrograph
Marinovum algicola	61.7 ± 3.3	115 ± 2	Fig. 7.5 a)
Marinovum algicola	61.7 ± 3.3	144 ± 2	Fig. 7.5 a)
ACR04	48.8 ± 8	151 ± 18	Fig. 7.5 b)
ACR04	48.8 ± 8	202 ±17	Fig. 7.5 b)
"*Ruegeria*" 198	42.6 ± 1	236 ± 9	Fig. 7.5 c)

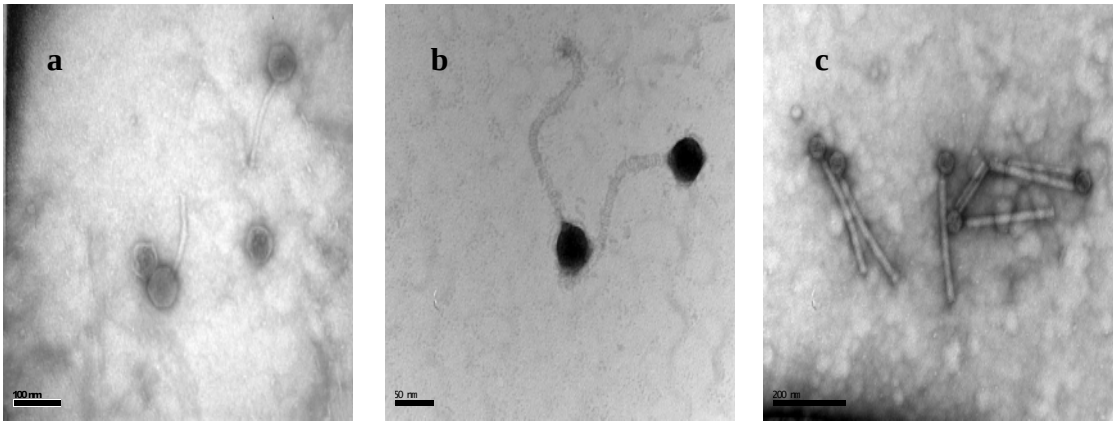


Figure 7.5 Electron micrographs of induced prophage from a) *Marinovum algicola,* b) ACR04 and c) "*Ruegeria*" sp. 198. Purified phage samples were stained with Uranyl acetate and examined by TEM. Images were taken using a Gatan camera and subsequently process with DigitalMicrograph™.

Interestingly, during analysis to determine the size of virion particles, samples from *Marinovum algicola* and ACR04 showed two distinct morphologies which suggested these two lysogens contain two prophage elements. Supporting this theory was the presence of two phage bands in CsCl gradients, Fig 7.6.

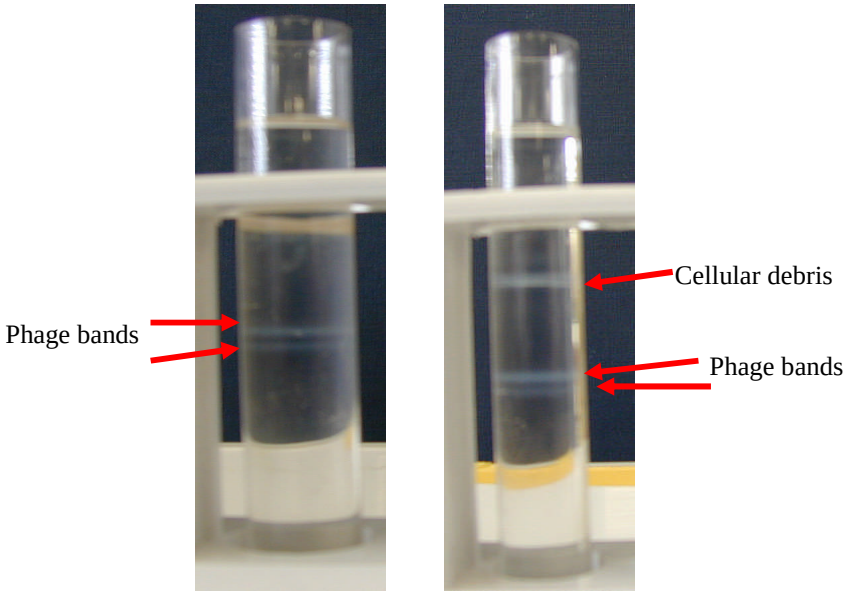


Figure 7.6 Caesium Chloride gradients of induced prophages from *Marinovum algicola* and ACR04. A 1 L mid-exponential phase culture of known lysogen was exposed to Mitomycin C (final concentration, 0.5 $\mu g ml^{-1}$) for 30 min then incubated for eight hours. Bacterial cells were removed by centrifugation and the phage particles purified by treatment with PEG (final concentration. 10%) overnight at 4 °C. The phage lysate was then further purified by CsCl isopycnic centrifugation. The results post-centrifugation are shown in the images above; two whitish phage bands were present in both tubes.

7.2.4 Pulsed field gel electrophoresis

To determine the genomes sizes of the induced prophages and to confirm the hypothesis that *M. algicola* and ACR04 contain two prophage elements, PFGE was carried out on CsCl-purified samples, see Fig. 7.7. Unfortunately, the samples from induced "*Ruegeria*" 198 did not produce clear bands and so the genome size remains unknown.

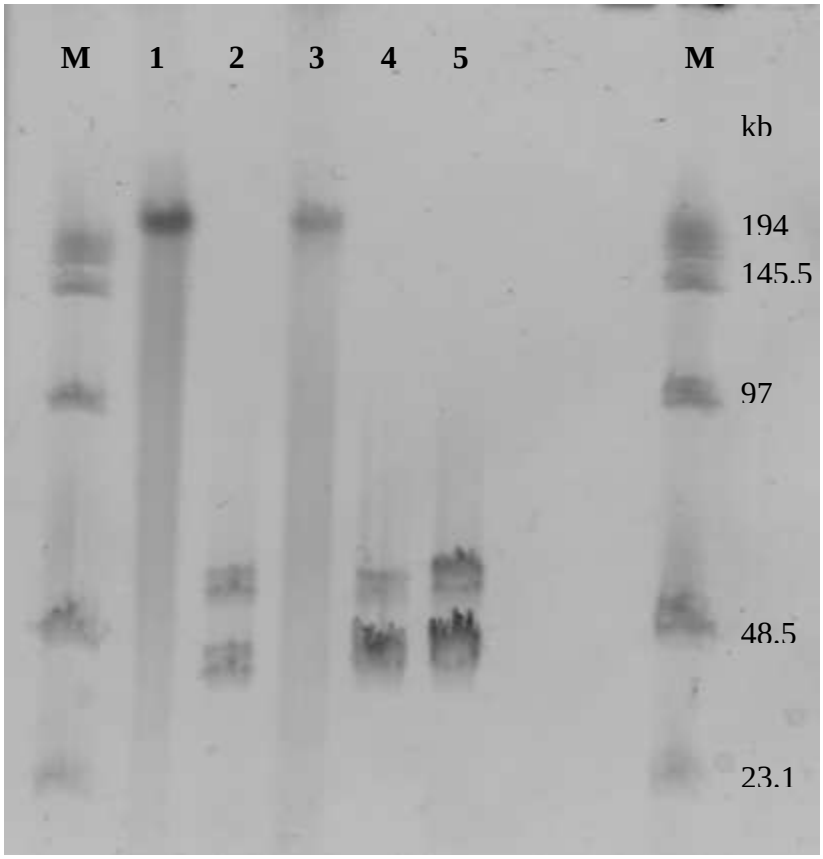


Figure 7.7 PFGE of lysogens and purified induced prophage. Lane 1 – *M. algicola,* 2 – induced *M. algicola* prophages, 3 – ACR04, 4 – induced ACR04 prophages, 5 – induced prophages from both *M. algicola* and ACR04 (two samples combined into one plug).

The results from the PFGE show that there are indeed two prophage elements present in *M. algicola* and ACR04. Table 7.4 shows the assigned names and genomes sizes for the five isolated temperate phages. It should be noted that virion particles found in TEM analysis have not been matched to a genome.

Table 7.4 List of temperate phages and the lysogen from which they were induced.

Prophage induced from	Name	Genome size
Marinovum algicola	vB_Ma_MWS1 (Marinovum Warwick Siphovirus)	43
Marinovum algicola	vB_Ma_MWS2	27.5
ACR04	vB_ACR4_RWS3 (Roseobacter Warwick Siphovirus)	46
ACR04	vB_ACR4_RSW4	33.5
"*Ruegeria*" 198	vB_R198_RSW5 (Ruegeria Warwick Siphovirus)	-

7.2.5 Host range

Purified samples of the five isolated temperate phages were tested against the Warwick Roseobacter culture collection; neither plaque assay nor spot tests displayed plaques. Four possible explanations can be deduced from this:

1. None of the species tested were susceptible
2. The induced prophages are not capable of infection due to loss of key genes
3. Some species were susceptible but the induced prophages lysogenize at near 100% efficiency and so plaques would not be observed
4. The species tested contained non-inducible (on exposure to Mitomycin C) prophage elements so were protected against re-infection, i.e. homoimmunity.

7.2.6 Restriction digest pattern

Extracted DNA from the prophages induced from *M. algicola* and ACR04 were digested with common restriction enzymes to determine their digest pattern, see Fig. 7.8. Unfortunately, the samples contain both temperate phages as efforts to fully segregate the phages by CsCl gradients proved unsuccessful.

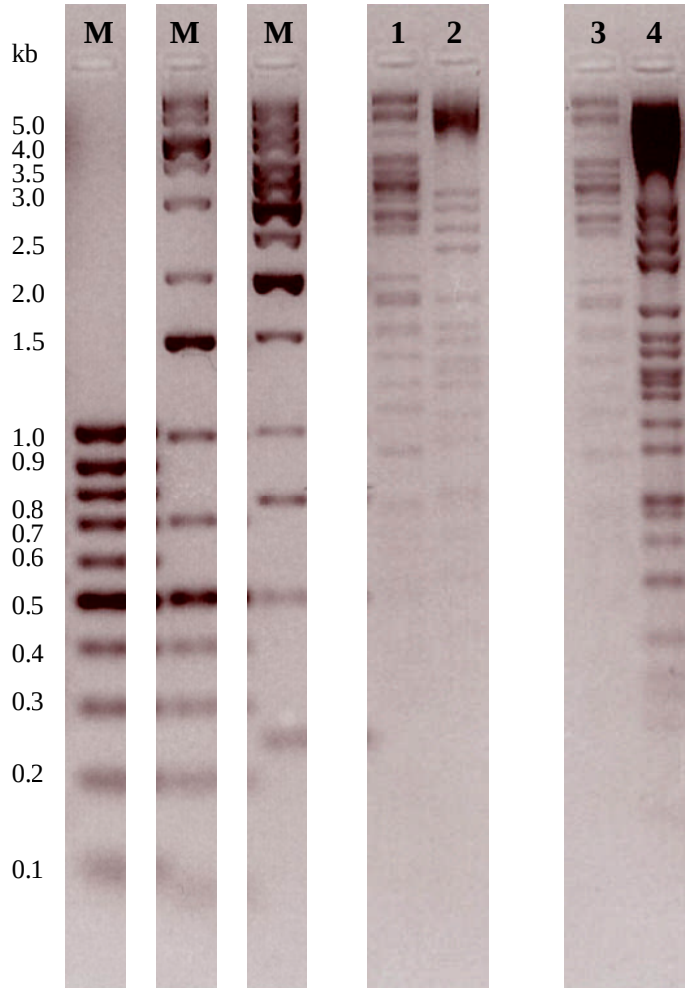


Figure 7.8 Restriction enzyme digest of prophages induced from *M. algicola* and ACR04. Lane 1 & 2 – *M. algicola* induced phages digested with *Hind*III and *Eco*RI respectively, Lane 3 & 4 – ACR04 induced phages digested with *Hind*III and *Eco*RI respectively.

7.3 Concluding comments

The results from the induction experiments show that 20% of the Warwick Roseobacter culture collection are lysogens. All are coastal isolates and though *Marinovum algicola* was isolated from the phycosphere of the dinoflagellate *Prorocentrum lima* (Lafay *et al.*, 1995), its symbiotic host is an estuarine species. None of the predicted prophages in *Rsv.* 217 or *Rsv. nubinhibens* were induced by Mitomycin C despite strong evidence supporting the presence of predicted prophages 4 and 6 in *Rsv.* 217 (see Table 7.2). However, it is still possible that they are actual prophages and not prophage relics. As much of what is known about prophage

induction is based on phage λ, the induction trigger for these prophages may yet have to be identified.

It is also interesting and slightly alarming to note that *Roseovarius nubinhibens* ISM was amongst those screened and no prophages were found to be induced. Contradictory to this observation, a study by Zhao *et al.* (2010) did find *Rsv. nubinhibens* produced VLPs in the lysate following Mitomycin C exposure. This may have been due to a loss of the prophage element during repeated rounds of culturing in the separate laboratories. Though many advantages to maintenance of prophages by lysogens were outlined in the introduction to this chapter, prolonged conservation of the prophage during the optimal growth conditions experienced in a laboratory may have proved too costly for the Warwick strain and so it may have been selected against. Another intriguing finding is that the "hidden" prophage induced by Zhao *et al.* (2010) is not one of the three predicted by the *in silico* analysis in Section 7.2.1. This illustrates the differences between laboratory strains of bacterial species and the impact of such changes on a given bacterium and its biological features.

In recent paper by Pradella *et al.* (2010), the extrachromosomal DNA content of four strains of *Marinovum algicola* were analysed. They found that this Roseobacter group species had unusually high numbers of extrachromosonal replicons, between 9 -12, of which one, a 34 kb replicon in *M. algicola* DG989, was suggested to be a linear plasmid-like prophage. Interestingly, the paper also found that the type strain, DSM10251, which was used in the induction experiments above, had a 46 kb plasmid which was absent in some DNA preparations which suggests that it is a prophage present in some of the laboratory DSM10251 culture. Due to their comparable sizes, it is possible that this plasmid represents the prophage, MWS1, identified in this study. No plasmid of similar size to MWS2 was identified in any of the strains examined by Pradella *et al.* (2010) which suggests that this prophage may be integrated into the main bacterial chromosome.

