Abstract

Title of Document: **ISOLATION, GENOMICS AND ECOLOGY OF BACTERIOPHAGES INFECTING MARINE ROSEOBACTERS**

Yuanchao Zhan, Doctor of Philosophy, 2017

Directed by: Dr. Feng Chen, Professor,
Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science

Viruses are the most abundant biological entities in seawater. They influence the population dynamics, genetic heterogeneity, and biogeochemical cycles in marine ecosystems. Isolation and characterization of viruses which infect specific hosts have greatly advanced our knowledge on the biological and ecological interactions between viruses and their hosts. *Roseobacter* is an important lineage of marine bacteria which are genetically diverse, abundant and ubiquitous in the ocean. Roseobacters can make up to 25% of bacterial communities in coastal environments and play an active role in the marine sulfur cycle. However, only few bacteriophages which infect marine roseobacters had been isolated at the time when I began my studies. To understand the types of bacteriophages that infect roseobacters and how they interact with their hosts, I devoted my research to isolation and characterization of the bacteriophages infecting roseobacters (roseophages hereafter).

In this dissertation, fourteen different phages infecting a marine strain,

Ruegeria pomeroyi DSS-3, are described in terms of their morphology, growth, genomics and global distributions. These 14 roseophages were divided into four different groups: ssDNA, CbK-like, Chi-like, and N4-like roseophages. Two ssDNA phages belongs to an unclassified group of *Microviridae*. They contain only four ORFs with a genome size of 4.2 kb, representing the smallest and of all known ssDNA phage isolates. Interestingly, the ssDNA roseophages fall into a large group of uncultivated viral sequences identified by viral metagenomics. The isolation of CbK-like roseophages uncovers a new type of *Siphoviridae* infecting a member of *Roseobacter* lineage, Prior to this work, CbK-like phages had only been reported in a freshwater bacterium *Caulobacter*. The two CbK-like roseophage genomes are highly mosaic, containing features from siphoviruses, podoviruses, gene transfer agents, integrases and a large number of tRNAs. Chi-like siphophages are another newly discovered group of roseophages. Five different Chi-like phages (*Siphoviridae*) were isolated from DSS-3. A resistant strain of *R. pomeroyi* DSS-3 was found during superinfection with Chi-like roseophage DSS3Φ1. Genome sequencing confirmed that the resistant strain contains the intact genome of DSS3Φ1. The ability to integrate phage genome into host chromosome confirms that DSS3Φ1 is a temperate phage. Five N4-like roseophages of DSS-3 were isolated. They belong to the phage N4 lineage in *Podoviridae*. Genomes of N4-roseophages are highly syntenic, sharing a very similar genomic arrangement.

The genomic conservation of N4-like phages allowed me to design N4-like phage specific primers based on their DNA polymerase genes. The primer set was used to PCR amplify the DNA pol gene of N4-like phages from 56 DNA samples to

investigate the diversity and distribution of N4-like phages in the Chesapeake Bay. Surprisingly, N4-like phage sequences were only detected in the winter samples collected over two years. Metagenomic recruitments also confirmed that N4-like phages appear to prevail in the cold environment, such as Organic Lake, a hypersaline lake in Antarctica, where the temperature is usually below -10 °C. According to metagenomic analyses, homologs of other DSS-3 phages (non-N4-like) are present in freshwater and marine habitats, Antarctica, human gut and feces, and coral-associated environments. This wide range distribution of roseophages seems to reflect the cosmopolitan nature of the *Roseobacter* clade. The discovery of different types of phages infecting a single strain and their wide distribution suggest that we are only seeing the tip of the iceberg of phages.

ISOLATION, GENOMICS AND ECOLOGY OF BACTERIOPHAGES INFECTING MARINE ROSEOBACTERS

by

Yuanchao Zhan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:

Professor Feng Chen, Chair
Professor Alison Buchan
Professor Russell T. Hill
Professor Harold J. Schreier
Professor Eric K. Wommack

Dedication

I dedicate this dissertation to my husband Mingming Yu, and my parents Wenbin Zhan and Yuanhong Wang for their unconditional love and support.

Acknowledgements

First and foremost, I would like to thank Dr. Feng Chen, my mentor of throughout Ph.D study, for spending enormous amounts of time with me on designing experiments, discussing results, and revising manuscripts and thesis chapters. While he has a high expectation for me, he always encourages me to think creatively and work independently. I am grateful to his patience, open mind, trust, and friendship. I gained so much from him, scientifically and personally. Six years have gone by so quickly as if it just happened yesterday. I still remember when Dr. Chen picked me up at the midnight of August 4, 2011 when I first landed on the BWI airport.

I also extend sincere thanks to my other committee members for their guidance and expertise: Dr. Alison Buchan for initiating the diversity study on N4-like phages and forging a productive collaboration; Dr. Eric Wommack for providing sequencing support and suggestion on viral metagenomic project; Dr. Hal Schreier for his great course and discussion on bacterial physiology; and Dr. Russell Hill for scientific discussions and encouragements that helped to steer my project and my academic career.

I greatly thank all the members of the Chen laboratory in the past six years. Specially, I thank Yongle Xu for instructing me on basic laboratory skills; David Marson for lively discussions on all types of topics from science, travel, history, to politics; Zhao Zhao for the company as a wonderful lab mate and roommate. I also want to thank my lab mates Mengqi Sun, Ana Sosa Morfin, Daniel Fucich for being helpful lab mates in so many ways. I'd love to thank Dr. Tsetso Bachvaroff for his kind

helps on command lines and bioinformatics. Samuel Major, Shadaesha Green, Ernest Williams, Daniela Tizabi, Daniel Chen and Heidi Li are thanked for proofreading my thesis.

I acknowledge the four-year fellowship from the Chinese Scholar Council; the financial support from the Ratcliffe Environmental Entrepreneurs Fellowship; and the funding supports from the Maryland Industrial Partnership Program.

Finally, I am forever in debt to my parents who raised me up and always supported me along the way. And to my husband, who always stands next to me and provides moral and emotional support in my life.

State of Contribution

Dr. Sijun Huang, from South China Sea Institute of Oceanology, Chinese Academy of Sciences, provided the TEM image and genome sequence of DSS3Φ1. The genome of DSS3Φ8 was sequenced by Dr. Sonja Voget, from the Institute for Microbiology and Genetics at the University of Göttingen, and Dr. Meinhard Simon, from the Institute for Chemistry and Biology of the Marine Environment at the University of Oldenburg. Dr. Eric Wommack and Daniel J. Nasko from the College of Earth, Ocean, and Environment at the University of Delaware, helped sequencing a pooled sample of roseophage using PacBio. The primers targeting on the DNA polymerase gene of N4-like phages were designed by Dr. Alison Buchan from the Department of Microbiology at the University of Tennessee.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 The need to isolate marine viruses

Marine viruses first entered microbial ecologists' view in 1989 when Bergh *et al.* used transmission electron microscopy to illustrate the abundance of marine viruses in seawater (Bergh *et al.*, 1989). It has been estimated that there are approximately 10^6 virus-like particles (VLPs) ml^{-1} in open oceans and 10^7 VLPs ml^{-1} in productive coastal waters (Suttle 2005). Since then, extensive studies have been conducted to understand the role of viruses in the marine environment. It is clear now that marine viruses are involved in shaping host population structure, mediating genetic exchange between hosts, and modulating trophic transfer in marine food webs (Fuhrman, 1999; Wommack and Colwell, 2000; Suttle, 2005, 2007; Breitbart, 2012). Culture-independent marine viral metagenomic studies suggest that viruses are the largest genetic repertoire in the ocean, and a great number of marine viruses await discovery (Angly *et al.*, 2006; Hurwitz and Sullivan, 2013; Brum *et al.*, 2015; Paez-Espino *et al.*, 2016). While the rapid advance in high throughput sequencing technology allows me to explore microbial diversity with deep coverage and relatively low cost, isolation and characterization of infectious viruses lags behind. Isolating viruses is an important step towards understanding their morphology, infectivity, genetics/genomics, ecology and evolution. It is also now evident that isolation of viruses helps in the interpretation of rapidly increasing virome databases.

1.1.1 Isolation of marine viruses continues to surprise us

It has been estimated that there are thousands of different viral genotypes in every liter of seawater (Edwards and Rohwer, 2005). Variable morphotypes of marine

bacteriophages (viruses that infect bacteria) have been observed under transmission electronic microscopy (Brum et al. 2013). Marine phages are generally divided into tailed and non-tailed forms. Tailed phages fall into the order *Caudovirales*, all of which are dsDNA viruses. *Caudovirales* consists of three families: *Myoviridae*, *Podoviridae* and *Siphoviridae* (Figure 1.1 a-c) (Ackermann, 1992). Non-tailed phages often have ssDNA genome and belong to the *Microviridae* and *Inoviridae* families. Phages of these types have recently been isolated from marine systems (i.e., the strait of Öresund (Holmfeldt *et al.*, 2013) and Arctic sea ice (Yu *et al.*, 2014)).

The discovery of archaeal viruses provides an example of how viral isolation contributes to our knowledge on the diversity and evolution of viruses. Isolated archaeal viruses have many different types of morphology, such as being bottle-shaped, spindle-shaped, droplet-shaped, or spherical (Figure 1.1 d) (Prangishvili, 2013). These unusual virion morphologies have not been observed in bacteriophages or viruses infecting eukaryotes (Pietilä *et al.*, 2014). Genomic sequencing reveals an extreme diversity of archaeoviruses making up fifteen viral families that have been assigned to 29 different archaeovirus isolates (Dellas *et al.*, 2014). The characterization of archaeoviruses helps us understand the taxonomy and evolution of viruses. For example, the major capsid proteins of archaeoviruses, STIV and HSTV-1, share structural similarities with the viruses infecting organisms from other domains of life (Snyder *et al.*, 2015). This observation infers the existence of ancient viral lineages before the separation of three life domains. Studies on the replication cycle of STIV demonstrate that a cellular protein of STIV is essential for the replication cycle of several eukaryotic viruses, such as HIV and Ebola (Carlton and Martin-

Serrano, 2007; Snyder *et al.*, 2013). Isolation of *Acidianus convivator* virus ATV allows the discovery that viral tails are developed outside host cells after viral particles are released (Häring *et al.*, 2005). Despite that only a limited number of archaeoviruses have been isolated, they have shown a greater phenotypic diversity compared to bacteriophages (Nasir *et al.*, 2014).

The discovery of giant viruses also expands our knowledge on virology. Mimivirus, which infects ubiquitous protozoa amoeba, was the first isolated giant virus (Scola *et al.*, 2003). Mimivirus was first thought to be a bacterium due to its large particle size (Figure 1.1 e), but was identified as a virus 10 years after its discovery (Scola *et al.*, 2003). More giant viruses have been isolated since then (Arslan *et al.*, 2011; Philippe *et al.*, 2013; Legendre *et al.*, 2014, 2015). Currently, giant viruses can be divided into four distinct families, which are *Mimiviridae*, *Pandoraviridae*, Pithovirus and Mollivirus (Abergel *et al.*, 2015). Genomes of giant viruses can be as complicated as genomes of bacteria or parasitic eukaryotes. Many signature cellular genes, including transcription signaling, protein translation and delivery mechanism have been found in giant viruses (Abergel *et al.*, 2015). Interestingly, virophages which replicate within the giant viruses were discovered (Figure 1.1 f) (La Scola *et al.*, 2008). A recent study found that virophages can integrate into giant virus genome, and protect its protozoan host from being infected by giant viruses infection (Fischer and Hackl, 2016). The cell-like genetic capacity encoded by giant viruses has challenged our current view on the evolution of viruses. Discovery of novel viruses through isolation and laboratory study will no doubt continue to surprise us.

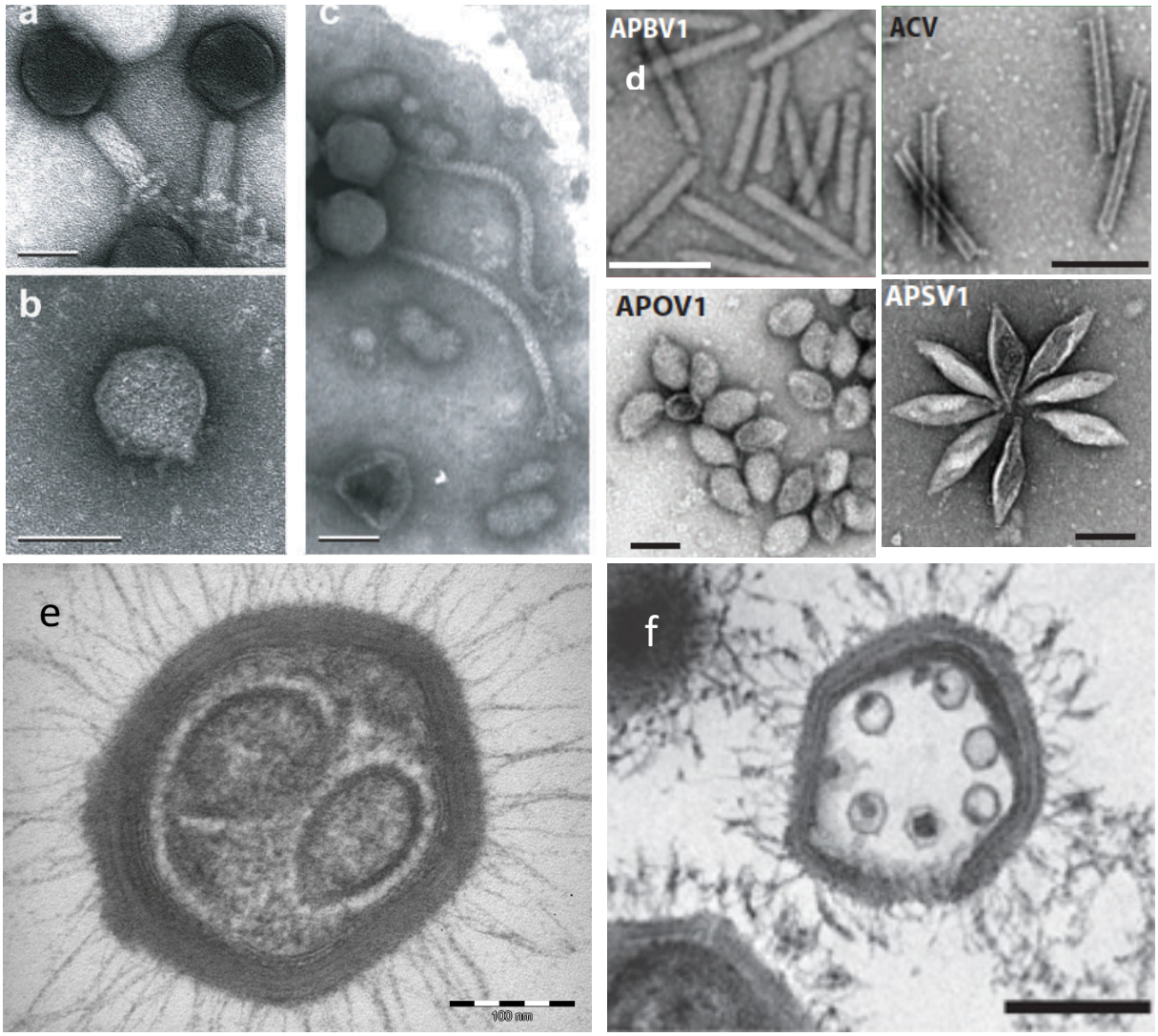


Figure 1.1 The diverse morphologies of marine viruses. (a-c) Typical morphology of marine bacteriophages: a) *Myoviridae*; b) *Podoviridae*; c) *Siphoviridae*. Scale bar = 50 nm. (d) Electron micrographs of archaeal virions collected form a coastal hot, nonacidic spring at Kagoshima in Japan. Scale bar = 100 nm. (e-f) Electron microscope image of e) mamavirus and f) mimivirus with virophage. Scale bar = 100 nm. Figures were adapted from Suttle 2007, Prangishvili 2013, and La Scola *et al.* 2008.

1.1.2 Marine bacteriophages with known genomes

Despite the ecological importance of marine viruses, the number of marine bacteriophages with known genomic sequences remains still limited, particularly compared to the genomes of marine bacteria (Figure 1.2). The first bacteriophage genome (ΦX174) which infects enterobacteria was sequenced in 1977 (Sanger *et al.*, 1977); twenty-two years later the first complete genome of a bacteriophage with marine origins (PM2), which infects the heterotrophic bacterium *Pseudoalteromonas* (Männistö *et al.*, 1999), was sequenced. The genome of the first marine cyanophage (P60) was sequenced in 2002. P60 is a phage which infects a marine *Synechococcus* WH7803 (Chen and Lu, 2002). The complete genome of the first *Prochlorococcus* phage was reported in 2003 (Sullivan *et al.*, 2003). In 2013, several phages which infect the marine bacterium *Pelagibacter ubique* (a cultivated representative of the ubiquitous and abundant SAR11) were isolated and sequenced (Zhao *et al.*, 2013).

By May 2015, a total of 130 complete genomes of marine phage isolates have been reported, although bacteriophage genomes assembled from metagenomes are available (Perez Sepulveda *et al.*, 2016). Less than 6% of bacteriophages with complete genomes were isolated from marine environments. Current studies on marine phages appear to be selective and focus on a few virus-host systems. For example, phages infecting cyanobacteria contribute more than 40% of all available marine phage genomes. The sequenced marine phages come from hosts belonging to 5 classes of bacteria, which only represents a minute fraction of marine bacterial communities.

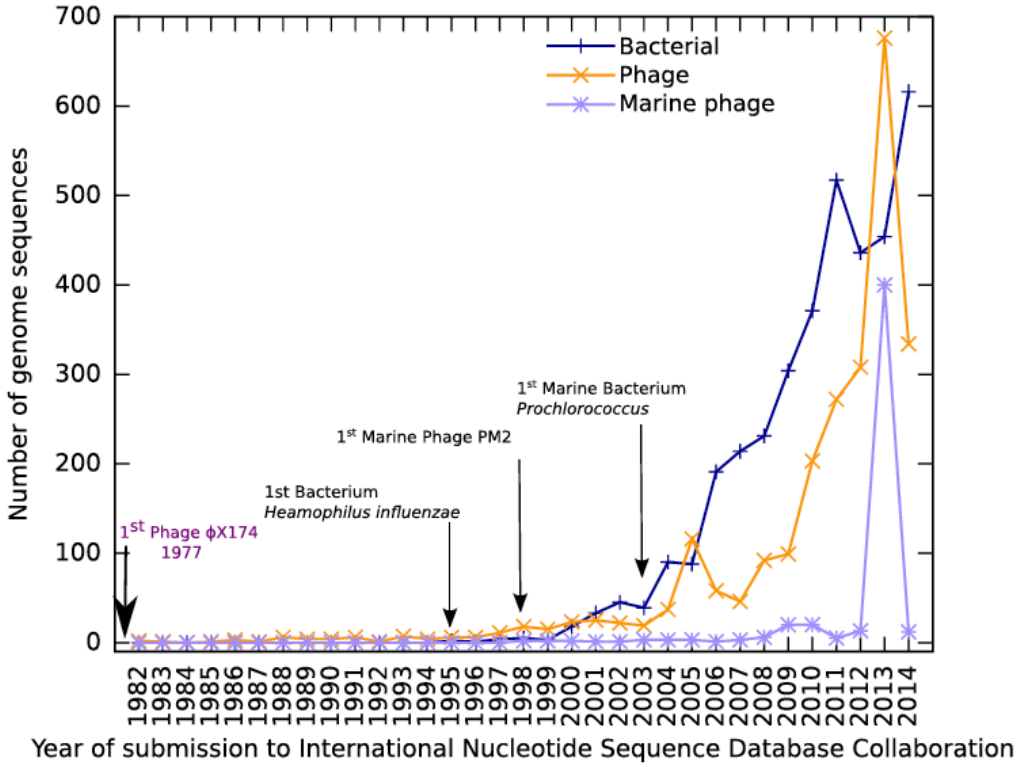


Figure 1.2 Number of bacterial, phage and marine phage genomes submitted per year to the INSDC (International Nucleotide Sequence Database Collaboration). Dates were extracted for all sequences within the EBI phage database (http://www.ebi.ac.uk/genomes/phage.html). Prophages are not included unless they have been specifically sequenced independently of their host bacterium. Phages were classified as marine if they were isolated from a marine environment. The peak of marine bacteriophages in 2013 included the bacteriophage genomes assembled from metagenomic databases. Adapted from Perez Sepulveda *et al.* 2016.

1.1.3 Viral isolation is beneficial to viral metagenomics

Viral isolation helps explain unknown metagenomic reads and unidentified viral populations. More than 60% of sequences in viral metagenomic databases remain unknown due to the small pool of viral references. For example, in the Pacific Ocean Virome, 87-93% of unknown reads do not match with any known viral taxa (Hurwitz and Sullivan, 2013). Part of this issue is due to the low number of sequenced phages. With the rapid development of next generation sequencing technology, viral metagenomics is at the forefront of viral diversity discovery. The number of identified viral populations has reached a plateau, according to the global marine viral metagenomic analysis (Brum *et al.*, 2015; Paez-Espino *et al.*, 2016). By using the large contigs (up to 100kb), Brum *et al.* were able to identify 5,476 pelagic viral populations in Tara Ocean Virome databases (Brum *et al.*, 2015), but only 39 have cultured representatives. A recent analysis based on assembled marine phages genome identified 617 genus-level operational taxonomic units (OTUs), among which only 17 OTUs contained representatives from known viruses (Nishimura *et al.*, 2017). The large proportion of unknown viral sequences is referred as the “phage dark matter” in viromics. Genome sequences of new viral isolates could uncover the mysterious sequences in the viral metagenomic databases. A good example is the discovery of four bacteriophages infecting ‘*Candidatus* Pelagibacter ubique’ HTCC1062, an isolated marine bacterium belonging to the ubiquitous SAR11 clade (Morris *et al.*, 2002). Genome sequences of these four phages were highly represented in marine viral metagenomes. Particularly, HTVC010P-like podoviruses were 2.5 times more abundant than all T4-like cyanophages combined, representing

a highly abundant phage group (Zhao *et al.*, 2010). Another example is phage HMO-2011 which infects another important strain of marine bacteria from the SAR116 clade. The genome sequence of HMO-2011 successfully explained up to 10%-25% of previous unknown reads in viral metagenomes (Kang *et al.*, 2013). The combination of phage genome sequences and metadata allows for global biogeographic studies of SAR11 and SAR116 phages. Therefore, in this era of high throughput sequencing of viral communities, the need for isolation and characterization of viruses in the natural environment is critical.

1.1.4 Viral isolation provides a system to study phage-host interaction

Viral isolation allows us to study the morphology, infection dynamics (i.e. one-step growth curve, burst size, etc.), host range, niche adaptation, genome evolution and phylogenetic relationship of viruses. Cyanobacteria, as important primary producers in the ocean, stimulate interest to study bacteriophages infecting them. More than 80 viruses which infect marine *Synechococcus* and *Prochlorococcus* have been isolated and characterized (Suttle and Chan, 1993; Waterbury and Valois, 1993; Wilson *et al.*, 1993; Lu *et al.*, 2001; Wang and Chen, 2008; Labrie *et al.*, 2013; Sullivan *et al.*, 2003). Currently, all the isolated cyanophages are dsDNA phages and belong to three well-established phage families (*Myoviridae*, *Podoviridae*, and *Siphoviridae*). Isolation of these cyanophages unveiled interesting virus-host strategies. For example, T7-like cyanophages tend to be very host specific, whereas T4-like cyanophages can cross infect both *Prochlorococcus* and *Synechococcus* (Weigele *et al.*, 2007; Sullivan *et al.*, 2003; Wang and Chen, 2008). A recent transcriptomic

study of the broad host-range cyanomyophage, Syn9, and its three distinct *Synechococcus* hosts, suggested that the broad host-range of Syn9 results from the effectiveness of host defense strategies, rather than the infectivity of the phage (Doron *et al.*, 2016). The genome sequence of cyanophages provides knowledge on their diversity, evolution and phage-host interaction. Sets of core genes have been established within cyanomyoviruses and cyanopodoviruses, respectively, inferring the genomic conservation within each group (Sullivan *et al.*, 2010; Huang *et al.*, 2015; Labrie *et al.*, 2013). In contrast, genomes of cyanosiphoviruses appear to be highly mosaic and variable (Huang *et al.*, 2012; Mizuno *et al.*, 2013a). The finding of cyanophage-encoded photosynthetic genes has raised many questions regarding the role of phages in regulating host photosynthetic activity (Millard *et al.*, 2004; Lindell *et al.*, 2005; Crummett *et al.*, 2016). Cyanobacterial resistance to viral infection occurs frequently (Avrani *et al.*, 2011), yet it is not easy to isolate phage mutants that overcome this resistance (Schwartz and Lindell, 2017).

In short, viral isolation is a fundamental step towards understanding the biological features of viruses and how they interact with their hosts. Genomic information of isolated phages is greatly needed to interpret many unknown 'viruses' in the vast viral metagenomic databases. The integration of viral genomics and viral metagenomics has become a powerful way to understand the diversity and interaction between viruses and microbes in the natural environment.

1.2 The marine *Roseobacter* lineage

The marine *Roseobacter* group serves as an important model organism for studying marine microbial ecology. All members of the *Roseobacter* lineage cluster form a distinct lineage within the *Rhodobacteraceae* family of the *Alphaproteobacteria* class. Based on the 16S rRNA gene phylogeny, the *Roseobacter* lineage is comprised of five deeply branching clades (Buchan *et al.*, 2005). Multiple studies based on concatenated core genes further support these five clades (Newton *et al.*, 2010; Luo and Moran, 2014; Simon *et al.*, 2017). To clarify, I use "*Roseobacter*" (instead of Roseobacter) to refer the lineage of *Roseobacter*, and "roseobacters" to refer the members and isolates within this lineage.

The *Roseobacter* lineage is one major clade of marine heterotrophic bacteria that was first found in the coastal waters of Georgia, in the south eastern U.S. (Gonzalez and Moran, 1997). This lineage accounts for a large proportion of the bacterioplankton community, comprising upwards of 20% of the coastal and 15% of mix-layer ocean bacterioplankton (Buchan *et al.*, 2005). Members of the *Roseobacter* lineage are distributed across diverse habitats, including Antarctic, Artic, biofilms, deep sea, sediment, hypersaline mat and reef environments (Allgaier *et al.*, 2003; Jonkers and Abed, 2003; Selje *et al.*, 2004; Gifford *et al.*, 2014; Wemheuer *et al.*, 2015). Roseobacters are commonly associated with microalgae, especially during phytoplankton blooms (Sass *et al.*, 2010; Laass *et al.*, 2014; Buchan *et al.*, 2014).

Roseobacters are considered to be ecological generalists with versatile metabolic features (Wagner-Döbler and Biebl, 2006; Buchan *et al.*, 2005). Some are capable of aerobic anoxygenic photosynthesis (AAnP) (Beja *et al.*, 2002; Allgaier *et*

al., 2003), oxidation of carbon monoxide (Todd *et al.*, 2007; Cunliffe, 2011), degradation of multiple organic sulfur compounds (Moran *et al.*, 2003; Lenk *et al.*, 2012), aromatic compounds (Gulvik and Buchan, 2013), and production of a variety of bioactive secondary metabolites (Geng *et al.*, 2008; Sonnenschein *et al.*, 2017). Several roseobacters have close symbiotic and pathogenic relationships with both vertebrates and invertebrates (Zan *et al.*, 2012; Patzelt *et al.*, 2013; Collins *et al.*, 2015). More importantly, roseobacterial members are able to degrade the phytoplankton produced osmolyte dimethylsulfoniopropionate (DMSP) using either the cleavage or the demethylation/dethiolation pathway (Moran *et al.*, 2004, 2012). Field studies have shown that members in the *Roseobacter* lineage can affect the fate of DMSP in nature (Howard *et al.*, 2008; Vila-Costa *et al.*, 2014; Varaljay *et al.*, 2015), inferring that the marine *Roseobacter* lineage is a key player in the global sulfur cycle.

Diverse functions seen in the *Roseobacter* lineage are reflected by their flexible genomes. A recent phylogenomic analysis revealed that marine genomes evolved from ancestral *Rhodobacteraceae* and diverged from their non-marine counterparts through gaining and lossing of genes, which improved their fitness in marine habitats (Simon *et al.*, 2017). By January, 2017, genomes of at least 69 marine roseobacters species have been sequenced (http://www.roseobase.org/index.html and http://www/arb-silva.de). Among them, 14 genomes are complete, and genome size ranges from 3.02 to 5.75 Mbp. Genes encoding diverse metabolic pathways, including nitrogen uptake, carbon utilization, sulfur degradation and secondary product production, have been found within

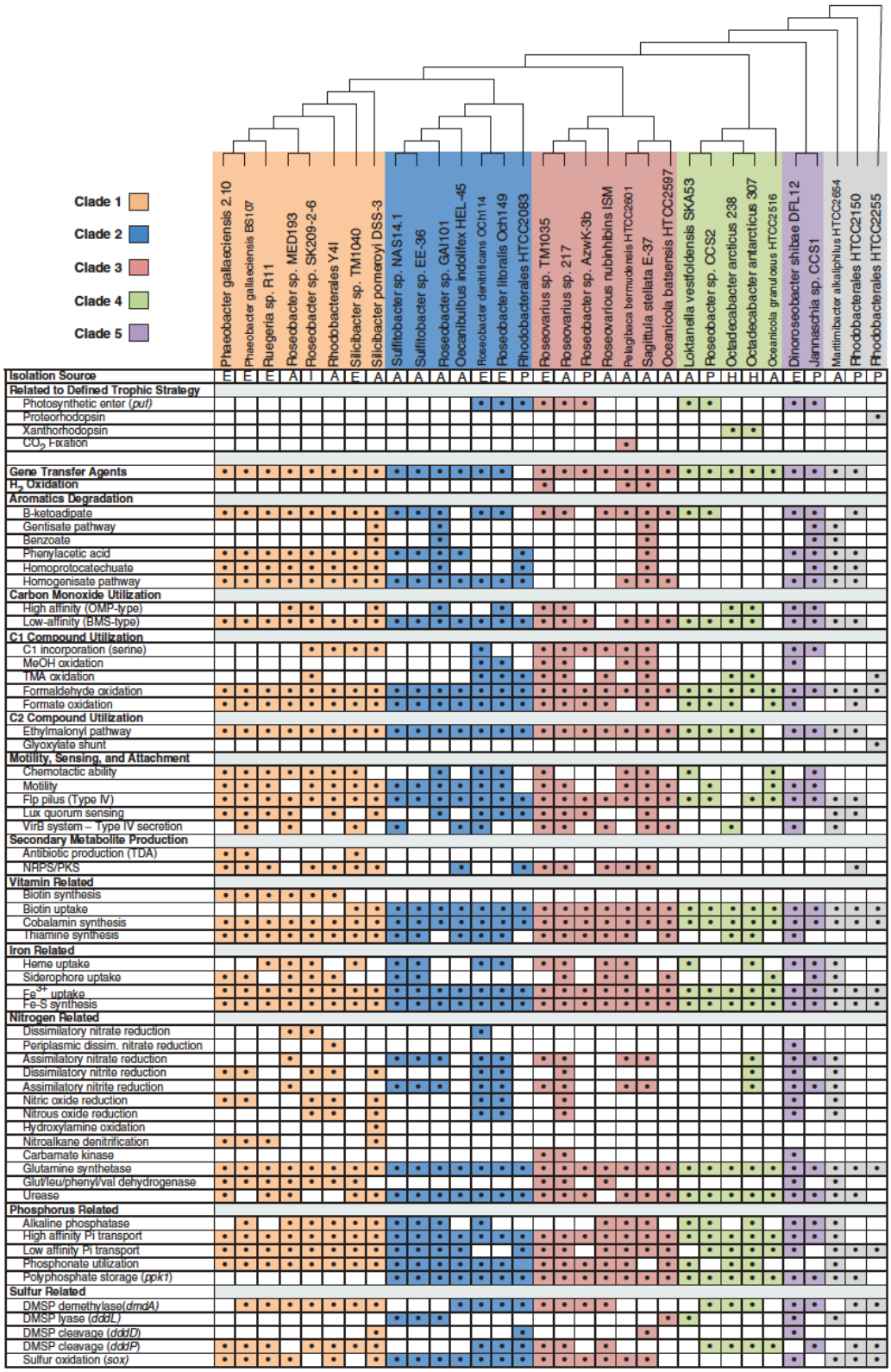


Figure 1.3 A matrix describing the presence of select genes or gene pathways in the

32 genomes (Newton *et al.* 2010.). A colored box containing a dot indicates the presence of the gene/pathway. An ultra-metric tree has been placed above the gene matrix for reference.

roseobacter genome (Figure 1.3) (Newton *et al.*, 2010; Luo and Moran, 2014; Simon *et al.*, 2017).

Interestingly, nearly all roseobacter genomes contain a complete or nearly complete gene transfer agent (GTA) operon (Figure 1.3) (Newton *et al.*, 2010; Luo and Moran, 2014). GTAs were first discovered in *Rhodobacter capsulatus* (RcGTA), another *Rhodobacteraceae* member(Lang and Beatty, 1999). GTAs are phage-like particles, encoding 15 to 17 genes. In contrast to bacteriophages, which package their own genomes, GTA particles usually package bacterial DNA (Lang *et al.*, 2012). It has been shown that GTA is able to mediate higher rates of gene transfer, compared with transformation and transduction (McDaniel *et al.*, 2010). The *Roseobacter* lineage is the only marine bacterial lineages, known to carry GTA genes (Luo and Moran, 2014). The production of GTA-like particles by using *R. pomeroyi*, *Roseovarius nubinhibens*, and *R. mobilis* has been demonstrated (Biers *et al.*, 2008; McDaniel *et al.*, 2010). Although *Roseobacter*-like GTA genes can be detected from estuarine waters (Zhao *et al.*, 2010), the abundance and global distribution of GTA particles remains unclear. The presence of GTA genes in roseobacter genomes and the ability of strains to produce GTA particles implies an important role for GTA in the *Roseobacter* lineage.

Recently sequenced genomes of pelagic roseobacters provides evidence for different strategies. The isolation of *Planktomarina temperate* RCA23 demonstrated the streamlined genomes of pelagic roseobacters and their possible adaptation machinery to oligotrophic environments (Voget *et al.*, 2015). Analyses based on genome comparison (Newton *et al.*, 2010), single-cell genomics (Luo *et al.*, 2014;

Zhang *et al.*, 2016), and metagenomics of uncultivable roseobacters (Luo *et al.*, 2012) seem to suggest that the current culture collection of the *Roseobacter* lineage may not represent the whole roseobacter community in the natural environment.

In summary, the members of marine *Roseobacter* are a dominant group of marine heterotrophic bacteria with versatile lifestyles, high genomic plasticity and important biogeochemical roles.

1.3 Bacteriophages of roseobacters

Compared to the knowledge gained on roseobacters over the last two decades, much less is known about the phages which infect this group of bacteria (hereafter refer as roseophage).

1.3.1 SIOΦ1, the first isolated roseophage

Phage SIOΦ1 is the first roseophage isolated from a marine roseobacter strain (Rohwer *et al.*, 2000). SIOΦ1, which infects *Roseobacter* SIO67, was isolated from the coastal seawaters of California. SIOΦ1 has an isometric capsid and a short tail, and belongs to the *Podoviridae* family. SIOΦ1 has a genome size of 39.9 kb, and its genome contains 30 open reading frames (ORFs) that can be divided into three modules: nucleotide synthesis and DNA replication, phosphate metabolism, and capsid structural proteins (Angly *et al.*, 2009). SIOΦ1 is closely clustered with a clone sequence from the Strait of Georgia (SOG) and is distinct from the T7 enterophages within *Podoviridae* (Labont *et al.*, 2009).

Four additional podoviruses were isolated from the same location using the same host *Roseobacter* SIO67, 12 years after SIOΦ1 was isolated (Angly *et al.*,

2009). These four new phages share a high degree of genomic similarity with SIOΦ1 (Angly *et al.*, 2009). This indicates that podoviruses of *Roseobacter* SIO67 are relatively stable in the natural environment.

1.3.2 Novel N4-like Roseophages

Since the first report of the SIOΦ1 roseophage in 2000, a few more roseophages have been discovered. In 2009, two novel phages, DSS3Φ2 and EE36Φ1 were isolated from the marine roseobacters, *Ruegeria pomeroyi* DSS-3 and *Sulfitobacter* sp. EE-36, respectively (Zhao *et al.*, 2009b). Phages DSS3Φ2 and EE36Φ1 were isolated from Baltimore Inner Harbor Pier V and belong to the phage family *Podoviridae*, with icosahedral capsids (~70 nm in diameter) and visible short tails (~26 nm long). The genome sequences of these two phages show that they are closely related to phage N4 that infects *E. coli* (Schito *et al.*, 1966). The isolation of DSS3Φ2 and EE36Φ1 marks the second case for both roseophage and N4-like phage. Prior to the discovery of DSS3Φ2 and EE36Φ1, phage N4 had been considered a genetic and taxonomic orphan for more than forty years, lacking any comparable counterparts.

In recent years, more N4-like roseophages were isolated in different laboratories. RLPΦ1 and RPPΦ1, were found from the coastal waters of the United Kingdom, which can specifically infect two strains of *Roseovarius* (Chan *et al.*, 2014). A N4-like phage which infects *Sulfitobacter sp.* strain 2047 was discovered from Raunefjorden, Norway (Ankrah *et al.*, 2014a). After that, three N4-like roseophages infecting *Dinoroseobacter shibae* DFL12 (Cai *et al.*, 2015; Ji *et al.*, 2015b; Li *et al.*,

2016) and two N4-like roseophages infecting *Roseobacter dinitrificans* OCh114 were further isolated and characterized (Li *et al.*, 2016).

Interestingly, all N4-like phages contain a large virion-encapsidated RNA polymerase gene (vRNAP). It has been reported that vRNAP in phage N4 (Zivin and Zehring, 1981) is packaged in viral particles and injected into host cells upon infection, and is responsible for early and middle transcription (Choi *et al.*, 2008). All the N4-like roseophages contain the vRNAP gene homologous to the vRNAP gene in phage N4. Therefore, it is postulated that N4-like roseophages use a similar transcription machinery as phage N4.

Genome sizes of known N4-like roseophages range from 73 to 75kb, similar to the genome size (70kb) of coliphage N4. In general, N4-like phages appear to contain a larger genome than the typical T7 like phages in the family *Podoviridae*. It is noteworthy that all N4-like roseophages share a similar genomic arrangement, suggesting the conserved evolution of this group of phages.

In the recent years, new N4-like phages, which infect other bacteria, such as *Vibrio*, *Pseudomonas*, *Salmonella*, and *Achromobacter* have been reported (Born *et al.*, 2011; Kulikov *et al.*, 2012; Fouts *et al.*, 2013; Gan *et al.*, 2013). These studies suggest that N4-like phages could infect a wide range of bacterial genera. However, we know little about the diversity and distribution of N4 phages in the natural environment. Most of our knowledge on the distribution of N4-like phage is based on metagenomic recruitment. Searching the Global Ocean Survey (GOS) metagenomic database using N4 roseophage DNA polymerase genes showed that roseobacter N4 phages are relatively abundant in coastal areas but are rare in open oceans (Zhao *et*

al., 2009b). A later study shows that N4 phages are present in the Antarctic Ocean, Saltern Sea and GOS metagenomes (Chan *et al.*, 2014).

1.3.3 Other lytic roseophages

Roseophage RDJLΦ1 and RDJLΦ2 which infect *Roseobacter denitrificans* OCh114 were isolated from the South China Sea (Zhang and Jiao, 2009; Liang *et al.*, 2016) and coastal seawater of Xiamen, China, respectively. Morphologically, these two phages belong to *Siphoviridae* family. The genomes of RDJLΦ1 and RDJLΦ2 exhibited 88% genome sequence similarity (Liang *et al.*, 2016). Both phages contain four genes that are highly homologous to the four genes found in the gene transfer agent (GTA) of *Rhodobacter capsulatus* (Huang *et al.*, 2011). The shared genetic features between a lytic phage and GTA suggest was proposed to common prophage ancestor (Huang *et al.*, 2011).

ΦCB2047-A and ΦCB2047-C are two lysogenic podoviruses, infecting *Sulfitobacter* sp. strain 2047 (Ankrah *et al.*, 2014b). They were isolated from an induced algal bloom mesocosm study in Raunefjorden, Norway. ΦCB2047-A and ΦCB2047-C were nearly identical at the nucleotide level. They shared 17 highly homologous genes with the uncultivated ΦEBPR podovirus (Ankrah *et al.*, 2014b). Integrases were found in both phage genomes, suggesting their temperate lifecycle. In contrast, these two phages did not carry well-characterized genes related with replication and nucleotide metabolism (Ankrah *et al.*, 2014b), which infers that ΦCB2047-A and ΦCB2047-C may heavily depend on host systems for production.

