5-2012

Diversity and Activity of Roseobacters and Roseophage

Charles Ryan Budinoff
cbud@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Charles Ryan Budinoff entitled "Diversity and Activity of Roseobacters and Roseophage." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Microbiology.

Alison Buchan, Major Professor

We have read this dissertation and recommend its acceptance:

Steven W. Wilhelm, Erik Zinser, Mark Radosevich

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Diversity and Activity of Roseobacters and Roseophage

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Charles Ryan Budinoff

May 2012

DEDICATION

I dedicate my dissertation to my maternal grandmother,
for her unwavering love and support.

ACKNOWLEDGEMENTS

Dr. Alison Buchan, your mentorship was essential for the completion of this dissertation and of my doctoral degree. I have learned many things from you over the past four years. Your expertise in bacterial diversity and your knowledge of molecular biology were of great value. But the most important lesson I learned was how to be a good mentor and leader. Your extraordinary interpersonal skills and astounding work ethic allow you to effectively lead a large university laboratory. In my opinion, these are two of the most important traits of a supervisor. Through leading by example, you undoubtedly earn the respect of fellow faculty and staff, but also of every student in the department. I am honored to have been in your lab and know that the examples you have set will guide me for the rest of my life.

Dr. Gary LeCleir, your friendship over the past decade cannot be summed in a few sentences. I will never be able to repay you for the help, advice, and laughter you gave to me, nor will I ever forget it. You are a person who never says ‘no’ to his friends. No matter how difficult the favor, I know you will be there with a big heart and a big smile. I hope to pay this forward by providing such guidance and friendship to others whenever I can.

Kellina Morris, thanks for being by my side this past year. You listened to all my griping, all my lapses in confidence, and never let me think about giving up. Knowing that my graduation would bring us one step closer to each other was the extra motivation I needed. Your support has allowed us to close our first chapter and will be the springboard for our future together.

Mary Hadden, thanks for keeping the lab in tip-top shape. I could not of had so many successful experiments without your hard work and devotion.

ABSTRACT

Bacteria of the *Roseobacter* lineage are dominant bacterioplankton in coastal systems and contribute significantly to secondary production in oceanic environments. Generalities of *Roseobacter* ecology, diversity, and distributions are known, but the intraspecific differences between species and their dynamics over short temporal periods is not well understood. Bacteriophage that infect Roseobacters ('roseophage') have the potential to shunt secondary production into the dissolved carbon pool and through the process of infection alter *Roseobacter* physiology. Despite their significance, little effort was made prior to the onset of this study to characterize roseophage. Using culture dependent and independent approaches, I describe the diversity and activity of Roseobacters and roseophage from two distinct coastal environments. Chapter 2 describes the development of an alternative method to enumerate viruses using epifluorescence microscopy that not only reduces sample processing costs, but also the total volume of sample required. A novel species of the *Roseobacter* lineage (*Marivita roseacus*) is proposed in Chapter 3. *M. roseacus* is unique in its needle-like morphology, forming long, relatively inflexible chains of cells. The *Marivita* genus is characterized by a distinct ecology, being closely associated with algae, resistant to grazing, and present in numerous marine and saline environments. Chapter 4 details the use of deep-amplicon sequencing (16S rDNA) to describe bacterial succession patterns during a mesocosm algal bloom, revealing the temporal dynamics of ~100 distinct phylotypes. A multivariate analysis showed that temporal portioning amongst the bacterial community was occurring at both high and low taxonomic levels. Chapter 5 details the isolation and genomic characterization of roseophage and describes their ecology using publically available metagenomic databases collected from throughout the world. Four distinct phage were isolated and sequenced including an N4-like strain, a novel Siphoviridae, and two temperate Podoviridae. The two temperate phage were practically identical at the nucleotide level, except for a 3000 bp putative replication module, which showed no homology between the two. Overall, this dissertation suggests that ecological partitioning within the *Roseobacter* lineage is occurring at and arguably below traditional species level taxonomic classifications and microdiversity amongst closely related marine bacteria is likely the norm rather than the exception.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 -
INTRODUCTION

A central goal of microbial ecology is to understand the roles that individual microbes play in Earth's ecosystems. Traditionally, questions surrounding marine microbes have been posed from an abiotic or "bottom up" perspective and examined population structure at broad taxonomic, temporal, and spatial scales. But new perspectives are emerging that take into account the biological, "top-down" controls that individuals within a community exert on each other (Miki and Jacquet 2008; Strom 2008) and the role that microscale habitats (Azam and Malfatti 2007) and microdiversity (Fuhrman 2009) have in shaping ecosystems. Many community interactions occur simultaneously in aquatic systems, including mortality, allelopathy, symbiosis, and nutrient partitioning. Such processes can contribute to defining a microorganism's environment and can influence niche differentiation and ultimately ecosystem processes. The microbial diversity within a single milliliter of water is astounding, not only at high taxonomic levels, where distinct protists, bacteria, and viruses are structuring food webs, but within species, where intraspecific genomic differences amongst closely related organisms can define the functional potential of individual populations. These concepts are the center of my research goals and the University of Tennessee afforded me the opportunity to pursue them. With Dr. Alison Buchan's extensive knowledge of marine heterotrophic bacteria and Dr. Wilhelm's experience with aquatic viruses, I chose to investigate the diversity and activity of a group of marine heterotrophic bacteria, Roseobacters, and their infecting bacteriophage, 'roseophage'. Bacteria of the *Roseobacter* lineage are highly abundant marine heterotrophs, comprising greater than 15% of the bacterioplankton in certain marine communities (see (Buchan et al. 2005) for a review of the Roseobacter lineage). They are found in many different marine habitats, including coastal waters, open oceans, the sea floor, sea ice, saline lakes, as well as in commensal relationships. Unlike other numerically dominant marine bacteria, Roseobacters are readily isolated into pure culture. Cultivated members have been shown to carry out many important biogeochemical processes including: the oxidation of organic and inorganic sulfur compounds, including growth with the algal osmolyte dimethylsufoniopropionate (DMSP) and lithoheterotrophy with sulfite and thiosulfate.

Prokaryotes and their infecting viruses, including Roseobacters and roseophage, are astoundingly diverse. Making sense of this diversity goes beyond simple categorical

classifications of species, extending into our understanding of the evolutionary forces behind diversification and how genetic differences shape microbial communities. Criteria to identify genetic groups of specific microbial populations, in terms of their evolutionary importance or ecological function, are not clear-cut. The innate complexity of natural environments underscores our need to utilize multiple approaches and technologies to investigate microbial diversity. For my dissertation, I utilized culture dependent and independent methods to examine how fine scale genetic differences among similar taxa relate to their distinct phenotypic, physiological, temporal, and spatial properties.

Quantifying the abundance of viral populations is one of the first steps towards understanding their significance to a particular ecosystem and the accuracy of these measurements is dependent on available tools. Viruses are extremely small in size, genetically diverse and use different nucleic acid-types as genomes (dsDNA, ssRNA, etc.), all of which make their quantification quite challenging. Epifluorescence microscopy (EFM) of viral populations utilizes nucleic acid binding fluorophores. Although not capturing all viruses (RNA-types and those with small genomes do not fluoresce well), these techniques have become a standard method in environmental microbiology (Suttle and Furman 2010). Enumeration by EFM requires the use of a special membrane filter to capture the virions from an aqueous solution, which is currently manufactured by only one company. Recently, supply shortages of these filters ceased sample processing in many research labs. A quote from Dr. Jed Furman (University of Southern California in Los Angeles) adds context to the issue; "A nightmare" (Torrice 2009). The second chapter of my dissertation sprung from this situation and details a method that utilizes an alternative filter membrane. Although the original filters are available again, my described method serves as a substitute that will undoubtedly be useful for certain research projects.

For the third chapter of my dissertation, I performed a systematic characterization of a common estuarine *Roseobacter* genus: *Marivita*. The process of describing a new species is an important part of microbiology and a valuable exercise for a microbial ecologist. My isolate, designated *M. roseacus*, displays a cell structure that is quite unique when compared to other *Marivita* species and Roseobacters for that matter. This unique morphology of *M. roseacus* is

enough to justify the study, but of additional interest is that *Marivita*-like ecotypes and strains are often closely associated with algal cells and have been shown to be resistant to grazing.

The fourth chapter of my dissertation investigates the temporal dynamics of bacterial diversity during an induced algal (*Emiliana huxleyi*) bloom using deep-amplicon sequencing of the 16S rRNA v3 region. One of our main goals for this chapter was to monitor *Roseobacter* population diversity and species succession at a high resolution during bloom development and demise. Multivariate statistical analysis of the sequence data allowed us to begin to elucidate the functional and temporal differences amongst ~100 individual bacterial 'species' or phylotypes. The chapter reinforces the metabolic diversity that is typically found within the major bacterial groups and suggests niche partitioning at the species and arguably the subspecies level.

The overall diversity and physiology of Roseobacters has been previously documented, but little is known about the bacteriophage capable of infecting the lineage. To fill this void, we set out to isolate and characterize roseophage. Using phylogenetically distinct *Roseobacter* strains we isolated roseophage from two coastal environments. This chapter of my dissertation focused on identifying the different types of phage capable of infecting Roseobacters, a detailed examination of their genetic structure, and their overall global distribution and ecology. A major goal of this work was to identify host-phage systems that can act as models for further studies and aid in the development of molecular tools to quantify roseophage within natural assemblages.

References

Azam, F., and F. Malfatti. 2007. Microbial structuring of marine ecosystems. Nat Rev Microbiol **5:** 782-791.

Buchan, A., J. M. Gonzalez, and M. A. Moran. 2005. Overview of the marine *Roseobacter* lineage. Appl Environ Microbiol **71:** 5665-5677.

Fuhrman, J. A. 2009. Microbial community structure and its functional implications. Nature **459:** 193-199.

Miki, T., and S. Jacquet. 2008. Complex interactions in the microbial world: underexplored key links between viruses, bacteria and protozoan grazers in aquatic environments. Aquat Microb Ecol **51:** 195-208.

Suttle, C., and J. Fuhrman. 2010. Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy**.** In *Manual of Aquatic Viral Ecology.* Edited by Wilhelm SW, Weinbauer MG, Suttle CA: ASLO; 2010: 145-153

Strom, S. L. 2008. Microbial ecology of ocean biogeochemistry: A community perspective. Science **320:** 1043-1045.

Torrice, M. 2009. Viral ecology research hit by filter shortage. *Science*Insider. *Science*Insider.

CHAPTER 2 -
A PROTOCOL FOR ENUMERATION OF AQUATIC VIRUSES BY EPIFLUORESCENCE MICROSCOPY USING ANODISC™ 13 MEMBRANES

A version of this chapter was originally published by Charles R. Budinoff, Star N. Loar, Gary R. LeCleir, Steven W. Wilhelm and Alison Buchan

Charles R. Budinoff, Star N. Loar, Gary R. LeCleir, Steven W. Wilhelm and Alison Buchan. "A protocol for enumeration of aquatic viruses by epifluorescence microscopy using Anodisc™ 13 membranes." *BMC Microbiol.* 11: 168 (2011): doi: 10.1186/1471-2180-11-168.

CRB developed the filtration procedures, coordinated the experimental design, performed the statistical analysis, and drafted the manuscript. SNL carried out the filtration of the samples and their microscopic enumeration. GRL participated in the experimental design, helped develop the filtration procedures, and helped to draft the manuscript. SWW participated in its design and coordination, and helped to draft the manuscript. AB participated in the design and coordination of the study, aided in the interpretation of the data, and helped to draft the manuscript. All authors read and approved the final manuscript.

Abstract

Background

Epifluorescence microscopy is a common method used to enumerate virus-like particles (VLP) from environmental samples and relies on the use of filter membranes with pore sizes <0.02 μm; the most commonly used protocols employ 25 mm Anodisc™ membranes with a built-in support ring. Other filters with small pore sizes exist, including the 13 mm Anodisc™ membranes without a support ring. However, the use of these membranes for viral enumeration has not been previously reported.