Unfortunately, due to the difficulties experienced during attempts to isolate pure samples of prophages from *Marinovum algicola* and ACR04, none of the induced prophages were sequenced. As a result, the genomic organisation (linear or circular) of the four prophages, in particular MWS1 and MWS2, could not be determined, nor could the mechanisms responsible for prophage repression and subsequent induction be elucidated. Instead only very basic characterisation into the five induced prophages were performed and so much future work must be carried out.

Chapter 8

General discussion

8.1 General discussion

The primary objective of this research was to isolate and characterise new bacteriophages which could infect members of an abundant group of marine bacterioplankton, the Roseobacter clade. When this project was started in 2006 there was only one lytic Roseobacter phage, SI01 (Rohwer *et al.*, 2000) and three lysogenic phages induced from *Silicibacter* sp. TM1040 (Chen *et al.*, 2006). However, these three phages were never shown to be fully functional phages capable of infection.

During the course of this research project, two new *Roseovarius* phages, RLP1 and RPP1 were isolated and characterised. Genome sequencing showed them to belong to the N4-like genus of Podoviruses. The two phages also proved to be atypical compared to previously isolated Roseobacter phages and other phages in general, as they required a solid culture condition for optimal infection. In addition five prophages, from three of the Roseobacter species in the Warwick culture collection, were found to be inducible upon exposure to the DNA-damaging agent Mitomycin C. The prophages were all identified as belonging to the *Siphoviridae* family by their morphology, however, it was not possible to confirm their infective potential as no susceptible hosts were found.

8.1.1 Genome sequencing of two new N4-like phages: implications for the N4-like genus

During the time span of this project a further four lytic N4-like phages were isolated elsewhere and had their genomes sequenced and published; DSS3Φ2 and EE36Φ1, which infect Roseobacter species *Ruegeria pomeroyi* DSS-3 and *Sulfitobacter* sp. EE-36 respectively (Zhao *et al.*, 2009) and LUZ7 and LIT1, phages of the opportunistic pathogen *Pseudomonas aeruginosa* (Ceyssens, 2009). One of the most intriguing results that has emerged from comparative studies of these seven genomes is that phages in the N4-like genus appear to have bipartite genomes consisting of a conserved core set of 23 genes (see Section 5.2.13 and Table 5.7) coupled with variable, often novel, peripheral assemblage of genes; this genomic arrangement is also observed in the T4 superfamily (Krisch and Comeau, 2008). As mentioned previously, the core genes seem to broadly fall into three categories: DNA metabolism/replication, transcription control and structural proteins. Among the T4-

like Myoviruses, the conserved core consists of genes with DNA replication/recombination and structural functions (Kirsch and Comeau, 2008). However, the number of core genes varies according to the subset of phages considered. For example, there are 90 common core genes when "true" T-even (T4), pseudo T-even (RB49) and schizo T-even (Aeh1) are compared, but this number of core genes falls to 24 when the Exo T-even (SPM2) phages are included (Kirsch and Comeau, 2008; Filée *et al.*, 2006). With the N4-like phages, the subdivisions below genus level remain unclear as none of the newly discovered phages appears to be significantly more related to N4 than to any of the others; indeed they all have around 25 N4-like genes. (LIT1 shares 27 genes, LUZ7 - 22, DSS3Φ2 - 26 and EE25P1 - 25.) It is of interest to note when comparing the functions that the core genes perform in the T4 superfamily and the N4-like genus, transcriptional control is also conserved in the latter. Enterobacteria phage N4 has long been considered an oddity in the phage world as it uses three different DNA-dependent RNA polymerases, one for each phase of gene expression. Consequently, it is not surprising to find this unusual mechanism of gene control to be conserved within the genus.

Having observed this apparent stable association of core genes within a genus, it is logical to next ask why this has occurred. Again, it is possible to look to the T4 superfamily for possible answers. Filée *et al.* (2006) suggested that phages gain an evolutionary advantage when they maintain large regions of conserved sequence. These can act as type of genetic glue maintaining the genetic cohesion via recombination within the most conserved sequences and mediating the swapping of nonconserved, hyperplastic sequences that they flank. Another more obvious reason is, in the categories mentioned above, the processes involve the coordination of various phage-encoded proteins instead of phage/host units. These conserved polypeptides form multiprotein complexes whose precise structural organizations are critical to their ability to function correctly. Domain or gene swapping within such intricate assemblies would probably result in complete loss of function, a lethal outcome. It should be noted however, that in the Roseobacter phages and LIT1 the positions of the N4 gp14 and gp22 have been rearranged. Without a function for gp14 and only a putative role for gp22 (homing endonuclease), the final outcome of such a change is uncertain though it is clearly non-lethal.

8.1.2 Shared peripheral genes found in Roseobacter phages

In addition to these core genes, the four N4-like Roseobacter phages also share a number of genes, see Table 5.6, and so it would appear that some of the peripheral genes are Roseobacter phage specific. Though many cannot be assigned a function at present, it seems likely that these gene products interact directly with host proteins/machinery specific to the Roseobacter group. This is exemplified in Chapter 6 where all of the proteins identified by mass spectrometry have a homologue among the RN4-like phages. Though five are encoded by the previously mentioned core set of genes, the remaining seven are likely to mediate binding to cell surface receptors specific to the Roseobacter clade. Intriguingly, gp16 and gp 65 are also host-like genes which suggests that they have been spread by LGT and that their maintenance confers a fitness increase on RN4-like phages possibly by increasing their binding affinity to host receptor.

Other host genes shared by the RN4-like phages are the ribonucleoside diphosphate reductase (*rnr*) and thioredoxin (*trx*) genes; as discussed in Section 6.2.8 it appears that *trx* is shared with marine podoviruses whilst *rnr* is more closely related to those found in Roseobacter spp. In addition, *trx* appears to be in close proximity to the DNA polymerase (DNAP) and helicase genes which has been suggested to form a DNA replication unit in marine T7-like phages (Hardies *et al.*, 2003). In that paper, the authors noted that the DNAP from the phage VpV262 lacked the thioredoxin-binding domain found in the T7 enzyme; this was concluded to be due to a distinct function and evolutionary origin. Like the VpV262 protein, the DNAP of RLP1 and RPP1 lacks a thioredoxin-binding domain, but unlike the *Vibrio* phage they have a clear phage not bacterial origin. Consequently, it appears likely that only the *trx* gene rather than the whole module has been subject to LGT.

Among the RN4-like phages there appear to be 33 shared peripheral genes and a further 24 genes are shared between the *Roseovarius* N4-like phages. There are only 29 unique ORFs or ORFans, out of a total of 340 predicted genes which demonstrates the high degree of relatedness between these four phages. However, due to the scarcity of sequenced marine phages it remains unclear if these peripheral RN4-like genes were shared by vertical (i.e. only within the N4 genus) or horizontal (between other marine Podoviruses such as with the *trx* gene) means.

8.1.3 Definitions of phage genera and the evolution of tailed phages

With the expansion in the number of genomes for members of the *Caudovirales* it is becoming increasingly obvious that these genomes are highly mosaic in nature (Hendrix *et al.*, 1999). It is clear that phages have undergone large amounts of horizontal gene exchange between seemingly unrelated phages and it is also equally obvious that the phage, and indeed the whole viral universe, can be categorised into a finite number of virion structure based lineages, see Section 1.3. This has led to debates amongst phage biologists as to whether or not phage genera actually exist or whether there is instead a continuum in which all tailed phages dip into a hypothetical melting pot to find the genome that works best. The mosaic model proposed by Hendrix *et al.*, in 1999 poses the best compromise to this problem. On a very simplistic level this model proposes that early phages have exchanged large chunks of genetic information prior to the demarcation of the now accepted supergroups. Fine tuning of host/environment specific genes between close relatives then followed, the consequence of which are phages with genomes created from a mixture of vertical and horizontal gene transfer events (Hendrix *et al.*, 1999).

The results from this study fit in well with this theory; the core genes of the *Roseovarius* phages isolated here appear to be derived from ancient phages thus accounting for the degree of homology and gene synteny found in the terrestrial (N4, LIT1 and LUZ7) and marine (RLP1, RPP1, DSS3Φ2 and EE36Φ1) phages, whilst the host-interactive genes such as *trx*, *rnr* (Section 5.2.8) and the structural proteins (Chapter 6) have been acquired from more recent lateral gene transfers from both bacterial and viral sources. Thus it appears the Linnaeus method of classification and idea of distinct phage genera, defined as a "group of species sharing certain common characters" (Statute 3.26 in the ICTV's Code of Virus Classification and Nomenclature, 2002), still largely holds true.

8.1.4 What is a phage species?

Due to the unusually high degree of similarity between RLP1 and RPP1 it is logical to question if they are examples of two species or merely two strains of one. However, as discussed previously, if definition of a viral genus is considered a difficult task then this difficultly is increased even more when considering species delimitations. Indeed in his paper "Concept of virus species" (1992) van Regenmortel

compared it to delimiting a mountain as "we cannot state with certainty where Mont Blanc starts and stops we do not claim it does not exist because its limits are unclear. We should not try to make absolutely clear distinctions where none exist." The crux of the problem lies in the need, in classification, to draw abstract taxon boundaries across the continuous range of genetic and phenotypic variability found in the phage and viral world.

When considering this problem it is useful to clarify what a strain is; virologists generally consider a viral strain to be a biological variant that is recognized due to its possession of unique phenotypic characters. These include a) biological properties such as a symptom or host range b) chemical or antigenic properties and c) the underlying genome sequence that is known to be correlated with the phenotypic uniqueness of the strain (van Regenmortel, 2007). Among eukaryotic viruses, antigenic properties and geographic locations can be considered, such as the sub-divisions below the species level found with HIV-1 and HIV-2 (Damon *et al.*, 2004). However, in the phage world classification according to location is haphazard at best due to the often stated hypothesis that "everything is everywhere" (Baas Becking, 1934; De Wit and Bouvier, 2006). This may be particularly applicable in the marine environment as conditions are in constant flux and a considerable degree of mixing due to currents occurs. Consequently, a particular phage found in location A may also be found in location B, as long as the host is too.

In the case of RLP1 and RPP1, evidence against their integration into one species includes the difference in the date and location of sample seawater (from which the phages were isolated), the host range (Chapter 3), their binding patterns (Chapter 4) and the presence of number of ORFans unique to each phage (Chapter 5). On the opposing side, the argument for integration is mainly based on the unusually high (95-100%) degree of similarity found in the majority of predicted genes. In the plant virus genus *Begomovirus*, 177 species were demarcated on the basis of a pair-wise sequence identity of less than 89% (Fauquet *et al.*, 2003. However, as mentioned in Section 6.2.12.1, other highly related phages with above 90% identity have been reported previously so precedence, at least in the phage world, does exist. Therefore, in this study RLP1 and RPP1 have been considered as two phage species, though perhaps in the future to avoid conflict they should instead be referred to as two isolates.

8.1.5 Control of planktonic and sessile phenotypes

In this study a physiological response comparable to that of marine Vibrios, described in Section 5.3, during a change in growth conditions was observed in the two Roseobacter species, *Rsv.* 217 and *Rsv. nubinhibens.* Unusually, this phenomenon was recognised during a phage-centric project. As discussed in Section 4.3 it is not surprising to find that a phage has exploited this bacterial response by binding to a receptor that is expressed during sessile/attached conditions but not during planktonic growth conditions as it too is exploiting the major advantage gained by a stationary lifestyle, a plentiful source of nutrients/prey. However, this study has only (tentatively) identified the cellular response, the physiochemical trigger and the control mechanism that precede the response can only be hypothesized.

In Section 4.3, the possibility of an AHL-based trigger was discussed and is an appealing candidate based on the results in the study by Bruhn *et al.* (2007). However, as discussed in Section 1.1.5 and in the review by Geng and Belas (2010), it seems likely that quorum sensing only mediates gene expression after a sessile state has been established and the molecular trigger remains unknown. In *Vibrio parahaemolyticus,* it was shown using *lux* and lateral flagella (*laf*) fusion proteins that increases in the viscosity of the medium, mediated by the addition of polymers such as polyvinylpyrrolidone, triggered luminescence in the fusion strains (Belas *et al.*, 1986). It is also possible that a change in physical parameters such as viscosity triggers a molecular switch as these two models are not mutually exclusive.

In his chapter on genetic control of bacterial adhesion, Silverman *et al.* (1984) described two types of control mechanisms that could regulate bacterial responses at surfaces: responsive and variable. In a responsive system, the bacterium senses an environmental signal and responds accordingly; this is the case in *Vibrio parahaemolyticus* where all individuals in culture are either swimming, using a polar flagellum or swarming, through use of multiple, lateral flagella. In the variable system, individuals in a bacterial population constantly switch between the two states allowing an equilibrium to form in which either the planktonic or the surface-bound state is favoured. As shown in Chapter 4, Figure 4.2, a limited amount of host lysis does occur in liquid cultures, which favours the idea that in the *Roseovarius*/phage systems investigated here, the variable model is present, as a minority of individuals remain susceptible in the planktonic state. In contrast, in the responsive system bacteria

display an “all or nothing” phenotype. As such, more doubt is placed on an AHL-based trigger as quorum sensing is typically a community response with all individuals participating concurrently. However, these conclusions are highly speculative as it is based on preliminary data only.

8.1.6 Comparison of agar plates and biofilms

It is tempting to equate the agar polysaccharide matrix present in a double layer agar plate to a biofilm. Indeed in his book on Bacteriophages and Biofilms (2010), Abedon uses phage plaques as a model of phage propagation as it occurs within biofilms. However, this conjecture was phage-centric as the mechanisms of phage plaque formation had been explored theoretically and (to a lesser extent) experimentally. As such, the author concluded much could be learnt by considering a plaque on an agar plate to be a simple representation of infection in a biofilm. In this project the other player, the bacterial lawn, is the focus.