1.3.4 Roseobacters contain prophages

Not only have lytic bacteriophages have been isolated from roseobacters, but many prophages have also been identified in roseobacter genomes. Prophage refers to the temperate phage that can integrate into the bacterial chromosome and enter the lysogenic cycle. Under certain circumstances, such as UV radiation, chemical exposure or other environmental stresses, the prophage can be induced and enter into lytic cycle to release the virulent phage particles. Both temperate phages and hosts benefit from the lysogenic cycle (Sime-Ngando, 2014). For phages, lysogenic cycle provides an opportunity to survive when the host cells are not abundant. For host cell, prophages affect the features of host in many ways, such as increasing growth (Edlin et al., 1975), providing pathogenesis (Waldor and Mekalanos, 1996) and adaptation to variable environments (Zeng et al., 2016), leading to expand the fitness of host cells (Feiner et al., 2015)

It has been reported that a wide range of marine bacterial isolates contain inducible prophages (Jiang and Paul, 1994, 1996, 1998; Weinbauer and Suttle, 1999; Stopar *et al.*, 2003). Based on the screening of 113 marine bacterial genomes, 43% of marine bacteria contain prophage-like elements in their genomes (Paul, 2008). These prophages are considered to be “the key to survival in the seas” due to their important roles in genetic exchanges of their hosts (Paul, 2008). However, the ecological roles of lysogenic infection remain underexplored compared to those of lytic infections (Howard-Varona *et al.*, 2017).

Prophages are commonly found in roseobacters (Chene *et al.*, 2006; Paul, 2008; Zhao *et al.*, 2010). After identifying prophage-like regions in *Silicibacter* sp.

TM1040 genome, Chen *et al.* successfully induced multiple prophages from roseobacters using mitomycin C (Chene *et al.*, 2006). A prophage was first identified *in silico* on the *Roseovarius nubinhibens* genome and later proved to be an active prophage via mitomycin C induction (Zhao *et al.*, 2010). In the natural environment, upon mitomycin C induction, the *Roseobacter*-associated phage integrase genes were detected at both genomic and transcriptomic levels (McDaniel *et al.*, 2008), suggesting the presence of lysogenized roseobacters in nature.

1.4 Scope of this dissertation

Although only a few phages infecting marine roseobacters were described at the beginning of my dissertation work, they have shown some interesting features indicating a possible different phage-host strategy compared to the better-studied cyanobacteria-cyanophage systems. What kind of viruses infect roseobacters? Are viruses infecting roseobacters different from those infecting marine *Synechococcus* or *Prochlorococcus*? Are N4-like phages a common type of phage infecting roseobacters? If so, are they genetically conserved, like T7 podoviruses? How are roseophages distributed in different marine habitats?

This dissertation is devoted to understanding the diversity, ecology and evolution of roseobacter phages through isolation and characterization of the phages infecting marine roseobacters. At the very beginning of my study, four roseobacter strains, *Roseovarius nubinhibens* ISM, *Silicibacter* sp. TM1040, *Sulfitobacter* sp. EE-36 and *Ruegeria pomeroyi* DSS-3, were used to isolate bacteriophages. However, bacteriophages were only isolated using *Ruegeria pomeroyi* DSS-3 as host strain.

Therefore, I decided to use *Ruegeria pomeroyi* DSS-3 as my model strain to study bacteriophages infecting marine roseobacter.

Hypothesis I: Diverse phages which infect a single strain are present in aquatic environments. Diverse roseophages exhibit distinct geographic distribution.

Only few roseophages have been isolated and described. Phage isolation permits the study of diversity and evolution of phages, and could provide insight into the potential interactions between bacteria and bacteriophages. *R. pomeroyi* DSS-3 has been referred as the model strain to study marine heterotrophic bacteria (Rivers *et al.*, 2014; Sun *et al.*, 2017; Moran *et al.*, 2004), with important biogeochemical features, such as degrading DMSP and utilizing carbon monoxide (Newton *et al.*, 2010). Two phages infecting *R. pomeroyi* DSS-3, DSS3Φ1 (Zhang and Jiao, 2009) and DSS3Φ2 (Zhao *et al.*, 2009a), have been isolated from Baltimore Inner Harbor in previous studies. These two DSS-3 phages are distinct from each other: one is a siphovirus and the other is N4-like podovirus. For this dissertation, I asked: how many different types of phage infecting *R. pomeroyi* DSS-3 can be isolated?

Hypothesis II: N4-like phages infecting various lineages of roseobacters are conserved at the genomic level. N4-like phages can be detected in the marine environments.

A high proportion of roseophages are N4-like phages. The coliphage N4 possesses a unique mode of infection that prompted decades of fundamental research to elucidate the molecular mechanisms that facilitate its propagation in its

host (*E.coli*) (Schito *et al.*, 1966). However, little is known about the distribution and genetic diversity of N4-like phage. With many N4-like phages being isolated in roseobacters and other bacterial lineages, we know now that N4-like phage can be present in different ecological habitats. All the known N4-like phages share many conserved genes, which allowed me to design specific PCR primers to explore the distribution and diversity of N4-like phage in the natural environment.

Hypothesis III: N4 phage can proliferate more effectively in cold environments than non N4 phages.

Based on preliminary studies of the geographic distribution of N4-like phage, marine N4 phages appear to be restricted to high latitudes and/or colder seasons. N4 phages, with their unique features, such as large burst size and viral encapsidated RNA polymerase, may allow them to thrive under cold environment. For the dissertation, I thought it would be interesting to compare the persistence of N4 phage and non-N4 phage under different temperatures. Since several different phages infecting *R. pomeroyi* DSS-3 have been isolated and characterized, they were used as reference phages to compare N4-like roseophages infecting the same host.

Chapter 2. A small and novel ssDNA bacteriophage infecting a marine bacterium unveils highly mosaic genomic evolution

2.1 Abstract

A lytic ssDNA phage, DSS3Φ22, infecting the marine bacterium *Ruegeria pomeroyi* DSS-3, was isolated from the Baltimore Inner Harbor. DSS3Φ22 contains only four open reading frames originating from diverse sources. A phylogeny based on the phage capsid gene shows that DSS3Φ22 is distantly related to known ssDNA phages and belongs to an unclassified ssDNA phage within the family *Microviridae*. The genome size of DSS3Φ22 is 4,248 kb, making it the smallest and simplest ssDNA phage among currently known ssDNA phage isolates. The GC content of DSS3Φ22 is high (58%) and similar to that of its host (64%). DSS3Φ22 lacks the spike protein commonly to ssDNA phages, suggesting that ssDNA phages can be more diverse than previously thought. Metagenomic analysis indicates that DSS3Φ22-like phages are widely distributed in diverse environments, ranging from human guts, coral reefs and deep ocean. Multi-origins of four DSS3Φ22 ORFs challenge the current view that point mutations are the principle driving force in the evolution of small ssDNA phages. The unique and highly mosaic genomic architecture of DSS3Φ22 indicates that horizontal gene transfer may be important in the genomic diversification of ssDNA phages.

2.2 Introduction

Viruses are the most abundant microbial entities in the marine environment. They are comprised mostly of bacteriophages. Viruses play pivotal roles in shaping host population structures, mediating gene transfer, and modulating biogeochemical cycling in the world's oceans (Suttle, 2005, 2007). Current knowledge on marine viruses is highly biased towards double-stranded DNA (dsDNA) viruses, and little is

known about the diversity of single-stranded DNA (ssDNA) or RNA viruses (Roux *et al.*, 2012b). Recent metagenomic studies have revealed that various marine environments harbor extensive previously unknown ssDNA viruses, suggesting a cosmopolitan distribution of ssDNA viruses (Tucker *et al.*, 2010; Labonté and Suttle, 2013b; Bryson *et al.*, 2015).

Currently, only a handful of marine ssDNA viruses have been isolated and characterized (Székely and Breitbart, 2016), including a few ssDNA viruses infecting marine diatoms (Kimura and Tomaru, 2013, 2015) and a few ssDNA phages infecting a marine *Bacteroidetes* of the genus *Cellulophaga* (Holmfeldt *et al.*, 2012, 2013). Most of our knowledge on lytic ssDNA phages is built on the phage family *Microviridae*. *Microviridae* contains two major subfamilies of ssDNA phages, *Bullavirinae* and *Gokushovirinae*, which infect *Escherichia coli* and *Chlamydia*, respectively (ICTV, 2015 release). The four ssDNA phages which infect the marine bacterium *Cellulophaga baltica* are distantly related to *Bullavirinae* and *Gokushovirinae* and belong to an unclassified cluster within the *Microviridae* (Holmfeldt *et al.*, 2013). Except for *Cellulophaga*, no ssDNA bacteriophages have been isolated from other phyla of marine bacteria.

Members of the *Roseobacter* lineage are abundant and widely distributed from nearshore to pelagic waters (Brinkhoff *et al.*, 2008; Buchan *et al.*, 2005; Wagner-Döbler and Biebl, 2006). They can contribute up to 20% of bacterial communities in the coastal ocean. Roseobacters are known to have diverse metabolic capabilities, including the ability to degrade dimethylsulfoniopropionate, an important sulfur compound in the marine environment (Newton *et al.*, 2010; Luo and Moran, 2014).

Currently, several lytic phages have been isolated from marine roseobacters (Rohwer *et al.*, 2000; Zhao *et al.*, 2009b; Zhan *et al.*, 2015), but no ssDNA phages infecting roseobacters have been reported.

Here, I report the isolation of a ssDNA bacteriophage, DSS3Φ22, which infects the marine roseobacter strain, *Ruegeria pomeroyi* DSS-3. Phage DSS3Φ22 is a novel ssDNA phage with the least number of coding genes and the smallest genome size of the known ssDNA phages. The morphology, genomic characteristics and potential origin of coding genes for DSS3Φ22 are described in this study. Furthermore, analysis of viral metagenome databases reveals that DSS3Φ22-like phages are widely present in the marine and human gut environments.

2.3 Method

2.3.1 Isolation and concentration of phage DSS3Φ22

R. pomeroyi DSS-3 was grown in ½ YTSS medium (4 g yeast extract, 2.5 g trypone and 20 g Crystal Sea per liter) at 28 °C. To isolate phage, a water sample was collected from the Baltimore Inner Harbor Pier V in January 2013. The water sample was filtered through 0.22 μm pore-size polycarbonate membrane filters (Millipore, USA). Two milliliters of filtered water were added to 20 ml of an exponentially growing *R. pomeroyi* DSS-3 culture. The mixture (DSS-3 culture and the filtered water) was incubated for 48 h at 28 °C. A plaque assay was performed to isolate and purify lytic viruses (Zhan *et al.*, 2016). Each lytic phage was purified at least three times using plaque assay.

To concentrate phage particles, one liter of phage lysate was precipitated by polyethylene glycol 8000. The phage concentrate was purified by iodixanol density gradient ultracentrifugation (OptiPrep^{TM}, Sigma-Aldrich, USA). The precipitated phage lysate was centrifuged at 41,000 rpm, at 4 °C for 12 hours using a SORVALL Discovery 1000 ultracentrifuge. Purified phages were dialyzed in TM buffer (50mM Tris-HCl, 2.5 mM $MgCl_2$, pH = 7.0) twice and stored at 4 °C.

2.3.2 Transmission electron microscopy

One drop (2 µl) of purified phage particles was adsorbed to a 200-mesh Formvar/carbon-coated copper grid and stained with 1% phosphotungstic acid (PTA) for one minute. The phage sample was examined using a FEI Tecnai T12 transmission electron microscope at the University of Maryland Baltimore.

2.3.3 Host range and one-step growth curve

Cross-infectivity of DSS3Φ22 was tested against five marine roseobacter strains: including *Roseovarius nubinhibens* ISM, *Silicibacter* sp. TM1040, *Sulfitobacter* sp. EE-36, *Dinoroseobacter shibae* DFL-12 and *Roseobacter denitrificans* OCh114. For each host strain, one drop of phage lysate was spotted onto a soft-agar overlay plate and incubated for 2 to 3 days at 28 °C. The formation of plaques was an indication of cross-infectivity (Rohwer *et al.*, 2000). The original host was also tested in parallel, as a positive control.

A one-step growth curve was determined using a previously reported method (Zhan *et al.*, 2016). Briefly, purified phages were added into an exponentially growing

DSS3 culture with an MOI = 0.1. After 20 min of adsorption, cells were pelleted, re-suspended and diluted 100-fold. Free-living bacteriophages were collected at different time points and quantified by qPCR (based on the phage capsid gene). Phage ssDNA was extracted by ZR-96 RNA Clean & Concentrator (Zymo Research, USA). The growth curve was conducted in triplicates. Three qPCR reactions were performed for each biological triplicate sample.

2.3.4 Confirmation of single stranded DNA nature of phage

Purified phage concentrate was treated with a combination of SDS (final concentration 1% w/v), proteinase K (final concentration 25 µg/ml) and EDTA (final concentration 5 µm) at 55 °C for 3 h. Phage DNA was extracted using phenol/chloroform/isoamyl alcohol (25:24:1), and precipitated with isopropanol (Steward and Culley, 2010).

To identify the nucleic acid type (DNA, RNA, ssDNA vs. dsDNA) of phage, aliquots (1 µl) of phage nucleic acids were digested with DNase I (final concentration 1 µg/ml), RNase A (final concentration 1 µg/ml), and S1 nuclease (final concentration 100 U), respectively, at 37 °C for 1 h. The treated nucleic acid was stained with 1X SYBR Gold and analyzed using gel electrophoresis.

2.3.5 Genome sequencing and annotation of phage

In order to sequence ssDNA phage, DNA was extracted as a replicative double-stranded form during phage infection using a QIAprep Spin Miniprep Kit (Qiagen, USA). Two hours after adding the DSS3Φ22, 2 ml of bacterial culture were

collected and centrifuged for 20 min at 12,000 rpm. The circular dsDNA was extracted by QIAprep Spin Miniprep Kit. Two different methods were used to sequence the phage genome: 1) The circular dsDNA was digested by EcoRI and BamHI, yielding two bands. The pBluescript SK+ plasmid was used to clone these two bands separately, and Sanger sequencing (ABI 3100 genetic analyzer, PE Applied Biosystems, USA) was performed to obtain the full genome sequence. 2) dsDNA was directly sequenced by Illumina MiSeq platform. Reads were assembled by the CLC Genomic Workbench.

Open reading frames (ORFs) were predicted using GeneMarkS, GeneMarkhmm and ORFfinder, then annotated by using BLASTp against the NCBI nonredundant databases with an E-value ≤ 1E-3. The sequence data was deposited in GenBank under accession number MF101922. The putative amino acid sequences of major capsid protein, replication initiator and peptidase were aligned to their closest homologs using Cluster W with default parameters. Maximum-likelihood phylogenetic trees were built using MEGA7 (Kumar *et al.*, 2016), with bootstrap value of 500.

2.3.6 SDS-PAGE and mass-spectrometry

Proteins from phage lysate were analyzed using SDS-PAGE with a Criterion XT^{TM} Precast Gel (Bio-Rad, USA). Protein extracted from each gel band were further analyzed by Electro Spray Ionization Mass Spectrometry (ESI-MS) at the University of Maryland Baltimore County. The protein structure of the major capsid protein was analyzed using I-TASSER (Zhang, 2008), with default settings.

2.3.7 Metagenomic recruitments

A reciprocal BLAST analysis was conducted to assess the distribution of DSS3Φ22 in the natural environment. Due to the nature of ssDNA viruses, I only recruited DSS3Φ22-like sequences from the metagenomic databases in which viral DNA was amplified by multiple displacement amplification. Reads were recruited to predict ORFs by using tBLASTn (E-value ≤ 1E-3). The recruited sequences were further confirmed to be DSS3Φ22-like phages as BLAST queries against the NCBI Virus RefSeq database.

2.4 Results and Discussion

DSS3Φ22 is a unique phage in the ssDNA phage family *Microviridae* based on the genomic and phylogenetic analyses (see the later sections). Henceforth, I refer to members of *Microviridae* as “microviruses”.

2.4.1 Phage morphology, growth and specificity

The bacterium *R. pomeroyi* DSS-3 was isolated from coastal waters of the Atlantic Ocean and has been referred to as a model strain for marine heterotrophic bacteria (Moran *et al.* 2004). It is a member of the Clade I of *Roseobacter* (Figure. 1.3). Several novel phages, including two N4-like phages and one CbK-like phage, have been isolated using DSS-3 as a host (Zhao *et al.*, 2009b; Zhan *et al.*, 2016). To identify other phages, this bacterium was employed to screen Inner Harbor water for novel phages. Phage forming plaques on *R. pomeroyi* DSS-3 was isolated and its properties were evaluated.

In order to determine the type of phage causing the plaques, DSS3Φ22 was examined by TEM using phosphotungsitic acid as a negative stain (Figure 2.1a). The DSS3Φ22 capsid was ca. 22 nm in diameter with no visible tail. This small size is on the small end of the range of characterized microviruses (22-34 nm) (Table 2.1).

To assay DSS3Φ22 replication, a culture of *R. pomeroyi* DSS-3 was inoculated at an MOI of 0.1 and phage populations monitored using qPCR of the capsid gene. Phage populations remain constant until three hours and then increased, indicating that DSS3Φ22 had a latent period of ca. 3 hours. The burst size was approximately 8 as calculated from the increase of phage population. DSS3Φ22 has a delayed latent period compared to the ssDNA viruses infecting *E. coli,* likely due to the slower growth rate of host DSS-3.

In order to understand the host range of DSS3Φ22, a cross-infecting test was conducted. DSS3Φ22 do not cross infect other five marine species tested (*Roseovarius nubinhibens* ISM, *Silicibacter* sp. TM1040, *Sulfitobacter* sp. EE-36, *Dinoroseobacter shibae* DFL-12 and *Roseobacter denitrificans* OCh114). Microviruses infecting *E. coli* and *Bacteroidetes* are known to restrict their infection to a single species (Bowes and Dowell, 1974; Holmfeldt *et al.*, 2007; Zhong *et al.*, 2015). However, several microviruses infecting *Chlamydia* are able to cross-infect different *Chlamydia* species (Read *et al.*, 2000a; Garner *et al.*, 2004)

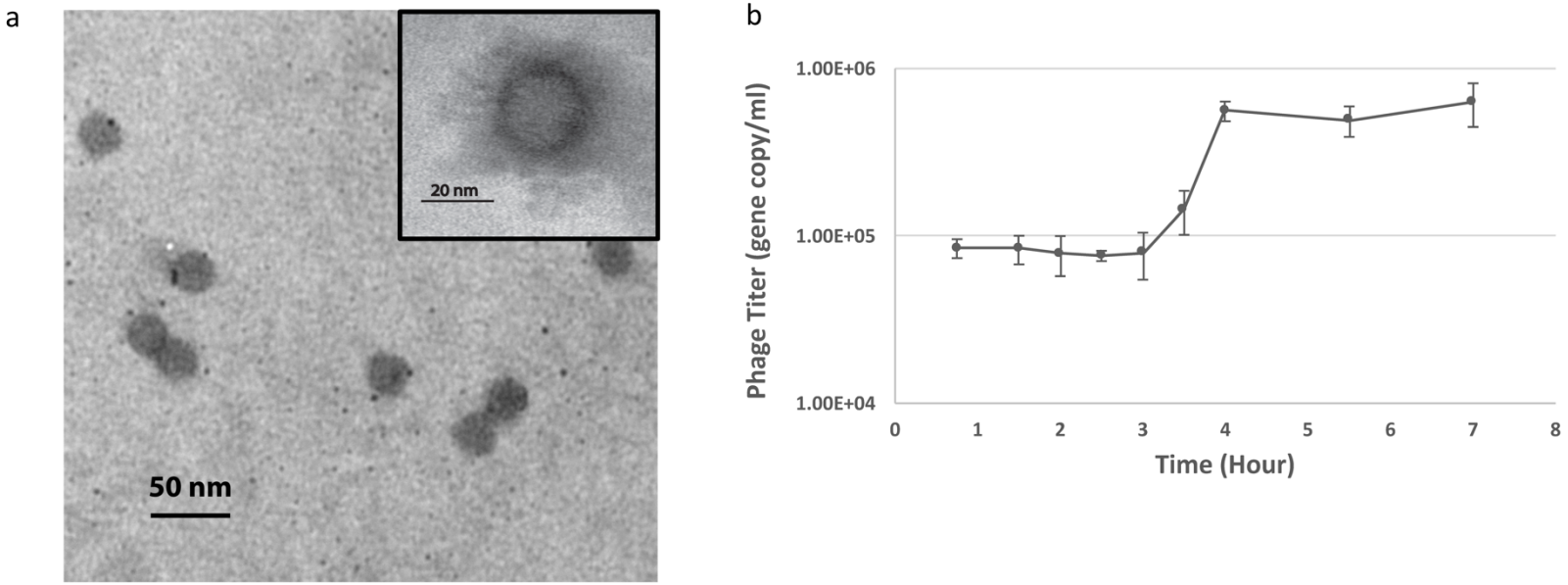


Figure 2.1 The morphology and one-step growth curve of DSS3Φ22. (a) Transmission electron microscopy image of DSS3Φ22. The image in the box provides a zoom-in view of phage particle (b) The one-step growth curve of DSS3Φ22.

Table 2.1 The summary of *Microviridae* isolates

Taxonomy	Phage name	Host	capsid size (nm)	Genome size (kb)	G+C content (%)	Number of ORFs	Reference
Bullavirinae	*Enterobacteria* phage alpha3	*E.coli* C	NA	6.09	45	10	Kodaira, 1992
	Escherichia phage phiX174	*E.coli* C	NA	5.39	45	11	Sanger, 1977
	Enterobacteria phage G4	*E.coli* C	NA	5.58	46	11	Godson, 1978
	Enterobacteria phage phiK	*E.coli* K-12	NA	6.09	45	10	Kodaira, 1996
	Enterobacteria phage ST-1	*E.coli* K-12	26	6.09	45	11	Bowes, 1994
	Enterobacteria phage MED1	*E.coli* SMQ-1277	*ca.* 25	5.39	45	11	Labrie, 2014
Gokushovirinae	phiMH2K	*Bdellovibrio bacteriovorus*	27	4.79	46	11	Brentlinger, 2002
	Chlamydia phage 1	*Chalmydia psittaci* (ducks)	22	4.88	37	12	Storey, 1989
	Chlamydia phage 2	*C. psittaci*; *C. abortus*	25	4.56	41	8	Liu, 2000
	Chlamydia phage CPAR39	*C. pneumoniae* AR39	NA	4.52	41	6	Read, 2000
	Chlamydia phage CPG1	*C. psittaci* (Guinea pig)	25	4.53	41	9	Hsia, 2000
	Chlamydia phage 3	*C. pecorum*	NA	4.55	41	8	Garner, 2004
	Chlamydia phage 4	*C. abortus*	NA	4.53	41	8	Sait, 2011
	Spiroplasma phage 4	*Spiroplasma melliferum* B63/G1	27	4.42	32	9	Renaudin 1987; Chipman, 1998
Unclassified	*Cellulophaga* phi 18:4	*Cellulophage baltica* #18	30-34	6.48	34	13	Holmfedlt, 2007, 2012, 2013
	Cellulophaga phi 12:2	*C. baltica* #12	29-33	6.45	35	13	Holmfedlt, 2007, 2012, 2013
	Cellulophaga phi 12a:1	*C. baltica* OL12a	28-32	6.48	34	13	Holmfedlt, 2007, 2012, 2013
	Cellulophaga phi 48:1	*C. baltica* NN016048	NA	6.48	34	13	Holmfedlt, 2007, 2012, 2013
	Roseophage DSS3Φ22	*R. pomeroyi* DSS-3	22	4.25	58	4	This Chapter

2.4.2 Genomic features

To determine the type of nucleic acid carried by DSS3Φ22, the nucleic acids from purified phage were extracted, and digested with S1 nuclease (specific to ssDNA and RNA), DNase I (Specific to both dsDNA and ssDNA) and RNase A (specific to RNA). After SYBR Gold staining, the phage genome could only be observed in RNase A treated samples, but not in S1 nuclease and DNase I treated samples. This result indicated that DSS3Φ22 is a ssDNA bacteriophage (Figure 2.2).

Further genomic sequencing, by both cloning and Illumina MiSeq, confirmed that DSS3Φ22 is a circular ssDNA phage with a genome size of 4,248 bp. Microviruses are among the smallest DNA viruses (Krupovic, 2013), with genomes ranging from 4.4 kb to 6.5 kb (Table 2.1). Although many microviruses which were assembled from metagenomic database carry a slightly smaller genome compared with my isolate, DSS3Φ22 has the smallest genome among isolated microviruses and also represents the smallest bacteriophage in terms of genome size ever isolated.

DSS3Φ22 is GC-rich (58% G+C) compared to other isolated microviruses. This high GC content is reflected in its codon selection and overall amino acid composition, *i.e.* guanine and cytosine are preferred at the third position for synonymous codons. The most frequent amino acids are Gly and Ala, both of which are encoded by GC rich codons and represent *ca*. 10% of total amino acids in translated ORFs. The GC-rich feature of DSS3Φ22 is consistent with the high GC content (64.2%) of its host, *R. pomeroyi* DSS-3 (Moran *et al.*, 2004).

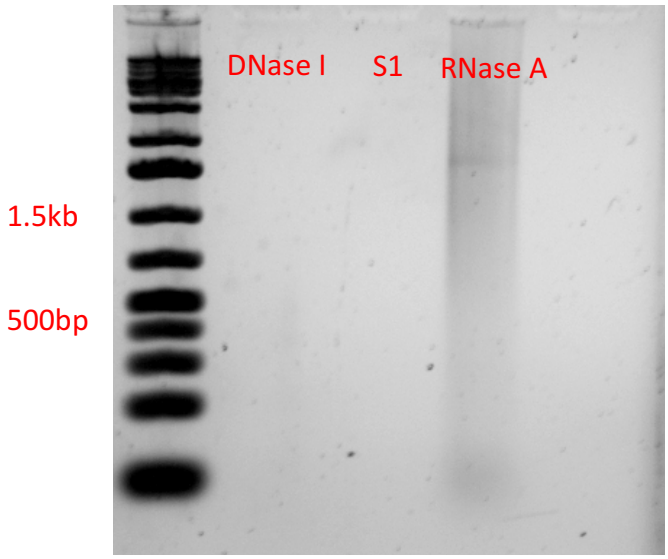


Figure 2.2 The agarose gel showing enzyme digestion of ssDNA phage DSS3Φ22. Nucleic acids of DSS3Φ22 were treated with DNase I (lane 2), S1 nuclease (lane 3), RNase A (lane 4). A GeneRuler 1kb plus DNA Ladder (Thermo Scientific) was used as molecular weight marker. The gel was stained by 1X SYBR Gold.

2.4.3 The simplest genome content for ssDNA phage

Consistent with its small genome size, DSS3Φ22 contains only four putative protein-coding sequences (CDS) encoding a replication protein, major capsid protein, M15_3 family peptidase, and a hypothetical protein (Figure 2.3). A 300 bp gap is present between the hypothetical gene and the capsid gene, which is atypical and was confirmed by multiple annotation strategies, although gaps with this size have been found in other microviruses (Brentlinger *et al.*, 2002; Sait *et al.*, 2011). DSS3Φ22 contains a different gene arrangement compared to other microviruses. The genome organization and coding content in microviruses are relatively conserved, particularly for those infecting *E.coli* Microviruses utilize overlapping reading frames to increase the protein coding capacity within their small viral genomes (Cherwa and Fane, 2011). Therefore, microvirus genomes usually contain 6 to 13 CDSs (Table 2.1). In contrast, DSS3Φ22 does not contain overlapping reading frames despite its smaller genome size. By far, DSS3Φ22 appears to contain the least number of CDSs among the isolated microviruses. DSS3Φ22 encodes three known genes responsible for viral replication, major capsid protein and lysis, and does not contain the genes related to the spike protein, DNA pilot protein and scaffolding protein commonly seen in microviruses (Sanger 1977; Brentlinger et al. 2012). The reduced genomes of microviruses have been found in the metagenomic assembled genomes recovered from soil, human and seawater (Tucker *et al.*, 2010; Roux *et al.*, 2012b; Labonté and Suttle, 2013a; Quaiser *et al.*, 2015), but no ssDNA phages with a reduced genome like DSS3Φ22 have been isolated. Meanwhile, in some metagenomic assembled

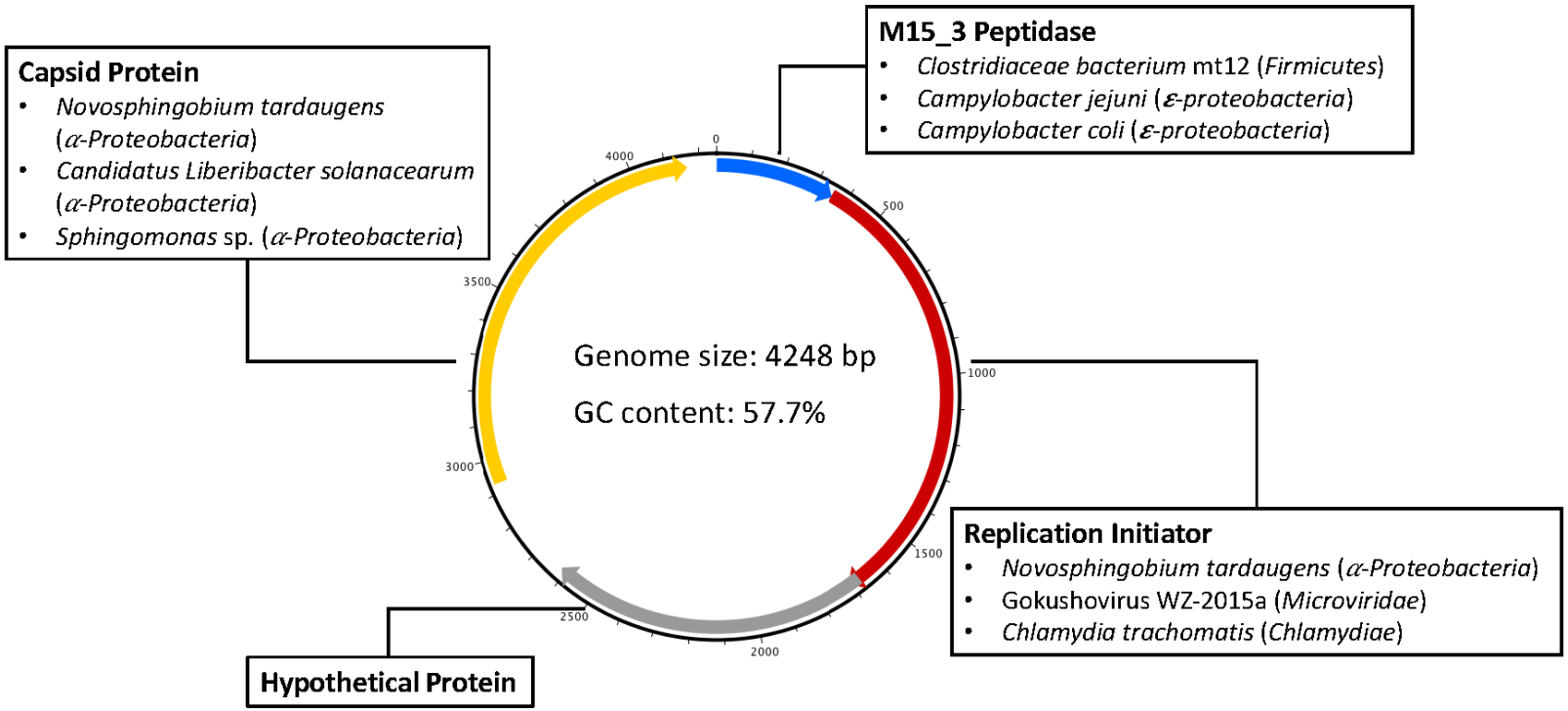


Figure 2.3 Genome map and annotation of DSS3Φ22. The top three hits from NCBI database are listed for each gene.

genomes from human gut, a similar genomic arrangement was observed when compare with that of DSS3Φ22.

The reduced genome size and minimized number of CDSs in DSS3Φ22 certainly places it in the class of smallest and simplest viruses. It is unclear how small microvirus genomes can be further reduced. The smallest ssDNA virus (Porcine circoviruses PCV1) has a genome size of 1,759 bp and only needs two CDSs (replication and capsid genes) to be a lytic virus (Finsterbusch and Mankertz, 2009). It will be interesting to determine whether the peptidase is essential for the lytic function and to understand the potential function of the hypothetical protein of DSS3Φ22.

2.4.4 ssDNA phage can be highly mosaic

Accumulation of point mutations has been thought to be more important than horizontal gene transfer (HGT) for the evolution of small ssDNA phages (Székely and Breitbart, 2016), mainly due to the fact that ssDNA phages have limited genome size and relatively high nucleotide mutation rate (Rokyta *et al.*, 2006; Sanjuán *et al.*, 2010). The DSS3Φ22 encoded genes appear to have multiple origins (Figure 2.3) indicating that HGT may be a dominant force in the evolution of its genome. The major capsid gene of DSS3Φ22 was distantly related to that of other known microvirus isolates (Figure 2.4). Capsid genes found in bacteria such as those in *Rhizobiales* and *Sphingomonadales* seem to share a common ancestor with DSS3Φ22. *Rhizobiales*, *Sphingomonadales* and *Roseobacter* are closely related lineages within the *Alphaproteobacteria* (Williams *et al.*, 2007). The presence of ssDNA capsid gene in

these bacterial genomes suggests a historical integration of ssDNA phage DNA within *Alphaproteobacteria* (see below). It is possible that ssDNA phages infecting *Proteobacteria* could be more closely related to each other, as seen from the ssDNA phages infecting *Cellulophaga*, *Chlamydia* and enterobacteria (Figure 2.4). Earlier studies have shown that microviruses appear to have narrow host ranges and sequences from different microviruses infecting the same host tend to cluster together (Székely and Breitbart, 2016).

A replication gene was identified next to a major capsid gene in *Novosphingobium tardaugens* NBRC 16725 (Figure 2.5). Also, a homolog of the DSS3Φ22 peptidase gene was found upstream of a capsid gene in *Sphingomonas sp.* 67-36 (Figure 2.5). The finding of these DSS3Φ22 homologs in bacterial genomes suggests that the two *Sphingomonadales* species may contain *Microviridae*-like prophages. Prophages attributed to *Microviridae* have been previously identified in *Bacteroides* genomes (Krupovic and Forterre, 2011). By analyzing assembled phage genomes from a peatlands metagenomic, *Microviridae*-associated prophages were also predicted in *Bacteroides* genomes (Quaiser *et al.*, 2015).

DSS3Φ22 contains the M15_3 family peptidase, which has not been previously reported in microviruses. The M15 family peptidase is widely distributed in bacteria, where it is involved in cell wall biosynthesis and metabolism (Bochtler *et al.*, 2004). In dsDNA phages, homologous peptidases cleave the amino acid bond of cell wall peptidoglycan (Loesser *et al.*, 1995). Having a peptidase enables phage to circumvent bacterial resistance afforded by the cell wall and adapt to new environmental conditions (Oliveira *et al.*, 2013).The best matches of the DSS3Φ22

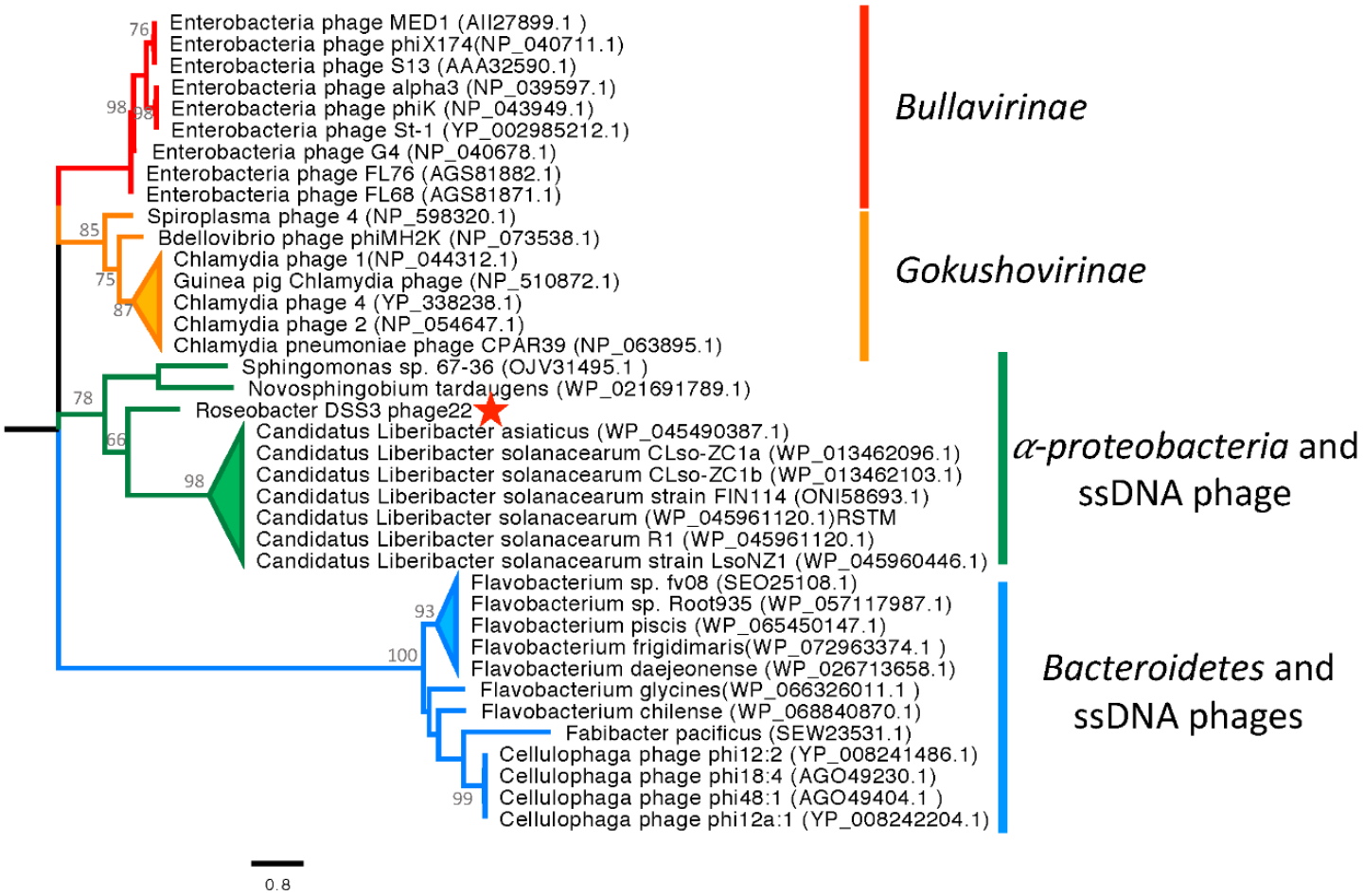


Figure 2.4 Maximum-likelihood analysis of DSS3Φ22 based on amino acid sequences of major capsid protein. Bootstrap replicates = 500.

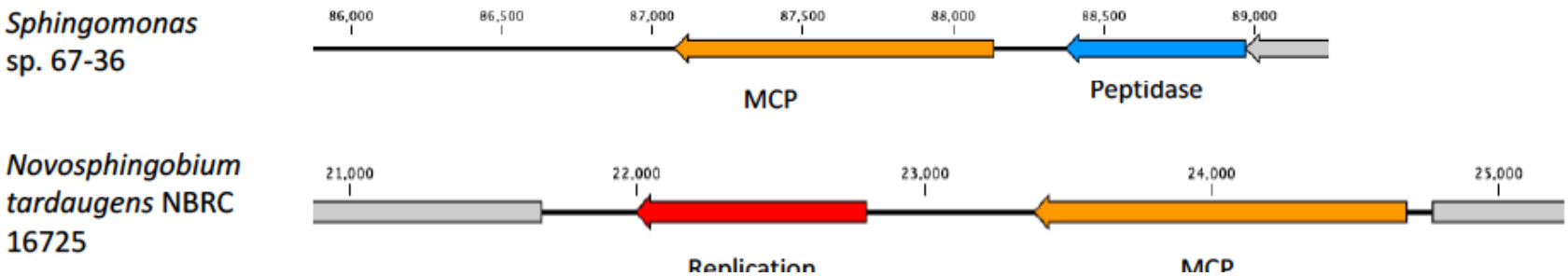


Figure 2.5 Genomic organization of the putative microvirus prophages residing in the genomes of *Sphingomonas* and *Novosphingobium*.

peptidase by BLAST are the peptidases from *Firmicutes* and *Proteobacteria* (Figure 2.3), suggesting that DSS3Φ22 may have acquired this gene from bacteria through HGT. Although the M15_3 family peptidases have not been found in known microvirus isolates, they were found in 11 metagenomic assembled microvirus genomes collected from human guts and lungs (Roux *et al.*, 2012b).

The highly mosaic nature of the DSS3Φ22 genome, conservation of genetic arrangement and sequence of DSS3Φ22 relatives in bacterial genomes and genetic/nucleotide diversity in environmental microvirus metagenomes suggest that HGT is also important for the genomic evolution and diversification of ssDNA phages. With more ssDNA phages being discovered, the role of HGT in microviruses can be better assessed.