Results

Here we describe a modified protocol for 13 mm Anodisc membranes that uses a custom filter holder that can be readily constructed in individual investigators' laboratories from commercially available Swinnex® filter holders. We compared VLP concentrations obtained from phage

lysates and seawater samples using both Anodisc membranes, as well as Nuclepore™ small pore-size membranes (0.015 or 0.030 µm). The 13 mm Anodisc membranes gave comparable estimates of VLP abundance to those obtained with the 25 mm Anodisc membranes when similar staining methods were employed. Both Nuclepore membranes typically gave an order of magnitude lower VLP abundance values for environmental samples.

Conclusions

The 13 mm Anodisc membranes are less costly and require smaller sample volumes than their 25 mm counterpart making them ideal for large-scale studies and sample replication. This method increases the options of reliable approaches available for quantifying VLP from environmental samples.

Background

Viruses are an important component of aquatic food webs. They contribute significantly to the mortality of marine microorganisms and consequently alter species composition and influence the flow of carbon and energy within an ecosystem [1]. As such, accurate and reproducible estimates of virus abundance from environmental samples are essential to our understanding of aquatic biology and biogeochemistry. The earliest estimates of virus-like particles (VLP) in aquatic samples relied on transmission electron microscopy (TEM) [2, 3]. However, the high cost, limited availability, and laborious nature of TEM quickly led investigators to switch to epifluorescence microscopy approaches [4-6] using Nuclepore™ track-etched polycarbonate membranes (pore sizes, 0.015 or 0.030 µm, Whatman North America) [4, 5, 7] and methods originally described for enumerating bacteria [8]. Due to slow flow rates, Nuclepore membranes were subsequently replaced by Anodisc™ inorganic (Al_2O_3) membranes (pore size 0.02 µm, Anodisc™, Whatman) (refer to Table 2.1) [9, 10]. Anodisc membranes are available in 13 and 25 mm diameters. The 25 mm membrane with a built-in support ring is commonly used to determine VLP abundances in natural systems and is recommended in several published protocols [11, 12]. However, the establishment of a protocol using the 13 mm membranes,

lacking a support ring, has the advantages of significantly reducing processing costs (by 50% or more; Table 2.1) and the amount of sample required.

Results and Discussion

A practical limitation of the 13 mm Anodisc membranes is the lack of a peripheral support ring to facilitate handling of the membranes. To alleviate this limitation, we constructed custom filter holders and used modifications of traditional protocols for enumeration of VLP. The feasibility of using Nuclepore filters for viral enumerations was also revisited using modified protocols to reduce filtration times. In part, our motivation to reevaluate the feasibility of Nuclepore membranes for VLP enumeration was prompted by production problems of Anodisc membranes [13], which have been subsequently resolved but serve as a reminder that the availability of alternate protocols would be useful.

Construction of custom filter holders for 13 mm Anodisc membranes

Filter towers were constructed using the inlet portion of a 13 mm Swinnex filter holder (Millipore, Billerica, MA) that was bonded to a makeshift funnel, the conical end of a 15 mL disposable centrifuge tube (Fig 1). The funnel was necessary as the inlet portion could only hold ~150 µL of liquid and the surface tension caused by the Luer-lock was too great to permit an even passage of liquid under vacuum. Briefly, the Luer-lock was cut off of the Swinnex fitting inlet to maximize the opening. Next, the tip of the 15 mL tube was removed and the end of the tube subsequently finely sanded so that when inserted into the inlet and assembled with the outlet it would not come in contact with the filter membrane. The two pieces were bonded using a cyanoacrylate-type glue and allowed to cure for 24 hours. For filtration, the inlet/funnel was screwed onto the outlet portion of the Swinnex, which was connected to a vacuum source. This filtration apparatus is inexpensive (<$20 USD) and in combination with a manifold, allows for high throughput filtration.

Enumeration of VLP using 13mm Anodisc membranes

Our protocol for preparing virus slides using 13 mm Anodisc membranes is based on that of Ortmann and Suttle (2009), with modifications of the staining procedure. Back-staining is the standard protocol for Anodisc 25 membranes and involves placing the membrane sample side up onto a drop of stain, incubating, then removing excess stain by either wicking [14] or applying vacuum [12]. However, back-staining is technically challenging due to the small size and absence of a support ring on the 13 mm membranes. Thus, samples were pre-stained prior to filtration. The detailed protocol is as follows: i) A virus sample was brought up to a final volume of 900 µL using 0.02-µm filtered diluent (AN media or seawater). ii) 100 µL of SYBR Gold (25X, 0.02 µm filtered) was added to the sample and then incubated for 15 min in the dark. iii) A backing filter (0.2 µm, polyethersulfone, Pall Corporation, Port Washington, NY) was placed onto the screen of the Swinnex outlet and overlaid with sterile MilliQ water (~2 mL). Vacuum pressure (5 in Hg) was applied to pull the water through and stopped immediately so not to dry out the filter. iv) The backing filter was overlaid with MilliQ (~2 mL) again and a 13 mm Anodisc placed on top of the water. v) The vacuum was then applied to pull the water through and sandwich the filters together. vi) With the vacuum still on, the modified Swinnex inlet (containing a gasket) was carefully screwed on and tightened with sufficient torque; excessive torque would crack the membrane and insufficient torque caused particles to be preferentially filtered towards the periphery of the membrane. vii) The sample was added to the center of the funnel. After all the liquid had visually disappeared, the vacuum was continued for an additional 30 seconds. viii) With the vacuum still on, the Swinnex inlet was carefully unscrewed, leaving the gasket and the two filters on the outlet. ix) The vacuum was cut and the three pieces (sandwiched filters and gasket) were removed as one and placed on Whatman (grade 4, qualitative) paper to dry for one min. x). Using forceps and a needle, the gasket was removed and the filters separated. xi) The Anodisc was mounted on a glass slide with anti-fade solution (50% glycerol, 50% PBS, 0.1% p-phenylenediamine). Filtration time was <5 min per mL. Parallel samples were also prepared with a post-stain rinse, where 500 µL of 0.02-µm filtered media or seawater was added to the funnel and pulled through with the vacuum.

Enumeration was performed on a Leica DMRXA using filter cube L5 (excitation filter BP 480/40, suppression filter BP 527/30). For each slide, 20 fields and at least 200 particles were counted. To calculate the concentration of virus particles ml-1, the average number of particles per field was multiplied by the dilution factor and microscope conversion factor and then divided by the volume of sample filtered (in ml). The microscope conversion factor was calculated as the filterable area of the membrane divided by the area of each individual field. Variance in the filterable area using the meniscus loading method for the 25 mm Anodisc filters and the Swinnex filter holders for the 13 mm filters was 18.38 (±0.115) and 9.61 (± 0.131), respectively.

Comparison of VLP counts using Anodisc membranes and evaluation of staining methods

VLP concentrations were determined from three sample types with both Anodisc membranes: a viral lysate of a marine cyanobacterium, open ocean surface seawater and coastal surface seawater. Three replicate slides were prepared for each sample type and method. Previous studies have recommended a rinse step following staining of Anodisc 25 mm membranes when processing natural samples with high organic matter content <(e.g. sediments, humic waters) to reduce background fluorescence [15]. Thus, we conducted a comparison of rinsing and no rinsing for both Anodisc membrane sizes across the three sample types. We also compared staining approaches (back- vs pre-) for the Anodisc 25 mm membranes. The cyanophage viral lysates gave indistinguishable VLP counts (ANOVA, $P > 0.05$) regardless of membrane diameter, staining and rinsing procedure. The two environmental samples showed variation among the methods tested that were due to the rinse step. Viral abundances determined using the two Anodisc membranes were significantly different (ANOVA, $P < 0.05$) when the post-rinse step was omitted. However, differences were not significant between the two membrane types when the post-rinse step was applied (ANOVA, $P > 0.05$) (Table 2.2). Replicate seawater samples had a higher coefficient of variation (5-30%) than phage lysates (5-10%). The higher variance amongst the replicate seawater samples is attributed to sample microheterogeneity. Microbial heterogeneity in natural aquatic samples is well known; bacteria and viruses have been shown to form aggregates or be in close association with organic particles [16, 17].

Discrepancies in VLP counts due to staining method and post-rinsing are most likely a reflection of differences in concentration and composition of viral communities (in terms of size and fluorescence) as well as organic material in the natural samples. For example, coastal environments and other highly productive systems typically contain a higher proportion of eukaryotic algae in the plankton then do oligotrophic systems, such as the open ocean [18]. Viruses that infect algae are routinely isolated and have been shown to be quite large in size (capsid, 100-220 nm) and contain large genomes [19, 20]. A higher proportion of smaller, less fluorescent viruses in the open ocean could contribute to lower VLP counts after post-rinsing. The issue of including a post-rinse in the processing of natural samples for VLP enumeration is environment dependent and beyond the scope of this report, which is designed to illustrate the comparability of sample processing with the 13 mm and 25 mm Anodisc membranes.

Analysis of Nuclepore membranes

The same samples described in the previous section were also processed using Nuclepore filters. Due to the low flow rate of Nuclepore membranes, filtering times have been traditionally quite long (> 1 hr). To maximize flow rates, existing protocols were modified. Specialized backing filters and filter holders were used and details are provided in the methods section. VLP enumeration from natural samples using Nuclepore membranes were generally an order of magnitude lower than parallel enumerations conducted using the Anodisc membranes (data not shown). Furthermore, analysis of Nuclepore filtrate subsequently passed through Anodisc membranes indicated VLP were passing through these membranes. Thus, Nuclepore membrane pore sizes were analyzed using scanning electron micrographs as described in the methods section. Pore sizes were consistent in membranes pre- and post-filtration. However, the pore sizes for Nuclepore 30 membranes were not uniform and ranged from 20 to 50 nm in size with the majority of pores being < 40 nm (78%)(Fig. 2.2B); the Nuclepore 15 membranes were also not uniform and ranged from 10 to 30 nm in size with the majority of pores being < 20 nm (69 %) (Fig. 2.2C).

Conclusions

Modifications of existing protocols allow the reliable use of Anodisc 13 membranes for enumeration of VLP using epifluorescence microscopy. In parallel studies, we found that Nuclepore filters (polycarbonate, 0.03 & 0.015 μm pore sizes) consistently yielded lower observable VLP. These low counts may be attributed to non-uniform pore sizes that were evident by scanning electron microscopy of these filters (Fig. 2.2). However, more rigorous parallel comparisons of the Nuclepore and Anodisc membranes are necessary to determine this conclusively. Differences in VLP abundance estimates between Anodisc 13 and 25 membranes were evident with environmental samples if a post-rinse step was not included in sample processing. While rinsing of membranes gave the most consistent results across the two Anodisc membranes, it may result in loss of enumeration of VLP depending upon the environment from which the sample was derived. Given the heterogeneity of natural virus populations, individual investigators will need to consider the issue of applying a post-rinse on a case-by-case basis.

Methods

Sample collection and preparation

Viral lysate was made using cyanophage S-PWM1, which infects Synechococcus sp. WH7803 (aka DC2) [21]. The lysate was filtered through a 0.2-μm Durapore™ filter and stored at 4° C – this filtered material served as the lysate standard. Open ocean water samples were collected from the Sargasso Sea (May 28, 2005; 36.343° N, 51.315° W) and coastal water samples were collected off the coast of Georgia, USA (Nov 18, 2007; 31.372° N, 80.561° W). Multiple seawater aliquots (2 mL) were uniformly distributed, fixed in 0.5% glutaraldehyde and frozen at -80° C at the start of this study to ensure reproducibility.

Enumeration of viruses using 25 mm Anodisc membranes

The protocol using 25 mm Anodisc membranes follows that published by Ortmann and Suttle (2009), with minor modifications. Briefly, filtration was performed on a Hoefer® filtration

manifold (Hoefer, Holliston, MA) without chimney weights. After the backing (0.45-µm pore-size cellulosics; MicroSep™, GE Water & Process Technologies, Trevose, PA) and the Anodisc filter were mounted on the filter stage with the vacuum on, the sample (final volume 1 mL) was applied to the top, forming a meniscus. The filter was back-stained by placement sample side up onto 100 µL of SYBR Gold stain (25 X concentration, Invitrogen, Carlsbad, CA) and incubated for 15 min followed by application of a vacuum to remove the stain. Samples were also prepared with a post-stain rinse of 850 µL of 0.02 µm filtered media or seawater. For direct comparison to the Anodisc 13 membranes, parallel samples were also pre-stained in a microcentrifuge tube prior to filtration. Filtration time using the above protocol was <5 min per mL of sample.