The experimental use of agar to entrap bacteria and simulate a biofilm was reported by Jouenne *et al.*, (1994). In their study they concluded that “artificial” immobilized-cell structures comprised of viable microorganisms entrapped in agar could serve as a simple *in vitro* model structure of natural biofilms. However, they also noted that alginate was more representative of biofilm EPS than agar. In this study, both agar and agarose were used, but not alginate. Nevertheless, agar has been used by biofilm researchers to simulate simplistic artificial biofilms and so parallels can be drawn. As such, it is not wishful thinking to believe that bacterial growth in a low percentage agar may in some way mimic growth in natural biofilms. Therefore, the plate-only receptor identified in Chapter 4 could be one of many proteins up-regulated and expressed when the bacterium is in a biofilm.

8.2 Future work and prospects

8.2.1 Phage-based studies

This investigation only represents an initial foray into the study of *Roseovarius* phages RLP1 and RPP1. There are several directions this can be taken further, however, two are of particular interest; definition of the genomic ends and identification of the “plate-only” phage receptor. Definition of the phage genomic ends should resolve the inconsistencies seen in the restriction enzyme digest patterns

observed in Section 5.2.3 and finally settle the question of gene order allowing for the identification of early, middle and late genes. As enterobacter phage N4 is thought to have a unique DNA replication mechanism (Ohmori *et al.*, 1998), elucidation of the ends of RLP1 and RPP1 would also provide data from which their replication mechanism could be determined and compared to that of other N4-like phages.

Once the "plate-only" phage receptor is identified, it could then be purified and used in many phage characterisation experiments such as establishment of the binding constant of the two phages and isolation of phage adhesins. Establishment of the bacterial cell surface receptor would also be key in many Roseobacter-based investigations as discussed in Section 7.2.2.

Another area of phage research would be phage transcriptional control. N4, LIT1 and LUZ7 have all been shown to have well defined early, middle and late transcription modules each mediated by a RNA polymerase in conjunction with various host proteins. For example, N4 early transcription is carried out by the viral RNA polymerase protein present in the capsid. However, before this can occur the host DNA gyrase is required to negatively supercoil the viral DNA so that 5-7 bp stem and 3nt loop hairpin structures are formed. An *E. coli* 177 amino acid single stranded DNA-binding protein (*EcoSSB*) then presents the hairpin-form promoter to the vRNAP for binding and transcription (Kazmierczak and Rothman-Denes, 2005). The *EcoSSB* is also involved in transcription elongation through template-recycling. Consequently, identification of the required host factors either through homology to known *E. coli* genes or by co-isolation with the various RNAPs used during infection would be a worthwhile project. With such information, the phage vRNAP could be potentially applied in Roseobacter related *in vitro* transcriptional assays.

8.2.2 *Roseovarius*-based studies

Further work could also be carried out with the two *Roseovarius* hosts once the phage receptor is identified. As alluded to previously, the sensor/trigger and genetic control mechanism behind cell surface change would be an extremely attractive area of research. Both *Rsv.* 217 and *Rsv. nubinhibens* were isolated as planktonic organisms, as such a pertinent area of research would be to investigate if in the environment they are found as surface-associated bacteria and/or in biofilms. It is well documented that Roseobacter species often dominate marine biofilms (Slightom

and Buchan, 2009) and as the relationship between biofilms and phages is currently a hot topic, it would be interesting to examine the roles of phages in marine biofilms perhaps using RLP1 and RPP1 as a model. Results from such studies could potentially be used in the future for the control of biofouling of submerged marine structures or biofilm related diseases e.g. juvenile oyster disease caused by *Rsv. crassostreae.*

Furthermore, as both the host strains for RLP1 and RPP1 were shown to metabolise DMSP (Gonzalez *et al.*, 2003; Schäfer *et al.*, 2005), it would be interesting to find out if their rates of DMSP degradation changes when in the planktonic and sessile state or indeed when infected by phage. Though neither phage appears to have any of the host genes involved in sulfur metabolism, this area of research is still in its early stages as many of the enzymes and genes involved the pathways may yet have to be discovered (see Section 1.1.6 for further details).

The reductionist's approach of studying individuals and their systems has taken much criticism of late as it often does not reflect the true total community effect observed in nature. As such metagene/transcript/proteomics projects are increasing in appeal. However, in phage biology as very little is known about phage genetic information, the designation of the billions of unexplored phages genes seems to be an insurmountable task. By studying individual host-phage systems in more detail, we may gain a greater understanding into many uncharacterised genes and molecular mechanisms present in other phages. Consequently, studies trying to isolate and characterise new phages should be greatly encouraged as without a proper understanding of the little details, the bigger picture may never become clear.

Bibliography

ABEDON, S. T. 2010. *Bacteriophages and biofilms: ecology, phage therapy, plaques*. (unpublished book)

ACKERMAN, H. W. 2009. Phage classification and characterization. *In:* CLOKIE, M. & KROPINSKI, A. M. (eds.) *Bacteriophages: Methods and Protocols.* 127-140. Humana Press.

ACKERMANN, H. W. 1999. Tailed bacteriophages: The order Caudovirales. *Advances in Virus Research, Vol 51.*135-201. San Diego: Academic Press Inc.

ACKERMANN, H. W. 2005. Classification of bacteriophages. *In:* CALENDAR., R. (ed.) *The Bacteriophages, 2nd Edition.* 8-16. Oxford: Oxford University Press.

ACKERMANN, H. W. & DEBOW, M. S. 1987. General properties of bacteriophages. *Viruses of prokaryotes.* Boca Raton, Florida: CRC Press.

ACKERMANN, H. W., DUBOW, M. S., JARVIS, A. W., JONES, L. A., KRYLOV, V. N., MANILOFF, J., ROCOURT, J., SAFFERMAN, R. S., SCHNEIDER, J., SELDIN, L., SOZZI, T., STEWART, P. R., WERQUIN, M. & WUNSCHE, L. 1992. The species concept and its application to tailed phages. *Arch Virol,* **124,** 69-82.

ANGLY, F., YOULE, M., NOSRAT, B., SRINAGESH, S., RODRIGUEZ-BRITO, B., MCNAIRNIE, P., DEYANAT-YAZDI, G., BREITBART, M. & ROHWER, F. 2009. Genomic analysis of multiple Roseophage SIO1 strains. *Environ Microbiol,* **11,** 2863-2873.

ANGLY, F. E., FELTS, B., BREITBART, M., SALAMON, P., EDWARDS, R. A., CARLSON, C., CHAN, A. M., HAYNES, M., KELLEY, S., LIU, H., MAHAFFY, J. M., MUELLER, J. E., NULTON, J., OLSON, R., PARSONS, R., RAYHAWK, S., SUTTLE, C. A. & ROHWER, F. 2006. The marine viromes of four oceanic regions. *Plos Biol,* **4,** 2121-2131.

ASHELFORD, K. E., DAY, M. J. & FRY, J. C. 2003. Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol,* **69,** 285-289.

BAAS BECKING, L. G. M. 1934. Geobiologie of inleiding tot de milieukunde, Van Stockum & Zoon, The Hauge.

BAIL, O. 1925. Der kolistamm 88 von gildemeister und herzberg. *Med Klin (Munich),* **21,** 1271-1273.

BARNODESS, J. J. & BECKWITH, J. 1995. bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *J Bacteriol,* **177,** 1247-1253.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* **315,** 1709-1712.

BAUMGARTNER, S., HOFMANN, K., CHIQUET-EHRISHMANN, R. & BUCHER, P. 1998. The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci,* **7,** 1626-1631.

BAYER, M. E., THUROW, H. & BAYER, M. H. 1979. Penetration of the polysaccharide capsule of *Escherichia coli* (Bi161-42) by bacteriophage K29. *Virology,* **94,** 95-118.

BECKNER, M. 1968. *The biological way of thought,* Berkeley, University of California Press.

BELAS, R., SIMON, F. & SILVERMANN, M. 1986. Regulation of lateral flagella gene transcription in *Vibrio parahaemolyticus*. *J Bacteriol,* **167,** 210-218.

BERGH, O., BORSHEIM, K. Y., BRATBAK, G. & HELDAL, M. 1989. High abundance of viruses found in aquatic environments. *Nature,* **340,** 467-468.

BOEHME, J., FISCHER, M. E., JIANG, S. C., KELLOGG, C. A., PICHARD, S., ROSE, J. B., STIENWAY, C., PAUL, J. H. (1993) Viruses, bacterioplanton, and phytoplankton in the southeastern gulf of Mexico: distribution and contribution to oceanic DNA pools. *Mar Ecol Prog Ser,* **97** 1 -10.

BOETTCHER, K. J., GEAGHAN, K. K., MALOY, A. P. & BARBER, B. J. 2005. *Roseovarius crassostreae* sp nov., a member of the *Roseobacter* clade and the apparent cause of juvenile oyster disease (JOD) in cultured Eastern oysters. *Int J Syst Evol Microbiol,* **55,** 1531-1537.

BOHANNAN, B. J. M. & LENSKI, R. E. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Evolution,* **53,** 292-295.

BORDET, J. 1925. Le problème de l'autolyse microbienne transmissible ou du bactèriophage. *Ann Inst Pasteur,* **39,** 711-763.

BORDET, J. & CIUCA, M. 1921. Evolution des cultures de coli lysogene. *Compt Rend Soc Bio,* **84,** 747.

BORSHEIM, K. Y. 1993. Native marine bacteriophages. *FEMS Microl Ecol,* **102,** 141-159.

BORSHEIM, K. Y., BRATBAK, G. & HELDAL, M. 1990. Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron-microscopy. *Appl Environ Microbiol,* **56,** 352-356.

BOSE, M. & BARBER, R. D. 2006. Prophage Finder: a prophage loci prediciton tool for prokaryotic genomes sequences. *In Silico Biol,* **6,** 223-227.

BOULANGER, P. & LETELLIER, L. 1988. Characterization of ion channels involved in the penetration of phage T4 DNA into *Escherichia coli* cells. *J Biol Chem,* **263,** 9767-9775.

BRAGG, J. G. & CHISHOLM, S. W. 2008. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *Plos One,* **3,** e3550.

BRATBAK, G., HELDAL, M., THINGSTAD, T. F. & TUOMI, P. 1996. Dynamics of virus abundance in coastal seawater. *FEMS Microbiol Ecol,* **19,** 263-269.

BREITBART, M., MIYAKE, J. H. & ROHWER, F. 2004. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett,* **236,** 249-256.

BREITBART, M. & ROHWER, F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol,* **13,** 278-284.

BREITBART, M., SALAMON, P., ANDRESEN, B., MAHAFFY, J. M., SEGALL, A. M., MEAD, D., AZAM, F. & ROHWER, F. 2002. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA,* **99,** 14250-14255.

BRINKHOFF, T., GIEBEL, H. A. & SIMON, M. 2008. Diversity, ecology, and genomics of the Roseobacter clade: a short overview. *Arch Microbiol,* **189,** 531-539.

BROUNS, S. J. J., JORE, M. M., LUNDGREN, M., WESTRA, E. R., SLIJKHUIS, R. J. H., SNIJDERS, A. P. L., DICKMAN, M. J., MAKAROVA, K. S., KOONIN, E. V. & VAN DER OOST, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science,* **321,** 960-964.

BROWN, M. V., SCHWALBACH, M. S., HEWSON, I. & FUHRMAN, J. A. 2005. Coupling 16S-ITS rDNA clone libraries and automated ribosomal

intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ Microbiol,* **7,** 1466-1479.

BRUHN, J. B., GRAM, L. & BELAS, R. 2007. Production of antibacterial compounds and biofilm formation by Roseobacter species are influenced by culture conditions. *Appl Environ Microbiol,* **73,** 442-450.

BRUSSAARD, C. P. 2004. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol,* **70,** 1506-1513.

BRUSSAARD, C. P., MARIE, D. & BRATBAK, G. 2000. Flow cytometric detection of viruses. *J Virol Methods,* **85,** 175-182.

BRUSSOW, H., CANCHAYA, C. & HARDT, W.-D. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev,* **68,** 560-602.

BUCHAN, A., GONZALEZ, J. M. & MORAN, M. A. 2005. Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol,* **71,** 5665-5677.

BUTLER, J. H. 2000. Atmospheric chemistry: better budgets for methyl halides? *Nature,* **403,** 260-261.

CAMPBELL, A. M. 2007. Bacteriophages *In:* FIELDS, B. N., KNIPE, D. M. K., HOWLEY, P. M. & GRIFFIN, D. E. (eds) *Fields virology 5th edition.* Lippincott Williams & Wilkins

CASJENS, S. R. 2008. Diversity among the tailed-bacteriophages that infect the Enterobacteriaceae. *Res Microbiol,* **159,** 340-348.

CEYSSENS, P. J. 2009a. Isolation and characterization of lytic bacteriophages infecting *Pseudomonas aeruginosa.* Katholieke Universiteit Leuven.

CEYSSENS, P. J., LAVIGNE, R., MATTHEUS, W., CHIBEU, A., HERTVELDT, K., MAST, J., ROBBEN, J. & VOLCKAERT, G. 2006. Genomic analysis of *Pseudomonas aeruginosa* phages LKD16 and LKA1: establishment of the phi KMV subgroup within the T7 supergroup. *J Bacteriol,* **188,** 6924-6931.

CEYSSENS, P. J., NOBEN, J. P., ACKERMANN, H. W., VERHAEGEN, J., DE VOS, D., PIRNAY, J. P., MERABISHVILI, M., VANEECHOUTTE, M., CHIBEU, A., VOLCKAERT, G. & LAVIGNE, R. 2009b. Survey of *Pseudomonas aeruginosa* and its phages: de novo peptide sequencing as a novel tool to assess the diversity of worldwide collected viruses. *Environ Microbiol,* **11,** 1303-1313.

CHARLSON, R. J., LOVELOCK, J. E., ANDREAE, M. O. & WARREN, S. G. 1987. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature,* **326,** 655-661.

CHEN, F., WANG, K., STEWART, J. & BELAS, R. 2006. Induction of multiple prophages from a marine bacterium: a genomic approach. *Appl Environ Microbiol,* **72,** 4995-5001.

CHEN, Y., GOLDING, I., SAWAI, S., GUO, L. & COX, E. C. 2005. Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *Plos Biology,* **3,** 1276-1282.