2.4.5 Missing attachment protein in DSS3Φ22

Currently, *Bullavirinae* (represented by ΦX174 which infects *E.coli*) and *Gokushovirinae* (represented by ΦMH2K which infects *Bdellovirbrio bacteriovorus*) are the two large subfamilies in *Microviridae*. The vast majority of ssDNA phage isolates belong to these two subfamilies. The principal difference between *Bullavirinae* and *Gokushovirinae* is the existence of two structural genes and the complexity of the major capsid protein (Székely and Breitbart, 2016). Members within *Bullavirinae* all encode a spike protein and external scaffolding protein, involved in attachment and viral entry (Jazwinski *et al.*, 1975). *Gokushovirinae* lack these two genes, but they encode a more complex major capsid protein compared to *Bullavirinae*. The genes coding for the capsid proteins of *Gokushovirinae* contain an

insertion loop (IN5) that forms a “mushroom-like” surface protrusion at each icosahedral threefold axis (Chipman *et al.*, 1998). This insertion loop is thought to be a relic of the spike protein and external scaffolding protein (Brentlinger *et al.*, 2002), and responsible for receptor recognition (Read *et al.*, 2000b).

The major capsid protein is the only annotated structural gene for DSS3Φ22. Only one band was seen on the protein gel (Figure 2.6) and was confirmed to be the major capsid protein by mass spectrometry. No spike protein and scaffolding protein were found in the DSS3Φ22 genome by *in silico* annotation. The multiple alignment of major capsid proteins among diverse microviruses revealed that DSS3Φ22 also lacks the IN5 insertion loop common to members within *Gokushovirnae* (Figure 2.7). Furthermore, a 3-D model of the DSS3Φ22 major capsid protein was constructed using and compared with the *Gokushovirinae* SpV4 protein VP1 (PDBID: 1KVP). The comparison of 3-D protein structures further confirms the absence of an extended loop in the major capsid protein of DSS3Φ22, which forms the “mushroom-like” protrusion in the *Gokushovirinae* homologs (Figure 2.8). These results suggest that DSS3Φ22 does not contain the attachment proteins described for *Bullavirinae* and *Gokushovirinae* representatives. It is noteworthy that the IN5 insertion loop and the spike protein are also not found in ssDNA phages infecting marine *Bacteroidates* (Holmfeldt *et al.*, 2013).Therefore, it is possible that the attachment protein might be provided by the hosts.

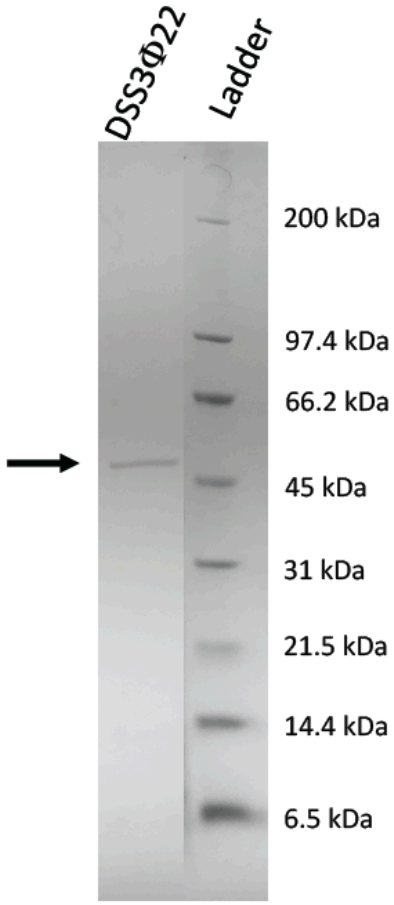
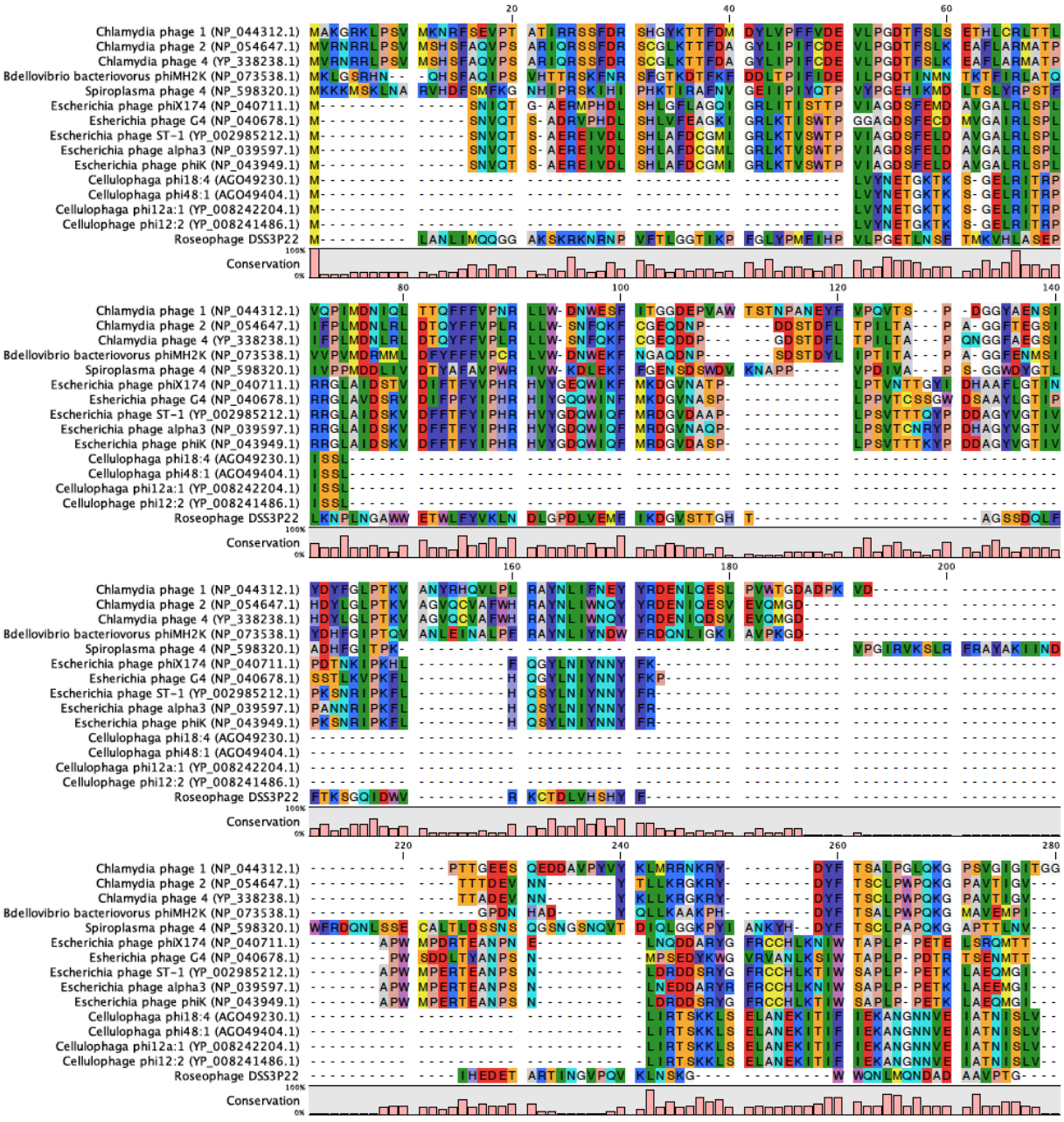
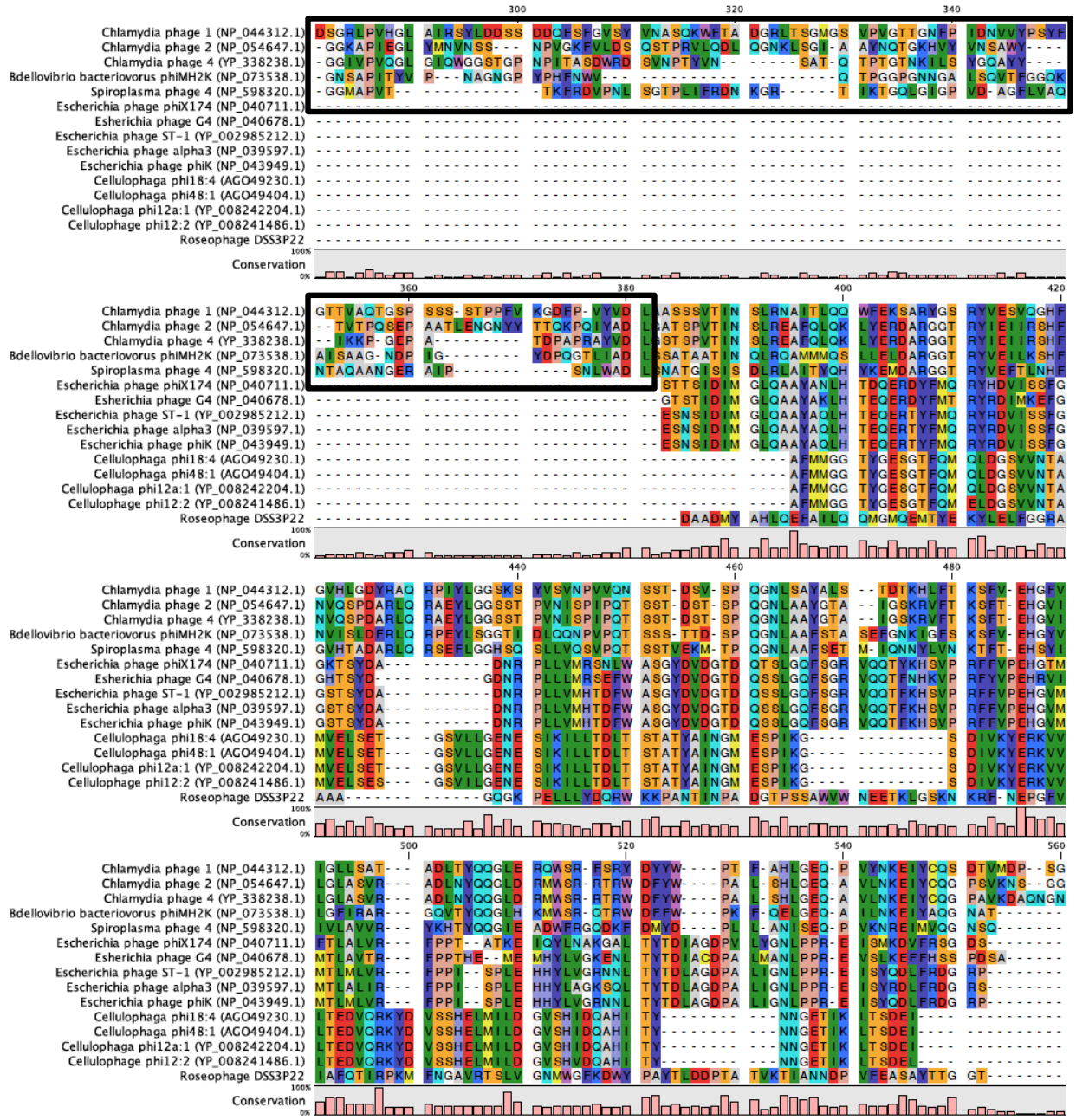


Figure 2.6 Visualization of the DSS3Φ22 structural protein separated by SDS-PAGE. The arrow indicates the only structural protein of DSS3Φ22.The molecular weights of standard proteins were listed next to the bands.


Chlamydia phage 1 (NP_044312.1)
Chlamydia phage 2 (NP_054647.1)
Chlamydia phage 4 (YP_338238.1)
Bdellovibrio bacteriovorus phiMH2K (NP_073538.1)
Spiroplasma phage 4 (NP_598320.1)
Escherichia phage phiX174 (NP_040711.1)
Esherichia phage G4 (NP_040678.1)
Escherichia phage ST-1 (YP_002985212.1)
Escherichia phage alpha3 (NP_039597.1)
Escherichia phage phiK (NP_043949.1)
Cellulophaga phi18:4 (AGO49230.1)
Cellulophaga phi48:1 (AGO49404.1)
Cellulophaga phi12a:1 (YP_008242204.1)
Cellulophage phi12:2 (YP_008241486.1)
Roseophage DSS3P22
Conservation
100%
0%
20
40
60
80
100
120
140
160
180
200
220
240
260
280


Chlamydia phage 1 (NP_044312.1)
Chlamydia phage 2 (NP_054647.1)
Chlamydia phage 4 (YP_338238.1)
Bdellovibrio bacteriovorus phiMH2K (NP_073538.1)
Spiroplasma phage 4 (NP_598320.1)
Escherichia phage phiX174 (NP_040711.1)
Esherichia phage G4 (NP_040678.1)
Escherichia phage ST-1 (YP_002985212.1)
Escherichia phage alpha3 (NP_039597.1)
Escherichia phage phiK (NP_043949.1)
Cellulophaga phi18:4 (AGO49230.1)
Cellulophaga phi48:1 (AGO49404.1)
Cellulophaga phi12a:1 (YP_008242204.1)
Cellulophage phi12:2 (YP_008241486.1)
Roseophage DSS3P22
Conservation

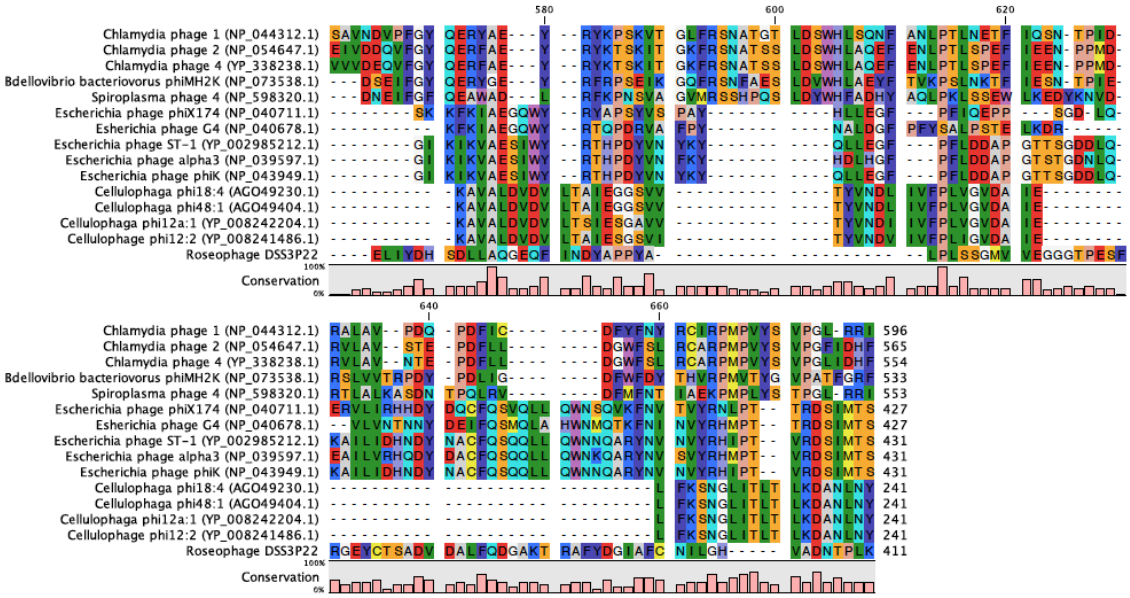


Figure 2.7 Multiple sequences alignment of major capsid proteins of *Microviridae* phages. Different colors represent different amino acids. The box indicates the insertion loop (IN5) which is specifically found in *Gokushovirinae*.

Figure 2.8 The predicted 3-D structure comparison of the major capsid protein of DSS3Φ22 alongside that of Sprioplasma phage 4 major capsid protein (shown in purple backbone trace, PDB ID: 1KVP). The arrow indicates the extended loop of Sprioplasma phage 4 that is missing from DSS3Φ22.

2.4.6 Evidence of DSS3Φ22 homologs in environmental metagenome

Based on the reciprocal Blast metagenomic search using DSS3Φ22-like sequences as query, sequences similar to the DSS3Φ22 genes can be found in a wide range of environments, ranging from deep sea, human gut and feces to coral-associated environments (Figure 2.9). The majority of the matches were to the capsid gene and replication initiator gene. I did not intend to compare the resultant sequences from different habitats due to the bias on ssDNA viruses by the multiple displacement amplifcation (MDA) (Roux *et al.*, 2016). Therefore the relative abundance of ssDNA virus from metagenomic search is not corelated with their abundance in the natural enviroment.

The major capsid gene has been widely used as a gene marker to study the diversity and evolution of *Microviridae* (Labonté and Suttle, 2013a). I conducted a phylogenetic analysis that includes all the microvirus isolates and hundreds of assembled microvirus genomes obtained from metagenomes. *Bullavirinae*, *Gokushovirinae* and three candidate subfamilies (*Alpavirinae* [Roux *et al.*, 2012b], *Aravirinae* [Quaiser *et al.*, 2015], *Pequenovirus* [Bryson *et al.*, 2015]) defined by assembled microvirus genomes were well separated based on the capsid gene phylogeny (Figure 2.10). Surprisingly, DSS3Φ22 clusters together within the candidate subfamily *Alpavrinae* (Figure 2.10). Currently, *Alpavirinae* only contains assembled ssDNA phage genomes from human gut and feces metagenomes (Roux *et al.*, 2012b), and no isolated phage has been reported for *Alpavirinae*. *Alpavirinae* appears to include multiple distinct lineages, and DSS3Φ22 forms a deep branch

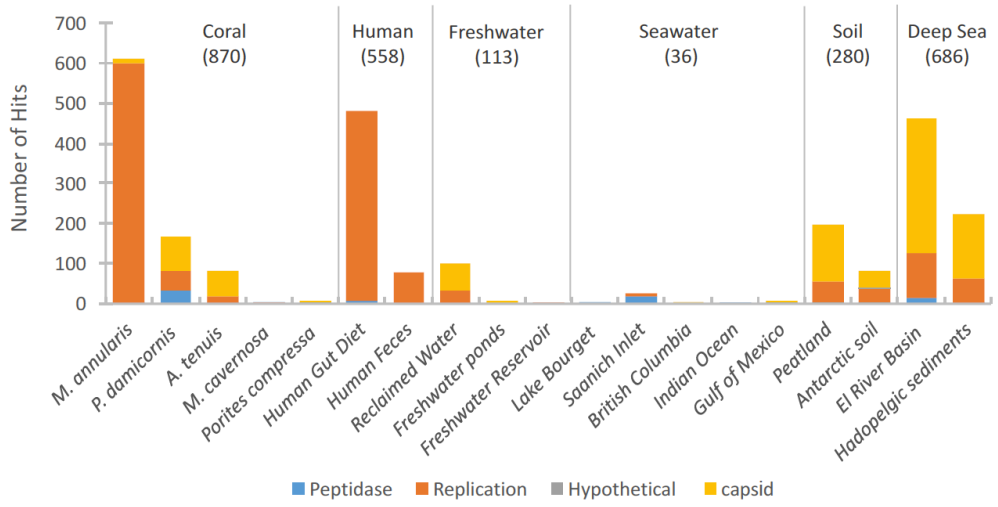


Figure 2.9 Abundance of DSS3Φ22-like sequences in different habitats based on viral metagenomic search.

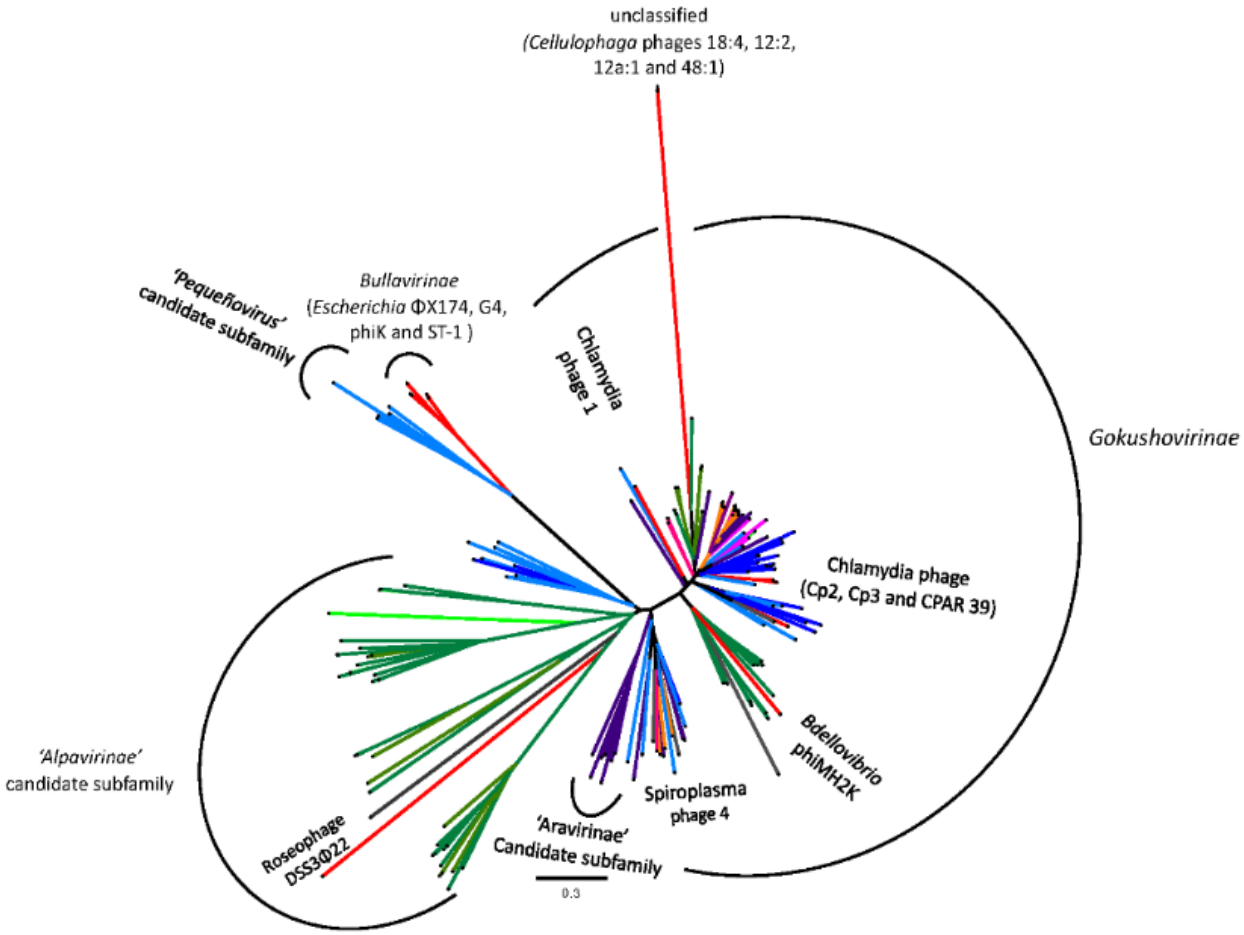


Figure 2.10 A maximum-likelihood phylogenetic relationship of known members of the *Microviridae* and the metagenomic assembled microviruses based on amino acid sequences of the major capsid gene. Bootstrap = 500. The

phages (red) included here are DSS3Φ22; Chlamydia phage 1 (NP_044312.1), 2 (NP_054647.1), and 4 (YP_338238.1); Spiroplasma phage 4 (NP_598320.1) and *Bdellovibrio bacteriovorus* phiMH2K (NP_073538.1); *Escherichia* phage phiX174 (NP_040711.1), G4 (NP_040678.1), phiK (NP_043949.1) and ST-1(YP_002985212.1); *Cellulophaga* phi18:4 (AGO49230.1), 12:2 (YP_008241486.1), 12a:1 (YP_008242204.1) and 48:1(AGO49404.1). The metagenomic assembled microviruses include: 19 assembled phage genomes from peatlands (purple) (KM589498-KM589516), 6 assembled phage genomes from seawater (pink) (KC131021-KC131025, HQ157198 and HQ157198), 22 assembled phage genomes from methane seep sediments (light blue) (KP0879360-KP087957), 18 assembled phage genomes from pelagic sediments (dark blue) (DRA000564), 81 assembled phage genomes from human gut (dark green), human feces (green), human lung (lime), fresh lake (orange), seawater (pink) (http://dx.doi.org/10.5061/dryad.8ht80), 2 assembled phage genomes from hot spring (Maroon) (NC_028994, NC_028993.1), and others (grey) (HQ2664138 and KM276541).

within this subfamily. A lack of resolution indicates that the number of isolated ssDNA phages is still very limited (Figure 2.10, red branches).

In general, microviruses have been discovered in diverse environments, including peatlands (Quaiser *et al.*, 2015), seep sediments (Bryson *et al.*, 2015), North Atlantic Ocean (Tucker *et al.*, 2010), deep sea (Yoshida *et al.*, 2013), coastal water (Labonté and Suttle, 2013a) and human-related related environments (Roux *et al.*, 2012b). Comparing the relative abundance of ssDNA viruses across different viral metagenomic databases can be challenging due to the lack of standardization and rapid improvement of sequencing technologies. Currently, nearly all of the marine viral metagenomic databases are biased for dsDNA viruses (Hurwitz and Sullivan, 2013; Brum *et al.*, 2015). Meanwhile, the use of MDA selects for small circular ssDNA templates, which obscures the relative abundance of ssDNA viruses (Kim and Bae, 2011; Marine *et al.*, 2014). The development of new metagenomic strategies, especially in library preparation, is necessary in order to fully understand the diversity of ssDNA viruses (Székely and Breitbart, 2016; Roux *et al.*, 2016).

2.5 Conclusion

I isolated a lytic phage, DSS3Φ22, which infects a marine roseobacter strain, *R. pomeroyi* DSS-3 and belongs to an unclassified genus within the *Microviridae*. Among the known microvirus isolates, DSS3Φ22 contains the fewest number of coding genes (4 ORFs) and has the smallest genome size (4.3 kb). The genomic arrangement of DSS3Φ22 is unique relative to known microvirus isolates and may be diagnostic for detecting prophage integration into bacterial genomes. The lack of a

spike protein and the presence of a host-like peptidase gene are other novel features compared to isolated ssDNA microviruses. Evidence for HGT of DSS3Φ22 genes suggest that ssDNA viruses can be highly mosaic and that gene transfer can play a substantial role in their evolution. The model of mutation-mediated evolution has been thought to be the major force diversifying microviruses, but may now be challenged as new microviruses are being discovered. The finding of this unclassified phage in *Microviridae* provides new insight into the biology, evolution and diversity of microviruses.

Current viral metagenomic databases are heavily biased towards dsDNA viruses due to the technological limitations of sequencing other genome types. Typical sequencing methods, such as Illumina and PacBio, heavily biased towards dsDNA than ssDNA. Several recent virome studies that considered ssDNA viruses have unveiled a tremendous diversity of microviruses across different environments (Tucker *et al.*, 2010; Labonté and Suttle, 2013a; Labonté and Suttle, 2013b; Yoshida *et al.*, 2013; Hopkins *et al.*, 2014 Quaiser *et al.*, 2015; Bryson *et al.*, 2015). With rapidly increasing ssDNA viromics, I predict novel lineages of assembled microviruses will continue to emerge. However, too few microvirus isolates are currently available. Our study demonstrated the value of phage isolation. I propose DSS3Φ22 as a potential model phage for infection of *Proteobacteria* by ssDNA viruses. Isolated phages not only provide important links to hosts, but also permit studying phage morphology, infectivity and genomics. The combination of viral genomics and viral metagenomics has become a powerful way to understand the diversity and interaction between viruses and microbes in the natural environment.

Chapter 3. A novel roseobacter phage possesses features of podoviruses, siphoviruses, prophages and gene transfer agent

3.1 Abstract

Bacteria in the *Roseobacter* lineage have been studied extensively due to their significant biogeochemical roles in the marine ecosystem. However, our knowledge on bacteriophages which infect roseobacters is limited. Here, I report a new bacteriophage, phage DSS3Φ8(a lytic siphovirus), which infects a marine roseobacter strain, *Ruegeria pomeroyi* DSS-3. Genomic analysis shows that DSS3Φ8 is most closely related to a group of siphoviruses, termed CbK-like phages, which infect the freshwater bacterium *Caulobacter crescentus*. DSS3Φ8 contains a smaller capsid and has a reduced genome size (146 kb) compared to other CbK-like phages (205-279 kb). DSS3Φ8 contains a DNA polymerase gene which is closely related to those of T7-like podoviruses. DSS3Φ8 also contains integrase and repressor genes, indicating its potential to enter the lysogenic cycle. In addition, four GTA (gene transfer agent) homologs were identified in the DSS3Φ8 genome. Genomic analysis suggests that DSS3Φ8 is a highly mosaic phage that inherited genetic features from siphoviruses, podoviruses, prophages and GTAs. This is the first report of CbK-like phages infecting marine bacteria. I believe phage isolation is a powerful tool that can lead to discovery of new phages and help interpret the vast numbers of unknown sequences in the viral metagenomics.

3.2 Introduction

Bacteria in the *Roseobacter* clade are abundant and widely distributed in the marine environment (Moran *et al.*, 2007; Gonzalez and Moran, 1997). The broad distribution of roseobacters in nature can be largely attributed to their diverse metabolic capabilities (Buchan *et al.*, 2005; Wagner-Döbler and Biebl, 2006; Brinkhoff *et al.*, 2008; Moran *et al.*, 2012). Currently, more than 69 genomes of different marine *Roseobacter* species have been sequenced, and analyses of representative genome sequences have generated new insight into ecological adaptation and evolution of the marine *Roseobacter* lineage (Luo and Moran, 2015).

Compared to the knowledge we gained on roseobacters, much less is known about bacteriophages which infect those bacteria (roseophage). Marine viruses, as the most abundant biological entities in the biosphere, play important roles in shaping host population structures, mediating genetic exchange between hosts, and modulating trophic transfer in marine food webs (Fuhrman, 1999; Wommack and Colwell, 2000; Suttle, 2005, 2007). Viral metagenomic studies have greatly expanded our knowledge of viral diversity in nature (Angly *et al.*, 2006; Hurwitz and Sullivan, 2013; Brum *et al.*, 2015), but the interpretation of data is often hampered due to the limitation of our knowledge of isolated viruses. Recent isolations of phages infecting two abundant marine bacteria are the great examples for linking the unknown phages in the metagenomic database with known phages (Zhao *et al.*, 2013; Kang *et al.*, 2013). In addition, for the isolated phages, a wealth of biological information such as morphology, lytic activity, host specificity, genomics, and evolution can be retrieved.

Currently, only a handful of phages which infect roseobacters have been isolated and characterized. Roseophage SIO1 was the first described phage which infects *Roseobacter* SIO67 (Rohwer *et al.*, 2000), and similar SIO1-like phages were found at the same site 10 years later (Angly *et al.*, 2009). Two N4-like phages infecting two different roseobacter strains (*Silicibacter pomeroyi* DSS-3 and *Sulfitobacter sp.* EE-36) were discovered, and they represent the first report of N4-like phages infecting marine bacteria (Zhao *et al.*, 2009b). Recently, several N4-like phages were isolated from different strains within the *Roseobacter* clade (Ankrah *et al.*, 2014a; Chan *et al.*, 2014; Ji *et al.*, 2015b), suggesting that N4-like phages may be a common type of phages infecting roseobacters. One siphovirus, phage RDJLΦ1, which infects *Roseobacter denitrificans* OCh114 was isolated from the South China Sea (Zhang and Jiao, 2009). Most recently, two temperate podoviruses infecting *Sulfitobacter sp.* strain 2047 were also described (Ankrah *et al.*, 2014b). These two temperate phages are almost identical at nucleotide level and were isolated from a mesocosm study in Raunefjorden.

Roseobacters also contain inducible prophages. Prophages have been induced from two strains, *Roseobacter* sp. TM1040 (Chene *et al.*, 2006) and *Roseovarius nubinhibens* ISM (Zhao *et al.*, 2010). Some bacterial strains (i.e. TM1040) may contain multiple prophages in their genomes (Paul, 2008). Another phage-like structure, the gene transfer agent (GTA), has been found in nearly all the roseobacter genomes (Zhao *et al.*, 2009a; Newton *et al.*, 2010; Luo and Moran, 2015). GTA is a phage-like entity that transfers random pieces (~4kb in size) of genome from donor cells to recipient cells (Lang and Beatty, 1999). Interestingly,

roseophage RDJLΦ1 contains four GTA-like genes in its genome (Huang *et al.*, 2011). Roseobacters are a group of highly diverse and adaptable bacteria. It is believed that phages, prophages and GTAs are the driving force for genomic diversification of marine *Roseobacter* lineage.

In this study, I report a new phage, phage DSS3Φ8, which infects a marine roseobacter, *R. pomeroyi* DSS-3. Phage DSS3Φ8 is a highly mosaic phage because it contains genetic elements of podoviruses, siphoviruses, prophages, and GTAs. To our knowledge, no phage similar to the very unusual DSS3Φ8 has been found in marine bacteria prior to this study.

3.3 Methods and Materials

3.3.1 Isolation of phage

Host stain *R. pomeroyi* DSS-3 was grown in YTSS medium (4 g yeast extract, 2.5 g trypone and 20 g Crystal Sea per liter) at 28 °C. Water samples, collected from Baltimore Inner Harbor Pier V in January 2012, were filtered through 0.22 μm polycarbonate membrane filters (Millipore, Bedford, MA, USA). Filtrate of 10 ml was added into 50 ml of exponentially growing bacterial cultures and incubated for two days. The DSS-3 culture was filtered through 0.22 μm polycarbonate membrane filter (Millipore, Bedford, MA, USA) to remove bacteria. 100 μl of this cell-free lysate was added to 500 μl of exponentially growing DSS-3 culture, and plated using plaque assay. Phage isolates were purified three times by plaque assay.

3.3.2 Transmission electron microscopy (TEM)

For TEM, one drop (2 µl) of purified phage lysate was adsorbed to the 200-mesh Formvar/carbon-coated copper grid and stained with 1% phosphotungstic acid (PTA) for one minute (Brenner and Horne, 1959). Samples were examined with a JEOL JEM2100F transmission electron microscope, at the University of Oldenburg.

3.3.3 Cross infection

Cross-infectivity of roseophage DSS3Φ8 was tested against five other marine roseobacter stains, including *Roseovarius nubinhibens* ISM, *Silicibacter* sp. TM1040, *Sulfitobacter* sp. EE-36, *Dinoroseobacter shibae* DFL-12 and *Roseobacter denitrificans* OCh 114. For each strain, 500 µl of exponentially growing cells was added to 5 ml top agar and poured on the plates. After the agar was solidified, one drop of purified phage lysate was spotted onto each plate, and incubated for 2–3 days at 28 °C (Angly *et al.*, 2009). The formation of plaques was assessed.

3.3.4 One-step growth curve

One-step growth curve was determined using a method described elsewhere (Hyman and Abedon, 2009). Generally, exponentially growing cultures of *R. pomeroyi* DSS-3 were incubated on ice for 20 min to synchronize the growth of the host. After the DSS-3 culture was recovered to room temperature, roseophage DSS3Φ8 was added into the culture at a multiplicity of infection (MOI) of 0.5. After 20 min attachment, cells were pelleted, re-suspended and diluted 100 times to avoid possible

secondary infection. An aliquot of the cell suspension was collected at different time points between 30 min and 12 h.

The number of phages was enumerated by using qPCR (Edelman and Barletta, 2003; Anderson *et al.*, 2011; Ankrah *et al.*, 2014c). Collected samples were filtered through 0.22 µm polycarbonate membrane filters (Millipore, Bedford, MA, USA) and DNA was extracted by Wizard® PCR Preps DNA Purification System (Promega Corporation, WI, USA). A set of PCR primers was designed based on the sequence of the phage DNA polymerase (Forward: ATGCTGCTCCGAACGTATCT, Reverse: ACTCGCCCTTCTTTTCCTTC). qPCR was conducted by Applied Biosystems®, 7500 Fast Real-time PCR system (ThermoFisher Scientific, New York, USA). qPCR reactions were performed in a 25 µl volume with 12.5 µl PerfeCTa® SYBR® Green SuperMix (Quanta Biosciences, Gaithersburg, MD, USA), 400 nM forward and reverse primers and 5 µl of template. The amplification program was set as: 95 °C for 10 min, 40 cycle of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s. Fluorescence measurements were conducted at 60 °C of each cycle. Standards were developed using DNA extracted from phage lysate with known titer (1E10 viral particles/ml) and 10-fold serial dilutions of the DNA (until 1E5 viral particles/ml) were used in the reactions. Standard curves were determined as the correlation between the log of gene copy numbers and the Ct (R^2 = 0.99).

3.3.5 DNA extraction

One-liter phage lysate was treated with DNase and RNase at a final concentration of 2 µg/ml (for both enzymes) at 4 °C to remove free DNA and RNA.

Phage particles in the supernatant were precipitated with polyethylene glycol 8000 (final 10% w/v), and centrifuged in an OptiPrep™ Density Gradient Medium for 12 h at 41,000 g. The visible viral band was extracted and dialyzed against TM buffer overnight at 4 °C. The purified phage particles were treated with a combination of sodium dodecyl sulfate (SDS, final concentration 1% w/v), proteinase K (final concentration 30 μg/ml) and EDTA (final concentration 5 μM) at 55 °C for 3 h. Phage DNA was then extracted using phenol/chloroform/isoamyl alcohol (25:24:1).

3.3.6 Sequencing and annotation

The genome of roseophage DSS3Φ8 was sequenced on an Illumina GAIIx sequencer. A 112 bp paired-end run resulted in 157,500 reads. The high quality reads were assembled by MIRA assembler, resulting in three contigs (139,090 bp, 2,108 bp and 4,767 bp in size). Primer walking with the Sanger sequencing method was used to close the circular phage genome. Sanger sequencing was performed using an ABI 3100 genetic analyzer (PE Applied Biosystems, Foster City, CA, USA) at the Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science.

The open reading frames (ORFs) were predicted using GeneMarkS (Besemer *et al.*, 2001) and tRNAscan-SE (Lowe and Eddy, 1996) was used to identify tRNA sequence. Translated ORFs were compared with known protein sequences using BLASTp against NCBI's non-redundant databases (E-value ≤ 1E-5 and identity ≥ 30%). The complete genome sequence was submitted to GenBank with the accession number KT870145.

Sequence alignment and phylogenetic analysis for both *pol*A and four concatenated GTA were performed using the CLC Genomic 7 program. The amino acids sequences of the *pol*A and the GTA genes were used for alignment. A maximum-likelihood method was used to construct the phylogenetic tree, with bootstrap value of 500.

3.3.7 Metagenomic recruitment

To explore the geographic distribution of DSS3Φ8, the phage genes were used as queries to search against the metagenomic databases in iMicrobe (http://data.imicrobe.us/). A reciprocal BLASTp analysis was conducted. Each of the recruited metagenomic reads was compared against the NCBI RefSeq database (June 2015), which contains all complete viral and bacterial genomes. The metagenomic reads were considered as DSS3Φ8 origin only if the DSS3Φ8 genes were the best hits. The relative abundance of each gene in every metagenomic database was calculated following the methods described elsewhere (Zhao *et al.*, 2013). The count for each recruited read was divided by the number of total reads in the database, and further divided by the size of the gene product. Samples were scaled using the mean of all samples.

3.4 Results and Discussion

3.4.1 Morphology and biological features

Roseophage DSS3Φ8 was isolated from Baltimore Inner Harbor, Baltimore, USA in January 2012. DSS3Φ8 has a prolate cylindrical head and a long, flexible,

non-contractile tail. The capsid size of DSS3Φ8 is *ca.* 100 nm long and 35 nm wide, and the tail length is *ca.* 300 nm (Figure 3.1 A). Based on its morphology, DSS3Φ8 is a member of the B3 morphotype of *Siphoviridae*, which comprises approximately 1% of all known phages (Ackermann, 2001). Genome sequence showed that DSS3Φ8 belongs to the phiCbK genus under *Siphoviridae*, which contains five B3 morphotype bacteriophages infecting the freshwater bacterium *Caulobacter crescentus* (Gill *et al.*, 2012). Phage CbK was first isolated in 1970, and has been an important genetic and cytological tool for studying cell cycle regulation, because it specifically infects the "swarmer" cell of its host (Agabian-Keshishian and Shapiro, 1970). Roseophage DSS3Φ8 is much smaller than previously characterized the five CbK-like phages in terms of capsid size and tail length.

The cross infectivity tests showed that DSS3Φ8 specifically infects its own host, *R. pomeroyi* DSS-3, but none of the other marine strains tested. The latent period of DSSΦ8 was about 2 hours, and it reached the growth plateau in 6 hours. The burst size of DSSΦ8 is *ca.* 120 (Figure 3.1 B).