Determination of filterable area for Anodisc membranes

The filterable area of the Anodisc membranes was determined by passage of a cell culture of the naturally pigmented bacterium Synechococcus sp. WH7803 through them. Digital images were analyzed with Adobe® Photoshop® CS4 (Adobe Systems Incorporated, San Jose, CA) to calculate the area containing pigmented cells. The data reported is a range of the averages obtained from triplicate filters.

Enumeration of viruses using Nuclepore membranes

As pre-stained black Nuclepore membranes with pore sizes of 15 and 30 nm are not commercially available, membranes were stained using 0.2% Irgalan Black (Acid black 107, Organic Dyestuffs Corporation, East Providence, RI) dissolved in 2% acetic acid as previously described [8], with the exceptions that staining time was reduced from 3 hours to 15 minutes and filters were used immediately. Polyester drain discs (Whatman), which are designed to improve flow rate and provide a flat surface to eliminate rupturing were used as backing filters. Filters were placed in 25 mm Swinnex filter holders for filtration and processed using the same reagents and solutions described for the Anodisc membranes. The filtration time required for the Nuclepore 15 and 30 membranes using the above protocol was <60 min and <10 min per mL, respectively.

SEM imaging of Nuclepore membranes

To assess whether the filtration protocol could be damaging or altering membrane pore size, scanning electron micrographs of the Nuclepore membranes were taken before and after filtrating media (0.02 μM filtered AN) or seawater (0.02 μM filtered Sargasso Sea water) using a LEO 1525 field emission scanning electron microscope (Carl Zeiss Inc., Thornwood, NY, USA). Avoiding lateral stress, the membranes were cut, mounted on a stub and viewed. No coating was applied so as to not obscure the pores. At least 3 regions of each filter were viewed and at least 50 pores measured from each filter. Filtration did not appear to damage the filters or change pore size. Initial attempts at preparing the filters for SEM did suggest that lateral stress (excessive stretching or twisting) of the membranes could drastically increase pore size (data not shown).

Statistical comparison of virus counts from the Anodisc membranes

The statistical software package SPSS was used to compare the VLP counts between the technical replicates (repeated-measures ANOVA, C.I. of 5%) and between the membrane types (2-tailed paired t test, C.I. of 5% or repeated-measures ANOVA, C.I. of 5%). Counts obtained from the individual fields of each slide were first evaluated using the Shapiro-Wilks test. Data sets that failed the Shapiro-Wilks test (having p-values < 0.05) were transformed using the Box-Cox transformation. The resulting transformed variables were consistent with a normal distribution. Mauchly's test of sphericity was performed and if the test was found to be significant (having p-values < 0.05) either the Huynh-Feldt (for epsilon values >0.75) or the Greenhouse-Geisser (for epsilon values <0.75) correction was applied.

Acknowledgements

This work was funded by the National Science Foundation (OCE-0550485 to AB and OCE-0825405 and OCE- 0851113 to SWW). The authors would like to thank J. Dunlap for assistance with SEM.

References

1. Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, Heldal M, Kimmance SA, Middelboe M, Nagasaki K, Paul JH, et al: **Global-scale processes with a nanoscale drive: the role of marine viruses.** *ISME J* 2008, **2:**575-578.
2. Bergh O, Børsheim KY, Bratbak G, Heldal M: **High abundance of viruses found in aquatic environments.** *Nature* 1989, **340:**467-468.
3. Proctor LM, Fuhrman JA: **Viral mortality of marine bacteria and cyanobacteria.** *Nature* 1990, **343:**60-62.
4. Hara S, Terauchi K, Koike I: **Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy.** *Appl Environ Microbiol* 1991, **57:**2731-2734.
5. Proctor LM, Fuhrman JA: **Mortality of marine bacteria in response to enrichments of the virus size fraction from seawater.** *Mar Ecol Prog Ser* 1992, **87:**283-293.
6. Suttle CA, Chan AM, Cottrell MT: **Infection of phytoplankton by viruses and reduction of primary productivity.** *Nature* 1990, **347:**467-469.
7. Suttle C: **Enumeration and isolation of viruses.** In *Handbook of Methods in Aquatic Microbial Ecology.* Edited by Kemp PF, Cole JJ, Sherr BF, Sherr EB. Boca Raton: CRC Press; 1993: 121-134
8. Hobbie JE, Daley RJ, Jasper S: **Use of nuclepore filters for counting bacteria by fluorescence microscopy.** *Appl Environ Microbiol* 1977, **33:**1225-1228.
9. Hennes KP, Suttle CA: **Direct counts of viruses in natural waters and laboratory cultures by epifluorescence microscopy.** *Limnol Oceanogr* 1995, **40:**1050-1055.
10. Tranvik L: **Effects of Colloidal Organic Matter on the Growth of Bacteria and Protists in Lake Water.** *Limnol Oceanogr* 1994, **39:**1276-1285.
11. Noble RT, Fuhrman JA: **Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria.** *Aquat Microb Ecol* 1998, **14:**113-118.

12. Ortmann A, Suttle C: **Determination of virus abundance by epifluorescence microscopy.** *Methods Mol Biol* 2009, **501:**87-95.
13. **Torrice M: Viral ecology research hit by filter shortage** [http://news.sciencemag.org/scienceinsider/2009/10/viral-ecology-r.html]
14. Patel A, Noble RT, Steele JA, Schwalbach MS, Hewson I, Fuhrman JA: **Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I.** *Nat Protoc* 2007, **2:**269-276.
15. Suttle C, Fuhrman J: **Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy.** In *Manual of Aquatic Viral Ecology.* Edited by Wilhelm SW, Weinbauer MG, Suttle CA: ASLO; 2010: 145-153
16. Simon M, Grossart HP, Schweitzer B, Ploug H: **Microbial ecology of organic aggregates in aquatic ecosystems.** *Aquat Microb Ecol* 2002, **28:**175-211.
17. Luef B, Neu TR, Peduzzi P: **Imaging and quantifying virus fluorescence signals on aquatic aggregates: a new method and its implication for aquatic microbial ecology.** *FEMS Microbiol Ecol* 2009, **68:**372-380.
18. Chisholm S: **Phytoplankton size.** In *Primary Productivity and Biogeochemical Cycles in the Sea.* Edited by Falkowski PG, Woodhead AD. New York: Plenum Press; 1992: 213-237
19. Monier A, Larsen JB, Sandaa RA, Bratbak G, Claverie JM, Ogata H: **Marine mimivirus relatives are probably large algal viruses.** *Virol J* 2008, **5:**12.
20. Wilson WH, Etten JL, Allen MJ: **The *Phycodnaviridae*: The story of how tiny giants rule the world.** In *Lesser Known Large dsDNA Viruses. Volume* 328. Edited by Etten JL: Springer Berlin Heidelberg; 2009: 1-42: *Current Topics in Microbiology and Immunology*.
21. Suttle CA, Chan AM: **Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics.** *Mar Ecol Prog Ser* 1993, **92:**99-109.

Appendix

Tables

Table 2.1. Specifications of Whatman membranes used in this study

Information obtained from Whatman North America.

Filter name	Part Number	Filterable Diameter (mm)	Pore Size (μm)	Flow rate[a]	Porosity (pores/cm^2)	Burst strength (psi)	Autoclavable	Cost per filter (USD)
Anodisc™ 13	6809-7003	13	0.02	4.9, 0.3	10^{10}	65-110	yes	2.08
Anodisc 25	6809-6002	21	0.02	4.9, 0.3	10^{10}	65-110	no	5.10
Nuclepore™ 15	110601	25	0.015	N/A, 0.002-0.04	10^{8}	>15	yes	1.84
Nuclepore 30	110602	25	0.03	N/A, 0.06-0.20	10^{8}	>15	yes	1.32

[a] water, air L/min/cm^2 @ 10 psi , 25° C.

Table 2.2. Comparison of back-staining and pre-staining of Anodisc™ membranes in VLP enumeration of three sample types.

Sample	Filter[a]	Staining method	Rinse	VLP[b]	CV[c]
Cyanophage lysate	Ano 25	Back	No	1.32×10^6 (0.08)	5.7
	Ano 25	Back	Yes	1.32×10^6 (0.10)	7.5
	Ano 25	Pre	No	1.63×10^6 (0.07)	4.5
	Ano 25	Pre	Yes	1.54×10^6 (0.15)	9.6
	Ano 13	Pre	No	1.29×10^6 (0.13)	10.1
	Ano 13	Pre	Yes	1.26×10^6 (0.07)	5.8
Sargasso Sea water	Ano 25	Back	No	9.59×10^5 (1.86)	19.4
	Ano 25	Back	Yes	1.66×10^5 (0.37)	22.5
	Ano 25	Pre	No	7.50×10^5 (1.30)	17.3
	Ano 25	Pre	Yes	1.75×10^5 (0.17)	9.7
	Ano 13	Pre	No	5.93×10^5 (1.15)	19.3
	Ano 13	Pre	Yes	2.28×10^5 (0.54)	23.5
Southeastern US coastal waters	Ano 25	Back	No	14.99×10^5 (0.45)	3.0
	Ano 25	Back	Yes	3.22×10^5 (1.06)	32.9
	Ano 25	Pre	No	4.41×10^5 (0.62)	13.9
	Ano 25	Pre	Yes	3.28×10^5 (0.35)	10.7
	Ano 13	Pre	No	2.58×10^5 (0.35)	13.7
	Ano 13	Pre	Yes	2.75×10^5 (0.41)	14.9

[a] Anodisc™ 25 mm (Ano 25) and 13 mm (Ano 13) membranes

[b] Average VLP abundance from triplicate filters along with the standard deviation

[c] The percent coefficient of variation from 3 replicate measures.

Figures

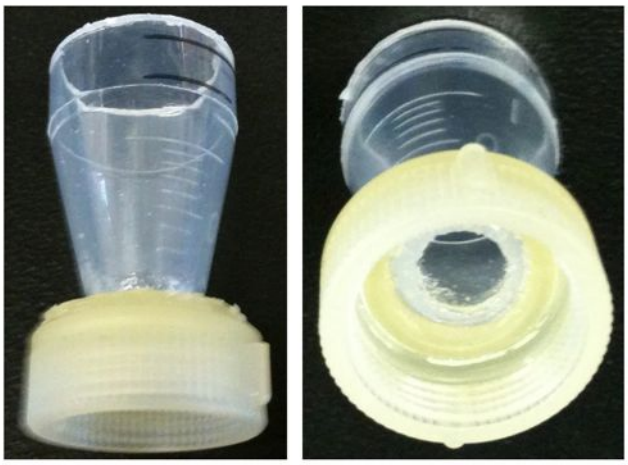


Figure 2.1. Custom-built 13 mm filter funnel

Funnel was assembled from a Swinnex® inlet bonded to the conical end of a 15 ml polypropylene tube.

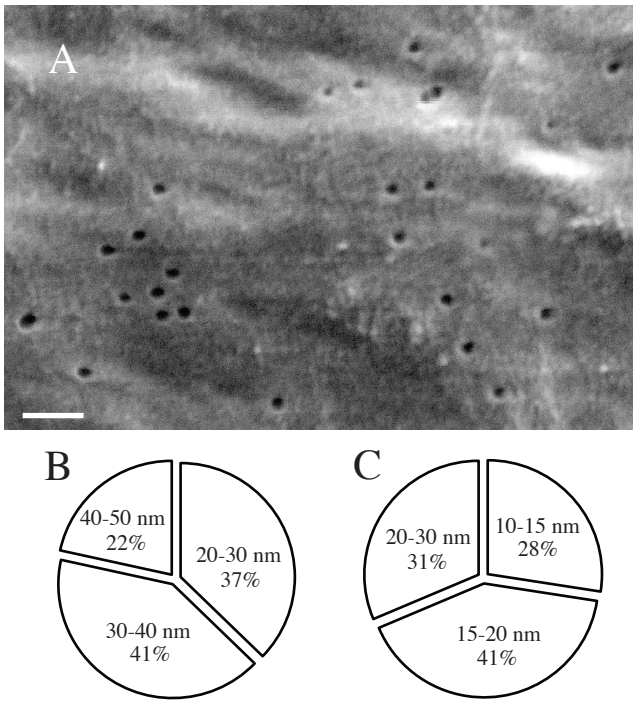


Figure 2.2. Pore size distribution of untreated Nuclepore™ filters determined by SEM analysis

(A) SEM image of Nuclepore™ 30 membrane. Scale bar is 200 nm. (B) Pore size range of Nuclepore 30 membrane. (C) Pore size range of Nuclepore 15 membrane.