CHOI, K. H., MCPARTLAND, J., KAGANMAN, I., BOWMAN, V. D., ROTHMAN-DENES, L. B. & ROSSMANN, M. G. 2008. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J Mol Biol,* **378,** 726-736.

CHOI, S., DUNAMS, D. & JIANG, S. C. 2010. Transfer of cholera toxin genes from O1 to non-O1/O139 strains by vibriophages from California coastal waters. *J Appl Microbiol,* **108,** 1015-1022.

CHOPIN, M.-C., CHOPIN, A. & BIDNENKO, E. 2005. Phage abortive infection in lactococci: variations on a theme. *Curr Opin Microbiol,* **8,** 473-479.

CLARK, A. J., INWOOD, W., CLOUTIER, T. & DHILLON, T. S. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J Micro Biol,* **311,** 657-679.

CLOKIE, M. R. J., SHAN, J., BAILEY, S., JIA, Y., KRISCH, H. M., WEST, S. & MANN, N. H. 2006. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol,* **8,** 827-835.

COCHLAN, W. P., WIKNER, J., STEWARD, G. F., SMITH, D. C. & AZAM, F. 1993. Spatial distribution of viruses, bacteria and chlorophyll a in neritic, oceanic and estuarine environments. *Mar Ecol Prog Ser,* **92,** 77-87.

CORBIN, B. D., MCLEAN, R. J. C. & ARON, G. M. 2001. Bacteriophage T4 multiplication in a glucose-limited *Escherichia coli* biofilm. *Can J Microbiol,* **47,** 680-684.

CORTEZ, D., FORTERRE, P. & GRIBALDO, S. 2009. A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol,* **10,** R65.

CULLEY, A. I., LANG, A. S. & SUTTLE, C. A. 2006. Metagenomic analysis of coast RNA virus communities. *Science,* **312,** 1795-11798.

CURSON, A. R. J., ROGERS, R., TODD, J. D., BREARLEY, C. A. & JOHNSTON, A. W. B. 2008. Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine alpha-proteobacteria and *Rhodobacter sphaeroides*. *Environ Microbiol,* **10,** 757-767.

D'HERELLE, F. 1922. Sur la prétendue production d'un principe lytique sous l'influence d'un antagonisme microbien. *Compt Rend Soc Bio,* **86,** 663-665.

DACEY, J. W. H. & WAKEHAM, S. G. 1986. Oceanic DMS: production during zooplankton grazing on phytoplankton. *Science,* **233,** 1314-1316.

DAMOND, F., WOROBEY, M., CAMPA, P., FARFARA, I., COLIN, G., MATHERON, S., BRUN-VÉZINET, F., ROBERTSON, D. L. & SIMON, F. 2004. Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. *AIDS Res Hum Retroviruses,* **20,** 666-672.

DANG, H.-Y. & LOVELL, C. R. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. *Appl Environ Microbiol,* **66,** 467-475.

DANG, H. & LOVELL, C. R. 2002. Seasonal dynamics of particle-associated and free-living marine *Proteobacteria* in a salt marsh tidal creek as determined using fluorescence *in situ* hybridization. *Environ Microbiol,* **4,** 287-295.

DAVYDOVA, E. K., SANTANGELO, T. J. & ROTHMAN-DENES, L. B. 2007. Bacteriophage N4 virion RNA polymerase interaction with its promoter DNA hairpin. *Proc Natl Acad Sci USA,* **104,** 7033-7038.

DE BOER, W. E., GOLTEN, C. & SCHEFFERS, W. A. 1975. Effects of some physical factors on flagellation and swarming of *Vibrio alginolyticus*. *Neth J Sea Res,* **9,** 197-213.

DE WIT, R. & BOUVIER, T. 2006. 'Everything is everywhere, but the environment selects': What did Baas Becking and Beijerinck really say? *Environ Microbiol.* **8,** 755-758.

DELBRÜCK, M. 1940. The growth of bacteriophage and lysis of the host. *J Gen Physiol,* **23,** 643-660.

DELONG, E. F., PRESTON, C. M., MINCER, T., RICH, V., HALLAM, S. J., FRIGAARD, N.-U., MARTINEZ, A., SULLIVAN, M. B., EDWARDS, R., BRITO, B. R., CHISHOLM, S. W. & KARL, D. M. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science,* **311,** 496-503.

DESNUES, C., RODRIGUEZ-BRITO, B., RAYHAWK, S., KELLEY, S., TRAN, T., HAYNES, M., LIU, H., FURLAN, M., WEGLEY, L., CHAU, B., RUAN, Y. J., HALL, D., ANGLY, F. E., EDWARDS, R. A., LI, L. L., THURBER, R. V., REID, R. P., SIEFERT, J., SOUZA, V., VALENTINE, D. L., SWAN, B. K., BREITBART, M. & ROHWER, F. 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature,* **452,** 340-U5.

DESTOUMIEUX-GARZÓN, D., DUQUESNE, S., PEDUZZI, J., GOULARD, C., DESMADRIL, M., LETELLIER, L., REBUFFAT, S. & BOULANGER, P. 2005. The iron-siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25: role of the microcin Val11-Pro16 beta-hairpin region in the recognition mechanism. *Biochem J,* **389,** 869-876.

DEVEAU, H., BARRANGOU, R., GARNEAU, J. E., LABONTE, J., FREMAUX, C., BOYAVAL, P., ROMERO, D. A., HORVATH, P. & MOINEAU, S. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol,* **190,** 1390-1400.

DINSDALE, E. A., EDWARDS, R. A., HALL, D., ANGLY, F., BREITBART, M., BRULC, J. M., FURLAN, M., DESNUES, C., HAYNES, M., LI, L. L., MCDANIEL, L., MORAN, M. A., NELSON, K. E., NILSSON, C., OLSON, R., PAUL, J., BRITO, B. R., RUAN, Y. J., SWAN, B. K., STEVENS, R., VALENTINE, D. L., THURBER, R. V., WEGLEY, L., WHITE, B. A. & ROHWER, F. 2008. Functional metagenomic profiling of nine biomes. *Nature,* **452,** 629-U8.

DOOLITTLE, M. M., COONEY, J. J. & CALDWELL, D. E. 1996. Tracing the interaction of bacteriophage with bacterial biofilms using fluorescent and chromogenic probes. *J Ind Microbiol,* **16,** 331-341.

DZIARSKI, R. 2004. Peptidoglycan recognition proteins (PGRPs)**. *Mol Immunol,* 40,** 877-886.

EARNSHAW, W. C. & HARRISON, S. C. 1977. DNA arrangement in isometric phage heads. *Nature,* **268,** 598-602.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res,* **32,** 1792-1797.

EDLIN, G., LIN, L. E. O. & KUDRNA, R. 1975. Lamda lysogens of *E. coli* reproduce more rapidly than non-lysogens. *Nature,* **255,** 735-737.

ETSON, C. M., HAMDAN, S. M., RICHARDSON, C. C. & VAN OIJEN, A. M. 2010. Thioredoxin suppresses microscopic hopping of T7 DNA polymerase on duplex DNA. *Proc Natl Acad Sci USA,* **107,** 1900-1905.

FAUQUET, C. M., BISARO, D. M., BRIDDON, R. W., BROWN, J. K., HARRISON, B. D., RYBICKI, E. P., STENGER, D. C. & STANLEY, J. 2003. Revision of taxonomic criteria for species demarcation in the family *Geminiviridae*, and an updated list of begomovirus species. *Arch Virol,* **148,** 405-421.

FILÉE, J., BAPTESTE, E., SUSKO, E. & KRISCH, H. M. 2006. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has

preserved a core genome with the viral replication and structural genes. *Mol Biol Evol,* **23,** 1688-1696.

FILIPPINI, M., BUESING, N., BETTAREL, Y., SIME-NGANDO, T. & GESSNER, M. O. 2006. Infection paradox: high abundance but low impact of freshwater benthic viruses. *Appl Environ Microbiol,* **72,** 4893-4898.

FISCHER, D. & EISENBERG, D. 1999. Finding families for genomic ORFans. *Bioinformatics,* **15,** 759-762.

FLOOD, J. A. & ASHBOLT, N. J. 1999. Virus-sized particles can be entrapped and concentrated one hundred fold within wetland biofilms. *Adv Envrion Res,* **3** 403-411.

FOKINE, A., LEIMAN, P. G., SHNEIDER, M. M., AHVAZI, B., BOESHANS, K. M., STEVEN, A. C., BLACK, L. W., MESYANZHINOV, V. V. & ROSSMANN, M. G. 2005. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc Natl Acad Sci USA,* **102,** 7163-7168.

FRASER, J. S., MAXWELL, K. L. & DAVIDSON, A. R. 2007. Immunoglobulin-like domains on bacteriophage: weapons of modest damage? *Curr Opin Microbiol,* **10,** 382-387.

FREIFELDER, D. 1987 *Molecular biology*. Jones & Barlett, Inc., Boston

FRIAS-LOPEZ, J., KLAUS, J. S., BONHEYO, G. T. & FOUKE, B. W. 2004. Bacterial community associated with black band disease in corals. *Appl Environ Microbiol,* **70,** 5955-5962.

FROST, L. S., LEPLAE, R., SUMMERS, A. O. & TOUSSAINT, A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Micro,* **3,** 722-732.

FUHRMAN, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature,* **399,** 541-548.

FUHRMAN, J. A. & SCHWALBACH, M. 2003. Viral Influence on aquatic bacterial communities. *Biol Bull,* **204,** 192-195.

GARZA, D. & SUTTLE, C. A. 1998. The effect of cyanophages on the mortality of *Synechococcus* spp. and selection for UV resistant viral communities. *Microb Ecol,* **36,** 281-292.

GENG, H. & BELAS, R. 2010. Molecular mechanisms underlying Roseobacter-phytoplankton symbioses. *Curr Opin Biotehcnol,* **21,** 332-338.

GENG, H., BRUHN, J. B., NIELSEN, K. F., GRAM, L. & BELAS, R. 2008. Genetic dissection of tropodithietic acid biosynthesis by marine Roseobacters. *Appl Environ Microbiol,* **74,** 1535-1545.

GILDEMEISTER, E. 1921. Über das d'Herellesche phänomen. *Berlin Klin Wschr,* **58,** 1355.

GIOVANNONI, S. J., TRIPP, H. J., GIVAN, S., PODAR, M., VERGIN, K. L., BAPTISTA, D., BIBBS, L., EADS, J., RICHARDSON, T. H., NOORDEWIER, M., RAPPE, M. S., SHORT, J. M., CARRINGTON, J. C. & MATHUR, E. J. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science,* **309,** 1242-1245.

GIVAN, A. L., GLASSEY, K., GREEN, R. S., LANG, W. K., ANDERSON, A. J. & ARCHIBALD, A. R. 1982. Relation between wall teichoic-acid content of *Bacillus subtilis* and efficiency of adsorption of bacteriophage Sp 50 and bacteriophage Phi 25. *Arch Microbiol,* **133,** 318-322.

GLEGHORN, M. L., DAVYDOVA, E. K., ROTHMAN-DENES, L. B. & MURAKAMI, K. S. 2008. Structural basis for DNA hairpin promoter

recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell,* **32,** 707-717.

GOERKE, C., PANTUCEK, R., HOLTFRETER, S., SCHULTE, B., ZINK, M., GRUMANN, D., BROKER, B. M., DOSKAR, J. & WOLZ, C. 2009. Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *J Bacteriol,* **191,** 3462-3468.

GOERKE, C., WIRTZ, C., FLUCKIGER, U. & WOLZ, C. 2006. Extensive phage dynamics in *Staphylococcus aureus* contributes to adaptation to the human host during infection. *Mol Microbiol,* **61,** 1673-1685.

GOLTEN, C. & SCHEFFERS, W. A. 1975. Marine *Vibrios* isolated from water along the Dutch coast. *Neth J Sea Res,* **9,** 351-364.

GONZALEZ, J. M. & MORAN, M. A. 1997. Numerical dominance of a group of marine bacteria in the α-subclass of the class *Proteobacteria* in coastal seawater. *Appl Environ Microbiol,* **63,** 4237-4242.

GONZALEZ, J. M., SIMO, R., MASSANA, R., COVERT, J. S., CASAMAYOR, E. O., PEDROS-ALIO, C. & MORAN, M. A. 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol,* **66,** 4237-4246.

GONZALEZ, J. M., COVERT, J. S., WHITMAN, W. B., HENRIKSEN, J. R., MAYER, F., SCHARF, B., SCHMITT, R., BUCHAN, A., FUHRMAN, J. A., KIENE, R. P. & MORAN, M. A. 2003. *Silicibacter pomeroyi* sp nov and *Roseovarius nubinhibens* sp nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int J Syst Evol Microbiol,* **53,** 1261-1269.

GOODSTADT, L. & PONTING, C. P. 2001. CHROMA: consensus-based colouring of multiple alignments for publication. *Bioinformatics,* **17,** 845-846.

GOUY, M., GUIDON, S. & GASCUEL, O. 2010 SeaView version 4: a multiplatform grphical user inferface for sequence alignment and phylogenetic tree building. *Mol Biol and Evol,* **27,** 221-224.

GRAM, L., GROSSART, H. P., SCHLINGLOFF, A. & KIØRBOE, T. 2002. Possible quorum sensing in marine snow bacteria: production of acylated homoserine lactones by *Roseobacter* strains isolated from marine snow. *Appl Environ Microbiol,* **68,** 4111-4116.

GRATIA, A. 1936. Des relations numeriques entre bacteries lysogenes et particules des bacteriophages. *Ann Inst Pasteur,* **57,** 652-676.

GUIXA-BOIXEREU, N., VAQUÉ, D., GASOL, J. M. & PEDRÓS-ALIÓ, C. 1999. Distribution of viruses and their potential effect on bacterioplankton in an oligotrophic marine system. *Aquat Microb Ecol,* **19,** 205-213.

HALLAM, S. J., PUTNAM, N., PRESTON, C. M., DETTER, J. C., ROKHSAR, D., RICHARDSON, P. M. & DELONG, E. F. 2004. Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science,* **305,** 1457-1462.

HAMMAD, A. M. M. 1998. Evaluation of alginate-encapsulated *Azotobacter chroococcum* as a phage-resistant and an effective inoculum. *J Basic Microbiol,* **38,** 9-16.