3.4.2 General genome features

DSS3Φ8 is a circular, double-stranded DNA virus. The genome size of DSS3Φ8 is 146,135 bp. In general, the genome of DSS3Φ8 is homologous to that of the CbK-like phages (see the later section), but the genome size of DSS3Φ8 is much smaller compared to the known CbK-like phages (205-279 kb) (Gill *et al.*, 2012). The G+C content of the phage genome is 56%, lower than that of other CbK-like phages

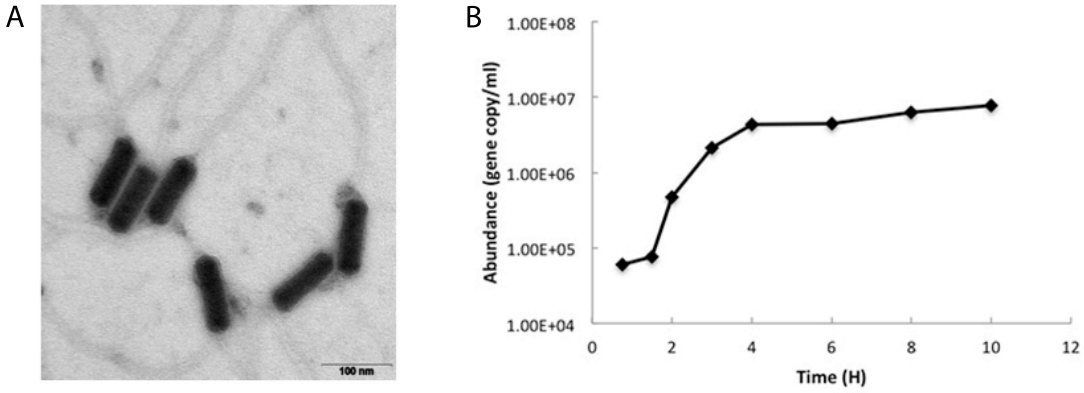


Figure 3.1 Morphology and growth curve of DSS3Φ8. (A) Transmission electron microscopy image of roseophage DSS3Φ8. Scale bar = 100 nm. (B) One-step growth curve of roseophage DSS3Φ8.

(62%-66%). The G+C content is 64 and 67% for *R. pomeroyi* DSS3 and *Caulobacter crescentus*, respectively.

A total of 229 predicted open reading frames (ORFs) was identified in the DSS3Φ8 genome, of which 106 ORFs had recognizable homologs in GenBank (Figure 3.2). Among the recognizable proteins, only 50 have described function, consistent with the idea that marine siphoviruses encode proteins that are under-represented in the database (Sullivan *et al.*, 2009).

DSS3Φ8 contains 34 tRNA genes coding 14 different amino acids. Most of tRNA genes found in phage DSS3Φ8 overlapped with those found in the host *R. pomeroyi* DSS3. The parallel possession of tRNA genes between DSS3Φ8 and host DSS3 is considered as a result of co-evolution of phage and host (Krakauer and Jansen, 2002; Bahir *et al.*, 2009). On the other hand, two tRNAs, $tRNA^{Lys/CTT}$ and $tRNA^{Gln/CTG}$, were only found in DSS3Φ8. The anticodon CTT and CTG account for more than 90% codon usage for Lys and Gln respectively for phages, however, these anticodon are the rarest codons in the host (Moran *et al.*, 2004). This finding supported previous findings that phages tend to carry tRNAs corresponding to codons that are used frequently by the phage genes while rare in the host genome (Bailly-Bechet *et al.*, 2007). Also, it is possible that tRNA genes present in phages facilitate the expression of the late phage gene encoding structural proteins as proposed from cyanophage Syn9 (Weigele *et al.*, 2007).

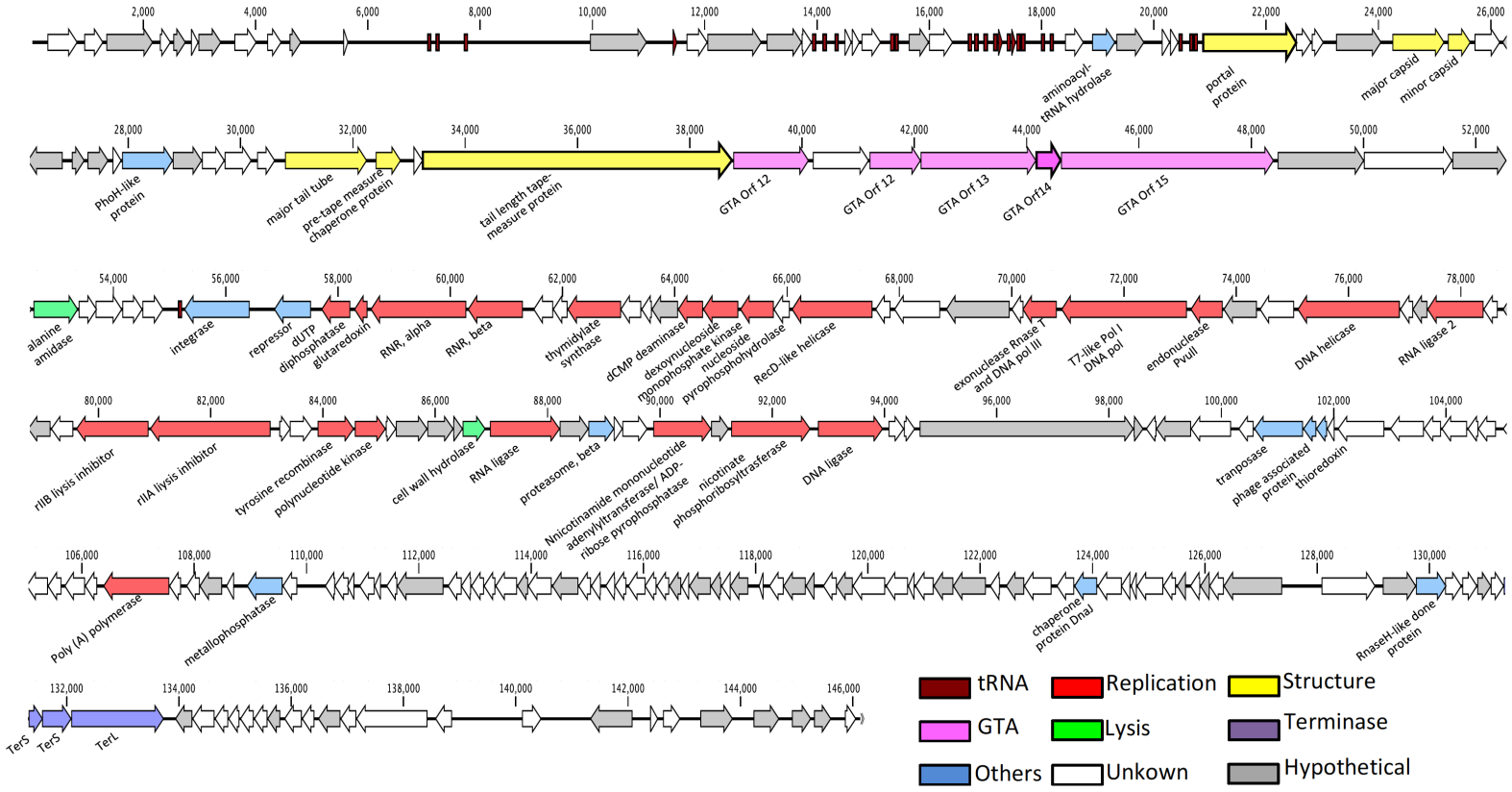


Figure 3.2 Genome map of roseophage DSS3Φ8. Gene features (unknown, hypothetical, tRNA genes, GTA and others) and genome modules (structure, lysis and DNA replication) are color-coded according to the legend below the figure

3.4.3 Homology between DSS3Φ8 and CbK-like phages

Genome wide comparison between DSS3Φ8 and the CbK-like phages reveals that they all share similar modules (Figure 3.3). Of the 106 characterized ORFs in the DSS3Φ8 genome, 24 were related to the CbK-like phages, including the genes involved in DNA metabolism, replication, structure and the genes with unknown functions. The phylogenetic analyses based on the DNA polymerase gene support the proposal that DSS3Φ8 is more closely related to the CbK-like phages than to other phages (Figure 3.4 B). It forms a deep branch of its own among the CbK-like podoviruses clade. Similar pattern was observed by using large subunit of terminase and major capsid protein as gene markers.

Although siphoviruses are usually highly mosaic in terms of their genomic evolution, siphoviruses with prolate capsid appear to maintain a conserved genomic architecture among them even though their genome sizes vary dramatically. For example, phage S-CBS2 (Sullivan *et al.*, 2009), a siphovirus with prolate capsid, infecting a marine *Synechococcus* strain, exhibits a genomic arrangement similar to P-SS2 (Sullivan *et al.*, 2009), a siphovirus with prolate head, infecting a marine *Prochlorococcus* strain. The genome size of phage S-CBS2, however, is only two thirds of that of phage P-SS2 (Huang *et al.*, 2012). In my case, DSS3Φ8 infecting a marine *Alphaproteobacterium*, *R. pomeroyi* DSS3, shared the highest genome homology with six siphoviruses CbK-like phages infecting a freshwater *Alphaproteobacterium*, *Caulobacter crescentus,* despite that the genome size of DSS3Φ8 is about half of the genome sizes of CbK-like phages. The high genetic homology between DSS3Φ8 and CbK-like phages is remarkable because they infect

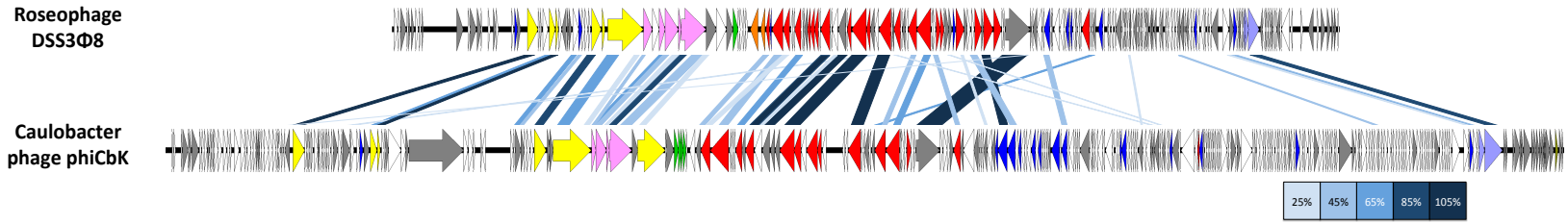
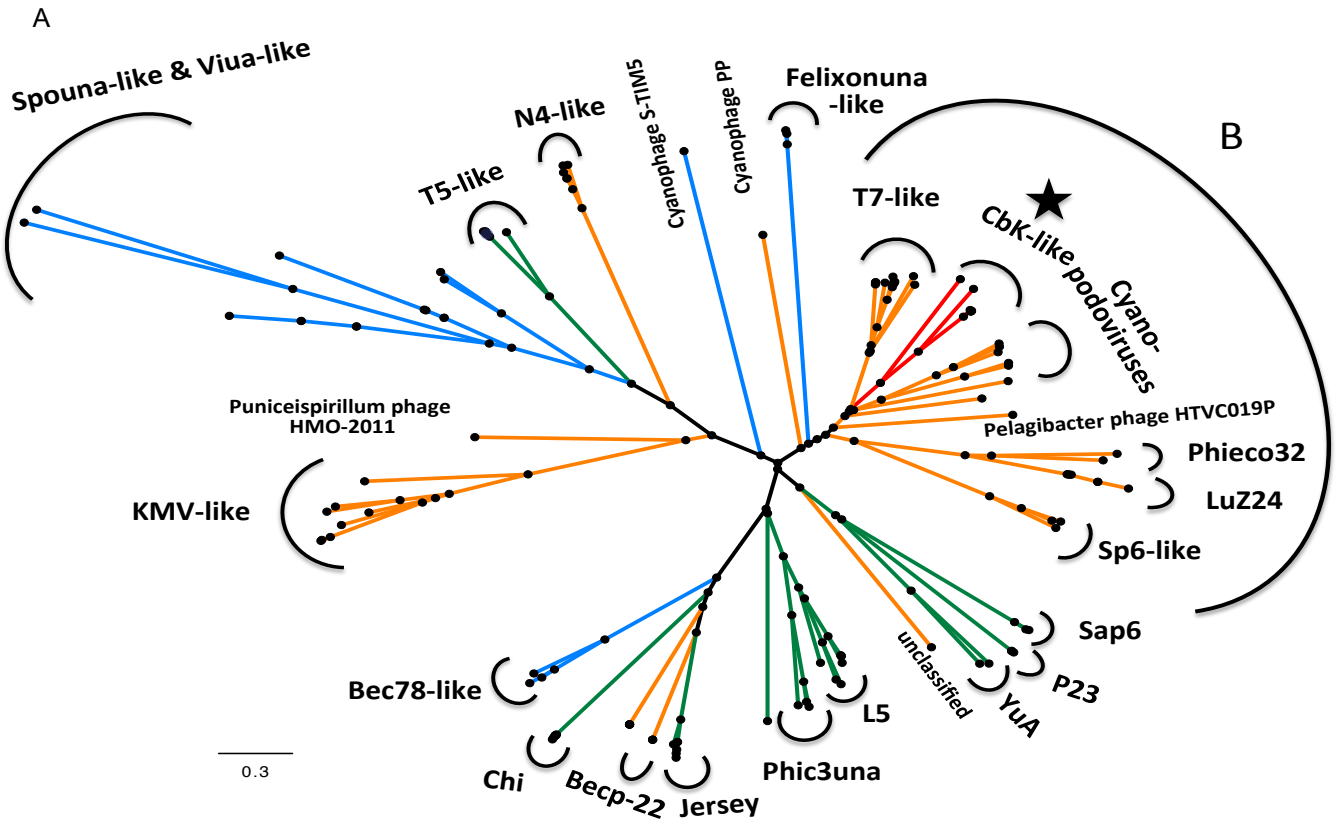


Figure 3.3 Genome wide comparison between roseophage DSS3Φ8 and *Caulobacter* phage phiCbK. White arrows represent ORFs for which no putative function can be attributed. Yellow, green, red and purple arrows represent structure, lysis, replication and packaging ORFs, respectively. Pink arrows stand for GTA structure, while orange arrows represent prophage-related ORFs. Grey color stands for hypothetical proteins. Related genes of these two phages are connected by blue shading. The color box corresponds to different amino acid identities. The figure was the created by Easyfig.


A
B
Spouna-like & Viua-like
N4-like
Cyanophage S-TIM5
Cyanophage PP
Felixonuna
-like
T5-like
T7-like
CbK-like
Cyano-
podoviruses
Puniceispirillum phage
HMO-2011
Pelagibacter phage HTVC019P
Phieco32
LuZ24
KMV-like
Sp6-like
Sap6
P23
YuA
unclassified
L5
Bec78-like
Phic3una
Chi
Becp-22
Jersey
0.3

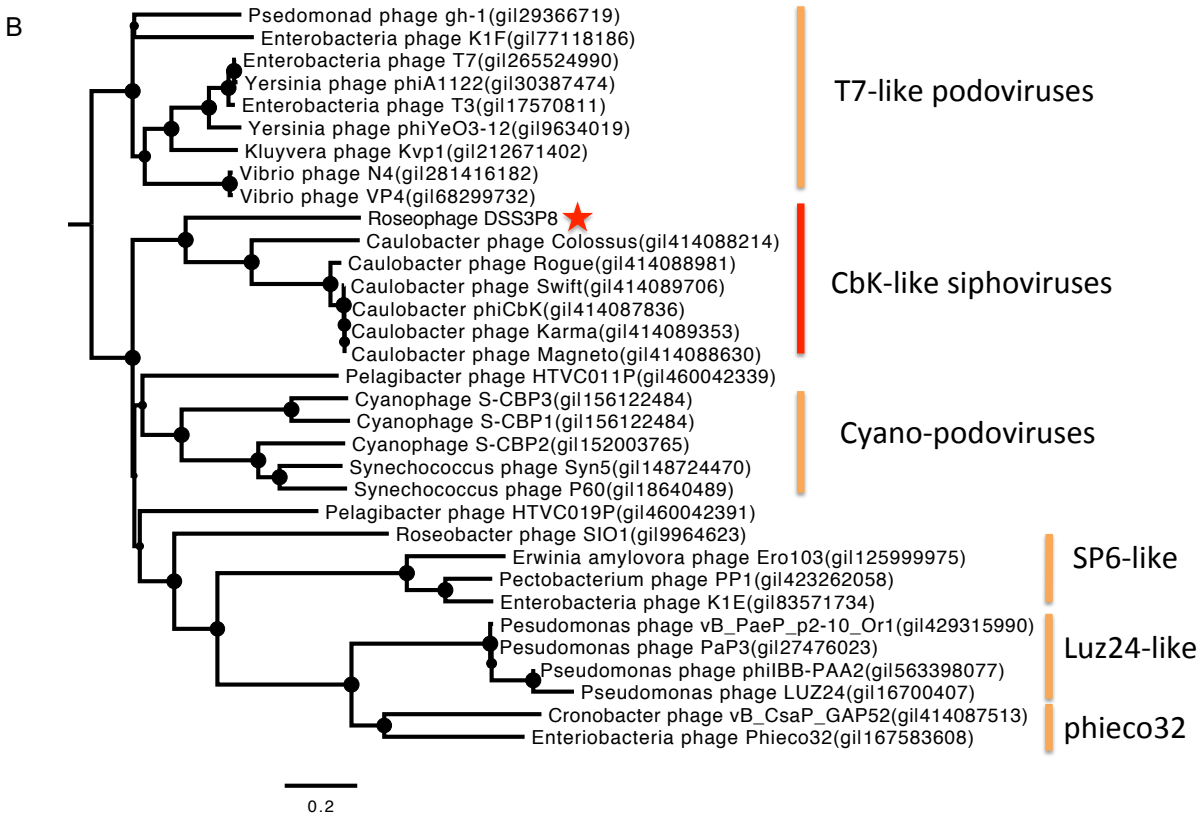


Figure 3.4 (A) The maximum-likelihood phylogenetic tree of DNA polymerase I of bacteriophages. Orange, blue and green colors represent *Podoviridae*, *Myoviridae* and *Siphoviridae*, respectively. Red color represents the CbK-like phages. Bootstrap = 500. (B) Zoom in of region B in Figure 3.4 A.

bacteria inhabiting two different environments (marine vs. freshwater). Together, the above studies tend to suggest that the genomes of siphoviruses with prolate capsids (B3 morphotype of *Siphoviridae*) are more conserved compared to the siphoviruses with no elongated heads. A recent survey of the global oceans showed that siphoviruses with prolate capsids can make up 48% of siphoviruses observed in the marine environment (Brum *et al.*, 2015). The relatively conserved genomes among the siphoviruses with prolate heads may allow me to identify a marker gene to explore the diversity of these type of siphoviruses in aquatic ecosystems.

3.4.4 Difference between DSS3Φ8 and CbK-like phages

Despite the conserved genome arrangement between DSS3Φ8 and the CbK-like phages, an obvious genome reduction of DSS3Φ8 was observed. DSS3Φ8 contains 229 ORFs, while 'our reference set ' CbK-like phages have 318 to 448 ORFs (Gill *et al.*, 2012). The majority of genes which are not shared between DSS3Φ8 and CbK-like phages, are those without any matches in the NCBI database. The structure genes of DSS3Φ8 are much smaller than those of phiCbK, especially for the tail related genes. For instance, the length of pre-tape measure gene of DSS3Φ8 is half that of phiCbK. The reduction of structure genes in DSS3Φ8 is consistent with the smaller head and shorter tail of DSS3Φ8 in comparison to the CbK-like phages. It appears that the loss of genes and the reduction of gene size both contribute to the reduction of the DSS3Φ8 genome.

For CbK-like phages, a lysis gene cassette coding for endolysin, holin and spanin was found between structure and replication gene cassettes. However, at the

same region, DSS3Φ8 only contains one lysis gene - peptidoglycan amidohydrolase (Figure 3.2 and Figure 3.3). On the other hand, DSS3Φ8 contains two hypothetical genes upstream to the endolysin gene, which are likely involved in the lysis of the host cell based on their location.

DSS3Φ8 contains a conserved replication system, which includes genes coding for DNA polymerase, helicase, ligase, recombinase and nucleotide metabolism. DSS3Φ8 and the CbK-like phages are siphoviruses, but they likely hijacked the DNA polymerase gene from T7-like podoviruses. Such a genetic recombination between siphoviruses and podoviruses was first reported in the CbK like phages (Gill *et al.*, 2012). Phylogenetic analysis based of the *polA* amino acid sequence showed that DSS3Φ8 and the CbK-like phages form their own cluster within the *Podoviridae* clade (Figure 3.4).

DSS3Φ8 and the CbK-like phages possess the helicase gene. But both of them lack the primase gene and the single strand binding protein gene. In the T7-like podovirus, DNA polymerase, helicase-primase, a single-stranded DNA binding protein, and exonuclease make up the conserved replication system (Huang *et al.*, 2015). Unlike the CbK-like phages, DSS3Φ8 contains a thioredoxin, which is a key protein to increase the processing rate of the DNA polymerase (Mark and Richardson, 1976; Huber *et al.*, 1987). Thioredoxin genes have been found in many podoviruses infecting marine bacteria (Pope *et al.*, 2007; Rohwer *et al.*, 2000; Chen and Lu, 2002)

3.4.5 Prophage

DSS3Φ8 encodes an integrase and transcriptional regulator (possible repressor) between the structure and replication modules, suggestive of a lysogenic potential of this phage. The closest match (E-value < 2.5E-51, 32% identity) to the DSS3Φ8 integrase gene is a putative integrase gene in roseobacter strain, *Ruegeria* sp. 6PALISEP08. Those two bacteria show 99% identity at the 16S rRNA genes. The genome of *R. pomeroyi* DSS-3 does not contain a prophage (Moran *et al.*, 2004). I searched for evidence of site-specific attachment sites (*att*P and *att*B) in the genomes of phage DSS3Φ8 and host DSS-3, but no identical sequences between phage DSS3Φ8 and host DSS-3 were found. The lack of *att*P-*att*B pairing between DSS3Φ8 and its host is unexpected as the *att*P-*att*B sites have been found in many bacteriophages, including siphoviruses infecting marine picocyanobacteria (Sullivan *et al.*, 2009; Huang *et al.*, 2012), and have been used to link phages with their potential hosts (Mizuno *et al.*, 2013b). One possible attachment site is a 63 bp sequence near the tRNA upstream of the integrase gene, which is 81% identical to the tRNA-Met4 gene in the host genome. Because of the conserved nature of tRNAs, they can become the integration sites for temperate phages (Campbell, 2003). Prophages are commonly found in roseobacter genomes, and more roseophages and their genomes are available now. It would be interesting to compare the genomic similarity between roseobacter prophages and roseophages to look for putative integration sites, further understanding the lysogenic nature of temperate phages infecting members of the *Roseobacter* clade.

3.4.6 The GTA genes

Several hallmark genes of DSS3Φ8 are homologous to genes 12, 13, 14 and 15 in RcGTA of the gene transfer agent found in *Rhodobacter capsulatus*. These four GTA related genes have been found in the CbK-like phages (Gill *et al.*, 2012), phage ΦJL001 infecting marine *Rhizobiales str.* JL001 (Lohr *et al.*, 2005), and roseophage RDJLΦ1 infecting *Roseobacter denitrificans* Och114 (Huang *et al.*, 2011). Both gp47 and gp49 in DSS3Φ8 were homologous to GTA gene 12, which encodes a glycoside hydrolase and contains the conserved superfamily DUF2460, a group of hypothetical proteins with apparent phage-derived regions of bacterial chromosomes. Interestingly, DSS3Φ8 carries two copies of RcGTA gene 12 in its genome, which share 34% amino acid identity and are separated by an unknown protein. This duplication of RcGTA gene 12 has not been seen in GTA. Also, the sizes of gp47 and gp49 products are larger than that of RcGTA gene 12, which is highly conserved between RcGTA and the phages containing the four GTA genes. gp50 is most closely related to the structure proteins detected by proteomic approaches, which could also be the product of RcGTA gene 13. gp51 contains a phage-related cell wall peptidase domain and was predicted to encode a hydrolase belonging to the NlpC/P60 superfamily (Lang and Beatty, 2001). gp 52 encoded a tail fiber protein has also been found in RcGTA as gene 15. Although all of the phages that carry RcGTA structures possess four genes, DSS3Φ8, the CbK-like phages and ΦJL001 have no sequence homology with RcGTA and RDJLΦ1 on gene 13 and 15. When DSS3Φ8, the CbK-like phages and ΦJL001 were aligned with RcGTA and other phages containing the four GTA genes, 200-430 bp and 380-490 bp insertions from products of gene 13 and gene 15, respectively, have been found.

GTA is a well-preserved genetic structure in *Alphaproteobacteria*, especially in the *Roseobacter* lineage. A phylogenetic analysis based on concatenated translated sequences of these four GTA genes showed that all roseobacters fall into their own clades (Figure 3.5), which are consistent with the phylogeny based on whole genome sequence analysis (Newton *et al.*, 2010). Interestingly, phage-derived four GTA genes belonged to three different clades, which were not closely related to those of their potential hosts. This result suggests a host-independent evolution of phage-encoded GTA genes. Despite the fact that DSS3Φ8 and the CbK-like phages infect different bacterial species, the GTA gene phylogeny clusters these phages together, suggesting that DSS3Φ8 and the CbK like phages acquired these four GTA genes from a common ancestor (Figure 3.5).

It is noteworthy that these four GTA genes are not found in several SIO-like phages and N4-like phages (podoviruses) that infect different marine roseobacter strains (Angly *et al.*, 2009; Zhan *et al.*, 2015). It is possible that these four GTA genes are common in siphoviruses infecting different strains, including phages ΦJL001, RDJLΦ1, and DSS3Φ8. These four GTA genes were also found in a few unpublished siphoviruses infecting roseobacters (e.g. KT253150, NC_026608 and KT266805). The finding of roseophages carrying the four GTA genes indicates that there is a long and tangled evolutionary relationship between GTAs, phages and even prophages (Lang *et al.*, 2012). Isolation and genome sequencing of more roseophages will provide new insight into the evolutionary relationship between roseophages, GTAs and the *Roseobacter* lineage.

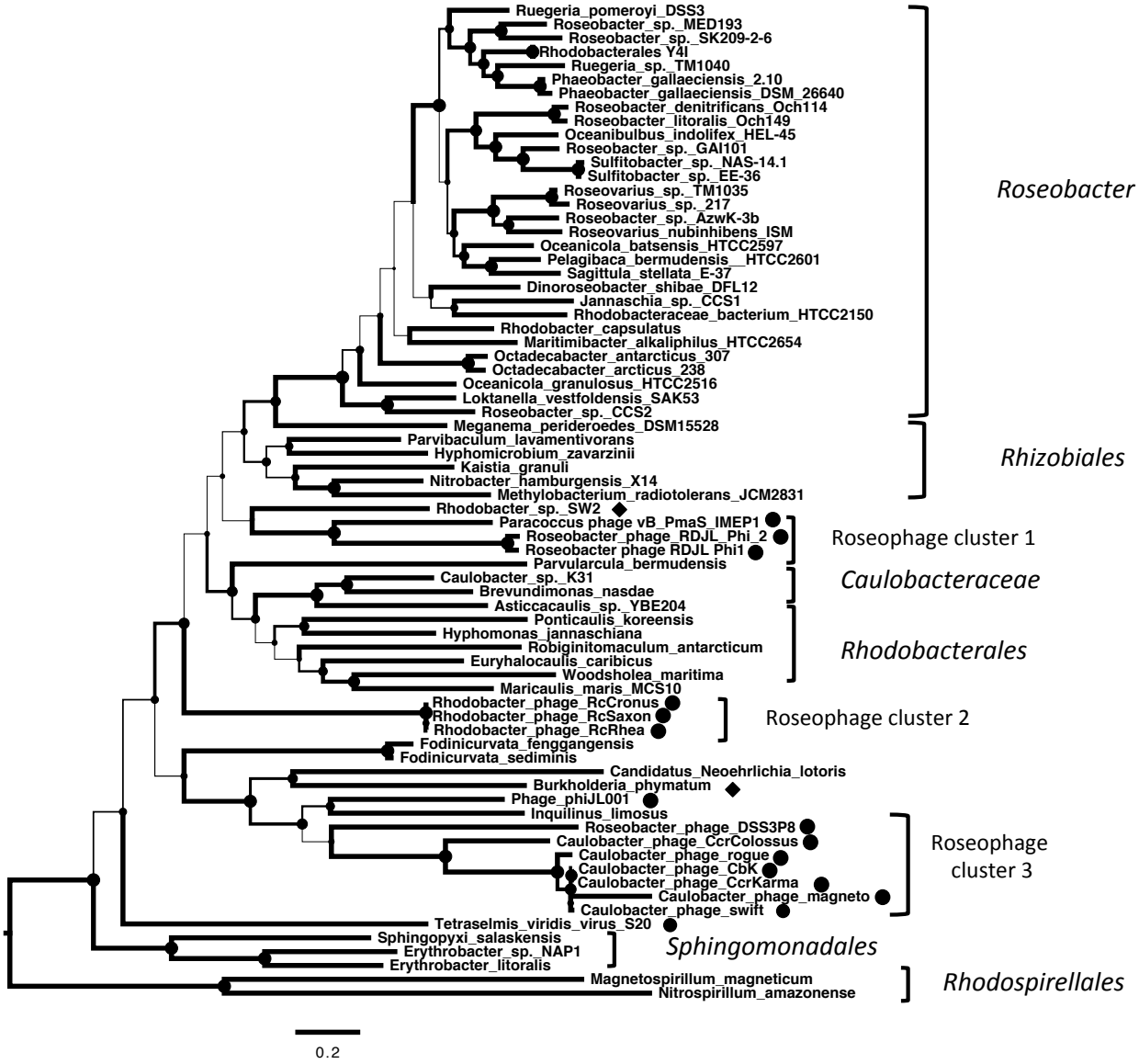


Figure 3.5 The maximum-likelihood phylogenetic tree of concatenated protein sequences of RcGTA-like genes 12-15 from bacteriophages and bacteria. The circle represents the phage-derived four GTA genes, while the diamond indicates the GTA gene cluster came from the bacteria. Bootstrap = 500. The width of the branch corresponds to the bootstrap value.

3.4.7 Environmental Distribution of DSS3Φ8

Metagenomic recruitments show that DSS3Φ8 reads were detected at the rates ranging from 10^{-10} to 10^{-7} per base pair in the viral metagenomic databases from iMicrobe by June 2015. The DSS3Φ8 homologs can be found in a wide range of aquatic environments, ranging from freshwater to open ocean (Figure 3.6). The highest recruitment rate came from the samples from Scripps Pier of Pacific Ocean Virome, where many SIO-like roseophages were isolated (Rohwer *et al.*, 2000; Angly *et al.*, 2009). DSS3Φ8 homologs were present in Antarctica, especially Organic Lake and Ace Lake, where abundant N4-like phages (Roseobacter N4-like phages) were observed (Chan *et al.*, 2014; Zhan *et al.*, 2015). DSS3Φ8 homologs were also recruited from other environments, such as chimney biofilm, stromatolite and whalebone. It appears that DSS3Φ8-like phages are widely spread in nature.

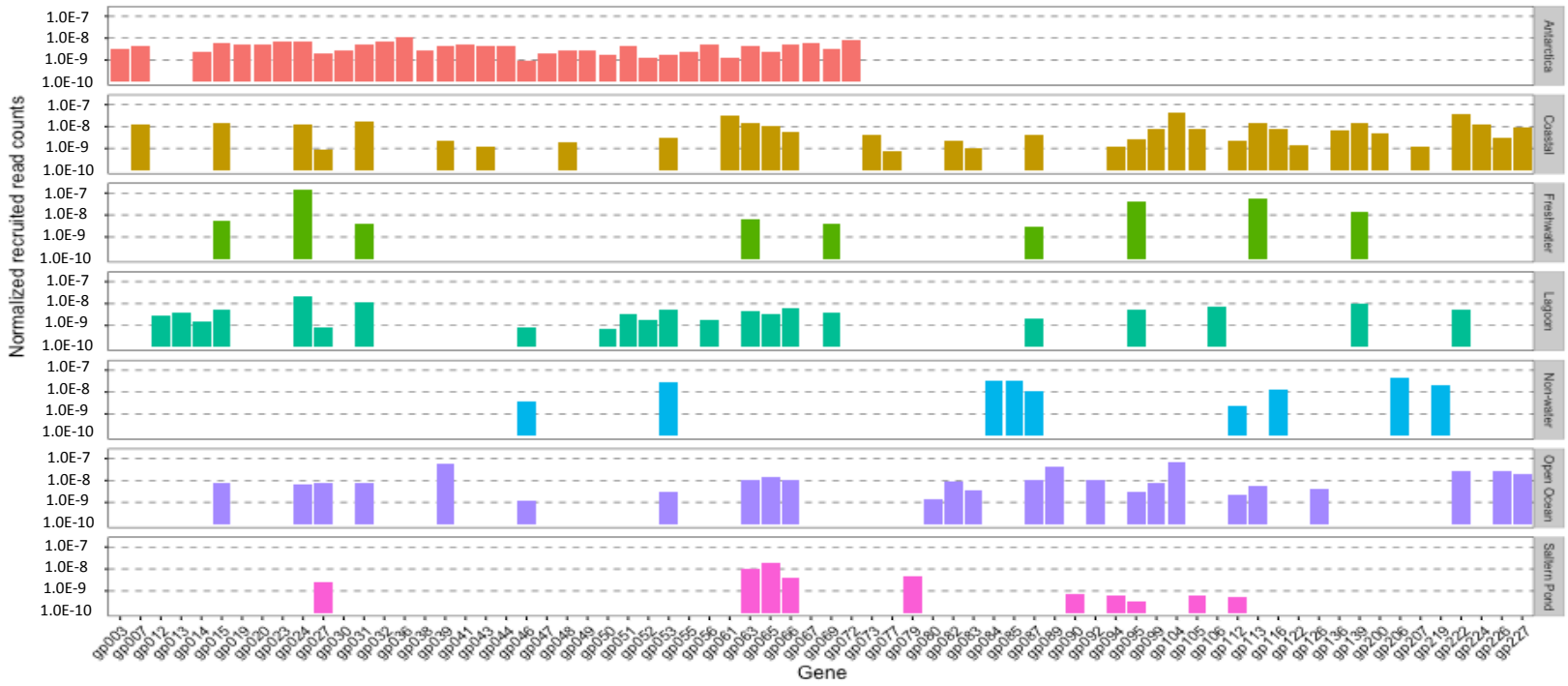


Figure 3.6 Rates of occurrences of DSS3Φ8 genes in various metagenomic databases. The absolute counts of retrieved reads were normalized against the data sizes of metagenomes and the average gene size.

3.5 Conclusion

Lytic phage DSS3Φ8 is a new member of marine roseophages, and represents the first CbK-like phage found in the marine environment. The genome sequence of DSS3Φ8 differs largely from all the known roseophages. DSS3Φ8 contains many features related to *Podoviridae, Siphoviridae*, prophage and GTA. Acquisition of T7-like DNA *pol*A gene suggests DSS3Φ8 is less dependent on host transcriptional machinery compared with other siphoviruses. On the other hand, the presence of integrase and repressor genes implies the lysogenic potential of DSS3Φ8. With its highly mosaic genome, DSS3Φ8 could serve as an ideal phage system to study the genomic evolution of phages and how the acquisition of different genetic elements (i.e. *pol*, *int*, or GTA genes) affect the phenotypic characteristics of phage.

Chapter 4. Distinct groups of bacteriophages infecting a single marine roseobacter strain *Ruegeria pomeroyi* DSS-3

4.1 Abstract

Roseobacter is an important lineage of marine bacteria, especially in coastal waters. Marine roseobacters have been well studied in terms of their ecological distribution, biogeochemical role, genetic diversity and genomic evolution. However, little is known about phages that infect marine roseobacters. Here, I report the isolation of 12 different bacteriophages that infect a single strain of marine roseobacters, *Ruegeria pomeroyi* DSS-3. A wide range of genomic and morphological features has been observed among these roseophages (bacteriophages infecting roseobacters). Comparative genome analysis delineated three diverse clusters, with few genes shared between the clusters. The four clusters divide these phage isolates into Chi-like, CbK-like, N4-like and ssDNA phages. Chi-like roseophages carry the integrase gene and are able to enter lysogenic cycle. CbK-like phages, with large genome sizes, possess highly mosaic genomes. N4-like phage appears to be a common phage group infecting *R. pomeroyi* DSS-3. The two novel ssDNA viruses isolated from DSS-3 represent the smallest and simplest marine phage ever isolated. Despite the great diversity among these roseophage clusters, a high degree of genome conservation within each cluster indicates that members within each cluster evolve slowly and maintain their genomic stability compared to other members across clusters. Analysis of codon usage, GC content and number of tRNA shows different adaptation strategies among these phages. The presence of distinct phage types (N4-like phages, novel ssDNA phages, CbK-like siphoviruses and temperate Chi-like phages) infecting a single host implies that new types of phages could exist for the phylogenetically diverse *Roseobacter* lineage.

4.2 Introduction

Viruses are part of the microbial life cycle. They interact with hosts frequently, regulate host population structure, mediate genetic exchanges, and effect the biogeochemical cycle (Suttle, 2005, 2007). The diversity of viruses can be seen at the phenotypic and genetic levels. In fact, different types of phages can infect the same host strain (Hatfull, 2010; Holmfeldt *et al.*, 2013). Genomic sequences of marine viruses have contributed greatly to our understanding of adaption, evolution and phage-host relationship.

Currently, phage genomes were only found in phages infecting eight of 29 bacteria phyla. Although in the marine environment thousands of bacteriophage genomes were assembled from metagenomes, only 130 complete genomes from marine phage isolates are available (by May 2015) (Perez Sepulveda *et al.*, 2016). These marine phages were isolated from 22 genera of bacteria (Perez Sepulveda *et al.*, 2016), which represent a small fraction of marine bacterial populations.

Metagenomic studies have revealed that marine viruses are extremely diverse (Hurwitz and Sullivan, 2013; Brum *et al.*, 2015; Paez-Espino *et al.*, 2016). According to recent global metagenomic surveys, a total of 1,075,761 viral protein clusters were observed (Brum *et al.*, 2015). However, 63-93% of surveyed sequences could not be annotated taxonomically or functionally (Hurwitz and Sullivan, 2013; Brum *et al.*, 2015). The large proportion of unknown sequences is fundamentally due to the limitation of our knowledge on reference genomes. Only 39 of these 5,476 populations contains cultured viral genomes as references (Brum *et al.*, 2015). The era of metagenomic databases calls for a significant increase in the isolation and

characterization of viruses in their natural environment. Large clusters of marine bacteria have been difficult to grow in laboratories for decades. Tremendous efforts have been made to cultivate some major groups of marine bacteria, including ubiquitous SAR11 and SAR116 clade (Giovannoni *et al.*, 1990; Morris *et al.*, 2002). Recently, the isolation of phages infecting SAR11 and SAR116 demonstrated that their phages are highly abundant in the ocean (Zhao *et al.*, 2013; Kang *et al.*, 2013). The isolation of SAR11 and SAR116 phages successfully explained up to 25% of previous unknown reads in metagenome.

The isolation of viruses is not only helpful to viral metagenomics, but also allows us to study the morphology, biology, and genomics of marine viruses. Phages infecting cyanobacteria (cyanophages) are probably the most well-studied marine phages. Many cyanophages have been isolated and characterized. Additionally their genomic sequences provide a comprehensive knowledge on their diversity, life-style, and evolution (Sullivan *et al.*, 2010; Sabehi *et al.*, 2012; Huang *et al.*, 2012; Labrie *et al.*, 2013; Huang *et al.*, 2015;). The discovery of cyanophage-encoded photosynthesis genes raised many questions regarding the role of phages on regulating host photosynthetic activity (Millard *et al.*, 2004; Lindell *et al.*, 2005; Thompson *et al.*, 2011).

Members of the *Roseobacter* lineage are abundant in the marine environment, representing about 20% of the bacterial cells in coastal waters and 2-8% in open ocean waters (Buchan *et al.*, 2005; Newton *et al.*, 2010). They have been isolated from various habitats including deep pelagic oceans, coastal sediments, and polar oceans (Brinkhoff *et al.*, 2008). Marine roseobacters can be abundant in the microbial

communities associated with algal blooms because roseobacters are able to degrade the dimethylsulfoniopropionate (DMSP) produced by algae (Moran *et al.*, 2004). The ability to metabolize sulfur compounds makes roseobacters a prominent player in global sulfur cycles. Currently, at least 69 genomes of marine roseobacters have been sequenced, and analyses of representative genome sequences have generated new insights into ecological adaptation and the evolution of the marine *Roseobacter* clade. Many roseobacter genomes contain intact prophages or phage-like gene transfer agents (Paul, 2008; Newton *et al.*, 2010) suggesting their close interaction with phages.

Currently, only a handful of phages that infect roseobacters have been isolated and characterized. In 2000, Rohwer et al. isolated the first roseophage, SIO1, which infects a marine *Roseobacter* SIO67 (Rohwer *et al.*, 2000). SIO1-like phages were isolated from the same region many years later (Angly *et al.*, 2009). The second report of roseophages was the discovery of N4 like phages, DSS3Φ2 and EE36Φ1, which infect *Ruegeria pomeroyi* DSS-3 and *Sulfitobacter sp.* EE-36, respectively (Zhao *et al.*, 2009b). More N4-like phages have been isolated from roseobacters since then (Chan *et al.*, 2014; Ankrah and Budinoff, 2014; Ji *et al.*, 2015a; Cai *et al.*, 2015; Li *et al.*, 2016), suggesting that this type of phage could be prevalent for roseobacters. Two siphoviruses, phage RDJLΦ1 and RDJLΦ2, which infect *Roseobacter denitrificans* OCh114 were isolated recently (Zhang and Jiao, 2009; Liang *et al.*, 2016). Two temperate podoviruses infecting *Sulfitobacter sp.* strain 2047 were also described (Ankrah *et al.*, 2014b). Most recently, a new type of phage, CbK-like phage, DSS3Φ8, was isolated infecting *R. pomeroyi* DSS-3 and found to be

highly mosaic (Zhan *et al.*, 2016). CbK-like phage has not been reported in the marine system prior to the isolation of DSS3Φ8. All of these studies indicate that the diversity of roseophages exceeds our expectations.