CHAPTER 3 -

MARIVITA ROSEACUS SP. NOV., OF THE FAMILY RHODOBACTERACEAE, ISOLATED FROM A TEMPERATE ESTUARY AND AN EMENDED DESCRIPTION OF THE GENUS *MARIVITA*

A version of this chapter was originally published by Charles R. Budinoff, John R. Dunlap, Mary Hadden, and Alison Buchan

Charles R. Budinoff, John R. Dunlap, Mary Hadden, and Alison Buchan. "*Marivita roseacus* sp. nov., of the family Rhodobacteraceae, isolated from a temperate estuary and an emended description of the genus *Marivita*" *Journal of General and Applied Microbiology.* 57: 4 (2011): *In press*.

CRB isolated strain CB1051, coordinated the experimental protocols, performed the taxonomic assays and drafted the manuscript. JRD performed electron microscopy. MH assisted with the taxonomic assays. AB participated in the coordination of the study, aided in the interpretation of the data, and helped to draft the manuscript. All authors read and approved the final manuscript.

Abstract

A gram-negative, non-motile, pigmented, rod-shaped and strictly aerobic bacterium ($CB1052^T$) was isolated from a temperate estuary. On the basis of 16S rRNA gene sequence similarity, strain $CB1052^T$ belongs to the α-3 subclass of the *Proteobacteria*, within the family *Rhodobacteraceae*, having the highest similarity to members of the genus *Marivita* (97.8%) of the *Roseobacter* lineage. Pylogenetic analysis showed $CB1052^T$ to be a distinct sister clade to *M. litorea* and *M. cryptomonadis* and DNA-DNA relatedness was quite low amongst the strains (<35%). Strain $CB1052^T$ cells are non-motile and display a needle-like filamentous form, where individual cells can become quite elongated (up to 15 um). Similar to *M. litorea* and *M. cryptomonadis*, $CB1052^T$ harbors aerobic anoxygenic photosynthesis genes. However, in contrast to other described *Marivita* species, strain $CB1052^T$ actively produces bacteriochlorophyll *a*. Further physiological features, including antibiotic sensitivities, differentiate strain $CB1052^T$ from the other members of the genus. Therefore, strain $CB1052^T$ is considered to represent a novel species of the genus *Marivita*, for which the name *Marivita*

roseacus sp. nov. is proposed, with the type strain CB1052^T (=DSM 23118^T =ATCC BAA 1914^T).

Introduction

The genus *Marivita* was formally described by Hwang *et al.* (2009) and at present is comprised of two recognized species, *Marivita cryptomonadis* and *M. litorea*. The genus is a member of the *Roseobacter* lineage, which is an abundant marine bacterial group known to mediate key biogeochemical processes (Brinkhoff et al., 2008; Buchan et al., 2005). The *Roseobacter* clade is well represented in many diverse marine habitats, but their abundance is often highest near algal blooms or in association with organic particles. Surface attachment and colonization are two defining traits of the *Roseobacter* lineage and are thought to be contributing factors for their success in natural environments (Slightom and Buchan, 2009). Another defining feature of Roseobacters is the presence of aerobic anoxygenic photosythesis (AAP) genes and the production of bacteriochlorophyll *a* (bchl *a*). Roseobacters capable of AAP are often found in association with other organisms such as algae (Allgaier et al., 2003; Green et al., 2004; Shiba, 1991), cyanobacteria (Hube et al., 2009) and invertebrates (Martinez-Garcia et al., 2007). In the present study, we formally describe a novel *Marivita* sp., represented by a unique needle-like filamentous morphology and the active production of bchl *a*.

Materials and Methods

Isolation and growth.

Strain CB1052^T was isolated from seawater collected July 2007 in the lower Chesapeake Bay (Station 724, 37°16'N, 76°09'W) as previously described (Zhao et al., 2009). The Chesapeake Bay is a highly productive system known to support the growth of various phytoplankton and algal species (Marshall et al., 2005). Strain CB1052^T appeared as a pinkish irregular-shaped colony. The colony was sub-cultured once on original isolation media and subsequently purified on 2% w/v artificial seawater (ASW) containing 0.25% yeast and 0.4% tryptone at room

temperature. ASW (2%) contained 230 mM NaCl, 5.3 mM KCl, 3.9 mM $CaCl_2$ x $2H_2O$, 0.1 mM H_3BO_3, 11.8 mM $MgSO_4$ x $7H_2O$, 11.2 mM $MgCl_2$ x $6H_2O$, 0.8 mM $NaHCO_3$, 5 mM NH_4Cl, 75 μM K_2HPO_4, and 10 mM Tris-HCl (pH 7.5). Filter-sterilized stock solutions of Tris-HCl, ammonia, phosphate, and carbonate were added to the autoclaved basal salt solution along with vitamins, iron, and trace metals prepared as previously described (Budinoff and Hollibaugh, 2007). Agar plates were made with Noble agar (Difco) at 1%. After purification the organism was frozen at -80 °C in 25% v/v glycerol.

Type strains used for comparative studies.

For comparative studies the type strains *Marivita cryptomonadis* (DSM 21340^T) and *Marivita litorea* (DSM 21329^T) were obtained from the DSMZ, Braunschweig, Germany.

Biochemical characterization.

Gram-staining, along with catalase and oxidase tests were performed according to standard methods (Colwell and Wiebe, 1970) after growth on Difco Marine Agar 2216 (MA 2216). Antibiotic sensitivity was tested on MA 2216 using BD BBL™ Sensi-Disc™ susceptibility discs. Fatty acid profiling (Microbial ID, Inc. USA) was outsourced and performed on cells grown on MA 2216 at 30 °C for 4 days. Polar lipids were extracted and separated according to Hanson and Lester (1980) with cells grown on MA 2216 at 30 °C for 4 days. Determination of the G+C content was carried out by the Identification Service of the DSMZ and Dr. B.J. Tindall, DSMZ, Braunschweig, Germany using cells grown on Difco Marine Broth 2216 (MB 2216). Gelatinase activity was assayed in MB 2216 supplemented with 12% gelatin (Difco). After 10 days, the tubes were placed at 4° C and if they remained liquefied were considered positive for gelantinase. Degradation of Tween 80 (lipolytic activity) and nitrate reductase activity were performed according to standard methods (Colwell and Wiebe, 1970) after growth on MB or MA 2216. The presence of poly-β-hydroxybutyrate (PHB) granules was determined as previously described (Ostle and Holt, 1982) after growth on MB 2216 using a Leica DMRXA

epifluorescence microscope with filter cube L5. Production of bchl *a* was determined on cells grown in the dark, as previously described (Biebl and Wagner-Döbler, 2006).

Morphological characterization.

Light microscopy (Nikon TE2000-U) was conducted to examine cells after Gram stain and for the determination of motility via the hanging drop method. To determine general morphology scanning electron microscopy (SEM) was performed on cells in mid-log and early stationary phase grown in MB 2216, as previously described (Rowe et al., 2008). Negative-stain transmission electron microscopy (Hitachi H-800) was used to assess the presence of flagellum on cells in mid-log phase grown on MB 2216.

Physiological characterization.

Salinity, pH, and temperature ranges, along with anaerobic growth and inorganic sulfur oxidation were tested with both solid and liquid ASW media. Media for pH range measurements did not contain carbonate, which at a pH >9 would cause a small amount of precipitate to form. Salinity was adjusted from 0-8% (increments of 0.5%) by altering the concentration of the individual salts in the basal medium. The pH was adjusted from 6-10 (increments of 0.5%) using 10 mM MES (pH 6-7), Tris (pH 7 -9), or CAPSO (pH 9-10). The temperature range for growth was determined at 4, 10, 20, 25, 28, 30, and 37 °C. Anaerobic growth was determined using an anaerobic chamber (COY Laboratories, USA) and supplementing media with 5 mM KNO_3. Assimilation of various carbon compounds was determined on agar plates supplemented with 3 mM of the substrate as the sole carbon source. Carbon compounds tested include; acetate, trehalose, sucrose, citrate, cellobiose, glucose, galactose, lactose, glucosamine, dimethylysulfonoproponate, ferulate, 4-hydroxybenzoate, glycolate, coumarate, vanillate, benzoate, quinate, glycerol, caffeate, gentisate, and protocatechuate. Oxidation of inorganic sulfur compounds was determined on media containing 10 mM acetate and 10 mM thiosulfate as previously described (González et al., 2003). All measurements were made after 2 consecutive

transfers. Incubation times varied depending on the conditions tested, ranging from 3-14 days. Growth was assessed in comparison with controls based on colony formation or by measuring changes in OD_{540}. Direct comparisons of salinity tolerance and pH range between 1052^T, *M. cryptomonadis* and *M. litorea* (Table 3.1) were performed as described by Hwang *et al.* (2009).

Phylogeny and genomic DNA-DNA hybridization.

Genomic DNA was extracted from an exponentially growing culture using the MasterPure™ Complete DNA and RNA Purification Kit, EPICENTRE® Biotechnologies; the cells were pre-treated with 25 mg lysozyme mL^{-1} for 60 min at 37 °C. The nearly complete 16S rRNA gene and the entire 16S-23S internal transcribed spacer region (ITS) were PCR amplified using primers 16S-27F (Delong et al., 1993) and 23S-139R (Kan et al., 2008). Sequencing of the products was accomplished using the primers 16S-27F, 16S-1522R (Giovannoni, 1991) and 23S-139R. The presence of an aerobic anoxygenic photosystem was determined by amplifying and sequencing a portion of the *pufM* gene, which encodes for a subunit of the photosynthetic reaction center, using primers pufM_uniF and pufM_WAW as previously described (Yutin et al., 2005). The sequence traces were combined into a single contig and their quality was assessed using the software package CodonCode Aligner. The contigs were compared to available GenBank database sequences using BLAST. Initial phylogenetic analysis of the 16S rRNA gene fragment was performed with the software package ARB using databases and alignments provided by SILVA (Pruesse et al., 2007). Sequences grouping with strain $CB1052^T$ were exported for further analysis. Detailed phylogenetic studies were performed using MODELTEST (Posada and Crandall, 1998), to determine the model of nucleotide substitution that best fit the data, and PAUP (Swofford, 2002) to infer and interpret phylogenetic trees. DNA-DNA hybridization was performed in triplicate following the method of (Ezaki et al., 1989) using photobiotin-labelled DNA probes and microdilution wells; hybridizations were carried out at 50°C.

Nucleotide sequence accession number.

The 16S rRNA gene sequence of strain CB1052^T has been deposited under GenBank accession number GU137308. The *pufM* gene sequence of strain CB1052^T has been deposited under GenBank accession number GU320721.