HARA, S., KOIKE, I., TERAUCHI, K., KAMIYA, H. & TANOUE, E. 1996. Abuncance of viruses in deep oceanic water. *Mar Ecol Prog Ser,* **145,** 260-277.

HARDIES, S. C., COMEAU, A. M., SERWER, P. & SUTTLE, C. A. 2003. The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. *Virology,* **310,** 359-371.

HARTL, D. L. & JONES, E. W. 2009. Genetics of bacteria and their viruses. *In:* HARTL, D. L. & JONES, E. W. (eds.) *Genetics: analysis of genes and genomes.* 7th ed.: Jones and Bartlett Publishers.

HATFULL, G. F., PEDULLA, M. L., JACOBS-SERA, D., CICHON, P. M., FOLEY, A., FORD, M. E., GONDA, R. M., HOUTZ, J. M., HRYCKOWIAN, A. J., KELCHNER, V. A., NAMBURI, S., PAJCINI, K. V., POPOVICH, M. G., SCHLEICHER, D. T., SIMANEK, B. Z., SMITH, A. L., ZDANOWICZ, G. M., KUMAR, V., PEEBLES, C. L., JACOBS, W. R., LAWRENCE, J. G. & HENDRIX, R. W. 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *Plos Genet,* **2,** 835-847.

HELLWEGER, F. L. 2009. Carrying photosynthesis genes increases ecological fitness of cyanophage *in silico. Environ Microbiol,* **11,** 1386-1394.

HENDRIX, R. W. 2002. Bacteriophages: evolution of the majority. *Theor Popul Biol,* **61,** 471-480.

HENDRIX, R. W. & CASJENS, S. R. 2006. Index of Viruses - Podoviridae. *In:* BÜCHEN-OSMOND, C. (ed.) *ICTVdB - The Universal Virus Database, version 4.* New York, USA: Columbia University.

HENDRIX, R. W., SMITH, M. C. M., BURNS, R. N., FORD, M. E. & HATFULL, G. F. 1999. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc Natl Acad Sci USA,* **96,** 2192-2197.

HINTON, D. M., PANDE, S., WAIS, N., JOHNSON, X. B., VUTHOORI, M., MAKELA, A. & HOOK-BARNARD, I. 2005. Transcriptional takeover by σ appropriation: remodelling of the σ70 subunit of *Escherichia coli* RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. *Microbiology,* **151,** 1729-1740.

HOLMFELDT, K., MIDDELBOE, M., NYBROE, O. & RIEMANN, L. 2007. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts. *Appl Environ Microbiol,* **73,** 6730-6739.

HONG, J. & ZHOU, X.-P. 2006. The universal system of virus taxonomy in the 8(th) ICTV report. *Virologica Sinica,* **21,** 84-96.

HOWARD, E. C., HENRIKSEN, J. R., BUCHAN, A., REISCH, C. R., BUERGMANN, H., WELSH, R., YE, W. Y., GONZALEZ, J. M., MACE, K., JOYE, S. B., KIENE, R. P., WHITMAN, W. B. & MORAN, M. A. 2006. Bacterial taxa that limit sulfur flux from the ocean. *Science,* **314,** 649-652.

HOWARD, E. C., SUN, S., BIERS, E. J. & MORAN, M. A. 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ Microbiol,* **10,** 2397-2410.

HUANG, C., ZHANG, J. & JIAO, N. Z. 2010. Phage resistance of a marine bacterium, *Roseobacter denitrificans* OCh114, as revealed by comparative proteomics. *Curr Microbiol.* **61,** 141-147.

HUBER, H. E., TABOR, S. & RICHARDSON, C. C. 1987. *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates. *J Biol Chem,* **263,** 16224-16232.

HUELSENBECK, J. P. & RONQUIST, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics,* **17,** 754-755.

HUTCHINSON, G. E. 1961. The paradox of the plankton. *Am Nat,* **144,** 741-771.

JACOB, F. & MONOD, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol,* **3,** 318-356.

JACQUET, S., HELDAL, M., IGLESIAS-RODRIGUEZ, D., LARSEN, A., WILSON, W. & BRATBAK, G. 2002. Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. *Aquat Microb Ecol,* **27,** 111-124.

JANNASCH, H. W. 1958. Studies on planktonic bacterial by means of a direct membrane filter method. *J Gen Micrbiol,* **18,** 608-620.

JIANG, S. C. & PAUL, J. H. 1994. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. *Mar Ecol Prog Ser,* **104,** 163-172.

JIANG, S. C. & PAUL, J. H. 1996. Occurrence of lysogenic bacteria in marine microbial communities as determined by prophage induction. *Mar Ecol Prog Ser,* **142,** 27-38.

JIANG, S. C. & PAUL, J. H. 1998. Gene transfer by transduction in the marine environment. *Appl Environ Microbiol,* **64,** 2780-2787.

JIANG, W., LI, Z., ZHANG, Z., BAKER, M. L., PREVELIGE, P. E. J. & CHIU, W. 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat Struct Biol,* **10,** 131-135.

JOHNSTON, A. W. B., TODD, J. D., SUN, L., NIKOLAIDOU-KATSARIDOU, M. N., CURSON, A. R. J. & ROGERS, R. 2008. Molecular diversity of bacterial production of the climate-changing gas, dimethyl sulphide, a molecule that impinges on local and global symbioses. *J Exp Bot,* **59,** 1059-1067.

JOUENNE, T., TRESSE, O. & JUNTER, G.-A. 1994. Agar-entrapped bacteria as an *in vitro* model of biofilms and their susceptibility to antibiotics. *FEMS Microbiol Lett,* **119,** 237-242.

KAZMIERCZAK, K. M., DAVYDOVA, E. K., MUSTAEV, A. A. & ROTHMAN-DENES, L. B. 2002. The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *EMBO J,* **21,** 5815-5823.

KAZMIERCZAK, K. M. & ROTHMAN-DENES, L. B. 2005. Bacteriophage N4. *In:* CALENDAR., R. (ed.) *The Bacteriophages, 2nd Edition.* Oxford: Oxford University Press.

KIENE, R. P., LINN, L. J. & BRUTON, J. A. 2000. New and important roles for DMSP in marine microbial communities. *J Sea Res,* 43**,** 209-224.

KIM, B. K., PARK, Y. D., OH, H. M. & CHUN, J. 2009. Identification and characterization of metagenomic fragments from tidal flat sediment. *J Microbiol,* **47,** 402-410.

KIRKWOOD, M., LE BRUN, N. L., TODD, J. D. & JOHNSTON, A. W. B. 2010. The *dddP* gene of *Roseovarius nubinhibens* encodes a novel lyase that cleaves dimethylsulfoniopropionate into acrylate plus dimethyl sulfide. *Microbiology,* **156,** 1900-1906.

KOWALEWSKY, S., DAMBACH, M., MAUCK, B. & DEHNHARDT, G. 2006. High olfactory sensitivity for dimethyl sulphide in harbour seals. *Biol Lett,* **2,** 106-109.

KRISCH, H. M. & COMEAU, A. M. 2008. The immense journey of bacteriophage T4--From d'Hérelle to Delbrück and then to Darwin and beyond. *Res Microbiol,* **159,** 314-324.

KRISTENSEN, D. M., MUSHEGIAN, A. R., DOLJA, V. V. & KOONIN, E. V. 2010. New dimensions of the virus world discovered through metagenomics. *Trends Microbiol,* **18,** 11-19.

KROPINSKI, A. M., PRANGISHVILI, D. & LAVIGNE, R. 2009. Position paper: The creation of a rational scheme for the nomenclature of viruses of *Bacteria* and *Archaea. Environ Microbiol,* **11,** 2775-2777.

KRUEGER, A. P. 1931. The sorption of bacteriophage by living and dead susceptible bacteria. *J Gen Physiol,* **14,** 493-516.

KRÜGER, D. H. & SCHROEDER, C. 1981. Bacteriophage-T3 and bacteriophage-T7 virus-host cell-interactions. *Microbiol Rev,* **45,** 9-51.

KWAN, T., LIU, J., DUBOW, M., GROS, P. & PELLETIER, J. 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci USA,* **102,** 5174-5179.

LABONTE, J. M., REID, K. E. & SUTTLE, C. A. 2009. Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine Podovirus DNA polymerase gene sequences. *Appl Environ Microbiol,* **75,** 3634-3640.

LABRIE, S. J., SAMSON, J. E. & MOINEAU, S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Micro,* **8,** 317-327.

LAVIGNE, R., CEYSSENS, P. J. & ROBBEN, J. 2009. Phage Proteomics: Applications of Mass Spectrometry. *In:* CLOKIE, M. & KROPINSKI, A. M. (eds.) *Bacteriophages: Methods and Protocols, Volume 2: Molecular and Applied Aspects.* Humana Press.

LAVIGNE, R., NOBEN, J. P., HERTVELDT, K., CEYSSENS, P. J., BRIERS, Y., DUMONT, D., ROUCOURT, B., KRYLOV, V. N., MESYANZHINOV, V. V., ROBBEN, J. & VOLCKAERT, G. 2006. The structural proteome of *Pseudomonas aeruginosa* bacteriophage phiKMV. *Microbiology,* **152,** 529-534.

LECOUTERE, E., CEYSSENS, P. J., MIROSHNIKOV, K. A., MESYANZHINOV, V. V., KRYLOV, V. N., NOBEN, J. P., ROBBEN, J., HERTVELDT, K., VOLCKAERT, G. & LAVIGNE, R. 2009. Identification and comparative analysis of the structural proteomes of phi KZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages. *Proteomics,* **9,** 3215-3219.

LEIMAN, P. G., CHIPMAN, P. R., KOSTYUCHENKO, V. A., MESYANZHINOV, V. V. & ROSSMANN, M. G. 2004. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell,* **118,** 419-429.

LENSKI, R. E. 1988. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. *Evolution,* **42,** 425-432.

LINDELL, D., JAFFE, J. D., COLEMAN, M. L., FUTSCHIK, M. E., AXMANN, I. M., RECTOR, T., KETTLER, G., SULLIVAN, M. B., STEEN, R., HESS, W. R., CHURCH, G. M. & CHISHOLM, S. W. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature,* **449,** 83-86.

LINDELL, D., JAFFE, J. D., JOHNSON, Z. I., CHURCH, G. M. & CHISHOLM, S. W. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature,* **438,** 86-89.

LINDELL, D., SULLIVAN, M. B., JOHNSON, Z. I., TOLONEN, A. C., ROHWER, F. & CHISHOLM, S. W. 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA,* **101,** 11013-11018.

LINGOHR, E. J., VILLEGAS, A., SHE, Y. M., CEYSSENS, P. J. & KROPINSKI, A. M. 2008. The genome and proteome of the Kluyvera bacteriophage Kvp1-another member of the T7-like Autographivirinae. *Virol J,* **5**, 122.

LOPEZ-BUENO, A., TAMAMES, J., VELAZQUEZ, D., MOYA, A., QUESADA, A. & ALCAMI, A. 2009. High diversity of the viral community from an Antarctic lake. *Science,* **326,** 858-861.

LOVELOCK, J. E., MAGGS, R. J. & RASMUSSE, R. A. 1972. Atmospheric dimethyl sulfide and natural sulfur cycle. *Nature,* **237,** 452-453.

LU, M.-J. & HENNING, U. 1994. Superinfection exclusion by T-even-type coliphages. *Trends Microbiol,* **2,** 137-139.

LU, M. J., STIERHOF, Y. D. & HENNING, U. 1993. Location and unusual membrane topology of the immunity protein of the *Escherichia coli* phage T4. *J Virol,* **67,** 4905-4913.

LWOFF, A., SIMINOVITCH, L. & KJELDGAARD, M. 1950. Induction de la production de bactériophages chez une bactérie lysogène. *Ann Inst Pasteur,* **79,** 815-859.

MANN, N. H., COOK, A., MILLARD, A., BAILEY, S. & CLOKIE, M. 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature,* **424,** 741.

MÄNNISTÖ, R. H., KIVELÄ, H. M., PAULIN, L., BAMFORD, D. H. & BAMFORD, J. K. H. 1999. The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. *Virology,* **262,** 355-363.

MARSDEN, R. L., McGUFFIN, L. J. & JONES, D. T. 2002. Rapid protein domain assignement from amino acid sequence using predicted secondary structure. *Protein Sci,* **11,** 2814 - 2824.

MARSHALL, K. C. 2006. Planktonic versus sessile life of prokaryotes. *In:* DWORKIN, M., FALKOW, S., ROSENBERG, E., SCHLEIFER, K.-H. & STACHEBRANDT, E. (eds.) *The Prokaryotes.* New York: Springer.

MARTINEZ, J. M., SCHROEDER, D. C., LARSEN, A., BRATBAK, G. & WILSON, W. H. 2007. Molecular dynamics of *Emiliania huxleyi* and cooccurring viruses during two separate mesocosm studies. *Appl Environ Microbiol,* **73,** 554-562.

MAYALI, X., FRANKS, P. J. S. & AZAM, F. 2008. Cultivation and ecosystem role of a marine Roseobacter clade-affiliated cluster bacterium. *Appl Environ Microbiol,* **74,** 2595-2603.

MCDANIEL, L., HOUCHIN, L. A., WILLIAMSON, S. J. & PAUL, J. H. 2002. Plankton blooms: lysogeny in marine *Synechococcus. Nature,* **415,** 496-496.

MCLAUGHLIN, M. R. & BALAA, M. F. 2006. Enhanced contrast of bacteriophage plaques in *Salmonella* with ferric ammonium citrate and sodium thiosulfate (FACST) and tetrazolium red (TZR). *J Microbiol Methods,* **66,** 318-323.

MCLEAN, R. J., CORBIN, B. D., BALZER, G. J. & ARON, G. M. 2001. Phenotype characterization of genetically defined microorganisms and growth of phage in biofilms. *Methods Enzymmol,* **336,** 163-174.

MILLARD, A. 2004. *An investigation of cyanophages infecting marine Synechococcus.* University of Warwick.