Here, I report the isolation of 12 roseophages that infect *R. pomeroyi* DSS-3. *R. pomeroyi* DSS-3 was chosen because it is a coastal strain with important biogeochemical features, such as its capability for degrading DMSP and utilizing carbon monoxide (Newton *et al.*, 2010). In addition, DSS-3 has been referred as the model strain to study marine heterotrophic bacteria (Moran *et al.*, 2004). The morphology, growth, host specificity, and genome sequences of these DSS-3 phages are described. The isolation of these roseophages provides a new insight into the diversity, evolution and phage-host interactions for marine roseobacters.

4.3 Methods and Materials

4.3.1 Phage isolation

Around 500 ml water were collected monthly from the surface water of the Baltimore Inner Harbor using a bucket from 2012 to 2013. Temperature and salinity were measured immediately after collection. Collected water was pre-filtered through 20 μm filter. Ten mL of water were added to 50 ml of *R. pomeroyi* DSS-3 in exponential growth phase (OD between 0.2 to 0.3). The host and water were then incubated at 28 °C for 24-48 hours. After an enrichment step, viral lysate was collected via centrifugation at 8000g for 20 minutes. The plaque assay was conducted following the described protocol. At least three plaques were picked from each of the plates. One of the three was further purified three times by plaque assay.

4.3.2 Phage morphology, growth curve and cross-infectivity

Viral morphology was examined using an FEI Tecnai T12 transmission electron microscope at University of Maryland, Baltimore. Samples were prepared following the procedures described elsewhere (Brenner and Horne, 1959). Briefly, one drop (2 µl) of purified phage lysate was adsorbed to the 200-mesh Formvar/carbon-coated copper grid and stained with 1% phosphotungstic acid (PTA) for one minute.

For the one-step growth curve, the number of phages was enumerated by qPCR, following a previous description (Zhan *et al.*, 2016). The qPCR primers were designed based on the sequence of the phage major capsid gene. The detailed sequences of qPCR primers were listed in Table 4.1. Standards were developed using DNA extracted from phage lysate with a known titer. Cross-infection of DSS-3 phages were tested against five other marine roseobacter stains: *Roseovarius nubinhibens* ISM, *Silicibacter* sp. TM1040, *Sulfitobacter* sp. EE-36, *Dinoroseobacter shibae* DFL-12 and *Roseobacter denitrificans* OCh 114, as previously described (Zhan *et al.*, 2016).

4.3.3 DNA extraction and sequencing

Bacteriophage genomes were sequenced using Illumina MiSeq and/or Pacific Biosciences (PacBio) shown in Table 4.2. For the phages sequenced by Illumina, 1 liter of lysis was concentrated and purified by ultracentrifugation following the protocol previously described (Zhan *et al.*, 2016). DNA was extracted by phenol/chloroform/isoamyl alcohol (25:24:1) and precipitated in isopropanol. Phage

Table 4.1 List of primers used for qPCR in one-step growth curve experiment

Primer Name	Forward Primer (5'-3')	Reverse Primer (5'-3')
DSS3Φ1_MCP	ACCTCTACGCGGTCATCTTC	ACTTGGCGATCTCAACAACC
DSS3Φ2_MCP	TGCTACGATTGGTCTGCAAG	ATCGAGGAGAACCCGATCTT
DSS3Φ8_MCP	CTGCAACTGACACCACCATC	ATACCAACGAGCCATTCGAC
DSS3Φ22_MCP	AAATGGGCATGCAGGAAATG	GATCGTACAGAAGAAGCTCAGG

DNA samples were sent to the BAS Lab at the Institute of Marine and Environmental Science, UMCES and sequenced by Illumina MiSeq platform. Sequences were assembled by the CLC Genomic Workbench, with at least 450x coverage. For bacteriophages sequenced by PacBio, individual lysate was collected from a 3-time purified phage plaque. The phage titer was determined by the epi- fluorescence microscopic count method. Bacteriophages were then pooled together based on their abundance. DNA of the "pooled phage community" was extracted and sequenced by PacBio. Phage contigs were assembled using the hierarchical genome-assembly process version 3 (HGAP). Since the barcode were not added into each pooled bacteriophage, PCR identification was needed to assign contigs to corresponding phages. The PCR primers were designed based on the sequence information in the assembled contigs.

4.3.4 Annotation, codon usage and genomic comparison

Open reading frames (ORFs) were predicted by using GeneMarkS (Besemer *et al.*, 2001), and were annotated using BLASTp against NCBI's non-redundant databases (E-value ≤ 1E-5 and identity ≥ 30%). The genome maps were generated in R using GenoplotR package (Guy *et al.*, 2010). tRNAs of representative phages (DSS3Φ1, DSS3Φ2, DSS3Φ8 and DSS3Φ22) were identified by using tRNAscan-SE (Lowe and Eddy, 1996). Codon usage of the host, *R. pomeroyi* DSS-3, and representatives of phages were analyzed using CodonW (http://codonw.sourceforge.net//culong.html). An "all-to-all" BLASTp was used to

calculate the percentage of genes shared between pairwise DSS-3 phages, with the cut off E-value ≤ 1E-5.

4.3.5 Core genome analysis, phylogeny and average nucleotide identity

The core genomes of N4-like roseophages and Chi-like phages were analyzed using previous described methods (Sullivan *et al.*, 2010; Huang *et al.*, 2015; Labrie *et al.*, 2013). A reciprocal best BLASTp was conducted to compare proteins within each group. Homologous relationships were assigned when BLASTp results met the cut off E-value ≤ 1E-5 and alignment region covered more than 50% of shorter sequence. Homologous pairs were then transiently grouped and clustered orthologous groups (COG) were further built. Details of phages used for this analysis are listed in Table 4.3 (Chi-like phages) and Table 4.4 (N4-like roseophages).

For phages included in core genome analysis, each core protein sequence was aligned using Clustal W (Thompson *et al.*, 1994), and further concatenated. A maximum-likelihood tree of the aligned concatenated proteins was constructed using MEGA7 (Kumar *et al.*, 2016). Bootstrap values were determined from 1,000 resampling events. The average nucleotide identity (ANI) was calculated from JSpecies Web Server (Richter *et al.*, 2016). The measurement of ANI was based on BLAST+ and MUMmer.

4.3.6 Lysogen formation assay

To test prophage formation for Chi-like DSS-3 phages, *R. pomeroyi* DSS-3 was infected by DSS3Φ1 with MOI = 10 and plated in a soft agar overlay. Resistant

strains that grew on the plate were picked and purified 3 times. A primer pair specific to DSS3Φ1 was used to test the presence of prophage sequence in the phage resistant strains. Wild-type DSS3 was used as a negative control, and DSS3Φ1 DNA served as the positive control.

4.3.7 Genomic sequence of the phage resistant strain.

One of the mutant colonies with a phage signal was selected and cultivated in liquid ½ YTSS medium. The resistant strain was named DSS-3-P1R. The genome of DSS-3-P1R was extracted using the UltraClean Microbial DNA Isolation Kit (MoBio). A PCR was conducted using DSS3-3-P1R DNA as a template to confirm the phage signal. The genome of DSS-3-P1R was sequenced using Illumina Miseq. CLC genomic workbench was used for quality control, read pairing and assembly. A total of 51 contigs (≥1 kb) were assembled. Among them, one 49 kb contig with a coverage of 780x was found. This contig shared 99% nucleotide sequence identity and 82% coverage with DSS3Φ1.

4.4 Results and Discussion

Twelve new bacteriophages were isolated using *R. pomeroyi* DSS-3 (referred as DSS-3 phages hereafter), from the Baltimore Inner Harbor between 2012 and 2013. Each phage was isolated from a different water sample (Table 4.2). The temperature of each water samples ranged from 7 to 31°C, and the salinity ranged from 5 to 14 psu (Table 4.2). Currently, 15 DSS-3 phages, including 3 previously published DSS-3 phages (Zhao *et al.*, 2009; Zhan *et al.*, 2015, Chapter 2), have been

Table 4.2 The isolation parameter and genomic feature for 15 completely sequenced DSS-3 phages.

Classification	Phage	Genome Size (kb)	GC content (%)	Number of ORFs	Number of tRNAs	Collection Date	Temperature (°C)	Salinity (psu)	Sequences	Reference
Podoviridae (N4-like)	DSS3Φ2	74.91	47.91	81	3	Jan. 2007	NA	NA	Sanger	Zhao *et al.*
Podoviridae (N4-like)	DSS3Φ12	74.68	47.89	85	3	NA	NA	NA	Illumina	this Chapter
Podoviridae (N4-like)	DSS3Φ13	74.83	50.77	79	4	Jul. 2012	29	10	PacBio	this Chapter
Podoviridae (N4-like)	DSS3Φ14	74.79	47.91	88	3	Sept. 2012	23	14	PacBio	this Chapter
Podoviridae (N4-like)	DSS3Φ17	74.68	47.93	83	3	Oct. 2012	20	8	Illumina/PacBio	this Chapter
Podoviridae (N4-like)	DSS3Φ21	74.67	47.91	82	3	Jun. 2012	24	5	Illumina/PacBio	this Chapter
Siphoviridae (Chi-like)	DSS3Φ1	59.60	64.13	82	0	NA	NA	NA	Illumina	this Chapter
Siphoviridae (Chi-like)	DSS3Φ7	59.57	64.12	84	0	NA	NA	NA	454	this Chapter
Siphoviridae (Chi-like)	DSS3Φ18	59.11	64.03	82	0	Nov. 2012	16	12	Illumina/PacBio	this Chapter
Siphoviridae (Chi-like)	DSS3Φ11	59.55	64.02	83	0	Feb. 2013	7	6	Illumina/PacBio	this Chapter
Siphoviridae (Chi-like)	DSS3Φ16	61.38	63.64	85	0	Aug. 2012	31	7	Illumina/PacBio	this Chapter
Siphoviridae (CbK-like)	DSS3Φ8	146.13	56.35	229	34	Feb. 2012	9	6	454	Zhan *et al.*
Siphoviridae (CbK-like)	DSS3Φ10	147.48	56.37	235	35	Mar. 2012	8	8	PacBio	this Chapter
Microviridae (unclassified)	DSS3Φ22	4.25	57.70	4	0	Jan. 2013	7	7	Sanger	Chapter 2
Microviridae (unclassified)	DSS3Φ15	4.25	57.70	4	0	May, 2012	16	12	Illumina	this Chapter

isolated from Baltimore Inner Harbor (Table 4.2). Cross infection experimentation indicates that these 15 DSS-3 phages all have a narrow host range, as they did not infect other strains tested.

4.4.1 Genomic features of DSS-3 phages

Complete genomes of the 12 new DSS-3 phages were sequenced and analyzed. Together with three previously described DSS-3 phages, these DSS-3 phages show a wide range of genome size, G+C contents and the number of tRNAs that they encode (Table 4.2). The genomes of fifteen DSS-3 phages range in size from 59 to 147 kb, encoding for 79 to 235 ORFs (Table 4.2).

The GC content of DSS-3 phages ranged from 47 to 64%. Except for the N4-like phages, the remaining DSS-3 phages have relatively high GC contents (56-64%), which are consistent with high GC content of host *R. pomeroyi* DSS-3 (64%) (Moran *et al.*, 2004), indicating that they could be more host dependent compared to the N4-like phages. A variety number of tRNAs were found among these fifteen DSS-3 phages (Table 4.2). Half of the isolates do not carry any tRNAs. The genomes of six DSS-3 phages contain 3 or 4 tRNAs. A high number of tRNAs have been found in genomes of DSS3Φ8 and DSS3Φ10, encoding 34 and 35 tRNAs, respectively. This high number of tRNAs is rare for phages. A maximum of 20 and 24 tRNAs have been found in phages infecting *E.coli* and marine *Cellulophaga*, respectively (Chithambaram *et al.*, 2014b; Holmfeldt *et al.*, 2013).

Comparative genomics delineated the thirteen DSS-3 phages into four different clusters (Figure 4.1 A). At least 78% of phage genes were shared within

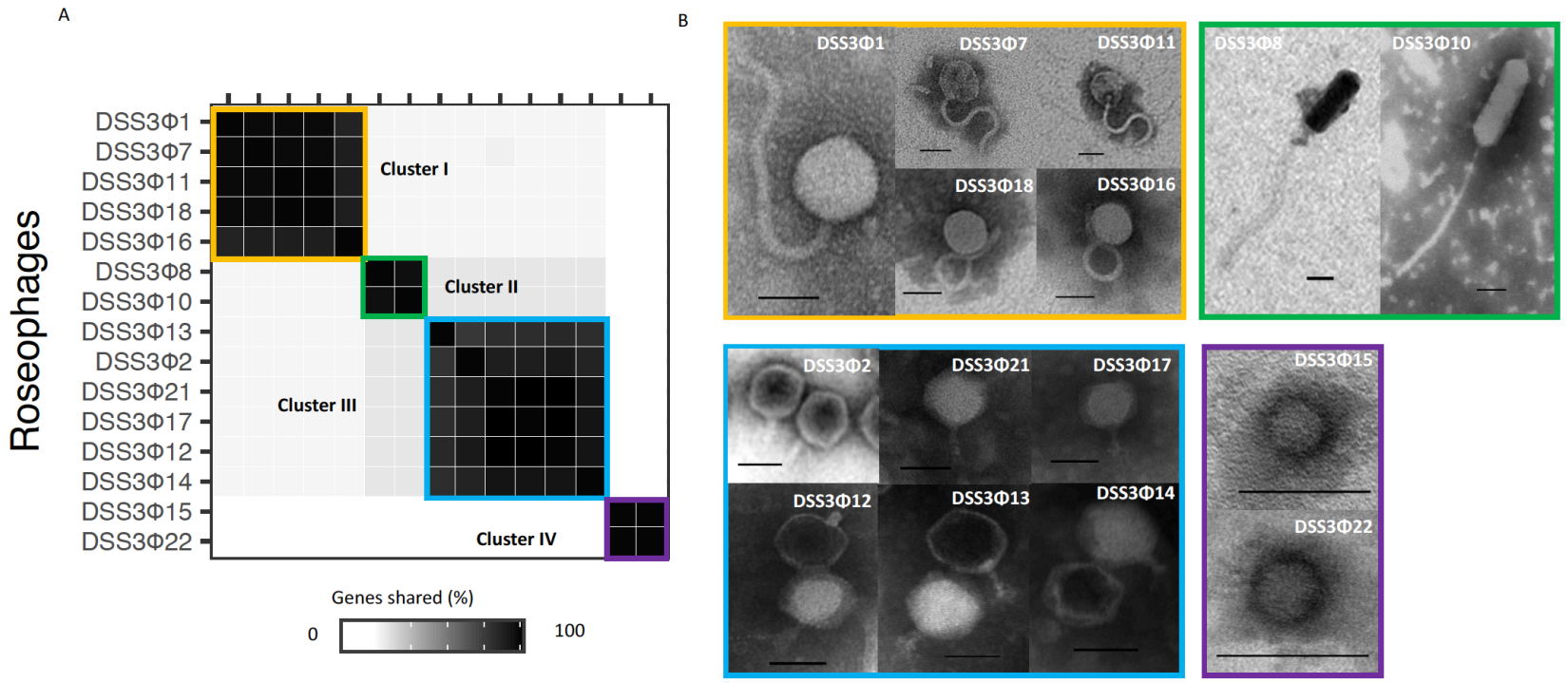
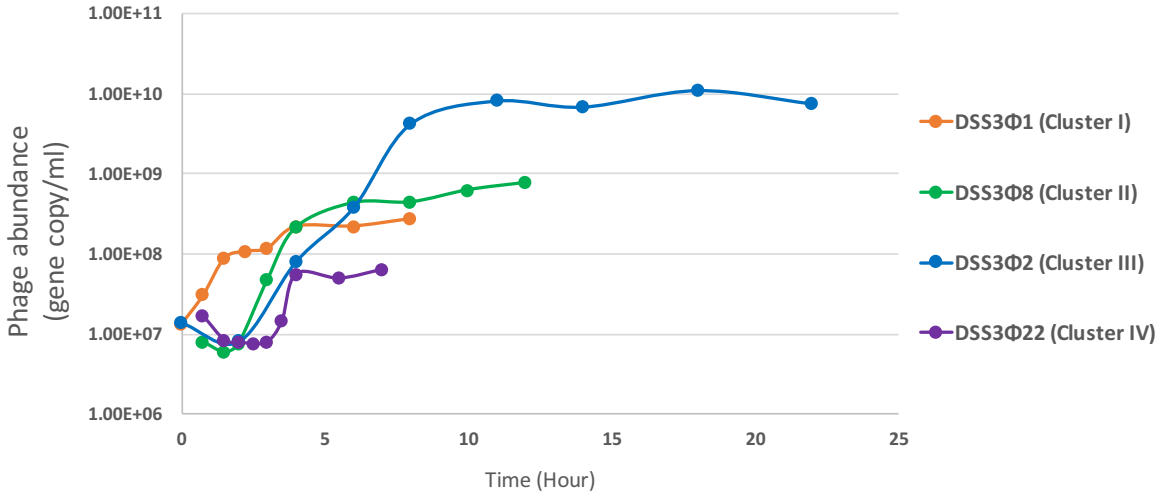


Figure 4.1 Genomic comparison and morphology of bacteriophages infecting *R. pomeroyi* DSS-3. (A) A heat map showing the percentage of shared genes between 13 DSS3 bacteriophages. (B) TEM images of 13 bacteriophages grouped based on gene comparisons. Scale bar = 50nm.

each cluster, while only a negligible number of genes (up to 8%) were shared between the clusters (Figure 4.1 A). Cluster I does not contain any previously characterized phages, and the phages in this cluster (DSS3Φ1, Φ7, Φ11, Φ16 and Φ18) have a long, flexible tail and an isometric head (Figure 4.1 B). Cluster I phages belong to *Siphoviridae* based on their morphological characteristics (Figure 4.1 B). Phage DSS3Φ10 displays an elongated head and a long tail, and clusters with a previously described CbK-like phage DSS3Φ8 (Zhan *et al.*, 2016) in Cluster II (Figure 4.1). Cluster III contains six DSS-3 phages, including DSS3Φ2, a N4-like phage (Zhao *et al.*, 2009b) (Figure 4.1 A). All members in Cluster III (DSS3Φ2, Φ12, Φ13, Φ14, Φ17, and Φ21) share a N4-like phage morphology, with a short non-contractile tail (Figure 4.1 B). Lastly, DSS3Φ15, with extreme small capsid and invisible tail, is closely related to ssDNA phage, DSS3Φ22 (see Chapter 2). DSS3Φ15 and DSS3Φ22 belong to Cluster IV (Figure 4.1)

4.4.2 Growth curves of DSS-3 phages

One-step growth curves show that phages in different clusters differ in growth patterns (Figure 4.2). DSS3Φ1, a Chi-like phage in Cluster I, has the shortest latent period (< 45 mins) and a burst size of 20. DSS3Φ8, a CbK-like phage in Cluster II, has a relatively longer latent period (ca. two hours), and a burst size of 130. DSS3Φ2, a N4-like phage in Cluster III has the largest burst size (~800) among the selected DSS-3 phages. The large burst size of N4-like roseophages has been reported (Zhao *et al.*, 2009b), with the largest burst size (~3,000) observed in coliphage N4 (Schito *et al.*, 1966). DSS3Φ22, a ssDNA phage in cluster IV has a burst size of 10, the



	DSS3Φ1 (Cluster I)	DSS3Φ8 (Cluster II)	DSS3Φ2 (Cluster III)	DSS3Φ22 (Cluster IV)
Latent period	< 45min	1.5 h	2 h	3h
Burst size	20	129	784	8

Figure 4.2 The one-step growth curves, latent period and burst size of selected phages from four clusters.

smallest among the selected DSS-3 phages. It is unclear why DSS3Φ22 has such a small burst size.

Together, different phenotypic and genomic characteristics seen in the above-mentioned DSS-3 phages imply that phages infecting a single host strain can be very diverse in nature.

4.4.3 Chi-like roseophages

DSS-3 phages in Cluster I all have an isometric head (~70 nm in diameter) and a long, flexible, non-contractile tail (~140 nm) (Figure 4.1 B). Genome sizes of Cluster I roseophages are similar, ranging from 59 to 61 Kb. The average G+C content of Cluster I phages is 64%, nearly identical to that of host (Table 4.2). Cluster I roseophages contain 82 to 85 predicted genes, only 22 of these predicted genes were assigned with putative functions. At least 86% genes were shared between members within Cluster I (Figure 4.1 B).

Manual annotation revealed that the Cluster I roseophages are closely related to the phages within Chivirus genus. Chivirus is a genus under *Siphoviridae* (ICTV 2015 release), and contains several phage isolates infecting *E.coli* (Kazaks *et al.*, 2012; Lee *et al.*, 2013;), *Salmonella* (Moreno Switt *et al.*, 2013)*, Burkholderia* (Lynch *et al.*, 2012) and *Xylella* (Ahern *et al.*, 2014) (Table 4.3). Members contained in this genus share a similar organization of four major functional modules and have a 15 kb highly divergent left arm of the genome (Ahern *et al.*, 2014). All the DSS-3 Chi-like phages shared 17 to 26 homologs with 12 other known Chi-like phages (Table 4.3). Phylogenetic analysis based on the concatenated 13 core genes of the 17 Chi-like

phages (5 DSS-3 phages and 12 other phages) showed that DSS3 Chi-like phages clustered with phage pCB2051-A, which infects a host *Loktanella* sp. (Figure 4.3).

Chi-like phages are generally thought to be lytic. Prior to this study, no integration genes have been found in Chiviruses, except for *Burkholderia* phages AH2 (Lynch *et al.*, 2012). Two Chi-like phage members, phage Sano and Salvo, were also used to create phage resistant isolates. Phage lysogens were not detected in these resistant isolates, further demonstrating the a lytic life cycle of these two Chi-like phages (Ahern *et al.*, 2014). Interestingly, all of the DSS-3 Chi-like phages contain an integrase gene, which is accompanied by a DNA binding protein and a helix-turn-helix domain containing protein. These genes constitute an integration-related module, located between the modules of head morphogenesis and DNA metabolism (Figure 4.4). The finding of this integration module infers that DSS-3 Chi-like phages may be able to convert to the lysogenic cycle.

To test this, a lysogenic formation assay was conducted. I was able to obtain many DSS3Φ1 resistant strains upon super-infection (MOI = 10). After careful purification, DSS3Φ1 specific signals were detected by PCR in around 30% of the mutant colonies. One of these PCR-positive colonies was picked up and confirmed to contain DSS3Φ1 prophage based on the genome sequencing. This study suggests that DSS-3 Chi-like phages can enter the lysogenic cycle by integrating their genomes into the host genome. The wild type of *R. pomeroyi* DSS-3 does not carry detectable prophages in its genome (Moran *et al.*, 2004), the impact of host DSS-3 being lysogenized by prophage is not known. It would be interesting to investigate the effect of phage-integration on the physiological behaviors of DSS-3.

Table 4.3 Bacteriophages related to roseophage Cluster I

Phage	Genome size (kb)	GC content (%)	Host	Classification	Accession	No. of homologs to DSS-3 phage cluster 1
Burkholderia phage AH2	58.0	61.3	*Burkholderia cenocepacia* C6433	*Siphoviridae*	NC_018283	26
Burkholderia phage BcepNazgul	57.4	60.6	*Burkholderia cepacia*	*Siphoviridae*	NC_005091	24
Xylella phage Salvo	55.6	63.0	*Xylella fastidiosa*	*Siphoviridae*	KF626668	22
Xylella phage Sano	56.1	62.4	*Xylella fastidiosa*	*Siphoviridae*	KF626665	22
Providencia phage Redjac	58.1	49.5	*Providencia stuartii* MRSN 2154	*Caudovirales*	NC_018832	17
Salmonella phage FSL SP-030	59.7	56.6	*Salmonella enterica*	*Siphoviridae*	NC_021779	26
Salmonella phage FSL SP-088	59.4	56.4	*Salmonella enterica*	*Siphoviridae*	NC_021780	26
Salmonella phage SPN19	59.2	56.5	*Salmonella* sp.	*Siphoviridae*	NC_019417	25
Enterobacteria phage Chi	59.4	56.5	*E coli*, *Salmonella*, *Serratia*	*Siphoviridae*	NC_021315	26
Enterobacter phage Enc34	60.4	51.1	*Enterobacter cancerogenus*	*Siphoviridae*	NC_019524	25
Loktanella phage pCB2051-A	56.9	55.0	*Loktanella* sp. CB2051	unclassified	NC_020853	24
Achromobacter phage phiAxp-1	62.0	60.1	*Achromobacter xylosoxidans* A22732	*Siphoviridae*	NC_029033.1	24

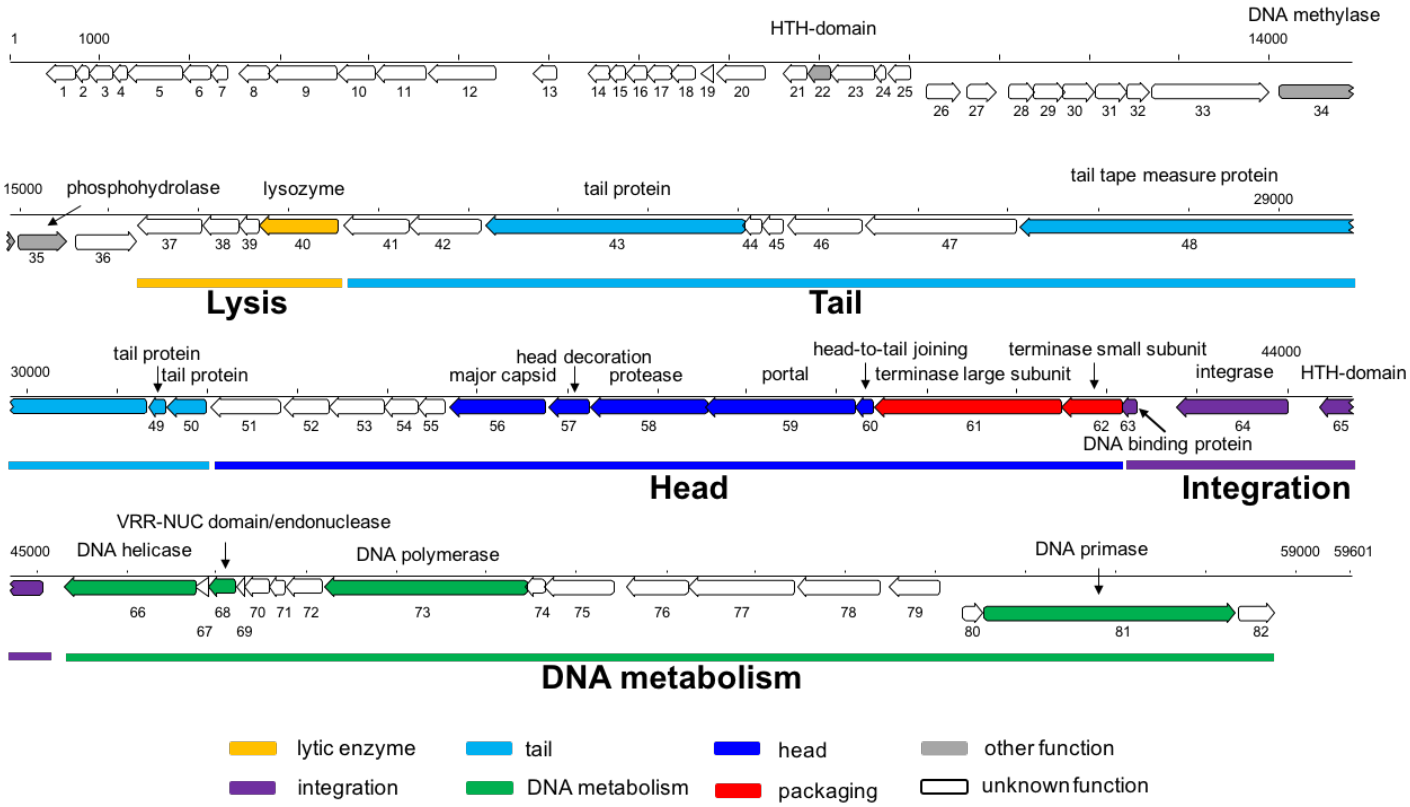


Figure 4.3 Genome map of roseophage DSS3Φ1. The arrows show the predicted genes and direction of transcription (left-pointing arrow: reverse, right-pointing arrow: forward). Colored arrows indicate genes with predicted function. Functional modules of lysis cassette, tail morphogenesis, head morphogenesis and DNA metabolism are indicated by yellow, blue, purple and green lines under the arrows, respectively.

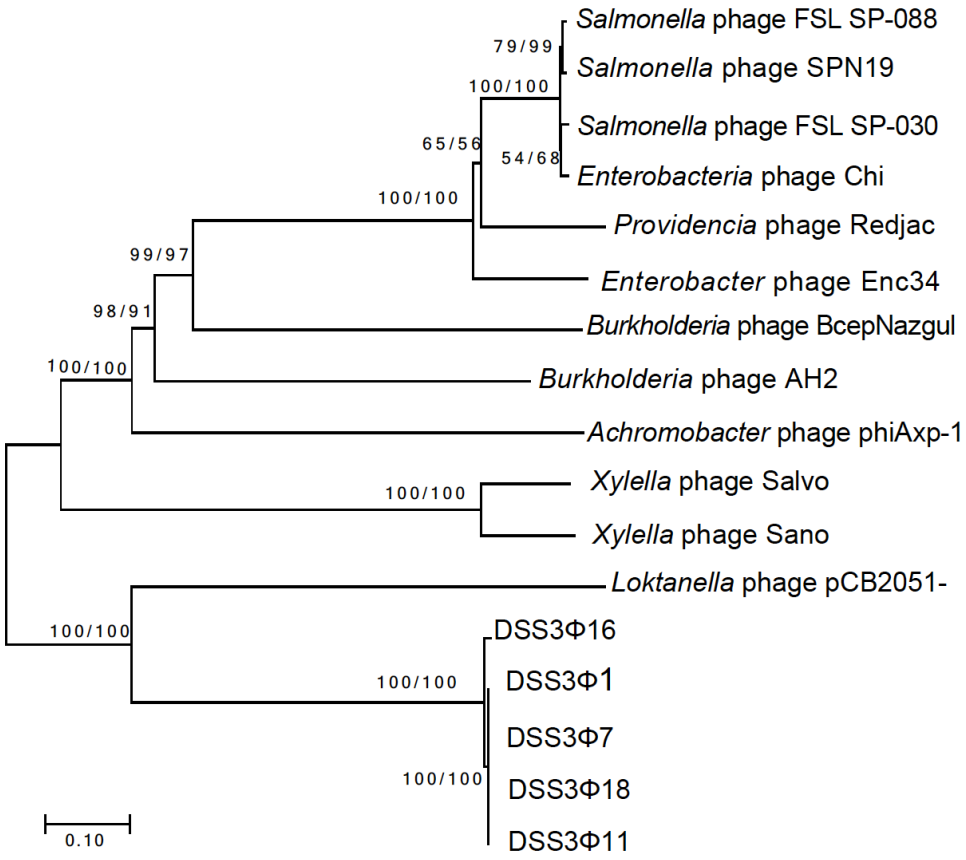


Figure 4.4 A maximum- likelihood and neighbor-joining phylogeny based on the 13 concatenated core genes of the 17 Chi-like phages. Bootstrap values (1000 testing replicates for both NJ and ML methods) showing on the node are in the order of ML/NJ.

4.4.4 N4-like roseophages

Five newly isolated phages (DSS3Φ12, Φ13, Φ14, Φ17and Φ21) within Cluster III have a capsid head size of 70 nm and a short tail length of 26nm (Figure 4.1 B). They are morphologically similar to a N4-like phage, DSS3Φ2, which was also isolated from DSS-3 (Zhao *et al.*, 2009b). All of the members within Cluster III shared at least 78% of the genes between each other (Figure 4.1 A). Prior to this study, 10 complete genomes of N4-like roseophages have been reported, and they infect seven different strains of roseobacters (Zhao *et al.*, 2009b; Chan *et al.*, 2014; Ankrah *et al.*, 2014a; Ji *et al.*, 2015; Cai *et al.*, 2015; Li *et al.*, 2016) (Table 4.4). It is noticeable that all these N4-like roseophages were isolated from the coastal environment, where marine roseobacters thrive (Newton *et al.*, 2010).

N4-like phages belong to a genus under *Podoviridae*. The genome sizes of N4-like roseophages ranged from 72.7 to 76.5 kb (Zhao *et al.*, 2009b; Ji *et al.*, 2015; Chan *et al.*, 2014; Cai *et al.*, 2015; Li *et al.*, 2016) (Table 4.4), which are larger than the genome size of podoviruses infecting *Roseobacter* SIO67 (Rohwer *et al.*, 2000) (39.9 Kb), and T7-like podoviruses infecting marine cyanobacteria (Huang *et al.*, 2015) (42.3-47.7 Kb). Genomes of N4-like roseophages are highly syntenic, sharing a very similar genomic arrangement (Figure 4.5).

A detailed genome-wide comparison was conducted and further identified 207 COGs across all 15 N4-like roseophages. Among the 207 COGs, a core-genome of 39 COGs shared by all 15 N4-like roseophages genomes was identified. The core genome mainly consists of genes involved in replication and structure (Figure 4.5, Table 4.5). This set of core genes contains two new core genes (gp61 and gp75 in DSS3Φ2) and one

Table 4.4 List of N4-like marine Roseophages

Phage	Host	Isolation location	Collection time	Genome size (kb)	GC content (%)	Number of ORFs	Number of tRNAs	Accession Number	References
DSS3Φ2	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	Jan-2007	74.6	47.9	81	3	FJ591093	Zhao *et al.*
ESS36Φ1	*Sulfitobacter sp.* EE-36	Baltimore Inner Harbor, USA	Jan-2007	73.3	47.0	79	3	FJ591094	Zhao *et al.*
phiCB2047-B	*Sulfitobacter sp.* 2047	Mesocosm Raunefjorden Norway	Jun-2008	74.5	43.0	77	15	HQ317387	Ankrah *et al.*
RLP1	*Roseovarius sp.* 217	Langstone Harbor, Hampshire, UK	Sep-2005	74.6	49.0	92	3	FR682616	Chan *et al.*
RPP1	*R. nubinhibens*	L4 sampling station, Plymouth, UK	Nov-1998	74.7	49.1	91	3	FR719956	Chan *et al.*
vB_DshP-R1	*D. shibae* DFL12	Baicheng Harbor, Xiamen, China	May-2012	75.0	49.3	86	2	KJ621082	Ji *et al.*
vB_DshP-R2C	*D. shibae* DFL12	Huangcuo station, Xiamen, China	Oct-2012	74.8	49.2	85	2	KJ803031	Cai *et al.*
DS-1410Ws-06	*D. shibae* DFL12	Sanya Bay, China	Oct-2014	76.5	50.0	77	0	KU885988	Li *et al.*
RD-1410W1-01	*R. denitrificans* OCh114	Sanya Bay, China	Oct-2014	72.7	49.5	77	0	KU885989	Li *et al.*
RD-1410Ws-07	*R. denitrificans* OCh114	Sanya Bay, China	Oct-2014	76.3	50.0	76	0	KU885990	Li *et al.*
DSS3Φ12	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	NA	74.7	47.9	85	3		this study
DSS3Φ13	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	Jul-2012	74.8	50.8	79	4		this study
DSS3Φ14	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	Sep-2012	74.8	47.9	88	3		this study
DSS3Φ17	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	Oct-2012	74.7	47.9	83	3		this study
DSS3Φ21	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	Jun-2012	74.7	47.9	82	3		this study

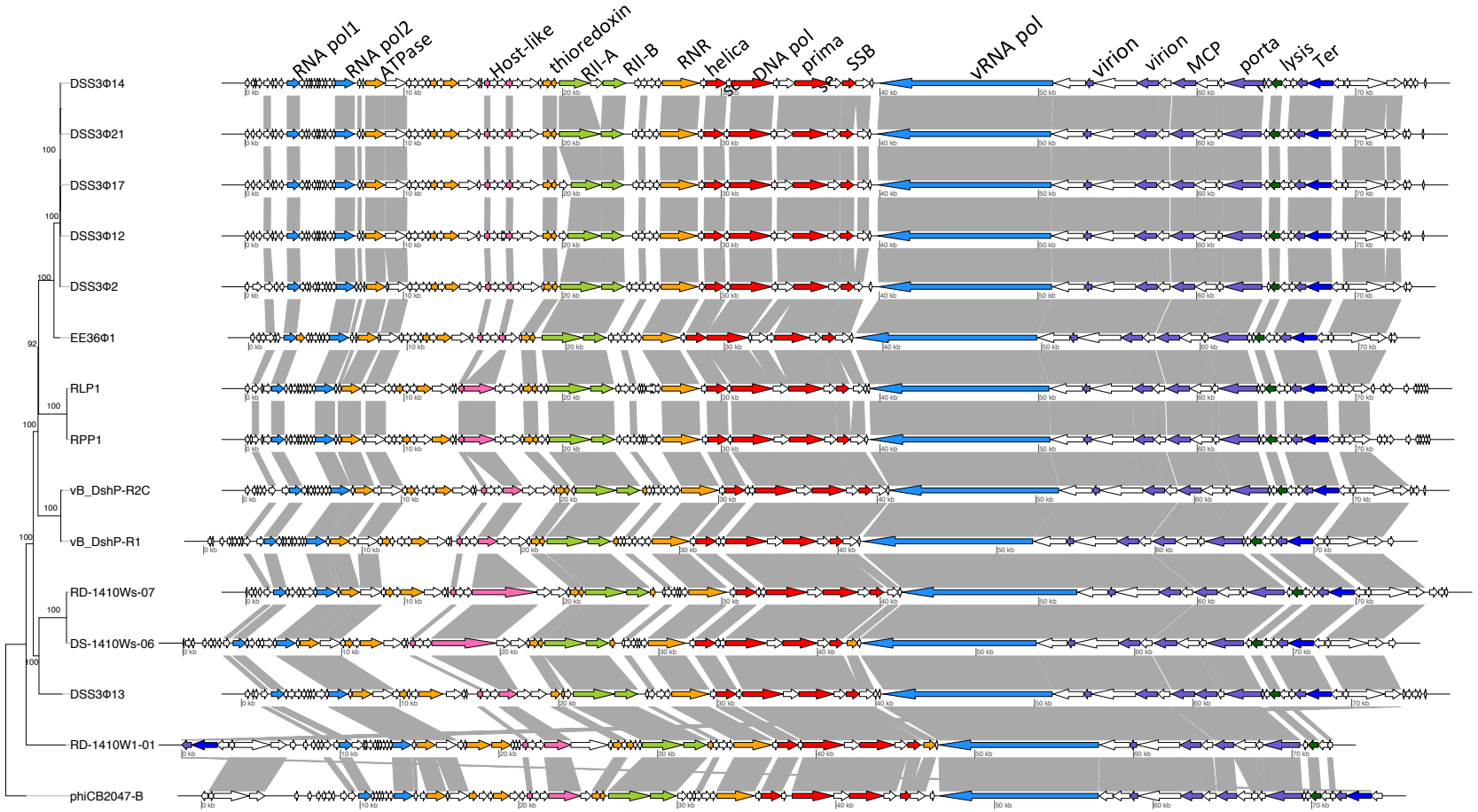


Figure 4.5 Alignment of the 15 N4-like roseophage genomes. Thirty-nine core genes are linked by gray shading. The arrows were colored based on the function of genes. The phylogeny based on these 39 core genes was plotted on the left side of the alignment.