Results and Discussion

When grown in liquid MA 2216 (28 °C, shaking), strain CB1052^T is non-motile, displays a needle-like filamentous morphology, where individual cells have a consistent width (~0.5 μm) but can become quite elongated (3-15 μm) and often attach to one another at their poles, forming relatively inflexible chains >100 μm in length (Fig. 3.1A). This morphology is similar to that described for strain F190-32, which shares a nearly identical 16S rRNA gene sequence (99.2%) (Sukoso et al., 1998). Roseobacters capable of forming elongated rods (>10 μm) are known, such as *Silicibacter lacuscaerulensis* (Petursdottir and Kristjansson, 1997). However, the formation of long chains has yet to be described for any characterized *Roseobacter* strain. The morphology of CB1052^T is distinct from the two previously described *Marivita* spp. that were reported to be motile by a single polar flagellum and rod-shaped at 1-3 μm in length (Hwang *et al.*, 2009). A morphological reassessment of *M. cryptomonadis* and *M. litorea* confirmed a smaller cell length than that of CB1052^T (Figs. 2 and 3), but indicated a lack of motility (via the hanging drop method) and the absence of polar flagella (via TEM), possibly indicating the transient expression of motility elements. Additionally, *M. cryptomonadis* produced tufts of polar fimbriae (Fig. 3.2A, Supp Figs. 3.4B, 3.4B) and *M. litorea* displayed non-prosthecate appendages (Fig. 3.2B, Supp Figs. 3.4A, 3.5A). These structures were absent from strain CB1052^T (Fig. 3.1B and Supp Fig. 3.5C). Additional isolates belonging to the genus *Marivita* (based on 16s rRNA gene similarity alone), such as DG1236, display another type of morphology: elongated rods that demonstrate cellular polarity and are capable of forming chains up to 20 μm (D. Green, personal communication). While the functional relevance of these differing morphologies is not known, it has been hypothesized that the varied cell shapes and surface appendages of roseobacters may contribute to their colonization success (Slightom and Buchan, 2009). Given the recognized

association of *Marivita* spp. with algal cultures (see below) cell morphology may be an important trait in defining the environmental niches of members of this genus. Interestingly, *Marivita* spp. have been shown to be resistant to grazing during mesocosm experiments (Lebaron et al., 2001), possibly indicating a selective advantage in natural systems when displaying a large morphotype or particular appendages.

Phylogenetic trees constructed with the 16S rRNA gene place strain CB1052^T within the genus *Marivita*. Refer to Hwang *et al.* (Hwang et al., 2009) for phylogenetic positioning of the genus within the *Rhodobacteraceae*. CB1052^T forms a distinct clade from those of *M. cryptomonadis* and *M. litorea* (Fig. 3.3), and groups strongly with strain F190-32 and an environmental clone from the Chesapeake Bay (EF471669). A total of five well-supported clades (bootstrap values >60%) are present within the *Marivita* genus. The majority of phylotypes are marine in origin, although there are a few clone sequences derived from saline lakes, such as the Salton Sea (clone SSW55N), Lake Kauhakō (clone K2-S-3) and the playa lakes of the Monegros Desert (data not shown, see GenBank AM085966) (Fig. 3.3). Clades I, II, and III are mainly comprised of strains that were isolated from unialgae cultures of raphidophytes and prasinophytes (Sukoso et al., 1998; Yoshikawa et al., 2001), dinophytes (Green et al., 2004; Jasti et al., 2005), and cryptophytes (Hwang et al., 2009). The specific mechanisms of these algal-bacterial associations are not known, but one study demonstrated that the addition of *Marivita* spp. to axenic *Chattonella* cultures resulted in an increase in total algal biomass and improved algal viability over extended incubation times (Sukoso and Sakata, 1996). *Marivita*-like sequences are often detected in marine environmental samples and were recently shown to make up a significant fraction of the *Roseobacter* population in a coastal upwelling system (Alonso-Gutiérrez et al., 2009). Furthermore, *Marivita* sequences (*e.g.* DQ890445) were detected in the abdominal setae tufts or hairs of marine mudshrimps that are known to feed primarily on detritus (Demiri et al., 2009).

Bchl *a* was detected in strain CB1052^T giving an absorption spectrum similar to that of strain F190-32 (Yoshikawa et al., 2001), showing the characteristic bchl *a* peak near 770 nm. Additionally, the *pufM* gene was successfully amplified and sequenced from CB1052^T (GU320721). In contrast, *M. cryptomonadis* and *M. litorea* were shown to contain AAP genes,

but did not produce bchl *a* under the culture conditions employed. Approximately ninety roseobacters have been formally described. Of these, only nineteen have been shown to produce bchl *a* and/or contain photosynthetic reaction center genes; nine of these species were isolated from algal-like samples or cultures (Table 3.2). However, not all described roseobacters have been assayed for bchl *a* production or the presence of AAP genes. Thus, the metabolic relevance of habitat type and the potential for AAP amongst roseobacters is poorly understood and is an intriguing avenue for further research.

The level of DNA-DNA relatedness between strain CB1052^T and *M. litorea* was 33.3±5.7% (29.9±5.3% reciprocal analysis) and between CB1052^T and *M. cryptomonadis* was 20.4±2.7 (19.0±2.9% reciprocal analysis). As these values are well below the 70% threshold generally accepted for species delineation (Stackebrandt and Goebel, 1994; Wayne et al., 1987), strain CB1052^T should be considered a novel species of the genus *Marivita*.

Results of biochemical analyses are summarized in the species description and in Table 3.1. The DNA G+C content of CB1052^T is 59.6 mol%. This value is within the range of G+C contents (58.6-61.0 mol%) observed for other members of this genus (Hwang et al., 2009; Sukoso et al., 1998). In comparison to other *Marivita* species, strain CB1052^T displayed the following notable physiological differences: negative for nitrate reductase activity and resistance to gentamycin. The three *Marivita* strains had similar fatty acid content and gave identical polar lipid patterns, showing the presence of phosphatidylcholine, phosphatidylethanolamine and diphosphatidglycerol. Strain CB1052^T was capable of utilizing dimethylsulfoniopropionate (DMSP) and several plant-related aromatic compounds (i.e. ferulate, 4-hydroxybenzoate, coumarate, and vanillate) as sole carbon sources, but was not capable of anaerobic growth or inorganic sulfur oxidation under the conditions tested. Biochemical and physiological reassessment of *M. cryptomonadis* and *M. litorea* led to some results contrary to those reported by Hwang *et al.* (2009). Most notably, both strains were capable of growth with acetate as the sole carbon source, tested positive for nitrate reductase, and did not contain PHB granules under growth conditions tested. Additionally, *M. litorea* was also capable of growth with glucose and lactate as sole carbon sources.

Description of Marivita roseacus sp. nov.

Marivita roseacus (ro.se.a'cus. L. fem. adj. *roseus*, rose-colored, rosy; L. fem. n. *acus*, needle; N.L fem. n. *roseacus* a rosy needle). Cells are non-motile rods, approximately 0.5 μm wide and 3-15 μm long. Capable of forming long chains (>100 μm). Gram-negative, catalase and oxidase positive, and strictly aerobic. Colonies are pinkish with a dendritic shape. Grows in sea salts at concentrations of 0.5-6% (w/v) (optimum, 2-3%), temperatures of 10-30 °C (optimum, 30 °C), and at pH of 6.5-9 (optimum, 7-8). Major cellular fatty acids include $C_{16:0}$, $C_{18:0}$, $C_{18:1}$ ω7c, $C_{12:1}$ 3-OH, iso-$C_{18:0}$, and 11-Methyl $C_{18:1}$ ω7c. Cells do not contain PHB granules. Utilizes acetate, citrate, cellobiose, glucose, galactose, lactose, glucosamine, DMSP, ferulate, 4-hydroxybenzoate, glycolate, coumarate, and vanillate as sole carbon sources. Produces bacteriochlorophyll *a*. Resistant to the antibiotics triple sulfa, clindamycin, and gentamycin but sensitive to chloramphenicol, carbenicillin, amikacin, ampicillin, streptomycin, penicillin, ciprofloxacin, tetracycline, and amthromycin. The type strain, CB1052^T (=ATCC BAA 1914^T =DSM 23118^T), was isolated from surface water of the lower Chesapeake Bay, U.S.A. The DNA G+C content of CB1052^T is 59.6 mol%.

Emended description of the genus Marivita, Hwang et al. (2009)

The description of the genus *Marivita* is as given by Hwang *et al.* (2009), with the following modifications. Cells are either non-motile or motile by means of a polar flagellum. All described species contain photosynthesis-related genes but production of bchl *a* is variable.

Acknowledgements

This work was funded by the National Science Foundation (OCE-0550485 to A.B.)

References

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

Alonso-Gutiérrez, J., Lekunberri, I., Teira, E., Gasol, J.M., Figueras, A., and Novoa, B. (2009) Bacterioplankton composition of the coastal upwelling system of 'Ría de Vigo', NW Spain. *FEMS Microbiol. Ecol.*, **70**, 161-173.

Biebl, H., and Wagner-Döbler, I. (2006) Growth and bacteriochlorophyll *a* formation in taxonomically diverse aerobic anoxygenic phototrophic bacteria in chemostat culture: Influence of light regimen and starvation. *Process Biochem.*, **41**, 2153-2159.

Biebl, H., Allgaier, M., Lunsdorf, H., Pukall, R., Tindall, B.J., and Wagner-Döbler, I. (2005a) *Roseovarius mucosus* sp. nov., a member of the *Roseobacter* clade with trace amounts of bacteriochlorophyll *a*. *Int. J. Syst. Evol. Microbiol.*, **55**, 2377-2383.

Biebl, H., Allgaier, M., Tindall, B.J., Koblizek, M., Lunsdorf, H., Pukall, R., and Wagner-Döbler, I. (2005b) *Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int. J. Syst. Evol. Microbiol.*, **55**, 1089-1096.

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

Buchan, A., González, J.M., and Moran, M.A. (2005) Overview of the marine *Roseobacter* lineage. *Appl. Environ. Microbiol.*, **71**, 5665-5677.

Budinoff, C.R., and Hollibaugh, J.T. (2007) Ecophysiology of a Mono lake picocyanobacterium. *Limnol. Oceanogr.*, **52**, 2484-2495.

Colwell, R., and Wiebe, W. (1970) "Core" characteristics for use in classifying aerobic, heterotrophic bacteria by numerical taxonomy. *Bull. Ga. Acad. Sci.*, **28**, 165-185.

Delong, E.F., Franks, D.G., and Alldredge, A.L. (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol. Oceanogr.*, **38**, 924-934.

Demiri, A., Meziti, A., Papaspyrou, S., Thessalou-Legaki, M., and Kormas, K.A. (2009) Abdominal setae and midgut bacteria of the mudshrimp *Pestarella tyrrhena*. *Cent. Eur. J. Biol.*, **4**, 558-566.

Ezaki, T., Hashimoto, Y., and Yabuuchi, E. (1989) Fluorometric deoxyribonucleic acid deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane-filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int. J. Syst. Bacteriol.*, **39**, 224-229.

Giovannoni, S. (1991) The polymerase chain reaction. In *Nucleic acid techniques in bacterial systematics*. Stackerbrandt, E., and Goodfellow, M. (eds), John Wiley and Sons, pp. 177-203.

González, J.M., Covert, J.S., Whitman, W.B., Henriksen, J.R., Mayer, F., Scharf, B., Schmitt, R., Buchan, A., Fuhrman, J.A., Kiene, R.P., and Moran, M.A. (2003) *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int. J. Syst. Evol. Microbiol.*, **53**, 1261-1269.

Green, D.H., Llewellyn, L.E., Negri, A.P., Blackburn, S.I., and Bolch, C.J.S. (2004) Phylogenetic and functional diversity of the cultivable bacterial community associated with the paralytic shellfish poisoning dinoflagellate *Gymnodinium catenatum*. *FEMS Microbiol. Ecol.*, **47**, 345-357.

Hanson, B.A., and Lester, R.L. (1980) The extraction of inositol-containing phospholipids and phosphatidylcholine from *Saccharomyces cerevisiae* and *Neurospora crassa*. *J. Lipid. Res.*, **21**, 309-315.

Hube, A.E., Heyduck-Soller, B., and Fischer, U. (2009) Phylogenetic classification of heterotrophic bacteria associated with filamentous marine cyanobacteria in culture. *Syst. Appl. Microbiol.*, **32**, 256-265.

Hwang, C.Y., Bae, G.D., Yih, W., and Cho, B.C. (2009) *Marivita cryptomonadis* gen. nov., sp nov and *Marivita litorea* sp nov., of the family *Rhodobacteraceae*, isolated from marine habitats. *Int. J. Syst. Evol. Microbiol.*, **59**, 1568-1575.

Jasti, S., Sieracki, M.E., Poulton, N.J., Giewat, M.W., and Rooney-Varga, J.N. (2005) Phylogenetic diversity and specificity of bacteria closely associated with *Alexandrium* spp. and other phytoplankton. *Appl. Environ. Microbiol.*, **71**, 3483-3494.