MILLARD, A., CLOKIE, M. R. J., SHUB, D. A. & MANN, N. H. 2004. Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. *Proc Natl Acad Sci USA,* **101,** 11007-11012.

MILLARD, A. D., GIERGA, G., CLOKIE, M. R. J., EVANS, D. J., HESS, W. R. & SCANLAN, D. J. 2010. An antisense RNA in a lytic cyanophage links psbA to a gene encoding a homing endonuclease. *ISME J.* (Epub ahead of print)

MILLARD, A. D., ZWIRGLMAIER, K., DOWNEY, M. J., MANN, N. H. & SCANLAN, D. J. 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ Microbiol,* 11**,** 2370-2387.

MILLER, R. V. 2005. Marine phages *In:* CALENDAR., R. (ed.) *The Bacteriophages, 2nd Edition.* Oxford: Oxford University Press.

MILLER, T. R. & BELAS, R. 2004. Dimethylsulfoniopropionate metabolism by *Pfiesteria*-associated *Roseobacter* spp. *Appl Environ Microbiol,* **70,** 3383-3391.

MILLER, T. R., HNILICKA, K., DZIEDZIC, A., DESPLATS, P. & BELAS, R. 2004. Chemotaxis of *Silicibacter* sp. strain TM1040 toward dinoflagellate products. *Appl Environ Microbiol,* **70,** 4692-4701.

MOLDOVAN, R., CHAPMAN-MCQUISTON, E. & WU, X. L. 2007. On kinetics of phage adsorption. *Biophys J,* **93,** 303-315.

MOLINEUX, I. J. 1991. Host-parasite interactions: recent developments in the genetics of abortive phage infections. *New Biol,* **3,** 230-236.

MOLINEUX, L. J. 2001. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Mol Microbiol,* **40,** 1-8.

MORAIS, M. C., CHOI, K. H., KOTI, J. S., CHIPMAN, P. R., ANDERSON, D. L. & ROSSMANN, M. G. 2005. Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol Cell,* **18,** 149-159.

MORAN, M. A., BUCHAN, A., GONZALEZ, J. M., HEIDELBERG, J. F., WHITMAN, W. B., KIENE, R. P., HENRIKSEN, J. R., KING, G. M., BELAS, R., FUQUA, C., BRINKAC, L., LEWIS, M., JOHRI, S., WEAVER, B., PAI, G., EISEN, J. A., RAHE, E., SHELDON, W. M., YE, W. Y., MILLER, T. R., CARLTON, J., RASKO, D. A., PAULSEN, I. T., REN, Q. H., DAUGHERTY, S. C., DEBOY, R. T., DODSON, R. J., DURKIN, A. S., MADUPU, R., NELSON, W. C., SULLIVAN, S. A., ROSOVITZ, M. J., HAFT, D. H., SELENGUT, J. & WARD, N. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature,* **432,** 910-913.

MORAN, M. A., GONZALEZ, J. M. & KIENE, R. P. 2003. Linking a bacterial taxon to sulfur cycling in the sea: studies of the marine *Roseobacter* group. *Geomicrobiol J,* **20,** 375-388.

NEVITT, G. A. 2008. Sensory ecology on the high seas: the odor world of the procellariiform seabirds. *J Exp Biol,* **211,** 1706-1713.

NEWTON, R. J., GRIFFIN, L. E., BOWLES, K. M., MEILE, C., GIFFORD, S., GIVENS, C. E., HOWARD, E. C., KING, E., OAKLEY, C. A., REISCH, C. R., RINTA-KANTO, J. M., SHARMA, S., SUN, S., VARALJAY, V., VILA-COSTA, M., WESTRICH, J. R. & MORAN, M. A. 2010. Genome characteristics of a generalist marine bacterial lineage. *ISME J*. **4,** 784-798.

NGUYEN, B. C., BELVISO, S., MIHALOPOULOS, N., GOSTAN, J. & NIVAL, P. 1988. Dimethyl sulfide production during natural phytoplanktonic blooms. *Mar Chem,* **24,** 133-141.

NOBLE, R. & FUHRMAN, J. 1997. Virus decay and its causes in coastal waters. *Appl Environ Microbiol,* **63,** 77-83.

NORDSTRÖM, K. & FORSGREN, A. 1974. Effect of protein A on adsorption of bacteriophages to *Staphylococcus aureus. J Virol,* **14,** 198-202.

OHMORI, H., HAYNES, L. L. & ROTHMAN-DENES, L. B. 1988. Structure of the ends of the coliphage N4 genome. *J Mol Biol,* **202,** 1-10.

PATTEE, P. A. 1966. Use of tetrazolium for improved resolution of bacteriophage plaques. *J Bacteriol,* **92,** 787-788.

PAUL, J. H. 2008. Prophages in marine bacteria: dangerous molecular times bombs or the key to survival in the seas? *ISME J,* **2,** 579-589.

PAUL, J. H., JIANG, S. C. & ROSE, J. B. 1991. Concentration of viruses and dissolved DNA from aquatic environments by vortex flow filtration. *Appl Environ Microbiol,* **57,** 2197-2204.

PAUL, J. H. & SULLIVAN, M. B. 2005. Marine phage genomics: what have we learned? *Curr Opin Biotechnol,* **16,** 299-307.

PEMBERTO, J. M. 1973. F116 - DNA bacteriophage specific for pili of *Pseudomonas aeruginosa* strain Pao. *Virology,* **55,** 558-560.

PRADELLA, S., PÄUKER, O. & PETERSEN, J. 2010. Genome organisation of the marine Roseobacter clade member *Marinovum algicola*. *Arch Microbiol,* **192,** 115-126.

PROCTOR, L. M. 1997. Advances in the study of marine viruses. *Microsc Res Tech,* **37,** 136-161.

PROCTOR, L. M. & FUHRMAN, J. A. 1990. Viral mortality of marine bacteria and cyanobacteria. *Nature,* **343,** 60-62.

RANDALL-HAZELBAUER, L. & SCHWARTZ, M. 1973. Isolation of bacteriophage lambda receptor from *Escherichia coli. J Bacteriol,* **116,** 1436-1446.

RAO, D., WEBB, J. S. & KJELLEBERG, S. 2006. Microbial colonization and competition on the marine *Alga ulva* australis. *Appl Environ Microbiol,* **72,** 5547-5555.

RAPSON, M. 2002. *The biology of Staphylococcus aureus specific bacteriophages and the assessment of their potential as therapeutic agents.* University of Warwick.

RESCH, A., FEHRENBACHER, B., EISELE, K., SCHALLER, M. & GOTZ, F. 2005. Phage release from biofilm and planktonic *Staphylococcus aureus* cells. *FEMS Microbiol Lett,* **252,** 89-96.

RITZ, D. & BECKWITH, J. 2001. Roles of the thiol-redox pathways in bacteria. *Ann Rev Microbiol,* **55,** 21-48.

ROHWER, F., SEGALL, A., STEWARD, G., SEGURITAN, V., BREITBART, M., WOLVEN, F. & AZAM, F. 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol Oceanogr,* **45,** 408-418.

ROSSMANN, M. G., MESYANZHINOV, V. V., ARISAKA, F. & LEIMAN, P. G. 2004. The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol,* **14,** 171-180.

ROUCOURT, B. & LAVIGNE, R. 2009. The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environ Microbiol,* **11,** 2789-2805.

RUSCH, D. B., HALPERN, A. L., SUTTON, G., HEIDELBERG, K. B., WILLIAMSON, S., YOOSEPH, S., WU, D. Y., EISEN, J. A., HOFFMAN, J. M., REMINGTON, K., BEESON, K., TRAN, B., SMITH, H., BADEN-TILLSON, H., STEWART, C., THORPE, J., FREEMAN, J., ANDREWS-PFANNKOCH, C., VENTER, J. E., LI, K., KRAVITZ, S., HEIDELBERG, J. F., UTTERBACK, T., ROGERS, Y. H., FALCON, L. I., SOUZA, V., BONILLA-ROSSO, G., EGUIARTE, L. E., KARL, D. M., SATHYENDRANATH, S., PLATT, T., BERMINGHAM, E., GALLARDO, V., TAMAYO-CASTILLO, G., FERRARI, M. R., STRAUSBERG, R. L., NEALSON, K., FRIEDMAN, R., FRAZIER, M. & VENTER, J. C. 2007. The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *Plos Biol,* **5,** 398-431.

SANTOS, S., CARVALHO, C., SILLANKORVA, S., NICOLAU, A., FERREIRA, E. & AZEREDO, J. 2009. The use of antibiotics to improve phage detection and enumeration by the double-layer agar technique. *BMC Microbiol,* **9,** 148.

SCHAEFER, J. K., GOODWIN, K. D., MCDONALD, I. R., MURRELL, J. C. & OREMLAND, R. S. 2002. *Leisingera methylohalidivorans* gen. nov., sp nov., a marine methylotroph that grows on methyl bromide. *Int J Syst Evol Microbiol,* **52,** 851-859.

SCHÄFER, H., MCDONALD, I. R., NIGHTINGALE, P. D. & MURRELL, J. C. 2005. Evidence for the presence of a CmuA methyltransferase pathway in novel marine methyl halide-oxidizing bacteria. *Environ Microbiol,* **7,** 839-852.

SCHOLL, D. & MERRIL, C. 2005. Polysaccharide-degrading phages. *In:* WALDOR, M. K., FRIEDMAN, D. I. & ADHYA, S. L. (eds.) *Phages: their role in bacterial pathogenesis and biotechnology.* Washington DC: ASM Press.

SCHOLL, D., ROGERS, S., ADHYA, S. & MERRIL, C. R. 2001. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*. *J Virol,* **75,** 2509-2515.

SCHUCH, R. & FISCHETTI, V. A. 2009. The secret life of the anthrax agent *Bacillus anthracis*: bacteriophage-mediated ecological adaptations. *PLoS One,* **4,** e6532.

SCHWARTZ, M. 1976. The adsorption of coliphage lambda to its host: effect of variations in the surface density of receptor and in phage-receptor affinity. *J Mol Biol,* **103,** 521-536.

SHARON, I., ALPEROVITCH, A., ROHWER, F., HAYNES, M., GLASER, F., ATAMNA-ISMAEEL, N., PINTER, R. Y., PARTENSKY, F., KOONIN, E. V., WOLF, Y. I., NELSON, N. & BEJA, O. 2009. Photosystem I gene cassettes are present in marine virus genomes. *Nature,* **461,** 258-262.

SHEA, T. B. & SEAMAN, E. 1984. Sp3 - a flagellotropic bacteriophage of *Bacillus subtilis*. *J Gen Virol,* **65,** 2073-2076.

SHIBA, T. 1991. *Roseobacter litoralis* gen. nov., sp. nov., and *Roseobacter denitrificans* sp. nov., aerobic pink-pigmented bacteria which contain bacteriochlorophyll a. *Syst Appl Microbiol,* **14,** 140-145.

SILVERMANN, M., BELAS, R. & SIMON, M. 1984. Genetic control of bacterial adhesion. *In:* MARSHALL, K. C. (ed.) *Microbial adhesion and aggregation.* Berlin: Springer.

SIMÓ, R. 2001. Production of atmospheric sulfur by oceanic plankton: biogeochemical, ecological and evolutionary links. *Trends Ecol Evol,* **16,** 287-294.

SLIGHTOM, R. N. & BUCHAN, A. 2009. Surface dolonization by marine Roseobacters: integrating genotype and phenotype. *Appl Environ Microbiol,* **75,** 6027-6037.

SOLIOZ, M. & MARRS, B. 1977. Gene transfer agent of *Rhodopseudomonas capsulata* - purification and characterisation of its nucleic acid. *Arch Biochem Biophys,* **181,** 300-307.

SPENCER, R. 1955. A marine bacteriophage. *Nature,* **175,** 690-691.

SULLIVAN, M. B., LINDELL, D., LEE, J. A., THOMPSON, L. R., BIELAWSKI, J. P. & CHISHOLM, S. W. 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol,* **4,** e234.

SUSSKIND, M. M. & BOTSTEIN, D. 1978. Molecular genetics of bacteriophage P22. *Microbiol Rev,* **42,** 385–413.

SUTTLE, C. A. 1993. Enumeration and Isolation of Viruses. *In:* KEMP, P. F., SHERR, B. F., SHERR, E. B. & COLE, J. J. (eds.) *Handbook of Methods in Aquatic Ecology.* Lewis Publishers.

SUTTLE, C. A. 2005. Viruses in the sea. *Nature,* **437,** 356-361.

SUTTLE, C. A. & CHAN, A. M. 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol,* **60,** 3167-3174.

SUZUKI, M. T., PRESTON, C. M., BÉJÀ, O., DE LA TORRE, J. R., STEWARD, G. F. & DELONG, E. F. 2004. Phylogenetic screening of ribosomal RNA gene-containing clones in Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey Bay *Microb Ecol,* **48,** 473-488.

SWINGLEY, W. D., SADEKAR, S., MASTRIAN, S. D., MATTHIES, H. J., HAO, J., RAMOS, H., ACHARYA, C. R., CONRAD, A. L., TAYLOR, H. L., DEJESA, L. C., SHAH, M. K., O'HUALLACHAIN, M. E., LINCE, M. T., BLANKENSHIP, R. E., BEATTY, J. T. & TOUCHMAN, J. W. 2007. The complete genome sequence of *Roseobacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. *J Bacteriol,* **189,** 683-690.

TESKE, A., BRINKHOFF, T., MUYZER, G., MOSER, D. P., RETHMEIER, J. & JANNASCH, H. W. 2000. Diversity of thiosulfate-oxidizing bacteria from marine sediments and hydrothermal vents. *Appl Environ Microbiol,* **66,** 3125-3133.

TÉTART, F., DESPLATS, C. & KRISCH, H. M. 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *J Mol Biol,* **282,** 543-556.

THINGSTAD, T. F. 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.

THINGSTAD, T. F. & LIGNELL, R. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol,* **13,** 19-27.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucl Acids Res,* **22,** 4673-4680.

TODD, J. D., CURSON, A. R. J., DUPONT, C. L., NICHOLSON, P. & JOHNSTON, A. W. B. 2009. The *dddP* gene, encoding a novel enzyme that converts dimethylsulfoniopropionate into dimethyl sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. *Environ Microbiol,* **11,** 1376-1385.