Table 4.5 Core genes shared by N4-like Roseophages

	DSS3 Φ2	EE36 Φ1	phiCB 2047-B	RLP 1	RPP 1	vB_DshP-R1	vB_DshP-R2C	DS-1410Ws-06	DS-1410W1-01	DS-1410Ws-07	DSS3 Φ12	DSS3 Φ13	DSS3 Φ14	DSS3 Φ17	DSS3 Φ21
hypothetical protein	gp03	gp04	gp66	gp02	gp03	gp78	gp06	gp08	gp12	gp05	gp04	gp04	gp48	gp04	gp04
RNA polymerase subunit I	gp06	gp07	gp65	gp05	gp6	gp76	gp08	gp09	gp14	gp06	gp08	gp05	gp52	gp8	gp08
RNA polymerase subunit II	gp16	gp15	gp58	gp15	gp16	gp66	gp17	gp16	gp25	gp13	gp19	gp15	gp64	gp19	gp19
hypothetical protein	gp18	gp16	gp57	gp17	gp18	gp65	gp18	gp17	gp26	gp14	gp20	gp16	gp65	gp20	gp20
putative AAA superfamily ATPase	gp20	gp18	gp55	gp18	gp19	gp64	gp19	gp18	gp27	gp15	gp22	gp17	gp67	gp22	gp22
N4 gp25-like protein	gp22	gp20	gp53	gp20	gp21	gp62	gp21	gp19	gp28	gp16	gp23	gp18	gp68	gp23	gp23
host-like protein	gp32	gp30	gp46	gp30	gp31	gp52	gp31	gp27	gp36	gp24	gp33	gp27	gp79	gp33	gp33
virion protein	gp35	gp33	gp43	gp31	gp32	gp49	gp34	gp30	gp39	gp27	gp37	gp30	gp82	gp36	gp36
hypothetical protein	gp39	gp36	gp41	gp35	gp36	gp45	gp38	gp34	gp42	gp31	gp40	gp33	gp86	gp39	gp39
thioredoxin	gp40	gp37	gp40	gp36	gp37	gp44	gp39	gp35	gp43	gp32	gp41	gp34	gp87	gp40	gp40
rIIA-like protein	gp41	gp38	gp39	gp39	gp40	gp43	gp41	gp36	gp47	gp33	gp42/43/44	gp36	gp88	gp41/gp42	gp41
rIIB-like protein	gp42	gp39	gp38	gp40	gp41	gp42	gp42	gp37	gp48	gp34	gp45	gp37	gp02	gp43	gp42
N4 gp14-like protein	gp45	gp42	gp34	gp43	gp44	gp38	gp46	gp40	gp50	gp37	gp47	gp39	gp04	gp45	gp44
ribonucleoside-diphosphate reductase	gp48	gp45	gp31	gp50	gp51	gp34	gp50	gp45	gp53	gp42	gp50	gp42	gp07	gp48	gp47
DNA helicase	gp50	gp47	gp29	gp52	gp53	gp32	gp52	gp47	gp55	gp44	gp52	gp44	gp09	gp50	gp49
DNA polymerase	gp52	gp48	gp27	gp54	gp55	gp29	gp55	gp50	gp57	gp47	gp54	gp46	gp11	gp52	gp51
N4 gp42-like protein	gp54	gp50	gp26	gp55	gp56	gp28	gp56	gp51	gp58	gp48	gp56	gp47	gp13	gp54	gp53
DNA primase	gp55	gp52	gp25	gp56	gp57	gp27	gp57	gp52	gp59	gp49	gp57	gp48	gp14	gp55	gp54
N4 gp44-like protein	gp56	gp53	gp24	gp57	gp58	gp26	gp58	gp53	gp60	gp50	gp58	gp49	gp15	gp56	gp55

single-strand DNA binding protein	gp57	gp54	gp23	gp58	gp59	gp25	gp59	gp54	gp61	gp51	gp59	gp50	gp16	gp57	gp56
hypothetical protein	gp58	gp55	gp22	gp59	gp60	gp24	gp60	gp55	gp62	gp52	gp60	gp51	gp17	gp58	gp57
virion-encapsulated RNA polymerase	gp60	gp57	gp20	gp62	gp62	gp22	gp62	gp57	gp64	gp54	gp62	gp54	gp19	gp60	gp59
hypothetical protein	gp61	gp58	gp18/ gp19	gp63	gp63	gp21	gp63	gp58	gp65	gp55	gp63	gp55	gp20	gp61	gp60
16.5kDa virion protein	gp62	gp59	gp17	gp64	gp64	gp20	gp64	gp59	gp66	gp56	gp64	gp56	gp21	gp62	gp61
N4 gp53-like protein	gp63	gp60	gp16	gp65	gp65	gp19	gp65	gp60	gp67	gp57	gp65	gp57	gp22	gp63	gp62
N4 gp54-like protein	gp64	gp61	gp15	gp66	gp66	gp18	gp66	gp61	gp68	gp58	gp66	gp58	gp23	gp64	gp63
N4 gp55-like protein	gp65	gp62	gp14	gp67	gp67	gp17	gp67	gp62	gp69	gp59	gp67	gp59	gp24	gp65	gp64
major capsid protein	gp66	gp63	gp13	gp68	gp68	gp16	gp68	gp63	gp70	gp60	gp68	gp60	gp25	gp66	gp65
N4 gp57-like protein	gp67	gp64	gp11	gp69	gp69	gp15	gp69	gp64	gp71	gp61	gp69	gp61	gp26	gp67	gp66
hypothetical protein	gp68	gp65	gp10	gp70	gp70	gp14	gp70	gp65	gp72	gp62	gp70	gp62	gp27	gp68	gp67
N4 gp59-like protein	gp69	gp66	gp09	gp71	gp71	gp13	gp71	gp66	gp73	gp63	gp71	gp63	gp28	gp69	gp68
lysis	gp71	gp69	gp06	gp74	gp74	gp10	gp74	gp68	gp75	gp65	gp73	gp66	gp31	gp71	gp70
hypothetical protein	gp73	gp71	gp04	gp76	gp76	gp08	gp76	gp70	gp77	gp67	gp75	gp68	gp33	gp73	gp72
virion protein	gp74	gp72	gp03	gp77	gp77	gp07	gp77	gp71	gp01	gp68	gp76	gp69	gp34	gp74	gp73
terminase subunit A	gp75	gp73	gp02	gp78	gp78	gp06	gp78	gp72	gp02	gp69	gp77	gp70	gp35	gp75	gp74
N4 gp69-like protein	gp76	gp74	gp01/ gp77	gp79	gp79	gp05	gp79	gp73	gp03	gp70	gp78	gp71	gp36	gp76	gp75
host-like protein	gp77	gp75	gp76	gp80	gp80	gp04	gp80	gp74	gp04	gp71	gp79	gp72	gp37	gp77	gp76
hypothetical protein	gp78	gp76	gp75	gp81	gp81	gp03	gp81	gp75	gp05	gp72	gp80	gp73	gp38	gp78	gp77
hypothetical protein	gp79	gp77	gp74	gp82	gp82	gp02	gp82	gp76	gp06	gp73	gp82	gp74	gp39	gp80	gp79

previously defined core gene (gp 43 in DSS3Φ2) which was not found in the five newly isolated DSS3 N4-like phages (Li *et al.*, 2016). These 39 core genes make up nearly 70% of the genome sizes of N4-like roseophages, while the accessory and unique genes only comprise a small portion of genome (average 19% and 6% respectively) (Figure 4.6 A). In contrast, core genes of other relative conserved marine phages, T7-like cyanopodophage and T4-like cyanomyophage, only account for 57% (Huang *et al.*, 2015) and 26% (Sullivan *et al.*, 2010), on average, of each genome size respectively. The high percentage of core genes infers the high level of genomic conservation of N4-like roseophages.

A cumulative curve of core genes showed that the number of cores genes settled after 11 genomes were sampled (Figure 4.6 B), indicating that the core gene set of N4-like roseophages was well defined by current isolates. Surprisingly, the total number of possible genes, a pan-genome, also approaches a plateau (Figure 4.6 C), which has not been observed in previously analyzed marine cyanopodoviruses (Huang *et al.*, 2015; Labrie *et al.*, 2013) and cyanomyoviruses (Sullivan *et al.*, 2010). The nearly saturated pan-genome reflects a constraint evolution force for N4-like roseophages.

A phylogenetic analysis based on core gene alignments was conducted. Generally, almost all of the N4-like phages infecting DSS3 were clustered together, forming a cohesive group (Figure 4.5). It is noteworthy that DSS3Φ13 is distantly related to other N4-like phages that infect the same host DSS-3. This result shows that N4-like phages infecting a same host are not necessarily monophyletic. Similarly, phage RD-1410Ws-07 and RD-1410Ws-01, infecting *R. denitrificans* OCh114, do not cluster together (Li *et al.*, 2016).

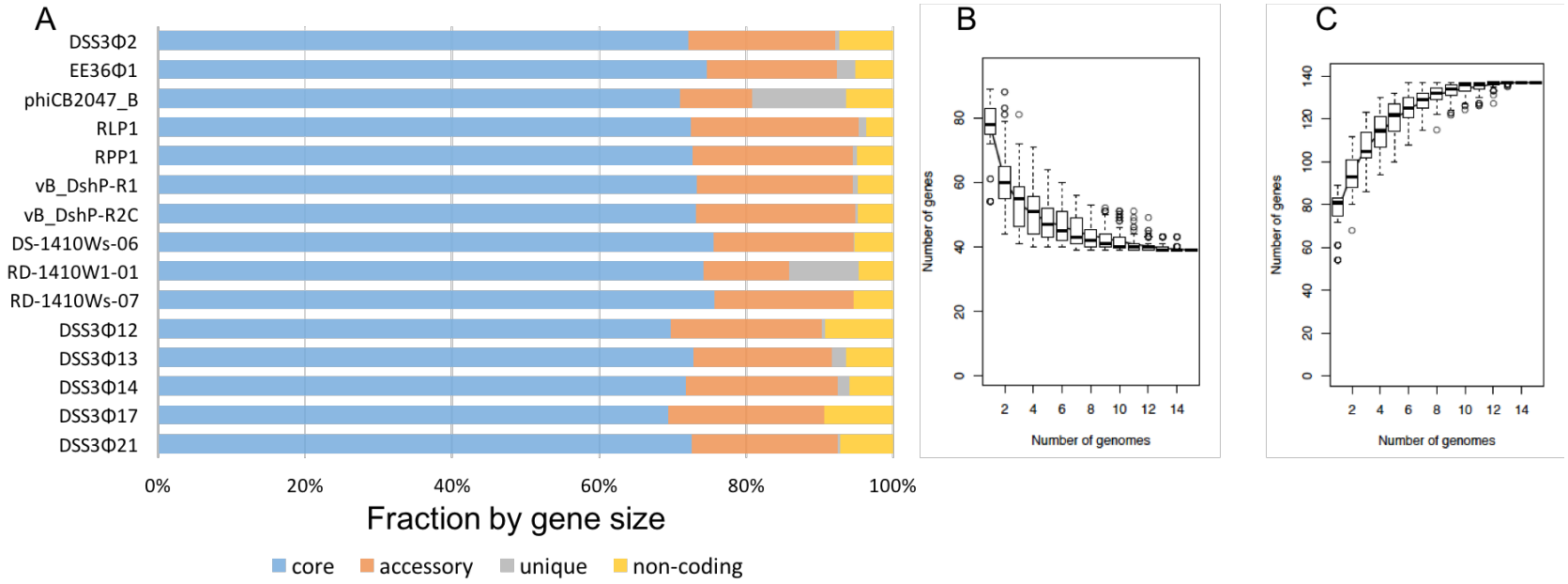


Figure 4.6 Core-genome and pan-genome analysis of N4-like roseophages. (A) Fraction of core, accessory and unique genes of each N4-like roseophage genome by gene size. A core (B) and pan-genomes (C) of N4-like roseophages. The core genome is the set of genes shared by all genomes, while the pan-genome is the total number of genes of genomes in the same subset. All possible combinations of genomes were analyzed. The average is represented by the line; the box indicates the lower and upper quantiles; and the dash line represents the confidence interval.

4.4.5 CbK-like phages

DSS3Φ10 is another CbK-like phage in Cluster II (Table 4.2). The genomic sequence of DSS3Φ10 is 97% identical to the genome of a previously described CbK-like phage, DSS3Φ8 (Zhan *et al.*, 2016). Both DSS3Φ8 and DSS3Φ10 were isolated from a winter sample. Similar to other CbK-like phages, both phages are highly mosaic, sharing genetic features with siphoviruses, podoviruses, GTA and prophages (Zhan *et al.*, 2016). Despite high genomic similarity, eight genes were not shared between DSS3Φ8 and DSS3Φ10. Seven of these eight genes code for the proteins without known function, while a gene coding for Chaperone DnaJ was only found in DSS3Φ8.

4.4.6 ssDNA phages

DSS3Φ15 is a small ssDNA phage isolated from a water sample collected in May, 2012 (Table 4.2). The genome sequence of this phage is identical to another ssDNA phage, DSS3Φ22, isolated from a water sample collected in January, 2013 (Table 4.6). The two identical ssDNA phages isolated from different seasons suggest that ssDNA phages infecting the same host are genetically stable (Rokyta *et al.*, 2006). Genome sizes of DSS3Φ15 and DSS3Φ22 are approximately 4.2 kb, encoding only four open reading frames. They represent the smallest and simplest ssDNA phage ever seen. Phylogeny based on the phage capsid genes shows that these two phages are distantly related to known ssDNA phages and belong to an unclassified ssDNA phage in the family of *Microviridae* (see Chapter 2).

4.4.7 tRNA, codon usage and phage-host interactions

The protein translation and reproduction of phages rely on host cellular machinery, hence it has been well accepted that phages need to fully utilize the host tRNAs pool and adjust their codon usage to accommodate bacterial codons (Krakauer and Jansen, 2002; Bahir *et al.*, 2009). Furthermore, it suggests that the tRNAs carried by phages correspond to the codons that are highly used by phage genomes (Bailly-Bechet *et al.*, 2007), compared to usage frequencies in host. Since DSS-3 phages have a wide range of tRNAs found in their genomes (Table 4.2), one phage from each dsDNA phage cluster (DSS3Φ1, DSSΦ2, DSS3Φ8) was selected to plot their codon usage against the host's codon usage (Figure 4.7) to study the codon adaptation strategies among different clusters.

For DSS3Φ1, a strong symmetrical distribution (slope = 0.9, R^2 = 0.8) was observed between phage and host codon frequencies, indicating that DSS3Φ1 and host genomes tend to use similar types of codons (Figure 4.7 A). The highly correlated codon usage is also consistent with the almost identical GC content between DSS3Φ1 and host (Table 4.2). DSS3Φ1 does not encode any tRNA genes. The lack of tRNA genes, similar codon usage, and GC content between DSS3Φ1 and its host suggests that DSS3Φ1 is highly dependent on the host system in terms of its translation and reproduction. As mentioned above, phage DSS3Φ1 can integrate into its host genome and enable a lysogenic relationship with the host. It has been proposed that temperate phages tend to have similar codon usage with host *E. coli* compared to virulent phages (Chithambaram *et al.*, 2014a; Bailly-Bechet *et al.*, 2007).

Therefore, it is possible that the Chi-like phages in Cluster I can increase their translation fitness by matching their codon usage and GC content with their hosts.

Thirty-four tRNAs has been found in DSS3Φ8. It has been proposed that the high number of tRNAs enable phages to broaden their host range (Bahir *et al.*, 2009; Bailly-Bechet *et al.*, 2007). For example, a high number of tRNAs have also been observed in cyanomyoviruses, which can cross infect hosts in different genera (Sullivan *et al.*, 2010). Meanwhile, when compared with another siphovirus DSS3Φ1, a phage without tRNAs, a slight difference in codon usage frequencies between DSS3Φ8 and host has been observed (Figure 4.7 B). This difference could be due to the insufficient co-evolution time, implying that the ability to infect *R. pomeroyi* DSS-3 could be a recent acquisition and *R. pomeroyi* DSS-3 may not be DSS3Φ8 original host. With a high number of tRNAs and non-symmetrical codon usage, it is possible that CbK-like roseophages are able to cross infect other hosts (beyond the strains I tested).

Phage DSS3Φ2, a N4-like phage in Cluster III, has three tRNA genes. The plot for DSS3Φ2 is highly scattered (Figure 4.7 C). The slope of codon usage frequencies between DSS3Φ2 and host is only 0.37, suggesting different codon usage frequencies between DSS3Φ2 and host. One interesting feature for N4-like roseophages is their low GC content (47-50%, Table 4.2), which is much lower compared to their host's GC content (64%). In general, phages share similar codon usage frequencies with host. However, it seems that the codon usage bias of DSS3Φ2 observed here breaks the paradigm. N4-like phages do not share a similar codon usage with host DSS-3. The N4 phage infecting *E. coli* encapsulates its own

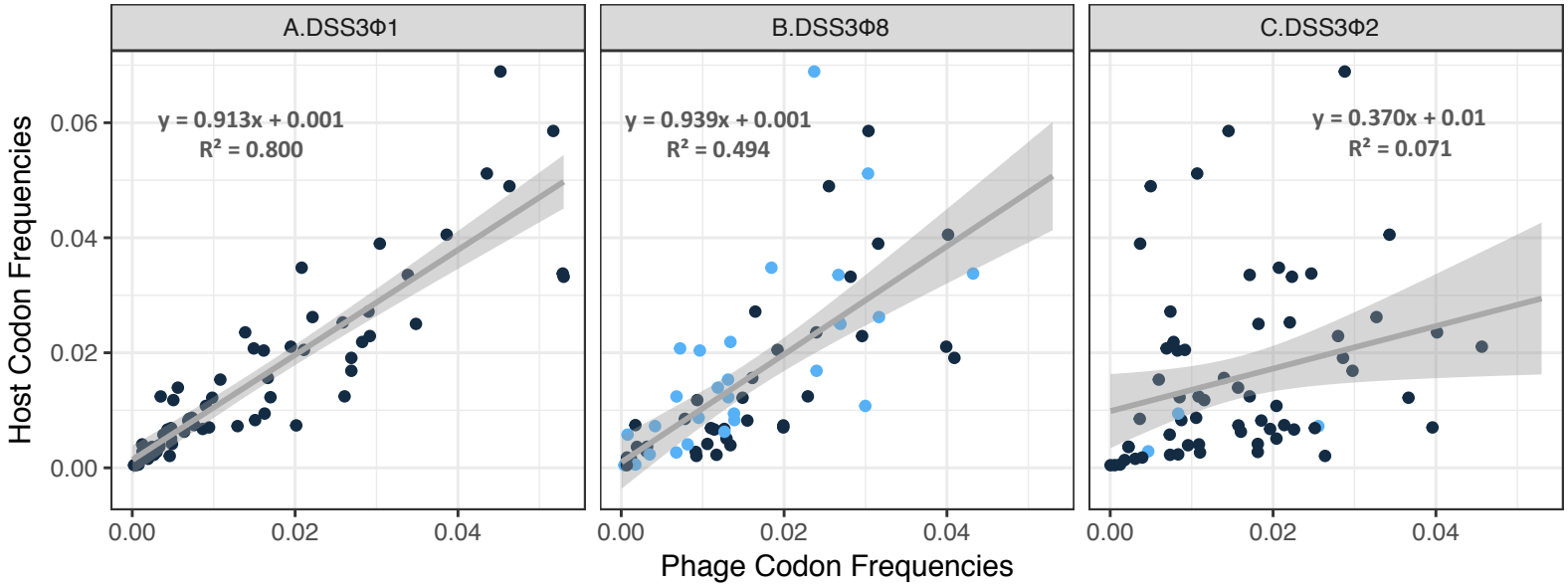


Figure 4.5 Codon usage frequencies of three dsDNA phages in different clusters compared with the codon frequencies of their host. The blue dot represents codon corresponding to tRNAs carried by phage.

viral-RNA polymerase in its capsid (Zhao *et al.*, 2009b), which initiates the early viral transcription (Lenneman and Rothman-Denes, 2015). It is possible that other N4-like roseophages also encapsulate their RNA polymerase, which make them less host-dependent and capable of by-passing the host codon usage restriction.

4.4.8 DSS-3 phages within each cluster are genetically stable

According to the comparative genomics, at least 78% of the genes were shared within each cluster. At the nucleotide level, the phage genomes within each cluster, except for DSS3Φ13 within N4-like cluster, were found to have 97.2-100% average nucleotide identity (Table 4.6). This high level of genomic conservation was also observed in other phages infecting different bacteria. For example, SIO1-like phages infecting *Roseobacter* SIO67 share 96.2%-98.4% nucleotide identity with each other after over a long ecological period (Angly *et al.*, 2009). Lytic 'phiKMV-like viruses' infecting *P. aeruginosa* shared a high nucleotide similarity (83%-97%) between each other (Ceyssens *et al.*, 2011). Similarly, mycobacteriophages also show a genetic conservation among the phages within each group (Hatfull, 2010). In our DSS-3 phage collection, N4-like podoviruses, CbK-like phages and ssDNA phages share high nucleotide sequence homology within each of their own groups, despite they were isolated from different times of year. This result supports the idea that marine bacteriophages can be genetically stable along an ecological time-period.

Table 4.6 Genomic similarities among the DSS-3 phages within each cluster.

	DSS3 Φ1	DSS3 Φ7	DSS3 Φ11	DSS3 Φ16	DSS3 Φ18	DSS3 Φ8	DSS3 Φ10	DSS3 Φ2	DSS3 Φ12	DSS3 Φ13	DSS3 Φ17	DSS3 Φ21	DSS3 Φ14	DSS3 Φ22	DSS3 Φ15
DSS3Φ1	100														
DSS3Φ7	99.81	100													
DSS3Φ11	99.81	99.94	100												
DSS3Φ16	99.81	97.16	97.04	100											
DSS3Φ18	99.99	100	100	100	100										
DSS3Φ8						100									
DSS3Φ10						97.73	100								
DSS3Φ2								100							
DSS3Φ12								99.71	100						
DSS3Φ13								**69.22**	**69.28**	100					
DSS3Φ17								99.71	99.97	**69.95**	100				
DSS3Φ21								99.37	99.52	**69.95**	99.63	100			
DSS3Φ14								99.49	99.64	**69.74**	99.65	100	100		
DSS3Φ22														100	
DSS3Φ15														100	100

Our study showed that phages across diverse families (*Podoviridae*, *Siphoviridae* and *Microviridae*) can infect the same host. However, the members within each family or cluster tend to maintain their genetic stability. Many new types of phages were isolated from one marine strain suggesting that we still know very little about the phages infecting diverse roseobacters in the marine environment.

Chapter 5. Novel N4 bacteriophages prevail in the cold biosphere

5.1 Abstract

Coliphage N4 is a lytic bacteriophage discovered nearly half century ago, and it was considered to be a 'genetic orphan' until very recently when several additional N4-like phages were discovered to infect non-enteric bacterial hosts. Interest in this genus of phage is stimulated by their unique genetic features and propagation strategies. To better understand the ecology of N4-like phages, I investigated the diversity and geographic patterns of N4-like phages by examining 56 Chesapeake Bay viral communities, using a PCR-clone library approach targeting a diagnostic N4-like DNA polymerase gene. Many new lineages of N4-like phages were found in the Bay and their genotypes shift from the lower to upper Bay. Interestingly, signature sequences of N4-like phages were only recovered from winter month samples, when water temperatures were below 4 °C. An analysis of existing metagenomic libraries from various aquatic environments supports the hypothesis that N4-like phages are most prolific in colder waters. In particular, a high number of N4-like phages were detected in Organic Lake, Antarctica, a cold and hypersaline system. The prevalence of N4-like phages in the cold biosphere suggests these viruses possess as-yet-to-be determined mechanisms that facilitate lytic infections under cold conditions.

5.2 Introduction

As the most abundant microbial form, viruses play important roles in shaping host population structures, mediating genetic exchange between hosts, and modulating trophic transfer in marine food webs (Fuhrman, 1999; Suttle, 2005, 2007). Marine viral metagenomic studies suggest that viruses encompass the largest genetic repertoire in the ocean (Angly *et al.*, 2006; Hurwitz and Sullivan, 2013). Yet, the function of ~70% of the putative genes identified in viral metagenomic studies remains unknown (Rosario and Breitbart, 2011). Recently, several novel phages infecting dominant marine bacteria have been isolated from the ocean (Zhao *et al.*, 2013; Kang *et al.*, 2013). Both cultivation techniques and molecular approaches suggest that a great deal of viruses in the sea await discovery.

Bacteriophage N4 was first isolated from sewer waters using *Escherichia coli* as a host nearly half decade ago (Schito *et al.*, 1966), and remained as a "genetic orphan" for decades, as no other genetically similar phages were characterized. Phage N4 has several unique features with regards to its morphology, physiology, and genome that have made it a focus of study. It has a 70 nm isocahedral capsid and its genome size is 70 kb. More remarkably, N4 is the only known phage that does not rely on host RNA polymerase for early transcription (Falco *et al.*, 1978; Davydova *et al.*, 2009). Instead, it contains a DNA-dependent virion-encapsidated RNA polymerase (vRNAP), which is co-injected with viral DNA into its host and initiates transcription. The phage also exhibits a lysis-inhibited infection cycle and extremely large burst size (ca. 3,000 phages per cell), suggestive of a novel mechanism of cell lysis regulation (Schito, 1974; Stojković and Rothman-Denes, 2007).

A recent, renewed interest in the ecology of N4 was prompted by the isolation of two new N4-like phages from a coastal estuary, which infect bacteria of the marine *Roseobacter* lineage (Zhao *et al.*, 2009b). Roseobacters are a widely distributed and abundant group of marine bacteria (Buchan *et al.*, 2005). Thus, the finding of roseobacter N4-like phages demonstrated this phage class is not restricted to *E. coli,* or other enterics, and suggested that the marine environment might be an important reservoir for this group of viruses. Subsequent studies have described the isolation and characterization of additional N4-like phages that infect a variety of bacterial species, including *Vibrio, Pseudomonas, Salmonella,* and *Achromobacter* (Fan *et al.*, 2012; Wittmann *et al.*, 2014; Born *et al.*, 2011; Kulikov *et al.*, 2012; Fouts *et al.*, 2013; Gan *et al.*, 2013; Moreno Switt *et al.*, 2013; Chan *et al.*, 2014). With the exception of *Enterobacter* phage EcP1 (NC_019485) which has a relatively small capsid and genome size, all N4-like phages characterized to date share a morphology, genome size, and genomic architecture similar to coliphage N4 (Chan *et al.*, 2014). Given the recent successes in isolating N4-like phages from phylogenetically diverse bacterial taxa, I was motivated to better understand the distributions of this phage group in natural systems. The conservative nature of N4-like genomes allowed the design of molecular tools diagnostic for group members, which facilitated the isolation-independent assessments of distribution and diversity of N4-like phages over space and time presented here.

5.3 Materials and methods

5.3.1 Sample collection and preparation

Chesapeake Bay viral community samples were collected in 2004 and 2005 from multiple stations in the bay during research cruises for the Microbial Observatories Viral Ecology project. Viral DNA was prepared following methods described elsewhere (Wang and Chen, 2004). Briefly, 50 L of water was filtered through A/E glass-fiber filters and then 0.45 μm pore-size polycarbonate filters. The viral fraction was then concentrated to ~500 ml by ultrafiltration through a 30 kDa cutoff filter cartridge. These concentrated viral communities (VCs) were further precipitated with polyethylene glycol 8000 powder at the final concentration of 100 g L^{-1}, and then re-suspended in SM buffer. Concentrated VCs were boiled to release viral DNA, which was used as DNA template for PCR.

A total of 56 archived viral DNA samples were analyzed for the presence of N4 phages. These samples were collected from nine different stations across the whole Bay during 2005 and subset of five of these stations during 2004 (Figure 5.1). Samples were collected four seasons (February [winter], May [spring], August [summer] and October [fall]).

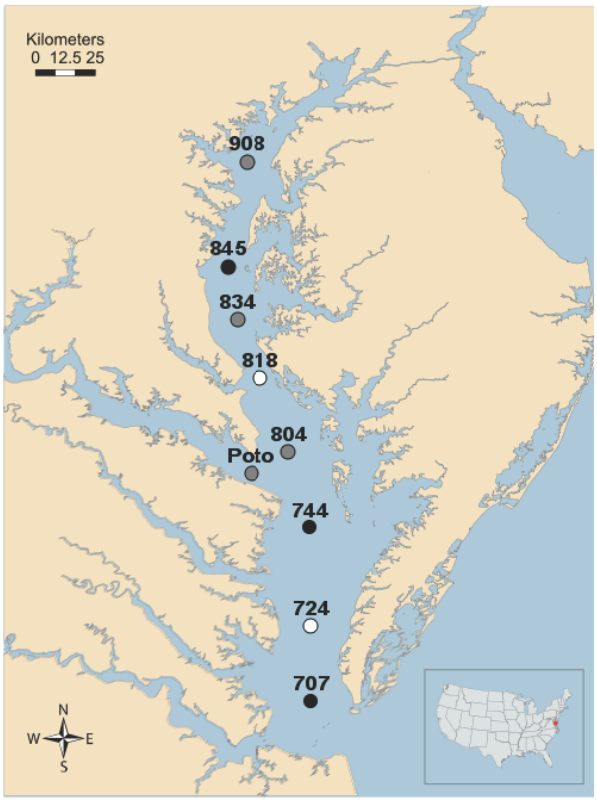


Figure 5.1 The locations of the winter Chesapeake Bay samples that were positive for N4-like specific DNA *pol* PCR products. The positive samples in both winter 2005 and 2004 are shown with a solid circle. The empty circle represents the samples that were only positive in one year. The samples at Stn. 908, 834, 804 and Poto which stands for Potomac river (shown in grey circle) were only available in 2005 but not in 2004. Figure was modified based on image by Tracy Saxby, Kate Boicourt, IAN Image Library (ian.umces.edu/imagelibrary/).

5.3.2 PCR primers and amplification

A set of degenerate primers were designed based on available DNA polymerase (DNA *pol*) gene of N4-like phages in the NCBI database through 2011, including *Enterobacteria* phage N4 (NC_008720), Roseophage DSS3Φ2 (NC_012697), Roseophage EE36Φ1 (NC_012696), *Roseovarius* Plymouth Podovirus 1 (FR719956), *Roseovarius* sp. 217 phage 1 (FR682616), *Enterobacter* phage EcP1 (NC_019485), *Pseudomonas* phage LIT1 (NC_013692), *Pseudomonas* phage LUZ7 (NC_013691) and *Sulfitobacter* phage pCB2047-B (HQ317387). The forward primer is N4-DNAP-F (5'-GGI ACI ATI ACI TTY TGY TGG-3') and the reverse primer is N4-DNAP-R (5'-RTA RTT ICC IGC RAA TYC YTG-3'). The expected PCR product size is *ca.* 400 bp. PCR conditions were optimized for template DNA and primer concentrations, annealing temperature, magnesium chloride concentration, and cycle number.

All the PCR reactions were performed in 25 µl volumes containing 1X reaction buffer, 2 mM $MgCl_2$, 200 µM dNTP, 2 µM of each primer, 1 U of Platinum *Taq* DNA Polymerase (Invitrogen, Carlsbad, CA, USA) and 4 µl DNA templates. The PCR reaction was performed with the following steps: denaturation for 10 min at 94 °C, 35 cycles of denaturation (30 s at 94 °C), annealing (1 min at 50 °C) and extension (1 min at 72 °C), with a final extension for 10 min at 72 °C. DNA of roseophage DSS3Φ2 was used as positive control, and a series of dilutions of positive control DNA were employed to determine the threshold of detection.

5.3.3 Clone library, sequencing and phylogenetic analysis

Five winter PCR amplicons, two from upper bay (84504, 84505), one from middle bay (83405) and two from lower bay (70704, 70705), were selected to construct clone libraries (Table 5.1). The PCR products were purified with the QIAquick® Gel Extraction Kit (QIAGEN, Venlo, The Netherlands) and cloned using the TOPO TA pcr4.0 Cloning Kit (Invitrogen, Carlsbad, CA, USA). Colonies were randomly picked from each clone library and sequenced using an ABI 3100 Genetic Analyzer (PE Applied Biosystems, Foster City, CA, USA). Homology searches (BLASTx) were performed against the NCBI database. If the top hit of the sequences was an annotated known N4-like phage from amino acid position 206 to 317 in the N4 phage DNA polymerase (NC_008720), I identified it as an N4-like phage sequence.

DNA *pol* partial gene sequences were translated into putative protein sequences and aligned by Clustal W (Larkin *et al.*, 2007) using default parameters. A neighbor-Joining (NJ) phylogenetic tree was built using MEGA 5.0 (Tamura *et al.*, 2011) and the *p*-distance model. Bootstrap values were determined from 1,000 resampling events. All sequences have been deposited in GenBank with accession numbers KM527952 - KM528122.

5.3.4 Diversity indices and metagenomic analysis

OTU (Operational Taxonomic Unit), Chao index, evenness, and Shannon Diversity index of each library were measured using the Mothur suite of programs (Schloss *et al.*, 2009) and were based on DNA sequences at 2% divergence.

Selected translated amino acid sequences from each subcluster were searched against the CAMERA portal (Sun *et al.*, 2011) by tBLASTn with the threshold of E-value ≤ 1E-10 and alignment length ≥ 80 amino acids. Three ocean metagenomic databases were chosen for N4-like phage recruitment, Broad Phage metagenomic database (http://www.broadinstitute.org/annotation/viral/Phage/Home.html), Global Ocean Survey (GOS) microbial metagenomic database (Yooseph *et al.*, 2007; Rusch *et al.*, 2007) and Pacific Ocean Virome database (Hurwitz and Sullivan, 2013). I also recruited N4-like phage from all CAMERA portal metagenomic libraries available by the end of 2013, including the Antarctica Aquatic Microbial Metagenome, Chesapeake Bay Virioplankton Metagenome, Chicken Cecum Microbiome, Wastewater Metagenome from Mallard Creek (all available at http://camera.crbs.ucsd.edu/) and the Human Microbiome Project (http://hmpdacc.org/catalog/). The sequences that were related to N4-like phages and fully overlapped with our clone sequences were included in a phylogenetic analysis.

5.4 Results and Discussion

5.4.1 Detection of N4-like phages in the Chesapeake Bay

Given that the first two marine N4-like phages were isolated from Chesapeake Bay waters, I sought to determine the prevalence and diversity of these phages over space and time in this coastal estuary. To do this, I screened a library of viral concentrates collected from the Chesapeake Bay over the course of two years (2004-2005) through the Microbial Observatory for Virioplankton Ecology Project

(http://www.virusecology.org/MOVE/Home.html) using a PCR primer set targeting conserved regions of the DNA polymerase (DNA *pol*) gene of N4-like phages. The DNA *pol* gene is one of the 14 conserved core genes of N4-like phages (Chan *et al.*, 2014) making it a valuable diagnostic marker for culture-independent surveys. A total of 56 viral samples, which cover four seasons and the full transect of the Bay, were tested. Twelve of 14 winter samples yielded a PCR product of the expected size, but none of viral samples in other seasons (spring, summer and fall) yielded positive results (Figure 5.2). These products were sequenced and confirmed to be N4-like DNA *pol* gene homologs.

The 12 PCR positive samples were collected during winter of two consecutive years and represent samples spanning the breadth of the Chesapeake Bay watershed (Figure 5.1) when water temperatures were below 4 °C. It is surprising that none of viral samples from spring, summer and fall were PCR positive. Although no quantitative measurements have been performed to compare the abundance of N4-like phages between the winter and other seasons, the end point PCR results obtained here imply that N4-like phages in other seasons are either absent or less than the current detection limitation. It is unlikely that negative PCR results occurred in other seasons were due to poor quality of these same viral DNA samples because control reactions targeting DNA *pol* of cyanobacterial podoviruses (Chen *et al.*, 2009) yielded positive results.

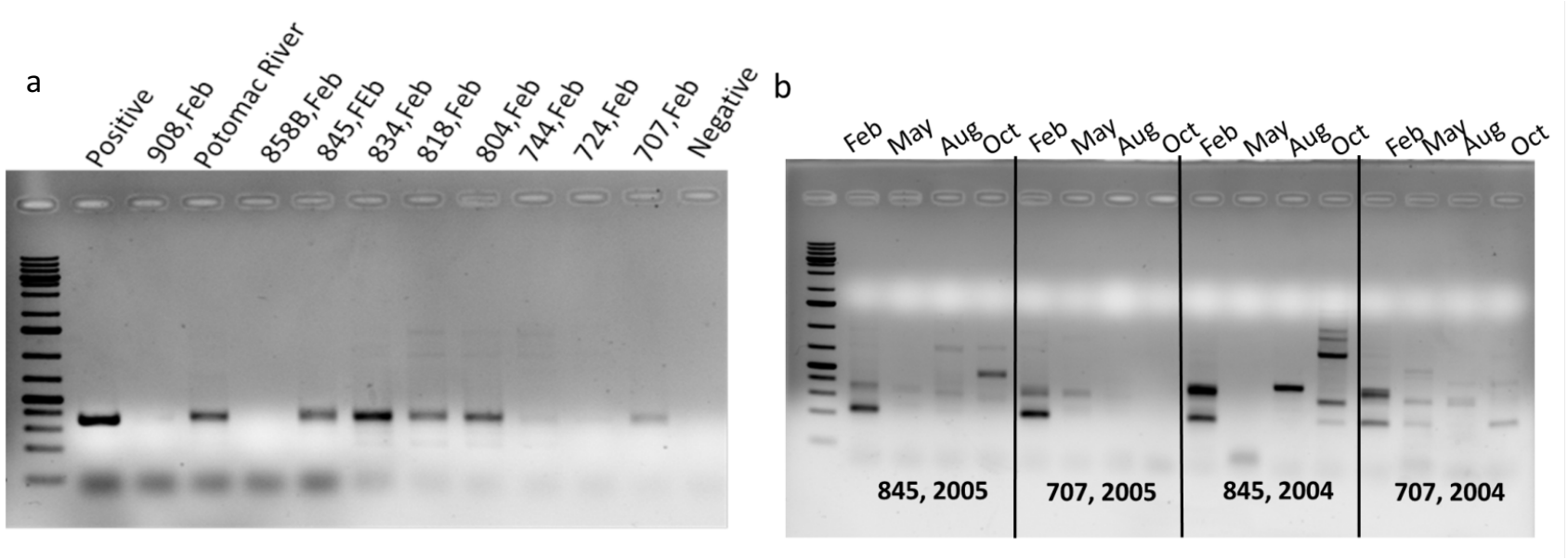


Figure 5.2 The amplification of N4-like phages in the Chesapeake Bay using degenerated primers. a) Samples were collected in February 2005, covering all the major stations along the bay. b) Samples were collected at stations 845 and 707 from 2004 and 2005. Samples from four seasons were included for testing.

Chesapeake Bay is a dynamic ecosystem. It has been reported that bacterial communities in the Bay exhibit distinct seasonal patterns (Kan *et al.*, 2006, 2007). N4-like phages are generally thought to have a narrow host range (Kulikov *et al.*, 2012). Indeed, the shift in bacterial populations and different bacterial metabolisms from cold to warm may be driving the distinct occurrence of N4-like phages observed in this study. However, an important caveat is that I cannot exclude the possibility that N4-like phages are present during other (spring, summer, fall) seasons. It is possible that the concentration of N4-like phages in the Chesapeake Bay during the warmer seasons is below the limit of detection of the PCR assay, which was determined to be approximately 100 viruses per ml using DNA isolated from the positive control roseophage DSS3Φ2. Furthermore, not all N4-like phages have been isolated from cold environments (Ji *et al.*, 2015b).

5.4.2 N4 phages are diverse in the Chesapeake Bay

Five of 12 positive amplicons from the Chesapeake Bay samples were selected for clone library analysis. Selected samples represent viral communities from the upper (station 845), middle (station 834) and lower bay (station 707) during winter 2005. Two additional samples collected from upper (station 845) and lower (station 707) samples from winter 2004 were chosen to assess annual variation. All the five samples were collected when the water temperature was between 1 and 4 °C, and represent a wide range of salinity and nutrient gradients (Table 5.1).

A total of 204 clones randomly selected from these five clone libraries were sequenced. Among them, 24 clones contained non-translatable or non-specific

sequences and were not included in subsequent analyses. The DNA *pol* phylogeny based on the phage isolates and clonal sequences divided N4-like phages into at least 15 distinct clusters with 80% amino-acid sequence identify cutoff for each cluster (Figure 5.3).

For characterized N4-like phages, the N4-like DNA *pol* phylogeny indicates general agreement with host phylogeny. For example, all five N4-like phages isolated on *E.coli* hosts fall into Cluster II. Similarly, all N4-like phages isolated from hosts fall into Cluster IX. N4-like phages that infect *Vibrio* spp. *Salmonella* spp., and *Pseudomonas* spp. also cluster with host species. Overall, the phylogeny of N4-like phages genotype based on DNA *pol* gene is concordant to that previously reported based on concatenated core genes of N4-like phages (Chan *et al.*, 2014). The congruency between DNA *pol* phylogeny and host taxonomy suggests N4 phages originated from a common ancestor and co-evolved with their hosts.

For the 180 Bay-derived N4 DNA *pol* sequences, the majority of fall into eleven clusters (Cluster I to X, and XVIII) (Figure 5.3). Among these eleven clusters, eight clusters do not contain any cultured representatives. Indeed, the vast majority of Bay clones do not cluster with N4-like DNA *pol* sequences that were obtained from cultured isolates. One environmental clone (845.04.08) does fall into Cluster II, which contains several N4-like phages infecting various *E. coli* strains, and another clone (845.05.29) is closely associated with N4-like phages that infect marine roseobacters. N4-like phages which infect *Vibrio*, *Salmonella*, and *Pseudomonas* spp. belong to clusters XI, XII, XIV, and XIX; no Chesapeake Bay clones mapped to these clusters.

Table 5.1 Environmental and community parameters at sampling stations

Station	Time	Temperature (°C)	Salinity (psu)	Bacterial Abundance* ($x10^5$ Cells/ml)	Viral Abundance* ($x10^5$ VLPs/ml)	NO_2^- &NO_3^+ (μM)
707	Feb, 2004	3.8	15.4	8.61	9.6	12.3
707	Feb, 2005	2.9	18.0	12.3	87.1	9.03
845	Feb, 2004	1.2	9.6	10.4	31.2	15.9
845	Feb, 2005	3.9	7.5	12.3	100.0	32.7
834	Feb, 2005	3.2	7.4	9.82	46.6	40.8

* The number of bacteria and viruses were enumerated by SYBR Gold staining, following the protocol previously described (Wang and Chen 2004).
+ Nitrite and Nitrate were determined using a Technicon Auto Analyzer II at the Horn Point Analytical Services Laboratory (www.hpl.umces.edu/services/as.html).