Kan, J., Evans, S.E., Chen, F., and Suzuki, M.T. (2008) Novel estuarine bacterioplankton in rRNA operon libraries from the Chesapeake Bay. *Aquat. Microbiol. Ecol.*, **51**, 55-66.

Labrenz, M., Lawson, P.A., Tindall, B.J., Collins, M.D., and Hirsch, P. (2005) *Roseisalinus antarcticus* gen. nov., sp. nov., a novel aerobic bacteriochlorophyll *a*-producing {alpha}-proteobacterium isolated from hypersaline Ekho Lake, Antarctica. *Int. J. Syst. Evol. Microbiol.*, **55**, 41-47.

Labrenz, M., Collins, M.D., Lawson, P.A., Tindall, B.J., Schumann, P., and Hirsch, P. (1999) *Roseovarius tolerans* gen. nov., sp. nov., a budding bacterium with variable bacteriochlorophyll *a* production from hypersaline Ekho Lake. *Int. J. Syst. Evol. Microbiol.*, **49**, 137-147.

Labrenz, M., Tindall, B., Lawson, P., Collins, M., Schumann, P., and Hirsch, P. (2000) *Staleya guttiformis* gen. nov., sp. nov. and *Sulfitobacter brevis* sp. nov., alpha-3-Proteobacteria from hypersaline, heliothermal and meromictic antarctic Ekho Lake. *Int. J. Syst. Evol. Microbiol.*, **50**, 303-313.

Lebaron, P., Servais, P., Troussellier, M., Courties, C., Muyzer, G., Bernard, L., Schafer, H., Pukall, R., Stackebrandt, E., Guindulain, T., and Vives-Rego, J. (2001) Microbial community dynamics in Mediterranean nutrient-enriched seawater mesocosms: changes in abundances, activity and composition. *FEMS Microbiol. Ecol.*, **34**, 255-266.

Mácian, M.C., Arahal, D.R., Garay, E., Ludwig, W., Schleifer, K.H., and Pujalte, M.J. (2005) *Thalassobacter stenotrophicus* gen. nov., sp. nov., a novel marine {alpha}-proteobacterium isolated from Mediterranean sea water. *Int. J. Syst. Evol. Microbiol.*, **55**, 105-110.

Marshall, H.G., Burchardt, L., and Lacouture, R. (2005) A review of phytoplankton composition within Chesapeake Bay and its tidal estuaries. *J. Plankton Res.*, **27**, 1083-1102.

Martinez-Garcia, M., Diaz-Valdes, M., Wanner, G., Ramos-Espla, A., and Anton, J. (2007) Microbial community associated with the colonial ascidian *Cystodytes dellechiajei*. *Environ. Microbiol.*, **9**, 521-534.

Ostle, A.G., and Holt, J.G. (1982) Nile Blue A as a fluorescent stain for poly-β-hydroxybutyrate. *Appl. Environ. Microbiol.*, **44**, 238-241.

Petursdottir, S.K., and Kristjansson, J.K. (1997) *Silicibacter lacuscaerulensis* gen. nov., sp. nov., a mesophilic moderately halophilic bacterium characteristic of the Blue Lagoon geothermal lake in Iceland. *Extremophiles*, **1**, 94-99.

Posada, D., and Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J., and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.*, **35**, 7188-7196.

Rathgeber, C., Yurkova, N., Stackebrandt, E., Schumann, P., Beatty, J.T., and Yurkov, V. (2005) *Roseicyclus mahoneyensis* gen. nov., sp. nov., an aerobic phototrophic bacterium isolated from a meromictic lake. *Int. J. Syst. Evol. Microbiol.*, **55**, 1597-1603.

Rowe, J.M., Dunlap, J.R., Gobler, C.J., Anderson, O.R., Gastrich, M.D., and Wilhelm, S.W. (2008) Isolation of a non-phage-like lytic virus infecting *Aureococcus anophagefferens*. *J. Phycol.*, **44**, 71-76.

Sass, H., Kopke, B., Rutters, H., Feuerlein, T., Droge, S., Cypionka, H., and Engelen, B. (2009) T*ateyamaria pelophila* sp. nov., a facultatively anaerobic alphaproteobacterium isolated from tidal-flat sediment and emended description of *Tateyamaria omphalii*. *Int J Syst Evol Micr*, ijs.0.013524-013520.

Shiba, T. (1991) *Roseobacter litoralis* gen. nov., sp. nov., and *Roseobacter denitrificans* sp. nov., aerobic pink-pigmented bacteria which contain bacteriochlorophyll *a*. *Syst. Appl. Microbiol.*, **14**, 140-145.

Slightom, R.N., and Buchan, A. (2009) Surface colonization by marine roseobacters: Integrating genotype and phenotype. *Appl. Environ. Microbiol.*, **75**, 6027-6037.

Stackebrandt, E., and Goebel, B.M. (1994) A Place for DNA-DNA Reassociation and 16s Ribosomal-RNA Sequence-Analysis in the Present Species Definition in Bacteriology. *Int. J. Syst. Bacteriol.*, **44**, 846-849.

Sukoso, and Sakata, T. (1996) Effect of co-existent bacteria on the growth of *Chattonella marina* in non-axenic culture. *Fisheries Sci.*, **62**, 210-214.

Sukoso, Iwamoto, K., Sakata, T., and Yoshikawa, T. (1998) Characteristics of filamentous bacteria co-existing with some marine microalgae. *Fisheries Sci.*, **64**, 65-70.

Suzuki, T., Mori, Y., and Nishimura, Y. (2006) *Roseibacterium elongatum* gen. nov., sp. nov., an aerobic, bacteriochlorophyll-containing bacterium isolated from the west coast of Australia. *Int. J. Syst. Evol. Microbiol.*, **56**, 417-421.

Suzuki, T., Muroga, Y., Takahama, M., and Nishimura, Y. (1999) *Roseivivax halodurans* gen. nov., sp. nov. and *Roseivivax halotolerans* sp. nov., aerobic bacteriochlorophyll-containing bacteria isolated from a saline lake. *Int. J. Syst. Evol. Microbiol.*, **49**, 629-634.

Swofford, D. (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.0 b10. *Sinauer, Sunderland, Massachusetts, USA*.

Van Trappen, S., Mergaert, J., and Swings, J. (2004) *Loktanella salsilacus* gen. nov., sp. nov., *Loktanella fryxellensis* sp. nov. and *Loktanella vestfoldensis* sp. nov., new members of the *Rhodobacter* group, isolated from microbial mats in Antarctic lakes. *Int. J. Syst. Evol. Microbiol.*, **54**, 1263-1269.

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., and Truper, H.G. (1987) Report of the ad-hoc-committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.*, **37**, 463-464.

Yoon, J.H., Kang, S.J., Park, S., Oh, K.H., and Oh, T.K. (2010) *Jannaschia seohaensis* sp. nov., isolated from a tidal flat sediment. *Int. J. Syst. Evol. Microbiol.*, **60**, 191-195.

Yoshikawa, T., Iwamoto, K., and Sakata, T. (2001) Pigment compositions and phylogenetic positions of filamentous bacteria coexisting in marine microalgal cultures. *Microbes and Environ.*, **16**, 59-62.

Yutin, N., Suzuki, M.T., and Beja, O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl. Environ. Microbiol.*, **71**, 8958-8962.

Zhao, Y.L., Wang, K., Budinoff, C., Buchan, A., Lang, A., Jiao, N.Z., and Chen, F. (2009) Gene transfer agent (GTA) genes reveal diverse and dynamic *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay. *ISME J.*, **3**, 364-373.

Appendix

Tables

Table 3.1. Selected differential characteristics between strain CB1052^T, *M. cryptomonadis* and *M. litorea*.

Taxa: 1, CB1052^T; 2, *M. cryptomonadis*; 3, *M. litorea*. +, Positive; -, negative or absent; +/-, weakly positive; NS, not sensitive; S, sensitive; tr, trace. With the exception of G+C content, all data were determined in this study.

Table 3.1

Characteristic	1	2	3
Cell size (μm):			
Width	0.50	0.40-1.2	0.30-0.9
Length	3.0-15.0	1.9-3.5	1.0-3.5
Colony color	Faint pink	Creamy	Creamy
Colony shape	Dendritic	Round	Round
Bchl[b]*a* production	+	-	-
Growth at:			
pH 10.0	-	+	-[a]
Salinity Tolerance			
10%	-	+	-[a]
Utilization of:			
Trehalose	-	+/-	-
Glycerol	+/-	+	+
Glycolate	+	+	-
Quinate	+/-	-	-
Nitrate reductase	-	+[a]	+[a]
Antibiotic sensitivity			
Ciprofloxacin	S	NS	S
Gentamycin	NS	S	S
Major fatty acids (>1%)			
$C_{16:0}$	1.8	2.4	2.8
$C_{18:0}$	3.5	1.9	2.6
$C_{18:1}$ ω7c	68.8	64.7	68.9
$C_{10:0}$ 3-OH	-	-	1.1
$C_{12:1}$ 3-OH	4.7	4.9	4.7
$C_{12:0}$ 3-OH	-	-	-
iso-$C_{18:0}$	12.6	12.7	10.9
11-Methyl $C_{18:1}$ ω7c	5.6	10.4	9.0
DNA G+C content	59.6	58.6	61.0

[a] Data is contrary to that reported in Hwang *et. al*. (2009)

b Bchl, Bacteriochlorophyll

Table 3.2. Species of the *Roseobacter* lineage showing evidence of aerobic anoxygenic photosynthesis

Species	Source	Bchl *a*	Reference
Roseobacter denitrificans	Coastal (*Enteromorpha* sp., Japan)	(+)	(Shiba, 1991)
Roseobacter litoralis	Coastal (seaweed, Japan)	(+)	(Shiba, 1991)
Roseovarius tolerans	Saline lake (Ekho Lake)	(+)[a]	(Labrenz et al., 1999)
Roseivivax halodurans	Saline lake (stromatolite, Lake Clifton)	(+)	(Suzuki et al., 1999)
Roseivivax tolerans	Saline lake (Charophytes Lake Clifton)	(+)	(Suzuki et al., 1999)
Sulfitobacter guttiformis	Saline lake (Ekho Lake)	(+)[a]	(Labrenz et al., 2000)
Loktanella vestfoldensis	Saline Lake (microbial mat, Ace Lake)	ND[b]	(Van Trappen et al., 2004)
Loktanella fryxellensis	Saline Lake (microbial mat, Lake Fryxell)	ND[b]	(Van Trappen et al., 2004)
Roseovarius mucosus	Algal culture (*Alexandrium* sp.)	(+)[a]	(Biebl et al., 2005a)
Dinoroseobacter shibae	Algal culture (*Alexandrium* sp.)	(+)	(Biebl et al., 2005b)
Thalassobacter stenotrophicus	Coastal (seawater, Mediterranean Sea)	(+)	(Mácian et al., 2005)
Roseisalinus antarcticus	Saline lake (Ekho Lake)	(+)	(Labrenz et al., 2005)
Roseicyclus mahoneyensis	Saline lake (Mahoney Lake)	(+)	(Rathgeber et al., 2005)
Roseibacterium elongatum	Coastal (tidal sand, Sharks Bay)	(+)	(Suzuki et al., 2006)
Tateyamaria pelophila	Coastal (tidal sediment, North Sea)	(+)	(Sass et al., 2009)
Marivita cryptomonadis	Algal culture (*Cryptomonas* sp.)	(-)[c]	(Hwang et al., 2009)
Marivita litorea	Coastal (seawater, Korea)	(-)[c]	(Hwang et al., 2009)
Marivita roseacus (CB1052^T)	Coastal (seawater, Chesapeake Bay)	(+)[a]	Present study
Jannaschia seohaensis	Coastal (tidal sediment, Yellow Sea)	(+)	(Yoon et al., 2010)

[a] Authors reported low level or variable production of bchl *a*.

[b] No bchl *a* production detected, however AAP gene sequences (*pufM*) were amplified from these strains

[c] Colonies are pink-pigmented on MA 2216 agar plates and genome contains *puf* genes.

ND = not determined.