TODD, J. D., ROGERS, R., LI, Y. G., WEXLER, M., BOND, P. L., SUN, L., CURSON, A. R. J., MALIN, G., STEINKE, M. & JOHNSTON, A. W. B. 2007. Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science,* **315,** 666-669.

TORRELLA, F. & MORITA, R. Y. 1979. Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, Oregon - ecological and taxonomical implications. *Appl Environ Microbiol,* **37,** 774-778.

TORSVIK, T. & DUNDAS, I. D. 1980. Persisting phage infection in *Halobacterium salinarium* str. 1. *J Gen Virol,* **47,** 29-36.

TOUSSAINT, A., GAMA, M. J., LAACHOUCH, J., MAENHAUTMICHEL, G. & MHAMMEDIALAOUI, A. 1994. Regulation of bacteriophage Mu transposition. *Genetica,* **93,** 27-39.

TRINGE, S. G., VON MERING, C., KOBAYASHI, A., SALAMOV, A. A., CHEN, K., CHANG, H. W., PODAR, M., SHORT, J. M., MATHUR, E. J., DETTER, J. C., BORK, P., HUGENHOLTZ, P. & RUBIN, E. M. 2005. Comparative metagenomics of microbial communities. *Science,* **308,** 554-557.

TUOMI, P., TORSVIK, T., HELDAL, M. & BRATBAK, G. 1997. Bacterial population dynamics in a meromictic lake. *Appl Environ Microbiol,* **63,** 2181-2188.

URBANCE, J. W., BRATINA, B. J., STODDARD, S. F. & SCHMIDT, T. M. 2001. Taxonomic characterization of *Ketogulonigenium vulgare* gen. nov., sp. nov. and *Ketogulonigenium robustum* sp. nov., which oxidize L-sorbose to 2-keto-L-gulonic acid *Int J Syst Evol Microbiol,* **51,** 1059-1070.

VAIRAVAMURTHY, A., ANDREAE, M. O. & IVERSEN, R. L. 1985. Biosynthesis of DMS and DMSP by *Hymenomonas carterae* in relation to sulfur source and salinity variations. *Limnol Oceanogr,* **30,** 59.

VALE, P. F. & LITTLE, T. J. 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. *Proc Biol Sci*. **277,** 2097-2103.

VAN REGENMORTEL, M. H. V. 1992. Concept of virus species. *Biodiversity Conserv,* **1,** 253-266.

VAN REGENMORTEL, M. H. V. 2007. Virus species and virus identification: past and current controversies. *Infect Genet Evol,* **7,** 133-144.

VAN VALEN, L. 1973. A new evolutionary law. *Evol Theory,* **1,** 1-30.

VANDECANDELAERE, I., NERCESSIAN, O., SEGAERT, E., ACHOUAK, W., FAIMALI, M. & VANDAMME, P. 2008. *Ruegeria scottomollicae* sp. nov., isolated from a marine electroactive biofilm. *Int J Syst Evol Microbiol,* **58,** 2726-2733.

VANDECANDELAERE, I., NERCESSIAN, O., SEGAERT, E., ACHOUAK, W., MOLLICA, A., FAIMALI, M. & VANDAMME, P. 2009. *Nautella italica*

gen. nov., sp. nov., isolated from a marine electroactive biofilm. *Int J Syst Evol Microbiol,* **59,** 811-817.

VENTER, J. C., REMINGTON, K., HEIDELBERG, J. F., HALPERN, A. L., RUSCH, D., EISEN, J. A., WU, D., PAULSEN, I., NELSON, K. E., NELSON, W., FOUTS, D. E., LEVY, S., KNAP, A. H., LOMAS, M. W., NEALSON, K., WHITE, O., PETERSON, J., HOFFMAN, J., PARSONS, R., BADEN-TILLSON, H., PFANNKOCH, C., ROGERS, Y.-H. & SMITH, H. O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science,* **304,** 66-74.

WAGNER-DOBLER, I. & BIEBL, H. 2006. Environmental biology of the marine *Roseobacter* lineage. *Ann Rev Microbiol,* **60,** 255-280.

WATERBURY, J. B. & WILLEY. M. 1989. Isolation and Growth of Marine Planktonic Cyanobacteria. In *Methods in Enzymology Vol. 167* Academic Press.

WATTS, S. F. 2000. The mass budgets of carbonyl sulfide, dimethyl sulfide, carbon disulfide and hydrogen sulfide. *Atmos Environ,* **34,** 761-779.

WEIGEL, C. & SEITZ, H. 2006. Bacteriophage replication modules. *FEMS Microbiol Rev,* **30,** 321-381.

WEINBAUER, M., WILHELM, S., SUTTLE, C. & GARZA, D. 1997. Photoreactivation compensates for UV damage and restores infectivity to natural marine virus communities. *Appl Environ Microbiol,* **63,** 2200-2205.

WEINBAUER, M. G. 2004. Ecology of prokaryotic viruses. *FEMS Microbiol Rev,* **28,** 127-181.

WEINBAUER, M. G., FUKS, D., PUSKARIC, S. & PEDUZZI, P. 1995. Diel, seasonal, and depth-related variability of viruses and dissolved DNA in the Northern Adriatic sea. *Microb Ecol,* **30,** 25-41.

WEINBAUER, M. G. & PEDUZZI, P. 1994. Frequency, size and distribution of bacteriophages in different marine bacterial morphotypes. *Mar Ecol Prog Ser,* **108,** 11-20.

WIGGINS, B. A. & ALEXANDER, M. 1985. Minimum bacterial density for bacteriophage replication: implications for significance of bacteriophages in natural ecosystems. *Appl. Environ. Microbiol.,* **49,** 19-23.

WIKOFF, W. R., LILJAS, L., DUDA, R. L., TSURUTA, H., HENDRIX, R. W. & JOHNSON, J. E. 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science,* **289,** 2129-2133.

WILHELM, S. W. & SUTTLE, C. A. 1999. Viruses and nutrient cycles in the sea. *Bioscience,* **49,** 781-788.

WILLIAMSON, S. J., RUSCH, D. B., YOOSEPH, S., HALPERN, A. L., HEIDELBERG, K. B., GLASS, J. I., ANDREWS-PFANNKOCH, C., FADROSH, D., MILLER, C. S., SUTTON, G., FRAZIER, M. & VENTER, J. C. 2008. The Sorcerer II global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *Plos One,* **3,** e1456.

WILSON, W. H., CARR, N. G. & MANN, N. H. 1996. The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp WH7803. *J Phycol,* **32,** 506-516.

WILSON, W. H., TARRAN, G. & ZUBKOV, M. V. 2002. Virus dynamics in a coccolithophore-dominated bloom in the North Sea. *Deep Sea Res,* **49,** 2951-2963.

WINGET, D. M. & WOMMACK, K. E. 2009. Diel and daily fluctuations in virioplankton production in coastal ecosystems. *Environ Microbiol,* **11,** 2904-2914.

WINTER, C., BOUVIER, T., WEINBAUER, M. G. & THINGSTAD, T. F. 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.

WOMMACK, K. E. & COLWELL, R. R. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev,* **64,** 69-114.

WOMMACK, K. E., HILL, R. T., KESSEL, M., RUSSEKCOHEN, E. & COLWELL, R. R. 1992. Distribution of viruses in the Chesapeake Bay. *Appl Environ Microbiol,* **58,** 2965-2970.

WOOD, S. R., KIRKHAM, J., MARSH, P. D., SHORE, R. C., NATTRESS, B. & ROBINSON, C. 2000. Architecture of intact natural human plaque biofilms studied by confocal laser scanning microscopy. *J Dent Res,* **79,** 21-27.

YIN, Y. & FISCHER, D. 2008. Identification and investigation of ORFans in the viral world. *BMC Genomics,* **9,** 24.

YOCH, D. C. 2002. Dimethylsulfoniopropionate: Its sources, role in the marine food web, and biological degradation to dimethylsulfide. *Appl Environ Microbiol,* **68,** 5804-5815.

YOSHIDA, H., KINOSHITA, K. & ASHIDA, M. 1996. Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. *J Biol Chem,* **271,** 13854-13860.

YUZENKOVA, J., NECHAEV, S., BERLIN, J., ROGULJA, D., KUZNEDELOV, K., INMAN, R., MUSHEGIAN, A. & SEVERINOV, K. 2003. Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage. *J Micro Biol,* **330,** 735-748.

ZEIDNER, G., BIELAWSKI, J. P., SHMOISH, M., SCANLAN, D. J., SABEHI, G. & BÉJÀ, O. 2005. Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ Microbiol,* **7,** 1505-1513.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruihn graphs. *Genome Res,* **18,** 821:829.

ZHANG, Y. Y. & JIAO, N. Z. 2009. Roseophage RDJL Phi 1, infecting the aerobic anoxygenic phototrophic bacterium *Roseobacter denitrificans* OCh114. *Appl Environ Microbiol,* **75,** 1745-1749.

ZHAO, Y., WANG, K., ACKERMANN, H. W., HALDEN, R. U., JIAO, N. Z. & CHEN, F. 2010. Searching for a "hidden" prophage in a marine bacterium. *Appl Environ Microbiol,* **76,** 589-595.

ZHAO, Y. L., WANG, K., JIAO, N. Z. & CHEN, F. 2009. Genome sequences of two novel phages infecting marine Roseobacters. *Environ Microbiol,* **11,** 2055-2064.

ZINDER, N. D. & LEDERBERBERG, J. 1952. Genetic exchange in *Salmonella*. *J Bacteriol,* **64,** 679-699.

ZOBELL, C. E. 1943. The effect of solid surfaces upon bacterial acitivity. *J Bacteriol,* **46,** 38-56.

ZUBKOV, M. V., FUCHS, B. M., ARCHER, S. D., KIENE, R. P., AMANN, R. & BURKILL, P. H. 2001. Linking the composition of bacterioplankton to rapid turnover of dissolved dimethylsulphoniopropionate in an algal bloom in the North Sea. *Environ Microbiol,* **3,** 304-311.

Appendix

Table A. 1 Table of the 32 sequenced Roseobacter genomes. Data taken from Newton *et al.*, 2010

Organism	Isolation source	Genome size (Mb)	Status	Genome completeness (%)
Dinoroseobacter shibae DFL 12	Prorocentrum lima, Bay of Tokyo	4.35	Closed	100
Jannaschia sp. CCS1	Bodega Head, USA	4.40	Closed	100
Roseobacter denitrificans Och 114	Enteromorpha linza, Australia	4.13	Closed	100
Ruegeria pomeroyi DSS-3	Coastal surface water, Georgia, USA	4.60	Closed	100
Ruegeria sp. TM1040	*Pfiesteria piscicda,* Chesapeake Bay, USA	4.15	Closed	100
Loktanella vestfoldensis SKA53	Surface water, North Atlantic	3.06	Draft	99
Maritimibacter alkaliphilus HTCC2654	10 m water, Sargasso Sea	4.53	Draft	99
Pelagibaca bermudensis HTCC 3601	10 m water, Sargasso Sea	5.43	Draft	98
Oceanibulbus indolifex HEL-45	10 m water, Sargasso Sea	4.11	Draft	100
Oceanibulbus batsensis HTCC2597	10 m water, North Sea	4.44	Draft	99
Oceanicola granulosus HTCC2516	10 m water, Sargasso Sea	4.04	Draft	100
Octadecabacter antarcticus 307	McMurdo Sound	4.89	Draft	100
Octadecabacter arcticus 238	Offshore water, Deadhorse, Alaska	5.39	Draft	96
Phaeobacter gallaeciensis 2.10	*Ulva lactuca,* Australia	4.16	Draft	100
Phaeobacter gallaeciensis BS107	*Pecten maximus*, Spain	4.23	Draft	100
Rhodobacterales bacterium HTCC2083	10 m coastal water, Oregon, USA	4.02	Draft	99
Rhodobacterales bacterium HTCC2150	Surface water, Oregon, USA	3.58	Draft	98
Rhodobacterales bacterium HTCC2255	10 m coastal water, Oregon, USA	4.81	Draft	96
Rhodobacterales bacterium Y4I	Coastal water, Georgia, USA	4.33	Draft	99
Roseobacter litoralis Och149	Seaweed	4.68	Draft	99
Roseobacter sp. AzwK-3b	Estuary, Monterey Bay	4.18	Draft	100
Roseobacter Sp. CCS2	Bodega Head, USA	3.50	Draft	99
Roseobacter Sp. GAI101	Coastal water, Georgia, USA	4.25	Draft	99
Roseobacter sp. MED193	1 m water, North West Mediterranean	4.65	Draft	100
Roseobacter sp. SK209-2-6	267 m water, Arabian Sea	4.56	Draft	100
Roseovarius nubinhibens ISM	Surface water, Caribbean Sea	3.67	Draft	100
Roseovarius sp. 217	L4 surface water, South East England	4.76	Draft	100
Roseovarius sp. TM1035	*Pfiesteria piscicda,* Chesapeake Bay, USA	4.21	Draft	100

Table A. 1 cont.

Organism	Isolation source	Genome size (Mb)	Status	Genome completeness (%)
Ruegeria sp. R11	*Delisea pulchra*, Australia	3.82	Draft	98
Sagittula stellata E-37	Coastal water, Georgia, USA	5.26	Draft	98
Sulfitobacter NAS-14.1	Coastal water, Georgia, USA	4.00	Draft	100
Sulfitobacter sp. EE-36	Surface water, North Atlantic	3.54	Draft	100

Table A.2 Identification and description of proteins in bands, seen in Fig. 6.3, analysed by MS.