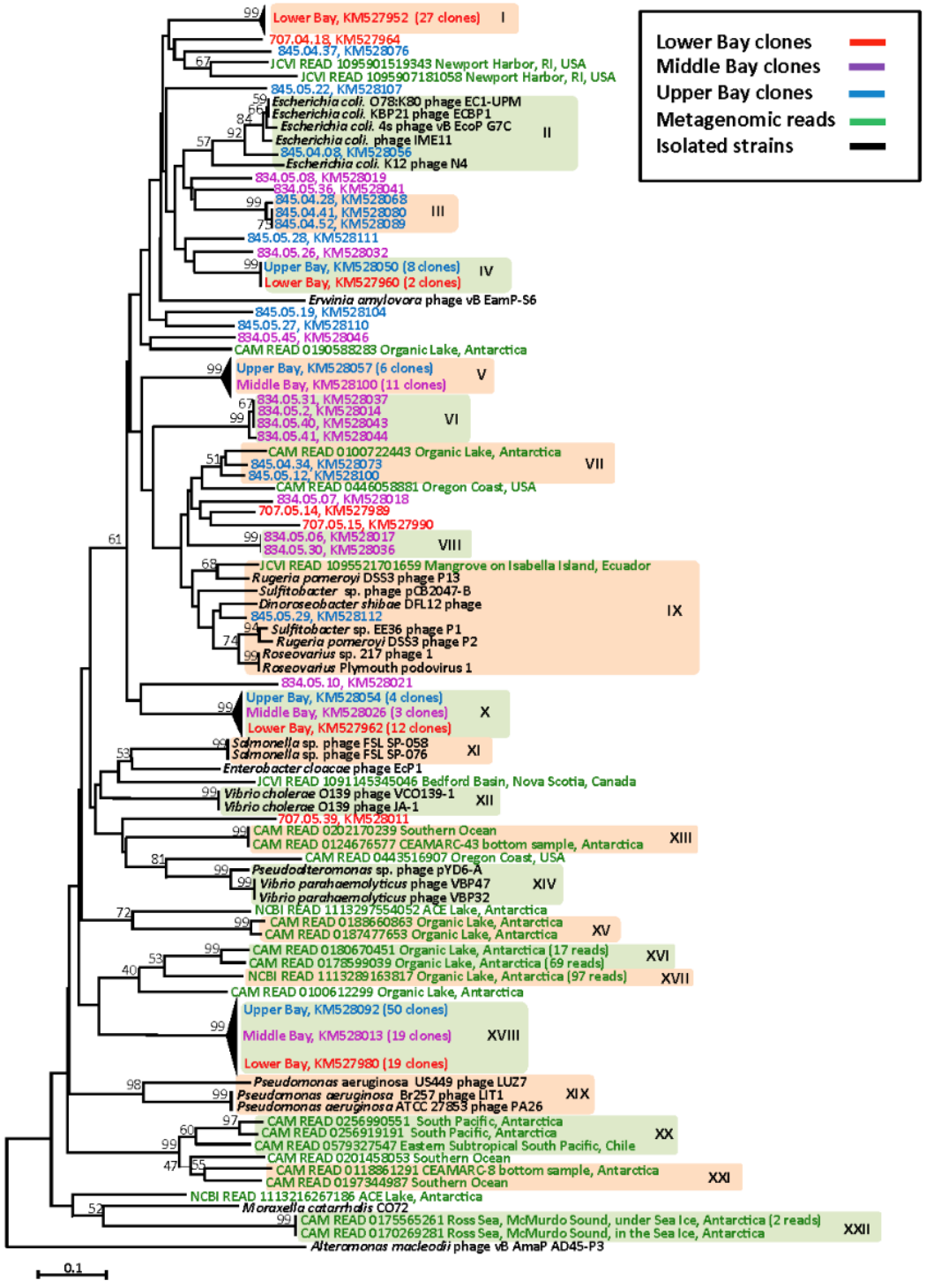


Figure 5.3 The neighbor-joining phylogenetic relationship of translated DNA *pol* gene sequences recovered from the clone libraries, known N4-like phages, and metagenomic databases. The tree is based on the aligned amino acid sequences (*ca.*

130 residues). The sequences of 24 N4-like phages and two outgroup sequences are shown in black. The sequences recovered from metagenomic searches, upper, middle and lower Chesapeake Bay are in green, red, purple and blue, respectively. The clusters were defined as sequences that share greater than 80% amino acid identity. The bootstrap value was calculated with 1000 replications. Bootstrap values ≥ 50% are shown at nodes. The scale bar represents 0.1 amino-acid substitutions.

The composition of N4-like phages differs between the lower and upper bay. Several clusters (I, III, VI, VII, and VIII) only contain clonal sequences from either the upper or lower bay, indicating that unique genotypes of N4-like phage occupy specific niches in the Chesapeake Bay. Meanwhile, some clusters (X and XVIII) have been detected in both upper and lower bay samples (Table 5.2 and Figure 5.4). In short, Chesapeake Bay viral communities contain diverse N4-like phage sequences in winter, and different N4-like phage genotypes were found in the mid-upper and lower Bay.

I believe that the number of environmental clones used in this study is sufficient to unveil the diversity of major populations of N4 phages as most of clone libraries were well sampled (Figure 5.5). High-throughput sequencing could potentially yield more clusters for minor N4 populations, but this is beyond the scope of this work. The presence of 11 distinct clusters of N4-like phages in the Chesapeake Bay alone suggests that N4-like phages can infect a much broader bacterial community than originally thought.

Table 5.2 Distribution of phylogenetic clusters and diversity indices among the five clone libraries.

Library Name	Cluster[+]											Un-clustered	Total clones	Number of OTUs[*]	Normalized OTUs[#]	Richness (Chao)	Evenness	Shannon diversity
	I	II	III	IV	V	VI	VII	VIII	IX	X	XVIII							
70704	18			2						4	4	1	29	5	0.17	5.00	0.69	1.11
70705	9									8	15	3	35	8	0.23	14.00	0.75	1.56
84504		1	3	6	4		1			3	25	1	44	11	0.25	13.50	0.69	1.66
84505				3	2		1		1	1	25	3	36	11	0.31	56.00	0.52	1.25
83405				1	11	4		2		3	10	5	36	19	0.53	32.20	0.88	2.60

Abbreviation: OTU, operational taxonomic unit.
[+] Clusters as defined in Figure 5.3.
[*] Number of OTUs was calculated based on DNA sequences at 2% divergence.
[#] The number of OTUs was normalized to the total number of clones from a given station.

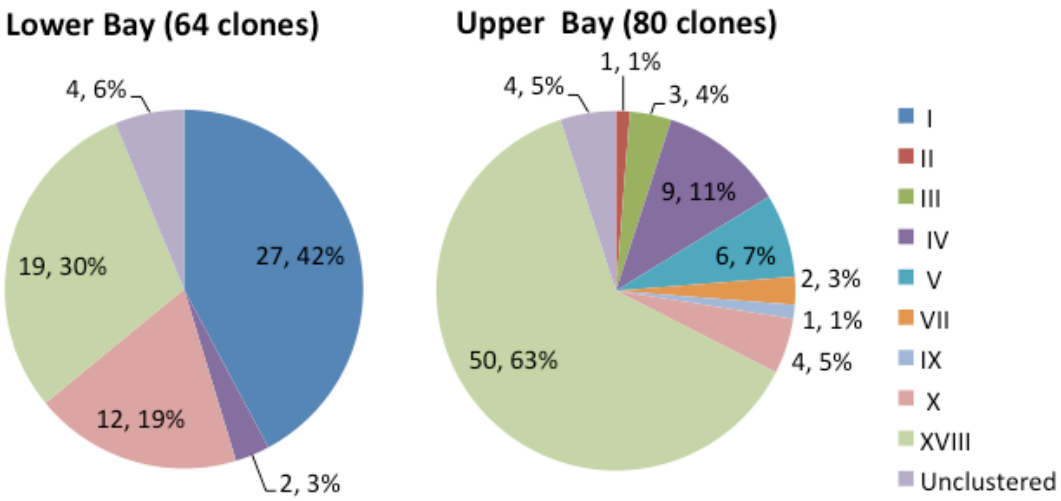


Figure 5.4 The comparison of proportion of DNA *pol* subclusters between the upper and lower Chesapeake Bay.

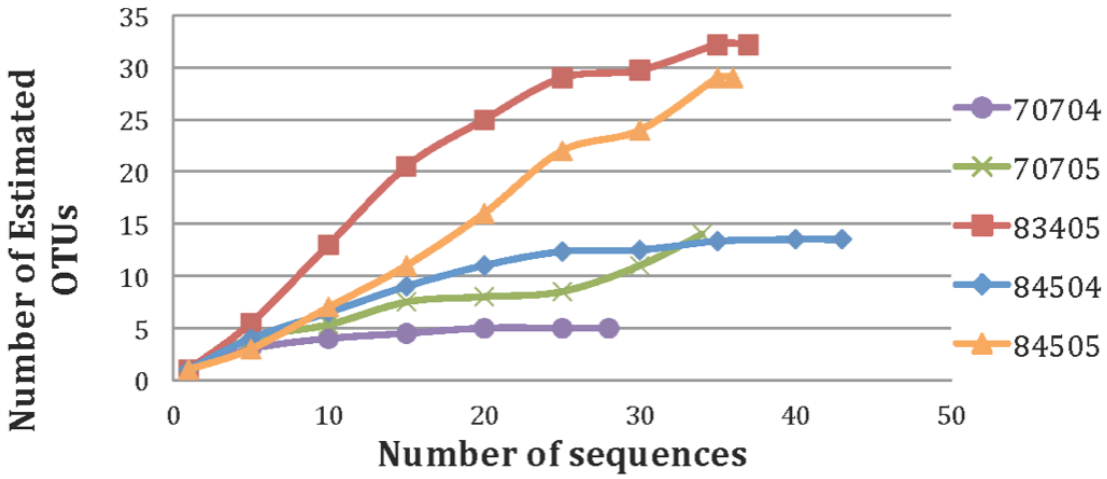


Figure 5.5 Rarefaction curve of five clone libraries.

5.4.3 Identification of N4-like phages in cold environments

The unexpected and apparent restriction of Chesapeake Bay N4-like phages to colder waters prompted a more global survey of their distribution and diversity. To that end, I surveyed available metagenomic libraries for N4-like DNA polymerase homologs. Representative DNA *pol* amino acid sequences from all subclusters were searched against the database with the threshold of E-value ≤ 1E-10 and alignment length ≥80 amino acids. The 206 homologous sequences identified were included in a DNA *pol* phylogenetic tree (Figure 5.3). These environmental sequences form seven unique clusters (XIII, XV, XVI, XVII, XX, XXI, and XXII) and the majority do not overlap with Chesapeake Bay clones nor isolated N4-like phages, suggesting a great diversity of N4-like phages exists in other freshwater and marine environments.

Consistent with my PCR-based survey on seasonally representative Chesapeake Bay samples, nearly all (97%) of the metagenome-derived DNA *pol* sequences were recovered from colder environments, including the Southern Ocean and sea ice (Ross Sea). Indeed, most (91%) of the metagenomic reads were retrieved from Organic Lake in Antarctica (Figure 5.6). The majority (97%) of Organic Lake hits (188 in total) were classified into two narrow clusters (XVI and XVII), suggesting a relatively limited number of potential hosts in the system. It is possible that a few major bacterial populations contribute to the production of these N4-like phages. Intriguingly, these two major clusters appear to be related to cluster XVIII, which contains the largest number of the Chesapeake Bay N4-like phage sequences. Together, my results suggest that N4-like phages can thrive in cold environments. For the purposes of this discussion, I define "cold" as water masses with

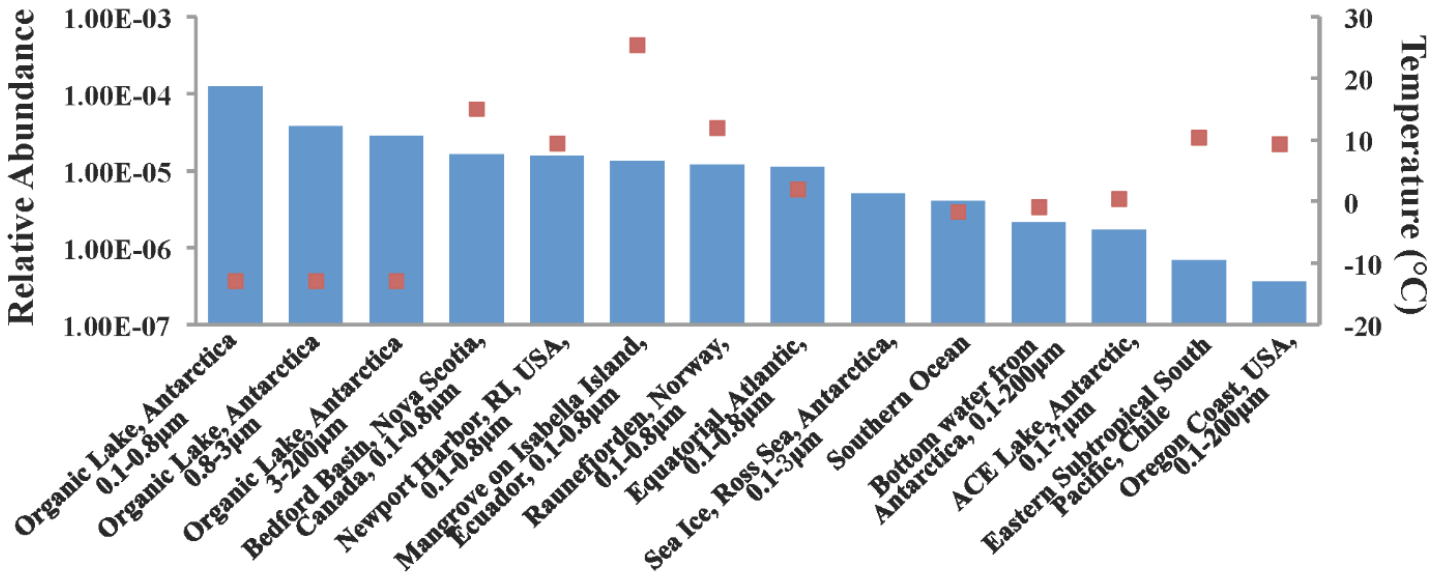


Figure 5.6 The occurrence of N4-like phages reads in metagenomic database (shown in blue bar) and corresponding temperature (shown in red spot). The temperature of sea ice, Ross Sea left as blank, since temperature is not provided by this metagenomic database.

temperatures ≤ 10 °C, and “warm” water masses as those with temperatures > 10 °C. When considering the metagenomic library analysis, N4-like phages in cold waters are significantly more abundant than those in warm waters (t-test, *p* value < 0.01). Unfortunately, there are not presently sufficient environmental data for the metagenomic libraries to delineate the relationships between the distribution of N4-like phages and environmental parameters.

Antarctica’s Organic Lake is a shallow, hypersaline lake and contains the highest concentration of dimethylsulfide reported in natural water environments (Franzmann *et al.*, 1987). Water temperatures are below -10 °C and salinity exceeds 160 psu (Yau *et al.*, 2013). Consequently, the abundance of bacteria and viruses are lower than in the oligotrophic ocean (both are less than 10^5 ml^{-1}), and the viral-to-bacterial ratio is approximately 1 (Franzmann *et al.*, 1987; Yau *et al.*, 2013). It is intriguing that a high number of N4-like phage sequences were recruited from this system despite the low viral abundance. It has been reported previously that members of the *Alphaproteobacteria* and *Gammaproteobacteria* classes dominate the bacterial composition of Organic Lake (Yau *et al.*, 2013) and may represent presumptive hosts for these phages.

The prevalence of N4-like phage from cold, saline waters suggests at least some of these phages may be adapted to environments with these characteristics. It might be hypothesized that the unique features of N4-like phages, especially the presence of an encapsulated RNA polymerase (Schito, 1973), may allow them to tolerate and replicate at cold (or even freezing) temperatures and under saline (or hypersaline) conditions. With viral RNA polymerase, N4-like phages are less host-

dependent compared to other phages (See Chapter 4). Therefore, it might be possible that N4-like phages can tolerate extreme environments, such as cold temperature and hypersaline conditions (See Chapter 6). Indeed, it has been previously reported that the RNA polymerase in *E.coli* phage N4 can tolerate high salt conditions (Murakami *et al.*, 2008). Further studies on cultured N4-like phage are necessary to ascertain whether cold tolerance is a defining feature of this phage class.

5.5 Conclusion

N4-like phages are a group of bacteriophages that are unique in terms of their genomes, phylogeny, taxonomy, and ecology. This study shows that diverse N4-like phages are present in nature and they exhibit dynamic seasonal and spatial variation. The distribution of N4-like phages in the marine environment appears to be restricted to high latitudes and/or colder seasons. The high recruitment of N4-like phages in the cold biosphere, such as Organic Lake of Antarctica, is unexpected based on prior successful cultivation efforts, in which N4-likephages were isolated only from more temperate environments. The cold selection of N4-like phages could be related to their unique features, including lysis inhibition, large burst size, and viral encapsidated RNA polymerase. However, the ecological role of N4 phages and the mechanisms involved in cold adaptation remains to be elucidated.

Chapter 6. Conclusion and Future Prospects

6.1 Isolation and genomics of roseophages

In this study, fourteen different phages infecting a marine strain, *R. pomeroyi* DSS-3, were isolated from Baltimore Inner Harbor. These phages exhibit distinct morphotypes including small non-tailed phages, large N4-like podoviruses, and various types of siphoviruses. No T7-like podoviruses and myoviruses were found in my phage collection. Although a few roseophages have been reported from other strains, the roseophages isolated through this dissertation work represent the most phenotypically diverse phages infecting marine roseobacters. Complete genomes of these 14 roseophages were sequenced and annotated. The genomic information shows that various genome types of roseophages interact with a single strain. Comparative genomics divided these 14 phages into four distinct phage groups (I to IV). Group I, II, III, and IV include five N4-like phages, two CbK-like phages, five Chi-like phages and two single-stranded phages, respectively (Table 6.1).

The isolation of the smallest ssDNA phages infecting roseobacters came as a surprise (Chapter 2, Table 6.1). In general, very few isolated ssDNA phages have been reported for marine bacteria even though many have been assembled in the marine metagenomic studies. Interestingly, the ssDNA roseophages (DSS3Φ22 and DS33Φ15) fall into a large group of unknown viral sequences as shown by viral metagenomics. The ssDNA roseophages only contain four ORFs with genome sizes of 4.2 kb. They represent the smallest and simplest ssDNA phage among currently known ssDNA phage isolates.

Table 6.1 Summary of all published bacteriophages infecting marine roseobacters.

Phage family	Phage	Host	Isolation Site	Genome Size (kb)	GC content (%)	Number of ORFs	Number of tRNAs	References
Podoviridae (N4-like)	DSS3Φ2	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	74.9	47.9	81	3	Zhao *et al.* (2009)
Podoviridae (N4-like)	DSS3Φ12	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	74.7	47.9	85	3	Chapter 4
Podoviridae (N4-like)	DSS3Φ13	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	74.8	50.8	79	4	Chapter 4
Podoviridae (N4-like)	DSS3Φ14	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	74.8	47.9	88	3	Chapter 4
Podoviridae (N4-like)	DSS3Φ17	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	74.7	47.9	83	3	Chapter 4
Podoviridae (N4-like)	DSS3Φ21	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	74.7	47.9	82	3	Chapter 4
Podoviridae (N4-like)	ESS36Φ1	*Sulfitobacter* sp. EE-36	Baltimore Inner Harbor, USA	73.3	47.0	79	3	Zhao *et al.* (2009)
Podoviridae (N4-like)	ΦCB2047-B	*Sulfitobacter* sp. 2047	Mesocosm Raunefjorden Norway	74.5	43.0	77	15	Ankrah *et al.* (2014a)
Podoviridae (N4-like)	RLP1	*Roseovarius* sp. 217	Langstone Harbor, Hampshire, UK	74.6	49.0	92	3	Chan *et al.* (2014)
Podoviridae (N4-like)	RPP1	*R. nubinhibens*	L4 sampling station, Plymouth, UK	74.7	49.1	91	3	Chan *et al.* (2014)
Podoviridae (N4-like)	vB_DshP-R1	*D. shibae* DFL12	Baicheng Harbor, Xiamen, China	75.0	49.3	86	2	Ji *et al.* (2015)
Podoviridae (N4-like)	vB_DshP-R2C	*D. shibae* DFL12	Huangcuo station, Xiamen, China	74.8	49.2	85	2	Cai *et al.* (2015)
Podoviridae (N4-like)	DS-1410Ws-06	*D. shibae* DFL12	Sanya Bay, China	76.5	50.0	77	0	Li *et al.* (2016)
Podoviridae (N4-like)	RD-1410W1-01	*R. denitrificans* Ch114	Sanya Bay, China	72.7	49.5	77	0	Li *et al.* (2016)
Podoviridae (N4-like)	RD-1410Ws-07	*R. denitrificans* Ch114	Sanya Bay, China	76.3	50.0	76	0	Li *et al.* (2016)
Siphoviridae (Chi-like)	DSS3Φ1	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	59.6	64.1	82	0	Chapter 4
Siphoviridae (Chi-like)	DSS3Φ7	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	59.6	64.1	84	0	Chapter 4
Siphoviridae (Chi-like)	DSS3Φ18	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	59.1	64.0	82	0	Chapter 4
Siphoviridae (Chi-like)	DSS3Φ11	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	59.6	64.0	83	0	Chapter 4
Siphoviridae (Chi-like)	DSS3Φ16	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	61.4	63.6	85	0	Chapter 4
Siphoviridae (Chi-like)	pCB2051-A	*Loktanella* sp. CB2051	Norwegian Sea, Arctic	56.9	55.0	77	0	unpublished
Siphoviridae (CbK-like)	DSS3Φ8	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	146.1	56.4	229	34	Zhan *et al.* (2016)/ Chapter 3
Siphoviridae (CbK-like)	DSS3Φ10	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	147.5	56.4	235	35	Chapter 4
Microviridae (unclassified)	DSS3Φ22	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	4.25	57.7	4	0	Chapter 2

Microviridae (unclassified)	DSS3Φ15	*R. pomeroyi* DSS-3	Baltimore Inner Harbor, USA	4.25	57.7	4	0	Chapter 4
Podoviridae (SIO1-like)	SIO1-1989	*Roseobacter* SIO67	Scripps Pier, CA, USA	39.9	46.2	32	0	Rohwer *et al.* (2000) Angly *et al.* (2009)
Podoviridae (SIO1-like)	SIO1-2001	*Roseobacter* SIO67	Scripps Pier, CA, USA	39.4	45.3	32	0	Angly *et al.* (2009)
Podoviridae (SIO1-like)	OS-2001	*Roseobacter* SIO67	Oceanside, CA, USA	38.2	44.8	32	0	Angly *et al.* (2009)
Podoviridae (SIO1-like)	SBRSIO67-2001	*Roseobacter* SIO67	Solana Beach CA, USA	38.1	44.2	32	0	Angly *et al.* (2009)
Podoviridae (SIO1-like)	MB-2001	*Roseobacter* SIO67	Mission Bay, CA, USA	38.2	44.1	32	0	Angly *et al.* (2009)
Podoviridae (SIO1-like)	P12053L	*Celeribacter* sp. IMCC12053	Yellow Sea, South Korea	35.9	46.1	56	0	Kang *et al.* (2012)
Siphoviridae	RDJLΦ1	*R. denitrificans* OCh114	South China Sea surface seawater	62.7	57.9	87	0	Huang *et al.* (2011)
Siphoviridae	RDJLΦ2	*R. denitrificans* OCh114	Wuyuan Bay, Xiamen, China	63.5	57.3	76	0	Liang *et al*. (2015)
Podoviridae	ΦCB2047-A	*Sulfitobacter* sp. 2047	Mesocosm study, Raunefjorden, Norway	40.9	58.8	73	0	Ankrah *et al.* (2014b)
Podoviridae	ΦCB2047-C	*Sulfitobacter* sp. 2047	Mesocosm study, Raunefjorden, Norway	40.9	59.0	73	0	Ankrah *et al.* (2014b)

The isolation of CbK-like roseophages adds another dimension of diversity for roseophages as this type of phage has not been reported in marine bacteria (Chapter 3, Table 6.1). CbK-like phages are a new group of siphoviruses first found in a freshwater bacterium *Caulobacter*. CbK-like phages appear to be a hybrid between podoviruses and siphoviruses. They contain the T7-like DNA polymerase gene but resemble a siphovirus. CbK-like roseophages are highly mosaic, containing the GTA-related genes, integrase and a large number of tRNAs. The genomic nature of CbK-like phages suggests that evolution of roseophage genomes can be complicated due to genetic exchange across phage families.

Five of 14 roseophages are Chi-like siphophages, belonging to another group of siphoviruses (Chapter 4, Table 6.1). Unlike CbK-like siphophages, Chi-like siphophages do not contain any tRNA or GTA-related genes. Upon superinfection with Chi-like roseophage DSS3Φ1, a resistant strain of DSS-3 was found to contain the intact genome of DSS3Φ1. This was confirmed by genome sequencing of resistant strain DSS-3-P1R. The ability to integrate phage genome into host chromosome confirms that DSS3Φ1 can perform lysogenic infection. Interestingly, the wild type DSS-3 does not contain prophage. With the lysogenized DSS-3 strain it is now possible for me to investigate the impact of lysogeny on *R. pomeroyi* DSS-3 (see Future Prospects).

Five N4-like roseophages were isolated in my study. Together with previous isolates, a total 15 N4-like roseophages have been isolated (Table 6.1). Genomes of N4-like roseophages are highly syntenic, sharing a very similar genomic arrangement. Comparative genomics show that all the N4-like roseophages share 39 core genes,

accounting for approximately 70% of their genome sizes. It is noteworthy that DSS3Φ13 is distantly related to the other N4-like roseophages infecting *R. pomeroyi* DSS-3, indicating that N4-like phages that infect the same host are not necessarily clustered together. With many N4-like roseophage genome sequences available, it is now possible to design more specific primers to study the diversity and distribution of N4-like roseophages (see Future Prospects).

All the roseophages I isolated were stored at 4 °C for future studies. Concentrated viral particles with concentrations *ca.* 1E12/ml were preserved in 1.5 ml Eppendorf tubes. Meanwhile, a large volume (*ca.* 50 ml) of viral lysates were also kept and are ready for reactivation.

It is evident that many different types of phages in one location can infect a single strain of roseobacter. Thus, I am only scratching the surface regarding the diversity of phages, and as a result, much more phage isolation work is needed in order to fully understand the diversity of phages infecting marine *Roseobacter*.

6.2 Ecological diversity and global distribution of marine roseophages

The genomic conservation of N4-like phages allows me to design PCR primers to detect N4-like phages in the natural environment. I investigated the distribution and diversity of N4-like phages in the Chesapeake Bay using 56 viral samples collected from the Microbial Observatories on Viral Ecology (MOVE) project in 2004 and 2005. Surprisingly, the PCR method only detected N4-like phage sequences in the Chesapeake Bay during the winter season. Further metagenomic recruitment based on the DNA polymerase gene of N4-like phages also confirms that N4-like phages

are relatively more abundant in the cold biosphere, especially in the Organic Lake, a hypersaline lake in Antarctica where the temperature is usually below -10 °C. Therefore, I hypothesize that N4-like phages could be more adaptive to the cold environment compared to non-N4-like phages. It is not clear whether the unique feature of N4-like phages, viral encapsidated RNA polymerase, allows them to succeed in the low temperature condition.

Metagenomic recruitment was also conducted for other non-N4-like DSS-3 phages. Although roseophages were less abundant compared with the phages infecting SAR11 and SAR116 (Zhao *et al.*, 2013; Kang *et al.*, 2013), homologs of DSS-3 phages can be found in a wide range of aquatic environments, ranging from freshwater to open oceans. For the CbK-like phages, the highest recruitment rate came from samples from Scripps Pier, where many SIO-like roseophages were isolated. CbK-like roseophages were also present in Antarctica, especially Organic Lake and Ace Lake, where abundant N4-like phage sequences were recovered. Homologs of DSS-3 ssDNA phages can be found in deep sea, human gut and feces, and coral-associated environments. DSS-3 ssDNA phages became the first culture representative of a candidate subfamily *Alpavirinae*, which only contains assembled ssDNA phage genomes from human gut and feces metagenomes (Roux *et al.*, 2012a). No isolated phage has been reported for this newly established phage subfamily *Alpavirinae,* which contains many environmental phage sequences.

My dissertation work shows that a single strain of roseobacters can be infected by various genotypes of phages and these genotypes are widely distributed in the

world's ocean. The global distribution of roseophages is consistent with the ubiquitous nature of the marine *Roseobacter* lineage.

6.3 Future prospects

6.3.1 Ecological relevance of N4-like roseophages

At the beginning of my thesis work, nine genome sequences of N4-like phages were available. These N4-like phages include two phages infecting *E. coli*, five infecting roseobacters and two infecting *Pseudomonas* (Zhan *et al.*, 2015). Because genome sequences are conserved among these N4-like phages, it allowed me to design a set of PCR primers based on the DNA polymerase gene to investigate the genetic diversity of N4-like phages in the natural environment. The results show that N4-like roseophage sequences only made up a small portion of the total N4-like environmental clones recovered from the Chesapeake Bay viral communities. On the other hand, I learned that N4-like phages infecting other bacteria such as *Vibrio*, *E. coli*, and *Pseudomonas* could be more abundant than N4-like roseophages. Currently, a total of 15 genomes of N4-like phages which infect different marine roseobacters are available, and these N4-like roseophages form a monophyletic clade based on the DNA polymerase gene phylogeny. It becomes possible now to design PCR primers specific for N4-like roseophages. The availability of N4-like roseophage specific primers will allow me to tackle a few important questions. For examples, how abundant are N4-like roseophages in the aquatic environment? How does the community of N4-like roseophages change over the spatial-temporal scale?

It appears that N4-like phage is a common type of phage interacting with marine roseobacters. Among 35 isolated roseophages, 15 of them are N4-like roseophages. The composition of the marine *Roseobacter* lineage can be investigated based on the GTA g5 gene (Zhao *et al.*, 2009a), 16S rRNA gene, or microbial metagenome (Lenk *et al.*, 2012). A set of PCR primers specific for N4-like roseophages will enable me to correlate the abundance and distribution of N4-like roseophages with roseobacters. This approach will address whether there is a co-variation between the *Roseobacter* lineage and N4-like roseophages in the natural environment.

6.3.2 Cold adaptation of N4-lik roseophages

The geographic distribution of N4-like phage (Chapter 5) suggests that marine N4-like phages appear to be prevalent in high latitude water and/or during cold seasons. *E. coli* phage N4, with viral encapsidated RNA polymerase, is able to initiate early transcription by itself. Due to the high genomic conservation of N4-like phages, it is expected that other N4-like phages have the same self-regulated transcription function as coliphage N4. Such a feature can make N4-like phages less host dependent and able to adapt to different environmental conditions. It would be interesting to study whether N4-like phages can proliferate more effectively in the cold environments than non-N4-like phages. Since several different phages infecting *R. pomeroyi* DSS-3 have been isolated and characterized, they can be used as reference phages to compare with DSS-3 N4-like phages.

Some preliminary data have been collected to compare the survival and reproduction of N4-like and non-N4-like phages infecting *R. pomeroyi* DSS-3 under different temperatures. Two N4-like phages, DSS3Φ2 and DSS3Φ13, were chosen. Meanwhile, DSS3Φ7, a temperate siphovirus, and DSS3Φ8, a CbK-like siphovirus, were selected as references. Currently, no T7-like podoviruses and myoviruses are available in our collection. Thus, I cannot include all types of phages as references. These four phages were kept at different temperatures and their infectivities were measured by plaque assay after 14 days. The recovery rate was calculated based on the ratio between number of plaques after incubation and the initial concentration of phages. The experiments were conducted in triplicates. The preliminary result shows that N4-like phages have a higher recovery rate at lower temperature compared to the reference phages (Figure 6.1).

I also tested how the phages replicate at low temperature. Host and selected phages were co-incubated at 28 °C (optimal temperature for host growth), 18 °C and 10 °C, in parallel. One-step growth curves of these phages were measured at the corresponding temperature. Unfortunately, replicates were not included in this experiments. The number of released phages was measured by qPCR targeting on the DNA polymerase gene of each selected phage (Figure 6.2). At 10 °C, no viral production was found for all the tested phages, which is probably because the host stops growing at 10 °C. The production of DSS3Φ8 (a CbK-like phage) was not even detected at 18 °C, suggesting that this phage is sensitive to the change of temperature. Both N4-like phages (DSS3Φ2 and DSS3Φ13) and siphovirus DSS3Φ7 were able to reproduce at 18 °C. Interestingly, the burst sizes of N4-like phages

(DSS3Φ2 and DSS3Φ13) increased dramatically at 18 °C compared to those at 28 °C. DSS3Φ7 has a much smaller burst size compared to the two N4-like phages. The extended latent period of N4-like phages at 18 °C may lead to the larger burst sizes (Table 6.2).

Having different types of bacteriophages infecting the same host strain enables me to compare the persistence and reproduction among different phages. Both persistence and reproduction data seem to support that N4-like phages infecting *R. pomeroyi* DSS-3 are able to proliferate more efficiently at low temperature compared to other non-N4-like phages. Ideally, this experiment should include a wider selection of reference phages, such as T7-like podoviruses or myoviruses. However, these types of phages infecting *R. pomeroyi* DSS-3 are not available. It may be too early to conclude that N4-like roseophages reproduce more effectively than other phage types in the cold environment. It will be important for future studies to include more N4-like roseophages and other reference phages. Another useful study will be to compare the infectivity of various phages in the same mixture under the same condition.

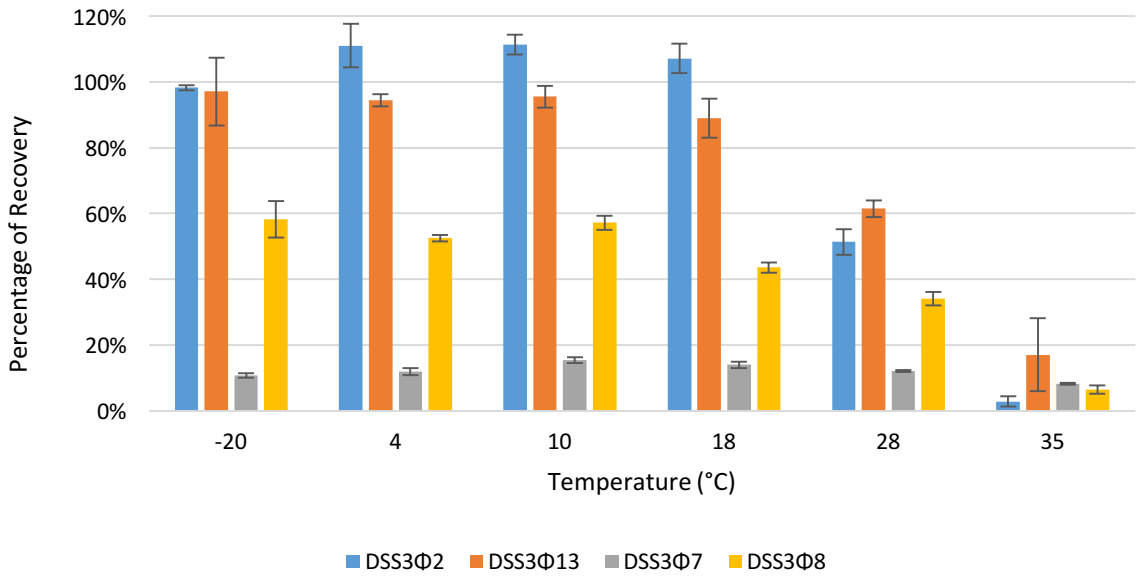


Figure 6.1 The recovery rate of four bacteriophages infecting *R. pomeroyi* DSS-3 after they were incubated at different temperatures for 14 days.

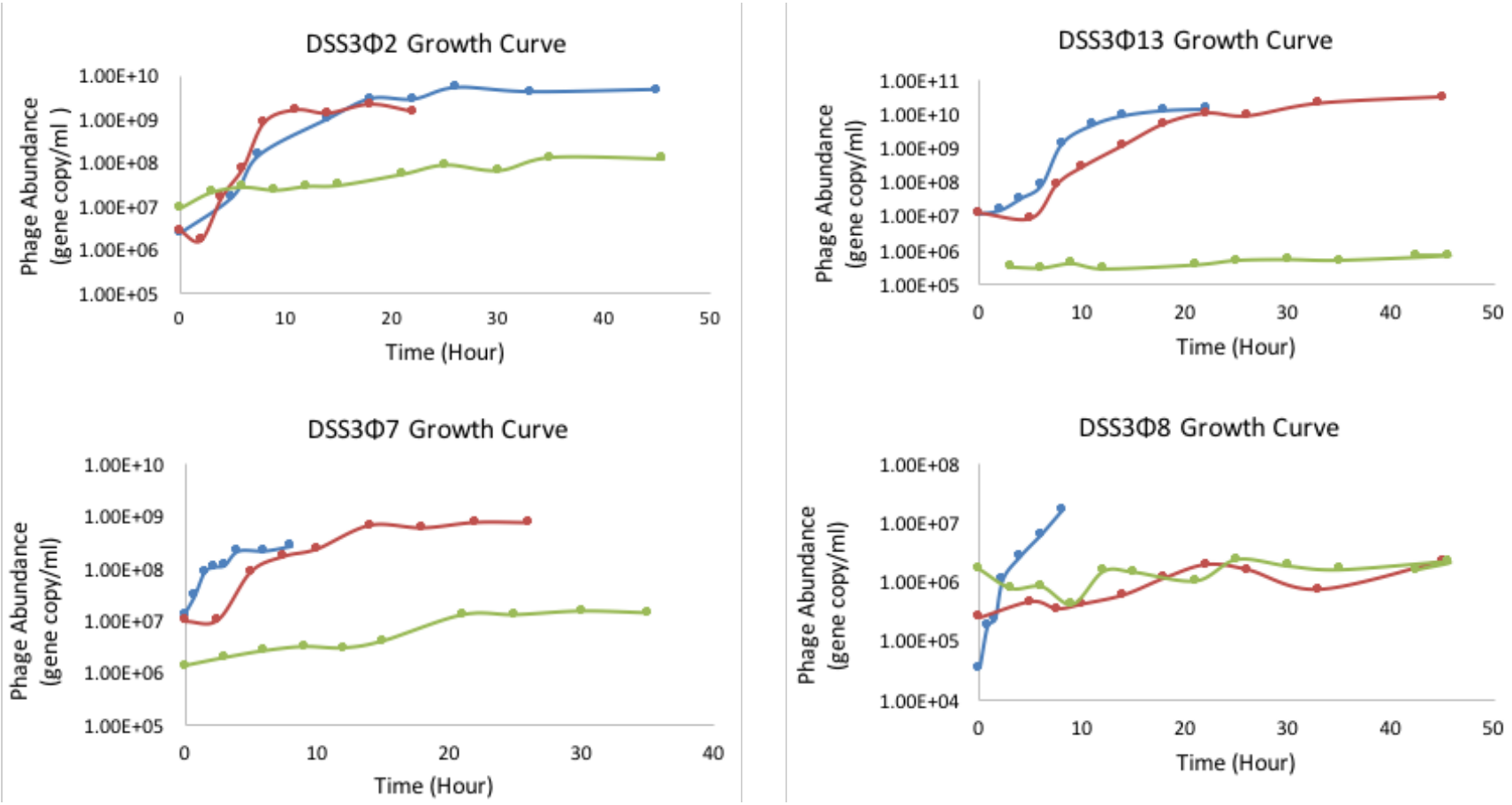


Figure 6.2 Growth curves of four different phages infecting *R. pomeroyi* DSS-3 at different temperature (28 °C blue; 18 °C, red; 10 °C, green).

Table 6.2 The burst sizes and latent period of selected phages growing at different temperature.

	DSS3Φ2		DSS3Φ13		DSS3Φ7		DSS3Φ8	
Temperature*	28 °C	18 °C	28 °C	18 °C	28 °C	18 °C	28 °C	18 °C
Latent period	2 h	5 h	2 h	5 h	< 45 min	2.5 h	< 45 min	> 5 h
Burst size	581	1764	1110	1668	20	75	183	7

*28 °C is the optimal temperature for *R. pomeroyi* DSS-3 growth.

6.3.3 Roseobacters and roseophages interaction

Superinfection of phage DSS3Φ1 (Chapter 4) yielded a phage-resistant strain DSS-3-P1R. Sequencing analysis confirms that phage DSS3Φ1 integrated its genome into DSS3-R1 chromosome, forming a prophage in the resistant strain.