Figures

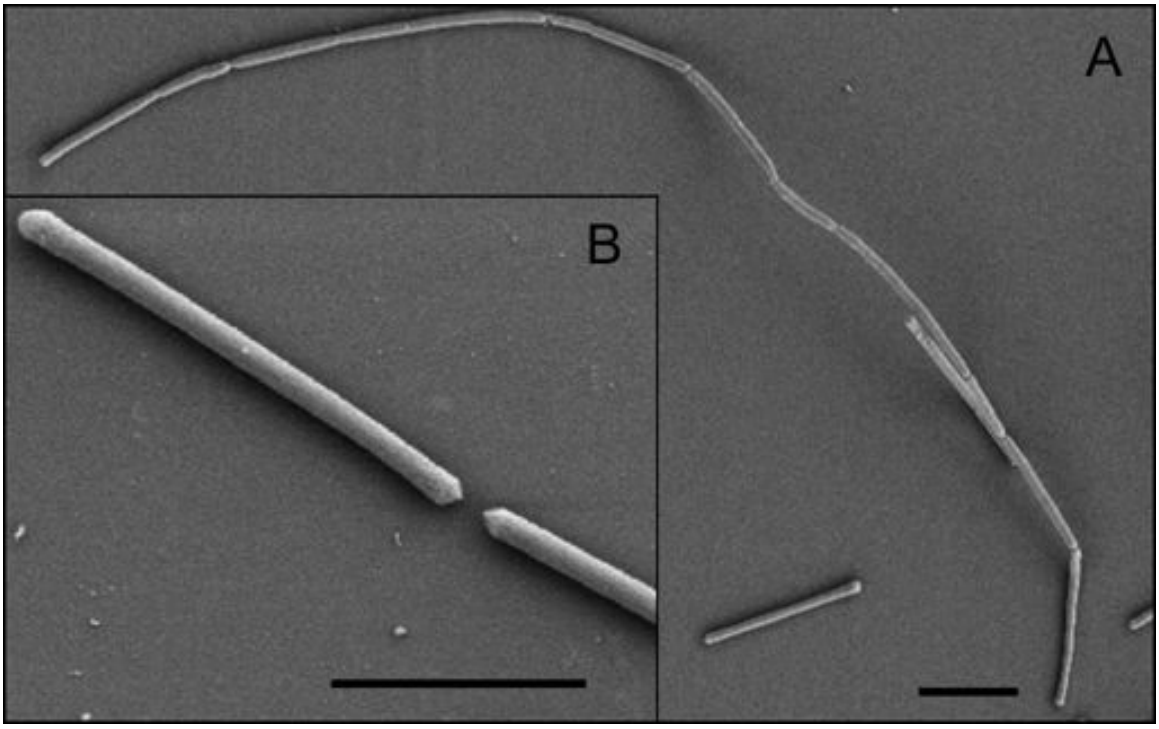


Figure 3.1. Scanning electron micrographs of CB1052^{T}. Scale bars are 3 µm.

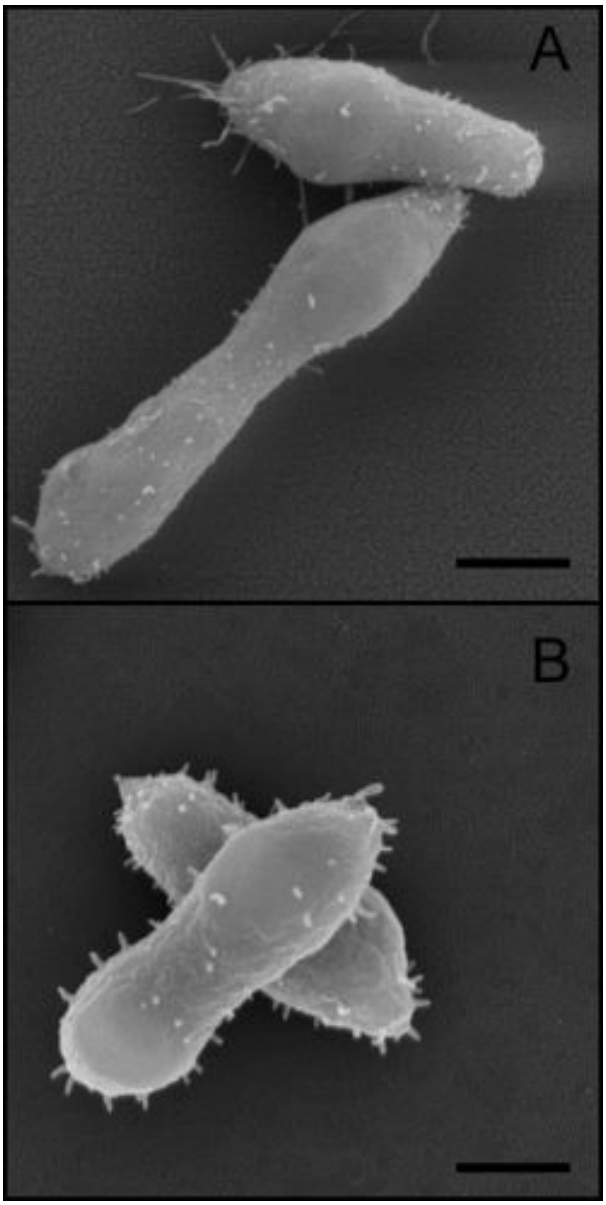


Figure 3.2. Scanning electron micrographs of (A) *M. cryptomonadis* and (B) *M. litorea.* Scale bars are 500 nm.

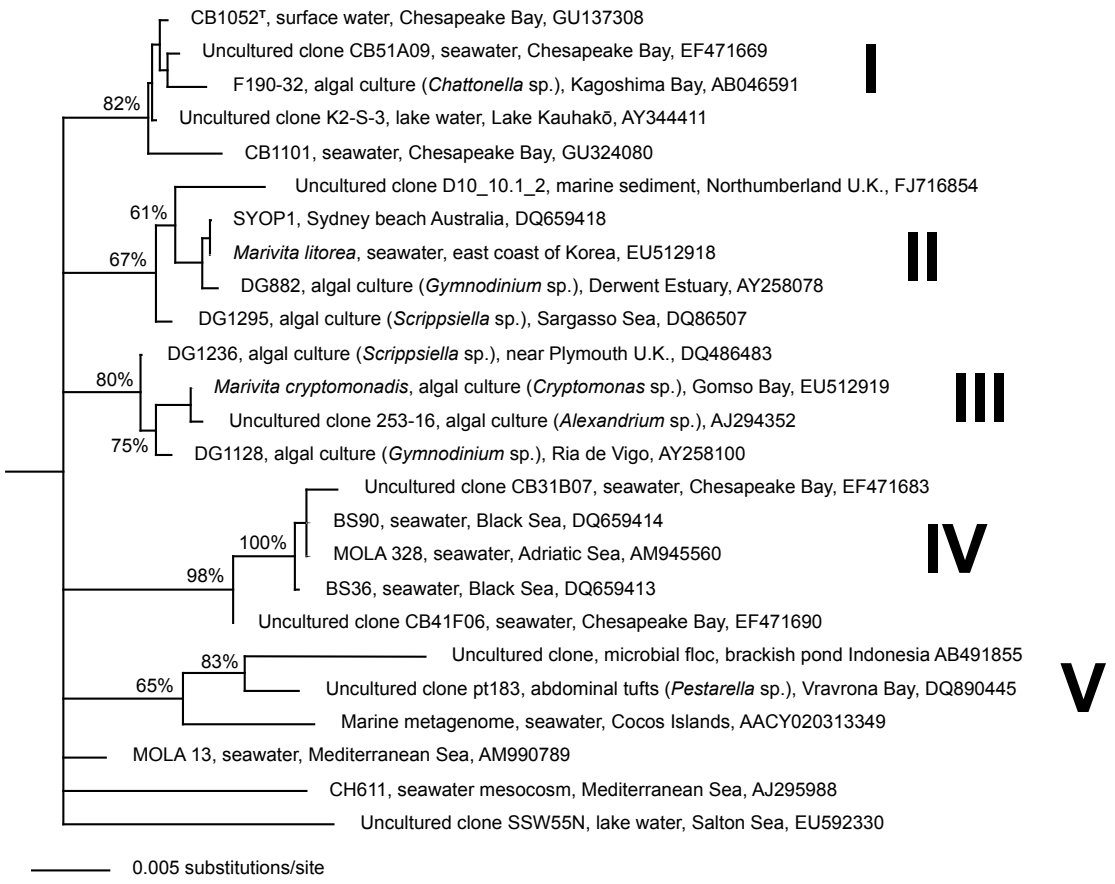


Figure 3.3. Neighbour-joining distance tree of the genus *Marivita*.

The tree was constructed using 1320 bp of the 16S rRNA gene in PAUP using the distance criterion coupled with a bootstrap analysis. Distance settings used; DNA substitution model was Tamura-Nei, all characters were resampled, 10,000 replicates were performed, the proportion of sites assumed to be invariable was 0.6908, and rates for the variable sites used a gamma shape parameter of 0.7109. Outgroup organisms (not shown) included *Rhodobacter sphaeroides* (DQ342321) and *Dinoroseobacter shibae* (AJ534211). Information listed: strain or clone ID, sample type (including coexisting organism, if applicable), source location, and GenBank accession number. Roman numerals denote well-supported clades discussed in text.

Supplemental Figures

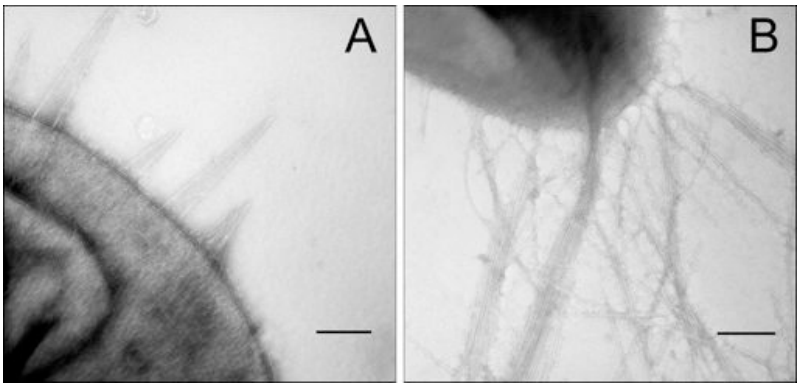


Figure 3.4. Negative stain transmission electron micrographs of (A) *M. litorea* and (B) *M. cryptomonadis*.

Scale bars are 250 nm.

Figure 3.5. Negative stain transmission electron micrographs of (A) *M. litorea*, (B) *M. cryptomonadis* and (C) CB1052^{T}.

Scale bars are 1 μm.

CHAPTER 4 –

HIGH PHYLOGENETIC RESOLUTION ANALYSIS OF BACTERIAL COMMUNITY COMPOSITION DURING A FJORD MESOCOSM STUDY DEMONSTRATES DIFFERENTIAL RESPONSES OF CLOSELY RELATED PHYLOTYPES

A version of this chapter will subsequently be published by Charles R. Budinoff, Gary R. LeCleir, Mary Hadden, Steven W. Wilhelm and Alison Buchan

CRB participated in the mesocosm experiment and sampling events, analyzed the sequencing data, performed the statistical analyses, and drafted the manuscript. MH performed PCR amplifications and barcoding reactions. GRL designed the PCR amplifications, barcoding reactions, and oversaw the sequencing of the samples. SWW participated in its design and coordination, and helped to draft the manuscript. AB participated in the design and coordination of the study, aided in the interpretation of the data, and helped to draft the manuscript.