Band	Protein	Description	No. of peptides	Mw/ kDa	*in silico* predicted Mw/kDa	Calculated Mw/kDa
RLP1						
B1	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	5	7.5	58.9	109
	gp28	Major capsid protein (MCP)	4	51.4		
B2	gp28	MCP	6	5.1	90.5	92.3
	gp16	Host-like protein	4	10.4		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	2	75		
B3	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	8	7.5	117.3	82.2
	gp28	MCP	5	51.4		
	gp30	N4 gp54-like structural protein	2	43.7		
	host contamination	ZP_01037873 hypothetical protein, similar to Silicibacter sp. TrichCH4B lipoprotein COG0614 ABC-type Fe3+-hydroxamate transport system, periplasmic component	2	14.7		
B4	gp28	MCP	8	51.4	58.9	70.3
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	7	7.5		
B5	gp28	MCP	12	51.4	73.8	63.3
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	8	7.5		
	host contamination	ZP_01037873 hypothetical protein, similar to Silicibacter sp. TrichCH4B lipoprotein COG0614 ABC-type Fe3+-hydroxamate transport system, periplasmic component	2	14.9		
B6	gp28	MCP	10	51.4	58.9	48.7
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	4	7.5		
B7	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	10	7.5	147.8	46.6
	gp28	MCP	7	51.4		
	gp25	N4 gp59-like, 94 kDa portal protein	7	88.9		

Table A.2 cont.

Band	Protein	Description	No. of peptides	Mw/ kDa	*in silico* predicted Mw/kDa	Calculated Mw/kDa
B8	gp28	MCP	8	51.4	176.2	44.7
	gp25	N4 gp59-like, 94 kDa portal protein	5	88.9		
	gp13	Shared with Roseobacter phages	3	35.9		
B9	gp28	MCP	11	51.4	120.8	44
	gp19	N4 gp67-like 30 kDa structural protein	3	25.1		
	host contamination	ZP_01036103 hypothetical protein, similar to Rsv. nubinhibens outer membrane porin	2	36.8		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	2	7.5		
B10	gp28	MCP	12	51.4	113.3	39.3
	gp19	N4 gp67-like 30 kDa structural protein	9	25.1		
	host contamination	ZP_01036103 hypothetical protein, similar to Rsv. nubinhibens outer membrane porin	2	36.8		
B11	gp28	MCP	11	51.4	183.7	37.2
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	4	7.5		
	gp25	N4 gp59-like, 94 kDa portal protein	4	88.9		
	gp13	Shared with Roseobacter phages	2	35.9		
B12	gp28	MCP	10	51.4	220.5	31
	gp25	N4 gp59-like, 94 kDa portal protein	6	88.9		
	gp13	Shared with Roseobacter phages	6	35.9		
	host contamination	ZP_01036103 hypothetical protein, similar to Rsv. nubinhibens outer membrane porin	2	36.8		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	2	7.5		
B13	gp28	MCP	13	51.4	177.1	25.6
	gp25	N4 gp59-like, 94 kDa portal protein	6	88.9		
	host contamination	ZP_01036103 hypothetical protein, similar to Rsv. nubinhibens outer membrane porin	4	36.8		
B14	gp25	N4 gp59-like, 94 kDa portal protein	13	88.9	228.3	24.9
	gp30	N4 gp54-like structural protein	9	43.7		
	gp28	MCP	7	51.4		
	host contamination	ZP_01036103 hypothetical protein, similar to Rsv. nubinhibens outer membrane porin	4	36.8		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	3	7.5		
B15	gp30	N4 gp54-like structural protein	13	43.7	228.3	23.1
	gp28	MCP	12	51.4		
	gp25	N4 gp59-like, 94 kDa portal protein	6	88.9		
	host contamination	ZP_01036103 hypothetical protein, similar to Rsv. nubinhibens outer membrane porin	2	36.8		

Table A.2 cont.

Band	Protein	Description	No. of peptides	Mw/ kDa	*in silico* predicted Mw/kDa	Calculated Mw/kDa
B15	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	2	7.5	228.3	23.1
B16	gp28	MCP	14	51.4	191.5	20.7
	gp30	N4 gp54-like structural protein	10	43.7		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	9	7.5		
	gp25	N4 gp59-like, 94 kDa portal protein	4	88.9		
B17	gp25	N4 gp59-like, 94 kDa portal protein	17	88.9	147.8	16.3
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	10	7.5		
	gp28	MCP	10	51.4		
B18	gp28	MCP	11	51.4	147.8	15.1
B18	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	7	7.5	147.8	15.1
	gp25	N4 gp59-like, 94 kDa portal protein	5	88.9		
B19	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	21	7.5	191.5	14
	gp28	MCP	11	51.4		
	gp25	N4 gp59-like, 94 kDa portal protein	8	88.9		
	gp30	N4 gp54-like structural protein	2	43.7		
B20	gp28	MCP	15	51.4	147.8	12.1
	gp25	N4 gp59-like, 94 kDa portal protein	11	88.9		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	3	7.5		
B21	gp28	MCP	14	51.4	147.8	10
	gp25	N4 gp59-like, 94 kDa portal protein	6	88.9		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	3	7.5		
RPP1						
B22	gp25	N4 gp59-like, 94 kDa portal protein	25	88.9	177.8	99.8
	gp28	MCP	3	51.4		
B24	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	22	75.1	126.5	86.1
	gp28	MCP	4	51.4		
B24	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	7	75.1	218.4	78
	gp64	Shared with Roseobacter phages, similar to gp230 in *Pseudomonas* phage 201Φ2-1	5	68.2		
	gp33	Shared with Roseobacter phages, putative cell wall hydrolase	1	75.1		
B25	gp28	MCP	17	51.4	51.4	62.5
B26	gp28	MCP	9	51.4	51.4	51.4
B27	gp14	Shared with Roseobacter phages	5	36	87.4	41.2
	gp28	MCP	2	51.4		
B28	gp14	Shared with Roseobacter phages	6	36	87.4	39.2
	gp28	MCP	4	51.4		

Table A. 2 cont.

Band	Protein	Description	No. of peptides	Mw/ kDa	*in silico* predicted Mw/kDa	Calculated Mw/kDa
B29	gp19	30 kDa protein	19	25.2	76.6	27.1
	gp28	MCP	3	51.4		
B30	gp68	Shared with Roseobacter phages	1	44.4	95.8	23.4
	gp28	MCP	1	51.4		
B31	gp28	MCP	2	51.4	73.7	12.3
	gp15	Shared with Roseobacter phages	1	22.3		
B32	gp32	16.5 kDa protein	2	16.1	23.3	10.4
	gp71	Shared with Roseobacter phages	1	7.2		
B33	gp65	Shared with Roseobacter phages	3	12.5	12.5	9.6
B34	gp16	Host-like protein	2	10.4	10.4	8.1

Table A.3 Full ProphageFinder results

Prophage	Length (kb)	Best BLAST hit	Phage	Accession number	E value
Rsv. 217 predicted prophage 1	13	hypothetical protein p25	Streptomyces phage VWB	NP_958267	0.005
		gp90	Mycobacterium phage corndog	NP_817941	0.004
		gp25	Klebsiella phage ΦKO2	YP_006605	5.E-28
		Gin G-segment invertase	Enterobacteria phage μ	NP_050655	1.E-27
		Cre cyclization recombinase	Enterobacteria phage P1	YP_006472	2.E-04
		ORF56	Lactococcus phage TP901-1	NP_112719	6.E-10
Rsv. 217 predicted prophage 2	6.6	gp3	Mycobacterium phage Rosebush	NP_817764	2.E-43
		gp5	Mycobacterium phage Rosebush	NP_817766	4.E-05
		GTP cyclohydrolase I family protein	Vibrio phage KVP40	NP_899371	2.E-04
		unknown, p65	Sinrhizobium phage PBC5	NP_542325	0.25
		hypothetical protein, p25	streptomyces phage VWB	NP_958267	0.086
Rsv. 217 predicted prophage 3	7.8	portal protein	Salmonella phage ST64B	NP_700377	8.E-16
		Pro-head protease	Salmonella phage ST64B	NP_700378	3.E-20
		major capsid protein precursor	Salmonella phage ST64B	NP_700379	1.E-49
		tail length tape measure protein	Enterobacteria phage HK022	NP_037676	0.001
		hypothetical protein	Pseudomonas phage D3112	NP_938257	0.006
Rsv. 217 predicted prophage 4	14.2	integrase	Enterobacteria phage KH97	NP_037720	5.E-05
		tail length tape measure protein	Xanthomonas phage XP10	NP_858965	5.E-07
		tail length tape measure protein	Enterobacteria phage HK022	NP_037676	5.E-05
		terminase large subunit	Enterobacteria phage HK97	NP_037698	2.E-44
		putative terminase (small subunit)	Burkholderia phage ΦE125	NP_536357	3.E-09

Table A.3 cont.

Prophage	Length (kb)	Best BLAST hit	Phage	Accession number	E value
Rsv. 217 predicted prophage 4	14.2	gp9	Enterobacteria phage HK022	NP_037670	0.15
		ORF19	Bacillus phage Φ105	NP_690803	5.E-07
		putative major capsid protein	Enterobacteria phage ΦP27	NP_543092	3.E-38
		putative prohead protease	Enterobacteria phage ΦP27	NP_543091	3.E-13
		gp3	Burkholderia phage Φ1026b	NP_945033	7.E-34
Rsv. 217 predicted prophage 5	17.1	putative helicase	Lactobacillus phage A2	NP_680515	2.E-24
		Mth		NP_878240	2.E-18
		integrase	Lactococcus phage TP901-1	NP_112664	6.E-21
		conserved hypothetical protein	Burkholderia cepacia phage BcepNazgul	NP_919008	1.E-23
		transferase	Streptococcus phage EJ-1	NP_945276	6.E-43
		hypothetical protein, p34	Xanthomonas phage XP10	NP_858981	6.E-06
		hypothetical protein, p35	Enterobacteria phage ΦP27	NP_543087	0.25
		Gin G-segment invertase	Enterobacteria phage μ	NP_050655	5.E-07
Rsv. 217 predicted prophage 6	30.5	cI repressor protein	Pseudomonas phage D3	NP_061565	2.E-08
		DNA transposition protein	Pseudomonas phage D3112	NP_938214	2.E-43
		gp5	Burkholderia phage Bcepμ	YP_024678	3.E-36
		hypothetical protein, p16	Enterobacteria phage μ	NP_050620	6.E-05
		ORF19	Vibrio phage VHML	NP_758912	2.E-06
		hypothetical protein, p26	Enterobacteria phage μ	NP_050630	8.E-05
		gp27	Burkholderia phage Bcepμ	YP_024700	1.E-05
		gp28	Burkholderia phage Bcepμ	YP_024701	2.E-95
		portal protein	Pseudomonas phage B3	YP_164068	4.E-98
		hypothetical protein, ORF33	Pseudomonas phage B3	YP_164069	2.E-34
		hypothetical protein, ORF35	Pseudomonas phage B3	YP_164071	4.E-13
		gp32	Burkholderia phage Bcepμ	YP_024705	1.E-51
		capsid protein	Pseudomonas phage B3	YP_164075	4.E-60
		unknown, p9	Enterobacteria phage SfV	NP_599041	7.E-04
		ORF53	Pseudomonas phage D3	NP_061549	0.19
		putative tape measure protein	Mycobacterium phage TM4	NP_569753	0.15
		hypothetical protein, p19	Yersinia phage PY54	NP_892065	0.004
		tail protein	Yersinia phage PY54	NP_892067	0.03
		gp20	Burkholderia phage Φ1026b	NP_945050	1.E-04
		gp21	Klebsiella phage ΦKO2	YP_006601	8.E-06

Table A.3 cont.

Prophage	Length (kb)	Best BLAST hit	Phage	Accession number	E value
Rsv. nubinhibens predicted prophage 1	7.4	putative portal protein	Enterobacteria phage ΦP27	NP_543090	7.E-14
		gp4	Burkholderia phage Φ1026b	NP_945034	7.E-14
		putative major capsid protein	Enterobacteria phage ΦP27	NP_543092	1.E-56
		hypothetical protein	Streptococcus pyogenes phage 315.5	NP_795643	8.E-04
		tail length tape measure protein	Enterobacteria phage HK022	NP_037676	6.E-09
		hypothetical protein	Pseudomonas phage D3112	NP_938257	3.E-05
Rsv. nubinhibens predicted prophage 2	9.1	similar to DNA helicase	Rhodothermus phage RM378	NP_835691	1.E-07
		gp179	Mycobacterium phage Bxz1	NP_818230	0.099
		gene 33 protein	Enterobacteria phage Sf6	NP_958209	0.22
		putative DNA cytosine methylase	Burkholderia phage ΦE125	NP_536413	1.E-04
		Res	Enterbacteria phage P1	YP_006476	0.001
		putative exonuclease	Staphylococcus phage K	YP_024504	0.001
Rsv. nubinhibens predicted prophage 3	16.9	gp12	Burkholderia phage Bcep1	NP_944320	2.E-35
		hypothetical protein, p19	Burkholderia phage Bcep22	NP_944247	0.38
		ORF19	Bacillus phage Φ105	NP_690803	2.E-14
		gp5	Mycobacteriuma phage Cjw1	NP_817455	2.E-04
		terminase	Pseudomonas phage D3	NP_061498	1.E-59
		putative portal protein	Klebsiella phage ΦKO2	YP_006584	2.E-24
		gp5	Enterobacteria phage N15	NP_046900	1.E-05
		putative major capsid protein	Burkholderia phage ΦE125	NP_536362	6.E-55
		ORF19	Pseudomonas phage D3	NP_061515	4.E-52
Roseobacter denitrificans predicted prophage 1	14.1	gp199	Mycobacterium phage Bxz1	NP_818250	2.E-39
		gp2	Burkholderia phage Bcepμ	YP_024675	1.E-02
		major head subunit	Enterobacteria phage μ	NP_050638	1.E-57
		putative protease protein	Enterobacteria phage μ	NP_050636	8.E-14
		flap endonuclease	Enterobacteria phage T5	YP_006958	3.E-06
		DNA polymerase	Enterobacteria phage T5	YP_006950	6.E-35
Roseobacter denitrificans predicted prophage 2	7.8	hypothetical protein	Pseudomonas phage D3112	NP_938257	1.E-04
		tail length tape measure protein	Enterobacteria phage HK022	NP_037676	0.004
		putative major capsid protein	Enterobacteria phage ΦP27	NP_543092	2.E-52
		putative prohead protease	Enterobacteria phage ΦP27	NP_543091	3.E-23
		portal protein	Salmonella phage ST64B	NP_700377	9.E-15