Lysogenic cycle is known to provide mutual benefits to both phage and host. Having a prophage may increase fitness or change the host behavior. For example, prophage can express superinfection exclusion proteins to prevent host cells from being infected by other phages (Sun *et al.*, 2006; Cumby *et al.*, 2012). Different mechanisms are involved in superinfection exclusion, including surface modification, blocking phage genome entry and interfering restriction-modification system (Samson *et al.*, 2013; Bondy-Denomy *et al.*, 2016). When mutant strain DSS-3-P1R was challenged by all the DSS-3 roseophages, it was resistant to almost all of the isolated roseophages. Interestingly, mutant strain DSS-3-P1R is resistant not only to the roseophages similar to DSS3Φ1, but also to the roseophages that are distantly related to DSS3Φ1, such as N4-like roseophages, CbK-like roseophages and even ssDNA phages. Apparently, carrying a prophage in the DSS-3-P1R genome inhibits further phage infection. It is not clear what mechanism lysogenic DSS-3-P1R uses to escape from phage infection. Is this the reason why DSS-3 is more vulnerable to phage infection? Such a defensive response in a bacterial system is interesting and merits further study.

Ironically, phage DSS3Φ16 is able to infect the mutant strain, DSS-3-P1R. The genome sequence of DSS3Φ16 is very similar to that of DSS3Φ1, except for a 2 kb extension on the 5’ end. No protein with known function has been identified in this 2

kb region. How does DSSΦ16 bypass the defense system of DSS-3-P1R and successfully infect both wild-type and mutant strains? Does any gene in the 2 kb region or single nucleotide change on tail protein allow DSS3Φ16 to bypass the resistance of DSS-3-P1R?

6.4 Significance of work

This dissertation research contributes greatly towards knowledge on the phenotypic and genotypic properties of bacteriophages that infect marine roseobacters. Discovery of novel phages is exciting, meanwhile, I also learn that members of the *Roseobacter* lineage likely involve a different phage-host interaction compared to another important and well-studied marine bacterial group - cyanobacteria. For example, lysogenic infection is rare for cyanobacteria, but temperature phages and prophages are common in roseobacters. N4-like podoviruses appear to be common for roseobacters, but no N4-like podoviruses have been reported for marine picocyanobacteria such as *Synechococcus* and *Prochlorococcus*. Instead, T7-like podoviruses and T4-like myoviruses are common for marine picocyanobacteria. Up to date, no T4-like myoviruses have been reported for roseobacters. Cyanobacteria and roseobacters occupy different ecological niches and have different microgeochemical roles. The "kill-the-winner" scheme may not be equally applied to all bacteria at the population level. Phage-host relationship may vary with living habitats (i.e. coastal estuary vs. open ocean), host trophic strategy (i.e. phototrophic vs. heterotrophic) and host genomic flexibility (i.e. reduced genome vs. large genome). My dissertation work further supports that interactions between

bacteria and phages are very complex at the population level. With more phages being isolated, detailed comparison across different virus-host systems is possible.

Bibliography

Abergel C, Legendre M, Claverie JM. (2015). The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol Rev* **39**: 779–796.

Ackermann HW. (1992). Frequency of morphological phage descriptions. *Arch Virol* **124**: 201–209.

Ackermann HW. (2001). Frequency of morphological phage descriptions in the year 2000. *Arch Virol* **146**: 843–857.

Agabian-Keshishian N, Shapiro L. (1970). Stalked bacteria: properties of deoxyribonucleic acid bacteriophage phiCbK. *J Virol* **5**: 795–800.

Ahern SJ, Das M, Bhowmick TS, Young R, Gonzalez CF. (2014). Characterization of novel virulent broad-host-range phages of *Xylella fastidiosa* and *Xanthomonas*. *J Bacteriol* **196**: 459–471.

Allgaier M, Felske A, Wagner-do I. (2003). Aerobic anoxygenic photosynthesis in *Roseobacter* clade bacteria from diverse marine habitats. *Appl Environ Microbiol* **69**: 5051–5059.

Anderson B, Rashid MH, Carter C, Pasternack G, Rajanna C, Revazishvili T, *et al.* (2011). Enumeration of bacteriophage particles. *Bacteriophage* **1**: 86–93.

Angly F, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, *et al.* (2006). The marine viromes of four oceanic regions. *PLoS Biol* **4**: e368.

Angly F, Youle M, Nosrat B, Srinagesh S, Rodriguez-Brito B, McNairnie P, *et al.* (2009). Genomic analysis of multiple roseophage SIO1 strains. *Environ Microbiol* **11**: 2863–2873.

Ankrah NYD, Budinoff CR, Wilson WH, Wilhelm SW, Buchan A. (2014a). Genome sequence of the *Sulfitobacter sp.* strain 2047 infecting lytic phage ΦCB2047-B. *Genome Annoucements* **2**: e00945-13.

Ankrah NYD, Budinoff CR, Wilson WH, Wilhelm SW, Buchan A. (2014b). Genome sequences of two temperate phages, ΦCB2047-A and ΦCB2047-C, infecting *Sulfitobacter sp.* strain 2047. *Genome Annoucements* **2**: e00108-14.

Ankrah NYD, May AL, Middleton JL, Jones DR, Hadden MK, Gooding JR, *et al.* (2014c). Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J* **8**: 1089–1100.

Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M. (2011). Distant Mimivirus relative with a larger genome highlights the fundamental features of *Megaviridae*. *Proc Natl Acad Sci USA* **108**: 17486–17491.

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

Bahir I, Fromer M, Prat Y, Linial M. (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol* **5**: 311.

Bailly-Bechet M, Vergassola M, Rocha E. (2007). Causes for the intriguing presence of tRNAs in phages. *Genome Res* **17**: 1486–1495.

Beja O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, *et al.* (2002). Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**: 630–633.

Bergh Ø, Børsheim KY, Bratbak G, Heldal M. (1989). High abundance of viruses found in aquatic environments. *Nature* **340**: 467–468.

Besemer J, Lomsadze A, Borodovsky M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* **29**: 2607–2618.

Biers EJ, Wang K, Pennington C, Belas R, Chen F, Moran MA. (2008). Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl Environ Microbiol* **74**: 2933–2939.

Bochtler M, Odintsov SG, Marcyjaniak M, Sabala I. (2004). Similar active sites in lysostaphins and D-Ala-D-Ala metallopeptidases. *Protein Sci* **13**: 854–861.

Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, *et al.* (2016). Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* **10**: 2854–2866.

Born Y, Fieseler L, Marazzi J, Lurz R, Duffy B, Loessner MJ. (2011). Novel virulent and broad-host-range *Erwinia amylovora* bacteriophages reveal a high degree of mosaicism and a relationship to *Enterobacteriaceae* phages. *Appl Environ Microbiol* **77**: 5945–5954.

Bowes JM, Dowell CE. (1974). Purification and some properties of bacteriophage ST-1. *J Virol* **13**: 53–61.

Breitbart M. (2012). Marine viruses: truth or dare. *Ann Rev Mar Sci* **4**: 425–448.

Brenner S, Horne RW. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochim Biophys Acta* **34**: 103–110.

Brentlinger KL, Hafenstein S, Novak CR, Fane BA, Borgon R, McKenna R, *et al.* (2002). *Microviridae*, a family divided: isolation, characterization, and genome sequence of MH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J Bacteriol* **184**: 1089–1094.

Brinkhoff T, Giebel HA, Simon M. (2008). Diversity, ecology, and genomics of the *Roseobacter* clade: a short overview. *Arch Microbiol* **189**: 531–539.

Brum JR, Ignacio-espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, *et al.* (2015). Patterns and ecological drivers of ocean viral communities. *Science* **348**: 1261498.

Bryson SJ, Thurber AR, Correa AMS, Orphan VJ, Vega Thurber R. (2015). A novel sister clade to the enterobacteria microviruses (family *Microviridae*) identified in methane seep sediments. *Environ Microbiol* **17**: 3708–3721.

Buchan A, Gonzalez JM, Moran MA. (2005). Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol* **71**: 5665–5677.

Buchan A, Lecleir GR, Gulvik CA, González JM. (2014). Master recyclers : features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12**: 686–698.

Cai L, Yang Y, Jiao N, Zhang R. (2015). Complete genome sequence of vB_DshP-R2C, a N4-like lytic roseophage. *Mar Genomics* **22**: 15–17.

Campbell A. (2003). Prophage insertion sites. *Res Microbiol* **154**: 277–282.

Carlton JG, Martin-Serrano J. (2007). Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. *Science* **316**: 1908–1912.

Ceyssens PJ, Glonti T, Kropinski NM, Lavigne R, Chanishvili N, Kulakov L, *et al.* (2011). Phenotypic and genotypic variations within a single bacteriophage species. *Virol J* **8**: 134.

Chan JZM, Millard AD, Mann NH, Schäfer H. (2014). Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like roseophage specific genes. *Front Microbiol* **5**: 506.

Chen F, Lu J. (2002). Genomic sequence and evolution of marine cyanophage P60 : a new linsight on lytic and lysogenic phages. *Appl Environ Microbiol* **68**: 2589–2594.

Chen F, Wang K, Huang S, Cai H, Zhao M, Jiao N, *et al.* (2009). Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. *Environ Microbiol* **11**: 2884–2892.

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.

Cherwa JE, Fane BA. (2011). *Microviridae* : Microviruses and Gokushoviruses. *eLs* John Wiley & Sons, Ltd: Chichester. 1–10.

Chipman PR, Agbandje-McKenna M, Renaudin J, Baker TS, McKenna R. (1998). Structural analysis of the spiroplasma virus, SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. *Structure* **6**: 135–145.

Chithambaram S, Prabhakaran R, Xia X. (2014a). Differential codon adaptation between dsDNA and ssDNA phages in *Escherichia coli*. *Mol Biol Evol* **31**: 1606–1617.

Chithambaram S, Prabhakaran R, Xia X. (2014b). The effect of mutation and selection on codon adaptation in *Escherichia coli* bacteriophage. *Genetics* **197**: 301–315.

Choi KH, Mcpartland J, Kaganman I, Bowman VD, Lucia B, Rossmann MG. (2008). Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J Mol Biol* **378**: 726–736.

Collins AJ, Fullmer MS, Gogarten JP, Nyholm S V. (2015). Comparative genomics of *Roseobacter* clade bacteria isolated from the accessory nidamental gland of *Euprymna scolopes*. *Front Microbiol* **6**: 123.

Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JBH. (2016). The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **499**: 219–229.

Cumby N, Edwards AM, Davidson AR, Maxwell KL. (2012). The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J Bacteriol* **194**: 5012–5019.

Cunliffe M. (2011). Correlating carbon monoxide oxidation with *cox* genes in the abundant marine Roseobacter clade. *ISME J* **5**: 685–691.

Davydova EK, Kaganman I, Kazmierczak KM, Rothman-Denes LB. (2009). Identification of bacteriophage N4 virion RNA polymerase - nucleic acid interactions in transcription complexes. *J Biol Chem* **284**: 1962–1970.

Dellas N, Snyder JC, Bolduc B, Young MJ. (2014). Archaeal viruses: diversity, replication, and structure. *Annu Rev Virol* **1**: 399–426.

Doron S, Fedida A, Hernández-Prieto MA, Sabehi G, Karunker I, Stazic D, *et al.* (2016). Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* **10**: 1437–1455.

Edelman DC, Barletta J. (2003). Real-time PCR provides improved detection and titer determination of bacteriophage. *Biotechniques* **35**: 368–375.

Edlin G, Lin L, Kudrna R. (1975). λ lysogens of *E. coli* reproduce more rapidly than non-lysogens. *Nature* **255**: 735–737.

Edwards RA, Rohwer F. (2005). Viral metagenomics. *Nat Rev Microbiol* **3**: 504–510.

Falco SC, Zivin R, Rothman-Denes LB. (1978). Novel template requirements of N4 virion RNA polymerase. *Proc Natl Acad Sci USA* **75**: 3220–3224.

Fan H, Fan H, An X, Huang Y, Zhang Z, Mi Z, *et al.* (2012). Complete genome sequence of IME11, a new N4-like bacteriophage. *J Virol* **86**: 13861.

Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol* **13**: 641–650.

Finsterbusch T, Mankertz A. (2009). Porcine circoviruses-small but powerful. *Virus Res* **143**: 177–183.

Fischer MG, Hackl T. (2016). Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature* **540**: 288–291.

Fouts DE, Klumpp J, Bishop-Lilly KA, Rajavel M, Willner KM, Butani A, *et al.* (2013). Whole genome sequencing and comparative genomic analyses of two *Vibrio cholerae* O139 Bengal-specific *Podoviruses* to other N4-like phages reveal extensive genetic diversity. *Virol J* **10**: 165.

Franzmann PD, Deperez PP, Burton HR, Van Den Hoff J. (1987). Limnolgy of Organic Lake, Antarctica, a meromictic lake that contains high concentrations of dimethyl sufphide. *Aust J Sci Res* **38**: 409–417.

Fuhrman JA. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541–548.

Gan HM, Sieo CC, Tang SGH, Omar AR, Ho YW. (2013). The complete genome sequence of EC1-UPM, a novel N4-like bacteriophage that infects *Escherichia coli* O78:K80. *Virol J* **10**: 308.

Garner SA, Everson JS, Lambden PR, Fane BA, Clarke IN. (2004). Isolation, molecular characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydophila pecorum*. *Virus Genes* **28**: 207–214.

Geng H, Bruhn JB, Nielsen KF, Gram L, Belas R. (2008). Genetic dissection of tropodithietic acid biosynthesis by marine Roseobacters. *Appl Environ Microbiol* **74**: 1535–1545.

Gifford SM, Sharma S, Moran MA. (2014). Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community. *Front Microbiol* **5**: 185.

Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez D, *et al.* (2012). The *Caulobacter crescentus* phage phiCbK: genomics of a canonical phage. *BMC Genomics* **13**: 542.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60–63.

Godson GN, Barrell BG, Staden R, Fiddes JC. (1978). Nucleotide sequence of bacteriophage G4 DNA. *Nature* **276**: 236–247.

Gonzalez JM, Moran MA. (1997). Numerical dominance of a group of marine bacteria in the alpha-subclass of the class *Proteobacteria* in coastal seawater. *Appl Environ Microbiol* **63**: 4237–4242.

Gulvik CA, Buchan A. (2013). Simultaneous catabolism of plant-derived aromatic compounds results in enhanced growth for members of the *Roseobacte*r lineage. *Appl Environ Microbiol* **79**: 3716–3723.

Guy L, Kultima JR, Andersson SGE, Quackenbush J. (2010). genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**: 2334–2335.

Häring M, Vestergaard G, Rachel R, Chen L, Garrett RA, Prangishvili D. (2005). Independent virus development outside a host. *Nature* **436**: 1101.

Hatfull GF. (2010). Mycobacteriophages: genes and genomes. *Annu Rev Microbiol* **64**: 331–356.

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.

Holmfeldt K, Odic D, Sullivan MB, Middelboe M, Riemann L. (2012). Cultivated single-stranded DNA phages that infect marine *Bacteroidetes* prove difficult to detect with DNA-binding stains. *Appl Environ Microbiol* **78**: 892–894.

Holmfeldt K, Solonenko N, Shah M, Corrier K, Riemann L, VerBerkmoes NC, *et al.* (2013). Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci USA* **110**: 12798–12803.

Hopkins M, Kailasan S, Cohen A, Roux S, Tucker KP, Shevenell A, *et al.* (2014). Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. *ISME J* **8**: 2093–2103.

Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J* **11***:* 1511–1520.

Howard EC, Sun S, Biers EJ, Moran MA. (2008). Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ Microbiol* **10**: 2397–2410.

Hsia RC, Ting LM, Bavoil PM. (2000). Microvirus of *Chlamydia psittaci* strain Guinea

pig inclusion conjunctivitis: isolation and molecular characterization. *Microbiology* **146**: 1651–1660.

Huang S, Wang K, Jiao N, Chen F. (2012). Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environ Microbiol* **14**: 540–558.

Huang S, Zhang S, Jiao N, Chen F. (2015). Comparative genomic and phylogenomic analyses reveal a conserved core genome shared by estuarine and oceanic cyanopodoviruses. *PLoS One* **10**: 1–17.

Huang S, Zhang Y, Chen F, Jiao N. (2011). Complete genome sequence of a marine roseophage provides evidence into the evolution of gene transfer agents in alphaproteobacteria. *Virol J* **8**: 124.

Huber HE, Tabor S, Richardson CC. (1987). *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates. *J Biol Chem* **262**: 16224–16232.

Hurwitz BL, Sullivan MB. (2013). The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* **8**: e57355.

Hyman P, Abedon ST. (2009). Practical methods for determining phage growth parameters. *Bacteriophages: Methods and Protocols*. Humana Press: Totowa, NJ, pp 175–202.

Jazwinski S, Lindberg A, Kornberg A. (1975). The gene H spike protein of bacteriophages phiX174 and S13. I. Functions in phage-receptor recognition and in transfection. *Virology* **66**: 283–293.

Ji J, Zhang R, Jiao N. (2015). Complete genome sequence of Roseophage vB _ DshP-R1, which infects *Dinoroseobacter shibae* DFL12. *Stand Genomic Sci* **10**: 6.

Jiang SC, Paul JH. (1996). Occurrence of lysogenic bacteria in marine microbial communities as determined by prophage induction. *Mar Ecol Prog Ser* **142**: 27–38.

Jiang SC, Paul JH. (1994). Seasonal and diel abundance of viruses and occurrence of lysogeny / bacteriocinogeny in the marine environment. *Mar Ecol Prog Ser* **104**: 163–172.

Jiang SC, Paul JH. (1998). Significance of lysogeny in the marine environment: studies with isolates and a model of lysogenic phage production. *Microb Ecol* **35**: 235–243.

Jonkers HM, Abed RMM. (2003). Identification of aerobic heterotrophic bacteria from the photic zone of a hypersaline microbial mat. *Aquat Microb Ecol* **30**: 127–133.

Kan J, Crump BC, Wang K, Chen F. (2006). Bacterioplankton community in Chesapeake Bay: predictable or random assemblages. *Limnol Oceanogr* **51**: 2157–2169.

Kan J, Suzuki MT, Wang K, Evans SE, Chen F. (2007). High temporal but low spatial heterogeneity of bacterioplankton in the Chesapeake Bay. *Appl Environ Microbiol* **73**: 6776–6789.

Kang I, Jang H, Oh H-M, Cho J-C. (2012). Complete genome sequence of *Celeribacter* bacteriophage P12053L. *J Virol* **86**: 8339–8340.

Kang I, Oh H-M, Kang D, Cho J-C. (2013). Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc Natl Acad Sci USA* **110**: 12343–12348.

Kazaks A, Dislers A, Lipowsky G, Nikolajeva V, Tars K. (2012). Complete genome sequence of the *Enterobacter cancerogenus* bacteriophage Enc34. *J Virol* **86**: 11403–11404.

Kim KH, Bae JW. (2011). Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl Environ Microbiol* **77**: 7663–7668.

Kimura K, Tomaru Y. (2015). Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. *Appl Environ Microbiol* **81**: 1120–1131.

Kimura K, Tomaru Y. (2013). Isolation and characterization of a single-stranded DNA virus infecting the marine diatom *Chaetoceros sp.* strain SS628-11 isolated from western Japan. *PLoS One* **8**: e82013.

Kodaira K, Nakano K, Okada S, Taketo A. (1992). Nucleotide sequence of the genome of the bacteriophage alpha 3: interrelationship of the genome structure and the gene products with those of the phages, phi X174, G4 and phi K. *Biochim Biophys Acta* **1130**: 277–88.

Kodaira K-I, Oki M, Kakikawa M, Kimoto H, Taketo A. (1996). The virion proteins encoded by bacteriophage phiK and its host-range mutant phi KhT: host-range determination and DNA binding properties. *J Biochem* **119**: 1062–1069.

Krakauer DC, Jansen VAA. (2002). Red queen dynamics of protein translation. *J Theor Biol* **218**: 97–109.

Krupovic M. (2013). Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Curr Opin Virol* **3**: 578–586.

Krupovic M, Forterre P. (2011). *Microviridae* goes temperate: microvirus-related proviruses reside in the genomes of *Bacteroidetes*. *PLoS One* **6**: e19893.

Kulikov E, Kropinski AM, Golomidova A, Lingohr E, Govorun V, Serebryakova M, *et al.* (2012). Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB_EcoP_G7C. *Virology* **426**: 93–99.

Kumar S, Stecher G, Tamura K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**: 1870–1874.

Laass S, Kleist S, Bill N, Drüppel K, Kossmehl S, Wöhlbrand L, *et al.* (2014). Gene regulatory and metabolic adaptation processes of *Dinoroseobacter shibae* DFL12T during oxygen depletion. *J Biol Chem* **289**: 13219–13231.

Labonté JM, Reid KE, Suttle CA. (2009). Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Appl Environ Microbiol* **75**: 3634–3640.

Labonté JM, Suttle CA. (2013a). Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. *Front Microbiol* **4**: 404.

Labonté JM, Suttle CA. (2013b). Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J* **7**: 2169–2177.

Labrie SJ, Frois-Moniz K, Osburne MS, Kelly L, Roggensack SE, Sullivan MB, *et al.* (2013). Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ Microbiol* **15**: 1356–1376.

Labrie SJ, Dupuis MÈ, Tremblay DM, Plante PL, Corbeil J, Moineaua S. (2014). A new *Microviridae* phage isolated from a failed biotechnological process driven by *Escherichia coli* . *Appl Environ Microbiol* **80**: 6992–7000.

Lang AS, Beatty JT. (1999). Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci USA* **97**: 859–864.

Lang AS, Beatty JT. (2001). The gene transfer agent of *Rhodobacter capsulatus* and 'constitutive transduction' in prokaryotes. *Arch Microbiol* **175**: 241–249.

Lang AS, Zhaxybayeva O, Beatty JT. (2012). Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Microbiol* **10**: 472–482.

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Lee JH, Shin H, Choi Y, Ryu S. (2013). Complete genome sequence analysis of bacterial-flagellum-targeting bacteriophage chi. *Arch Virol* **158**: 2179–2183.

Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, *et al.* (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a

pandoravirus morphology. *Proc Natl Acad Sci USA* **111**: 4274–4279.

Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, *et al.* (2015). In-depth study of *Mollivirus sibericum*, a new 30,000-year-old giant virus infecting *Acanthamoeba*. *Proc Natl Acad Sci USA* **112**: E5327–E5335.

Lenk S, Moraru C, Hahnke S, Arnds J, Richter M, Kube M, *et al.* (2012). *Roseobacter* clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. *ISME J* **6**: 2178–2187.

Lenneman B, Rothman-Denes L. (2015). Structural and biochemical investigation of bacteriophage N4-encoded RNA polymerases. *Biomolecules* **5**: 647–667.

Li B, Zhang S, Long L, Huang S. (2016). Characterization and complete genome sequences of three N4-like Roseobacter phages isolated from the South China Sea. *Curr Microbiol* **73**: 409–418.

Liang Y, Zhang Y, Zhou C, Chen Z, Yang S, Yan C, *et al.* (2016). Complete genome sequence of the siphovirus Roseophage RDJLΦ2 infecting *Roseobacter denitrificans* OCh114. *Mar Genomics* **25**: 17–19.

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

Liu BL, Everson JS, Fane B, Giannikopoulou P, Vretou E, Lambden PR, *et al.* (2000). Molecular characterization of a bacteriophage (Chp2) from *Chlamydia psittaci*. *J Virol* **74**: 3464–3469.

Loesser MJ, Wendlinger G, Scherer S. (1995). Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes. *Mol Microbiol* **16**: 1231–1241.

Lohr JE, Chen F, Hill RT. (2005). Genomic analysis of bacteriophage ΦJL001: insights into its interaction with a sponge-associated alpha-proteobacterium. *Appl Environ Microbiol* **71**: 1598–1609.

Lowe TM, Eddy SR. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.

Lu J, Chen F, Hodson RE. (2001). Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus spp.* in river estuaries. *Appl Environ Microbiol* **67**: 3285–3290.

Luo H, Löytynoja A, Moran MA. (2012). Genome content of uncultivated marine *Roseobacters* in the surface ocean. *Environ Microbiol* **14**: 41–51.

Luo H, Moran MA. (2014). Evolutionary ecology of the marine *Roseobacter* clade.

Microbiol Mol Biol Rev **78**: 573–587.

Luo H, Moran MA. (2015). How do divergent ecological strategies emerge among marine bacterioplankton lineages ? *Trends Microbiol* **23**: 577–584.

Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA. (2014). Evolutionary analysis of a streamlined lineage of surface ocean Roseobacters. *ISME J* **8**: 1428–1439.

Lynch KH, Stothard P, Dennis JJ. (2012). Comparative analysis of two phenotypically-similar but genomically-distinct *Burkholderia cenocepacia*-specific bacteriophages. *BMC Genomics* **13**: 223.

Männistö RH, Kivelä HM, Paulin L, Bamford DH, Bamford JKH. (1999). The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. *Virology* **262**: 355–363.

Marine R, McCarren C, Vorrasane V, Nasko D, Crowgey E, Polson SW, *et al.* (2014). Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* **2**: 3.

Mark DF, Richardson CC. (1976). *Escherichia coli* thioredoxin: a subunit of bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA* **73**: 780–784.

McDaniel L, Breitbart M, Mobberley J, Long A, Haynes M, Rohwer F, *et al.* (2008). Metagenomic analysis of lysogeny in Tampa Bay: implications for prophage gene expression. *PLoS One* **3**: e3263.

McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. (2010). High frequency of horizontal gene transfer in the oceans. *Science* **330**: 50.

Millard A, Clokie MRJ, Shub DA, Mann NH. (2004). Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proc Natl Acad Sci USA* **101**: 11007–11012.

Mizuno CM, Rodriguez-Valera F, Garcia-Heredia I, Martin-Cuadrado AB, Ghai R. (2013a). Reconstruction of novel cyanobacterial siphovirus genomes from Mediterranean metagenomic fosmids. *Appl Environ Microbiol* **79**: 688–695.

Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. (2013b). Expanding the marine virosphere using metagenomics. *PLoS Genet* **9**: e1003987.

Moran MA, Belas R, Schell MA, González JM, Sun F, Sun S, *et al.* (2007). Ecological genomics of marine roseobacters. *Appl Environ Microbiol* **73**: 4559–4569.

Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, *et al.* (2004). Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the

marine environment. *Nature* **432**: 910–913.

Moran MA, González JM, Kiene RP. (2003). Linking a bacterial taxon to sulfur cycling in the sea : studies of the marine *Roseobacter* group. *Geomicrobilogy J* **20**: 375–388.

Moran MA, Reisch CR, Kiene RP, Whitman WB. (2012). Genomic insights into bacterial DMSP transformations. *Ann Rev Mar Sci* **4**: 523–542.

Moreno Switt AI, Orsi RH, den Bakker HC, Vongkamjan K, Altier C, Wiedmann M. (2013). Genomic characterization provides new insight into *Salmonella* phage diversity. *BMC Genomics* **14**: 481.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, *et al.* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.

Murakami KS, Davydova EK, Rothman-Denes LB. (2008). X-ray crystal structure of the polymerase domain of the bacteriophage N4 virion RNA polymerase. *Proc Natl Acad Sci USA* **105**: 5046–5051.

Nasir A, Forterre P, Kim KM, Caetano-Anollés G. (2014). The distribution and impact of viral lineages in domains of life. *Front Microbiol* **5**: 1194.

Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, *et al.* (2010). Genome characteristics of a generalist marine bacterial lineage. *ISME J* **4**: 784–798.

Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, *et al.* (2017). Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* **2**: e00359-16.

Oliveira H, Melo LDR, Santos SB, Nobrega FL, Ferreira EC, Cerca N, *et al.* (2013). Molecular aspects and comparative genomics of bacteriophage endolysins. *J Virol* **87**: 4558–4570.

Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, *et al.* (2016). Uncovering Earth's virome. *Nature* **536**: 425–430.

Patzelt D, Wang H, Buchholz I, Rohde M, Gröbe L, Pradella S, *et al.* (2013). You are what you talk: quorum sensing induces individual morphologies and cell division modes in *Dinoroseobacter shibae*. *ISME J* **7**: 2274–2286.

Paul JH. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2**: 579–589.

Perez Sepulveda B, Redgwell T, Rihtman B, Pitt F, Scanlan DJ, Millard A. (2016). Marine phage genomics: the tip of the iceberg. *FEMS Microbiol Lett* **363**: 1–8.

Philippe N, Legendre M, Doutre G, Poirot O, Lescot M, Arslan D, *et al.* (2013). Pandoraviruses: amoeba viruses with genomes Up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **341**: 281–286.

Pietilä MK, Demina TA, Atanasova NS, Oksanen HM, Bamford DH. (2014). Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends Microbiol* **22**: 334–344.

Pope WH, Weigele PR, Chang J, Pedulla ML, Ford ME, Houtz JM, *et al.* (2007). Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a 'horned' bacteriophage of marine *Synechococcus*. *J Mol Biol* **368**: 966–981.

Prangishvili D. (2013). The wonderful world of archaeal viruses. *Annu Rev Microbiol* **67**: 565–585.

Quaiser A, Dufresne A, Ballaud F, Roux S, Zivanovic Y, Colombet J, *et al.* (2015). Diversity and comparative genomics of *Microviridae* in *Sphagnum*- dominated peatlands. *Front Microbiol* **6**: 375.

Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, *et al.* (2000a). Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res* **28**: 1397–1406.

Read TD, Fraser CM, Hsia RC, Bavoil PM. (2000b). Comparative analysis of *Chlamydia* bacteriophages reveals variation localized to a putative receptor binding domain. *Microb Comp Genomics* **5**: 223–231.

Renaudin J, Pascarel MC, Bové JM. (1987). Spiroplasma virus 4: nucleotide sequence of the viral DNA, regulatory signals, and proposed genome organization. *J Bacteriol* **169**: 4950–4961.

Richter M, Rosselló-Móra R, Glöckner FO, Peplies J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32**: 929–931.

Rivers AR, Smith CB, Moran MA. (2014). An updated genome annotation for the model marine bacterium *Ruegeria pomeroyi* DSS-3. *Stand Genomic Sci* **9**: 1–9.

Rohwer F, Segall A, Steward G, Seguritan V, Breitbart M, Wolven F, *et al.* (2000). The complete genomic sequence of the marine phage roseophage SIO1 shares homology with nonmarine phages. *Limnol Oceanogr* **45**: 408–418.

Rokyta DR, Burch CL, Caudle SB, Wichman HA. (2006). Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J Bacteriol* **188**: 1134–1142.

Rosario K, Breitbart M. (2011). Exploring the viral world through metagenomics. *Curr Opin Virol* **1**: 289–297.

Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S, *et al.* (2012a). Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS One* **7**: e33641.

Roux S, Krupovic M, Poulet A, Debroas D, Enault F. (2012b). Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. *PLoS One* **7**: e40418.

Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, *et al.* (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **4**: e2777.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, *et al.* (2007). The *Sorcerer II* Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: e77.

Sabehi G, Shaulov L, Silver DH, Yanai I, Harel A, Lindell D. (2012). A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proc Natl Acad Sci USA* **109**: 2037–2042.

Sait M, Livingstone M, Graham R, Inglis NF, Wheelhouse N, Longbottom D. (2011). Identification, sequencing and molecular analysis of Chp4, a novel chlamydiaphage of *Chlamydophila abortus* belonging to the family *Microviridae*. *J Gen Virol* **92**: 1733–1737.

Samson JE, Magadán AH, Sabri M, Moineau S. (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**: 675–687.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, *et al.* (1977). Nucleotide sequence of bacteriophage ΦX174 DNA. *Nature* **265**: 687–695.

Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. (2010). Viral mutation rates. *J Virol* **84**: 9733–9748.

Sass H, Köpke B, Rütters H, Feuerlein T, Dröge S, Cypionka H, *et al.* (2010). *Tateyamaria pelophila* sp. nov., a facultatively anaerobic alphaproteobacterium isolated from tidal-flat sediment, and emended descriptions of the genus *Tateyamaria* and of *Tateyamaria omphalii*. *Int J Syst Evol Microbiol* **60**: 1770–1777.

Schito GC. (1974). Development of coliphage N4: utrastructural studies. *J Virol* **13**: 186–196.

Schito GC. (1973). The genetics and physiology of coliphage N4. *Virology* **55**: 254–265.

Schito GC, Rialdi G, Pesce A. (1966). Biophysical properties of N4 coliphage. *Biochim Biophys Acta* **129**: 482–490.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Schwartz DA, Lindell D. (2017). Genetic hurdles limit the arms race between *Prochlorococcus* and the T7-like podoviruses infecting them. *ISME J* 1–16 (doi: 10.1038/ismej.2017.47).

Scola B La, Audic S, Robert C, Jungang L, Lamballerie X, Drancourt M, *et al.* (2003). A giant virus in amoebae. *Science* **299**: 2033–2033.

La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, *et al.* (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* **455**: 100–104.

Selje N, Simon M, Brinkhoff T. (2004). A newly discovered *Roseobacter* cluster in temperate and polar oceans. *Nature* **427**: 445–448.

Sime-Ngando T. (2014). Environmental bacteriophages: viruses of microbes in aquatic ecosystems. *Front Microbiol* **5**: 1–14.

Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, *et al.* (2017). Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. *ISME J* **11**: 1483–1499.

Snyder JC, Bolduc B, Young MJ. (2015). 40 Years of archaeal virology: expanding viral diversity. *Virology* **479–480**: 369–378.

Snyder JC, Samson RY, Brumfield SK, Bell SD, Young MJ. (2013). Functional interplay between a virus and the ESCRT machinery in archaea. *Proc Natl Acad Sci USA* **110**: 10783–10787.

Sonnenschein EC, Nielsen KF, D 'Alvise P, Porsby CH, Melchiorsen J, Heilmann J, *et al.* (2017). Global occurrence and heterogeneity of the *Roseobacter*-clade species *Ruegeria mobilis*. *ISME J* **11**: 569–583.

Steward GF, Culley AI. (2010). Extraction and purification of nucleic acids from viruses. *Man Aquat Viral Ecol* 154–165.

Stojković EA, Rothman-Denes LB. (2007). Coliphage N4 *N*-Acetylmuramidase defines a new family of murein hydrolases. *J Mol Biol* **366**: 406–419.

Stopar D, Cerne A, Zigman M, Poljsak-Prijatelj M, Turk V. (2003). Viral abundance and a high proportion of lysogens suggest that viruses are important members of the microbial community in the Gulf of Trieste. *Microb Ecol* **45**: 249–256.

Storey CC, Lusher M, Richmond SJ. (1989). Analysis of the complete nucleotide sequence of Chp1, a phage which infects avian *Chlamydia psittaci*. *J Gen Virol* **70**:

3381–3390.

Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, *et al.* (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* **12**: 3035–3056.

Sullivan MB, Krastins B, Hughes JL, Kelly L, Chase M, Sarracino D, *et al.* (2009). The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ Microbiol* **11**: 2935–2951.

Sullivan MB, Waterbury JB, Chisholm SW. (2003). Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**: 1047–1051.

Sun S, Chen J, Li W, Altintas I, Lin A, Peltier S, *et al.* (2011). Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Res* **39**: 546–551.

Sun X, Göhler A, Heller KJ, Neve H. (2006). The *ltp* gene of temperate *Streptococcus thermophilus* phage TP-J34 confers superinfection exclusion to *Streptococcus thermophilus* and *Lactococcus lactis*. *Virology* **350**: 146–157.

Sun Y, Powell KE, Sung W, Lynch M, Moran MA. (2017). Spontaneous mutations of a model heterotrophic marine bacterium. *ISME J* **11**: 1713-1718.

Suttle CA. (2007). Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.

Suttle CA. (2005). Viruses in the sea. *Nature* **437**: 356–361.

Suttle CA, Chan AM. (1993). Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar Ecol Prog Ser* **92**: 99–109.

Székely AJ, Breitbart M. (2016). Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiol Lett* **363**: fnw027.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, *et al.* (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon

metabolism. *Proc Natl Acad Sci USA* **108**: E757–E764.

Todd JD, Rogers R, Li YG, Wexler M, Bond PL, Sun L, *et al.* (2007). Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* **315**: 666–669.

Tucker KP, Parsons R, Symonds EM, Breitbart M. (2010). Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J* **5**: 822–830.

Varaljay VA, Robidart J, Preston CM, Gifford SM, Durham BP, Burns AS, *et al.* (2015). Single-taxon field measurements of bacterial gene regulation controlling DMSP fate. *ISME J* **9**: 1677–1686.

Vila-Costa M, Rinta-Kanto JM, Poretsky RS, Shulei S, Kiene RP, Moran MA. (2014). Microbial controls on DMSP degradation and DMS formation in the Sargasso Sea. *Biogeochemistry* **120**: 295–305.

Voget S, Wemheuer B, Brinkhoff T, Vollmers J, Dietrich S, Giebel H-A, *et al.* (2015). Adaptation of an abundant *Roseobacter* RCA organism to pelagic systems revealed by genomic and transcriptomic analyses. *ISME J* **9**: 371–384.

Wagner-Döbler I, Biebl H. (2006). Environmental biology of the marine *Roseobacter* lineage. *Annu Rev Microbiol* **60**: 255–280.

Waldor MK, Mekalanos JJ. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**: 1910–1914.

Wang K, Chen F. (2004). Genetic diversity and population dynamics of cyanophage communities in the Chesapeake Bay. *Aquat Microb Ecol* **34**: 105–116.

Wang K, Chen F. (2008). Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* **10**: 300–312.

Waterbury JB, Valois FW. (1993). Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* **59**: 3393–3399.

Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, *et al.* (2007). Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* **9**: 1675–1695.

Weinbauer MG, Suttle CA. (1999). Lysogeny and prophage induction in coastal and offshore bacterial communities. *Aquat Microb Ecol* **18**: 217–225.

Wemheuer B, Wemheuer F, Hollensteiner J, Meyer FD, Voget S, Daniel R. (2015). The green impact: bacterioplankton response towards a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches. *Front Microbiol* **6**: 1–13.

Williams KP, Sobral BW, Dickerman AW. (2007). A robust species tree for the *Alphaproteobacteria*. *J Bacteriol* **189**: 4578–4586.

Wilson WH, Joint IR, Carr NG, Mann NH. (1993). Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus sp.* strain WH7803. *Appl Environ Microbiol* **59**: 3736–3743.

Wittmann J, Dreiseikelmann B, Rohde M, Meier-Kolthoff JP, Bunk B, Rohde C. (2014). First genome sequences of *Achromobacter* phages reveal new members of the N4 family. *Virol J* **11**: 14.

Wommack KE, Colwell RR. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69–114.

Yau S, Lauro FM, Williams TJ, DeMaere MZ, Brown MV, Rich J, *et al.* (2013). Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *ISME J* **7**: 1944–1961.

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, *et al.* (2007). The *Sorcerer II* global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol* **5**: e16.

Yoshida M, Takaki Y, Eitoku M, Nunoura T, Takai K. (2013). Metagenomic analysis of viral communities in (hado) pelagic sediments. *PLoS One* **8**: e57271.

Yu Z, Chen X, Shen Q, Zhao D, Tang B, Su H, *et al.* (2015). Filamentous phages prevalent in *Pseudoalteromonas spp.* confer properties advantageous to host survival in Arctic sea ice. *ISME J* **9**: 871–881.

Zan J, Cicirelli EM, Mohamed NM, Sibhatu H, Kroll S, Choi O, *et al.* (2012). A complex LuxR-LuxI type quorum sensing network in a roseobacterial marine sponge symbiont activates flagellar motility and inhibits biofilm formation. *Mol Microbiol* **85**: 916–933.

Zeng Z, Liu X, Yao J, Guo Y, Li B, Li Y, *et al.* (2016). Cold adaptation regulated by cryptic prophage excision in *Shewanella oneidensis*. *ISME J* **10**: 2787–2800.

Zhan Y, Buchan A, Chen F. (2015). Novel N4 bacteriophages prevail in the cold biosphere. *Appl Environ Microbiol* **81**: 5196–5202.

Zhan Y, Huang S, Voget S, Simon M, Chen F. (2016). A novel roseobacter phage possesses features of podoviruses, siphoviruses, prophages and gene transfer agents. *Sci Rep* **6**: 30372.

Zhang Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40.

Zhang Y, Jiao N. (2009). Roseophage RDJLΦI, infecting the aerobic anoxygenic

phototrophic bacterium *Roseobacter denitrificans* OCh114. *Appl Environ Microbiol* **75**: 1745–1749.

Zhang Y, Sun Y, Jiao N, Stepanauskas R, Luo H. (2016). Ecological genomics of the uncultivated marine Roseobacter lineage CHAB-I-5. *Appl Environ Microbiol* **82**: 2100–2111.

Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, *et al.* (2013). Abundant SAR11 viruses in the ocean. *Nature* **494**: 357–360.

Zhao Y, Wang K, Ackermann H-W, Halden RU, Jiao N, Chen F. (2010). Searching for a 'hidden' prophage in a marine bacterium. *Appl Environ Microbiol* **76**: 589–595.

Zhao Y, Wang K, Budinoff C, Buchan A, Lang A, Jiao N, *et al.* (2009a). Gene transfer agent (GTA) genes reveal diverse and dynamic *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay. *ISME J* **3**: 364–373.

Zhao Y, Wang K, Jiao N, Chen F. (2009b). Genome sequences of two novel phages infecting marine roseobacters. *Environ Microbiol* **11**: 2055–2064.

Zhong X, Guidoni B, Jacas L, Jacquet S. (2015). Structure and diversity of ssDNA *Microviridae* viruses in two peri-alpine lakes (Annecy and Bourget, France). *Res Microbiol* **166**: 644–654.

Zivin R, Zehring W. (1981). Transcriptional map of bacteriophage N4. *J Mol Biol* **152**: 335–356.