Abstract

Microbial ecologists are beginning to utilize high-throughput pyrosequencing on a routine basis. The wealth of data obtained via such sequencing attempts has provided insight into the general distribution and richness of bacterial communities in the world's oceans. These investigations are typically of low resolution though, as many bacterial species are lumped together into broad taxonomic classes prior to analysis. By not examining deep-amplicon data sets at finer phylogenetic levels we risk overlooking valuable information concerning the functional potential of the community. To increase the phylogenetic separation of these data sets we coupled traditional automated workflows with robust tree building using samples obtained during a two week long mesocosm experiment that examined the dynamics of a coccolithophorid algal bloom (Bergen, Norway). Using the 16S rRNA v3 hypervariable region as our marker, we were able to separate over 20,000 bacterial sequences into distinct phylotypes at or near the species level. *Roseobacter*-like sequences were prevalent, accounting for 30%. Other bacteria included Flavobacteria (32%) γ-proteobacteria (20%), Verrucomicrobia (1.4%), and Actinobacteria (1.2%). The high resolution of the data set allowed us to chart the relative contribution of the individual community members throughout the bloom, revealing biodiversity patterns at both high and low taxonomic levels that showed a dynamic relationship amongst closely and distantly related phylotypes. Using physiological and other information about the nearest relative of the

obtained reads, we were able to construct a theoretical representation of the possible ecological niches or competitive responses of the individual phylotypes. Additionally, we validated the abundance of the distinct phylotype sequences using quantitative PCR (qPCR) with established ecotypes. Overall, the pyrosequencing data strongly correlated with the abundance trends observed with qPCR. Lastly, our analysis also allowed a comparison of two different nutrient regimes used to induce the mesocosm bloom, a phosphorous replete and a phosphorous deplete treatment. In general, the two bacterial communities were highly similar and the overall differences between the two were mainly a result of the less abundant taxa, including micro-clades of certain dominate phylotypes.

Introduction

Coccolithophorid algal blooms, particularly those of the species *Emiliania huxleyi,* are an important component of the world's oceans, contributing significantly to carbon and sulfur cycling at regional scales (Brown and Yoder 1994), especially at higher latitudes. Coccolithophores are ubiquitous microalgae named for their characteristic calcium carbonate scales called coccoliths, which surround the surface of individual cells. Coccoliths are resistant to dissolution leading to their accumulation in seafloor sediments, acting as a long-term carbon sink (Westbroek et al. 1993). These algae also create large amounts of the osmolyte dimethylsulfoniopropionate (DMSP) during bloom events (Malin et al. 1993; Matrai and Keller 1993), which can be converted to the gas dimethylsulfide (DMS) through a variety of processes. DMSP and DMS are remarkably versatile substrates that are rapidly acted upon by bacterioplankton (Kiene and Linn 2000; Kiene et al. 2000). Bacteria can break down DMSP by using pathways that either liberate DMS gas or utilize a demethylation step that results in sulfur incorporation into biomass. Marine inputs of DMS are a significant source of sulfur for the atmosphere, where it oxidizes and provides cloud condensation nuclei which could affect regional albedo and influence local climate (Charlson et al. 1987). Marine algae are also a consistent supply of other forms of dissolved organic matter and nutrients that contribute to sustaining a large microbial population. The microbial food web can control the flow of this

carbon either through remineralization, and the subsequent release of CO_2 into the atmosphere, or by predation, where energy in transferred up the trophic ladder. These processes can influence the amount of carbon exported through sedimentation to the seafloor.

Given the major role bacteria play in transforming the organic matter produced by algae and the subsequent impact theses transformations can have on biogeochemical cycling, it is important that we are able to quantify the rates of microbial driven processes. This includes not only transformations of organic and inorganic matter, but also of the mechanisms behind microbial mortality. Central to this idea is determining the composition of the microbial communities associated with bloom events. Bacteria are one of the most genetically diverse domains of life making them extremely versatile and adaptable. Understanding the role(s) that individual bacterial species play in bloom formation, peak, and eventual collapse, will lead to a better understanding of the forces controlling the cycling of climate influencing compounds.

Bacteria associated with coccolithophorid blooms and DMSP degradation in marine surface waters are phylogenetically diverse and highly abundant. Representative taxa include α-*proteobacteria*, γ-*proteobacteria*, and *Flavobacteria*. Initial molecular diversity studies were performed with methods such as clone library sequencing and fluorescence in situ hybridization (FISH). FISH allows for the quantification of particular bacterial groups, but without *a priori* knowledge of community membership the probes used are often phylogenetically broad, where the bacteria are lumped into mostly family level classifications (Malmstrom et al. 2004; Øvreås et al. 2003; Vila et al. 2004; Zubkov et al. 2001). Clone libraries from bulk samples (González et al. 2000) or of sorted cells (Mou et al. 2005) have supplied actual sequence data, thus increasing taxonomic resolution. For example, clone library sequencing of naturally occurring coccolithophorid blooms and of DMSP and DMS amended microcosms have indicated certain phylotypes or species are associated with bloom events or after substrate addition (González et al. 2000; Pinhassi et al. 2005; Zubkov et al. 2002). But due to the cost and labor of sequencing individual clones these libraries have usually been small in overall size, preventing any temporal or statistically significant analysis of diversity and abundance. In general, only a couple of the most abundant members are monitored in the system due to low sequencing depth. This short fall

is not reserved to just coccolithophorid blooms but a component of many studies of natural bacterial communities.

Deep-amplicon sequencing of variable regions in the 16S rRNA gene, along with metagenomic and metatranscriptomic approaches, are now often applied to marine systems. These high-throughput 2nd generation sequencing technologies have the ability to generate hundreds of thousands of sequences, revealing a high resolution of gene diversity and abundance. A limited number of researchers have applied these approaches to marine algal blooms (Andersson et al. 2010; Gilbert et al. 2008; Mitra et al. 2010) and seawater amendment experiments using DMSP (Vila-Costa et al. 2010). These studies, like most 2nd generation sequencing efforts, examined diversity patterns at broad taxonomic scales and did not present a thorough analysis of species level changes within the systems studied. Although the variable regions of the 16S rRNA gene have been shown to have high discriminatory power, to at least the genus level and arguably to the species level (Chakravorty et al. 2007; Liu et al. 2008), deep-amplicon studies rarely go beyond family level classifications, grouping many genera and species together in the analysis. This is a direct result of the reliance on automated sequence processing pipelines, specifically the taxonomic classifiers employed. Automated 16S rDNA sequence classifiers are limited by the small reference libraries they use for species identification, which are based on cultured organisms. Thus, depending on the environment in which the sequences were obtained, automated classifiers may leave the majority of sequences unidentified at the genus level (Liu et al. 2008) and for some bacterial lineages at the family level. In this study, we manually classify our 16S rDNA sequences to the species or near-species level by building a custom reference sequence database. This provided a high-resolution description of the bacterial communities associated with a mesocosm-based algal bloom.

Mesocosm experiments have been conducted on the western coast of Norway for over 20 years. Mesocosms designed to monitor the development and eventual collapse of algal blooms have been used to address specific questions and test a variety of hypotheses related to chemical, physical, and biological factors. Their utility to marine scientists lies mainly in their accessibility, reliability and reproducibility. Initial studies focused on the influence of nutrient loading on the growth and species composition of phytoplankton (Egge and Aksnes 1992; Jacobsen et al. 1995),

as well as determining the causes of algal bloom collapse (Bratbak et al. 1993). Norwegian phytoplankton communities stimulated with nutrients are often characterized by species succession (Jacobsen et al. 1995; Larsen et al. 2001; Paulino et al. 2008). Phytoplankton taxa are typically comprised of picocyanobacteria, diatoms (*Skeletonema* sp., *Thalassiosira* sp.), and pico- and nanoeukaryotic algae, which include cryptophytes, and members of the Prymnesiophyceae (*Emiliana huxleyi*, *Phaeocystis pouchetti*) and Prasinophyceae (*Micromonas* sp., *Pyramimonas* sp.). Succession patterns and overall dominance of the phytoplankton groups within a mesocosm experiment depends largely on the added nutrients and the composition of the seedling water mass.

Nutrient additions to Norwegian coastal waters typically result in initial increases of diatom and pico- and nanoeukaryotic algae (not including *E. huxleyi*) (Jacobsen et al. 1995; Nejstgaard et al. 1997) with no change or slight decreases in picocyanobacterial numbers (Muller et al. 2005; Paulino et al. 2008). It is theorized that the larger eukaryotic algae, in particular diatoms, outcompete the picocyanobacteria under nutrient rich conditions. Diatoms have been shown to generally have higher assimilation rates of nitrate than picocyanobacteria and nano- and picoeukaryotic algae, especially under colder temperatures (Lomas and Glibert 2000; Reay et al. 1999) and often dominate natural Norwegian coastal waters (Larsen et al. 2004) and mesocosms (Jacobsen et al. 1995). For diatoms to become dominate in these systems, silicate must be plentiful, and low concentrations will limit diatom growth and allow succession by other algae (Egge and Aksnes 1992). Eventually the abundance of the succeeding nano- and picoeukaryotic algae declines too, likely a result of mortality via protistian grazing and virus infection. Microzooplankton increase during the first few days of a typical mesocosm bloom and heterotrophic protists have been shown to reach maximal abundance during the height of the early nano- and picoeukaryotic algae peak and their largest biomass a few days after algal decline (Nejstgaard et al. 1997). Types of grazers in the mesocosms include various ciliates, choanoflagellates and dinoflagellates. Direct estimates of protistian grazing on *Micromonas* spp. during a mesocosm bloom showed that the bulk of algal mortality was a result of grazing (Evans et al. 2003). Estimates of bacterivory during these types of blooms are not available, but given the protistan species composition it is reasonable to assume that bacteria are being grazed upon

as well. Microzooplankton grazing is well known to produce particulate organic matter through the digestion of phyto- and bacterioplankton, which is subsequently colonized and further degraded by bacteria. Microzooplankton are also an important component in the diet of copepods during these blooms. Copepod abundance quickly increases after the rise in microzooplankton, further contributing to particulate organic matter accumulation (Nejstgaard et al. 1997).

Zooplankton grazing has been shown to release organically bound metals that were contained within prey cells (Hutchins and Bruland 1994; Sato et al. 2007) and bacteria from Norwegian coastal waters and mesocosm experiments have been implicated in the sequestration of metals after grazing events and are thought to influence metal bioavailability and toxicity. (Heldal et al. 1996; Muller et al. 2003; Muller et al. 2005). This nutrient regeneration stage of the bloom is thought to trigger the rapid growth of *E. huxleyi*, leading to another algal succession event. Phytoplankton grazing continues at this point, but the availability of organic and inorganic nutrients are too great and *E. huxleyi* concentrations grow substantially. Instead, mesocosm blooms of *E. huxleyi* are often terminated by viral lysis (Bratbak et al. 1996; Jacquet et al. 2002). Lysis of *E. huxleyi* produces a variety of cellular debris and a plethora of dissolved organic matter (including DMSP and DMS) that is immediately available to the bacterial community.

Studies examining bacterial diversity associated with Norwegian mesocosm blooms have been performed (Castberg et al. 2001; Løvdal et al. 2008; Øvreås et al. 2003). These studies have increased our knowledge of the microbial responses occurring in these experiments and have revealed that the bacterial community changed through the course of the bloom. But due to (the before mentioned) methodological constraints in measuring microbial biodiversity, we still lack a detailed picture of bacterial species composition and succession events. A recent mesocosm experiment designed to investigate the role of stress and viral infection on coccolithophores (Kimmance et al. 2011) allowed us an opportunity to determine the composition of the bacterial community from bloom formation to collapse through the pyrosequencing of the v3 region of the 16S rRNA gene. We present this data in the context of observed phytoplankton succession patterns and allude to the ecophysiology of the representative bacteria.

Materials and Methods

Sample collection, phytoplankton abundance, and DNA extraction

The mesocosm experiment was carried out in Raunefjorden, located on the western coast of Norway at the Marine Biological Field Station, Espeland, from June 2nd through June 25th 2008. Specific details of the experiment including mesocosm set-up, nutrient additions, and sampling regime are described in detail elsewhere (Kimmance et al. 2011). Briefly, mesocosm enclosures (11 m^3) were subject to daily nutrient additions in a N:P ratio of 15:1 (1.5 µmol L^{-1} $NaNO_3$ and 0.1 µmol L^{-1} KH_2PO_4) to the P-replete enclosures and at a ratio of 75:1 (1.5 µmol 128 L^{-1} $NaNO_3$ and 0.02 µmol L^{-1} KH_2PO_4) to the P-deplete enclosures. Samples were taken every six hours for a nutrient concentrations (N and P), phytoplankton and bacterial abundance were determined with flow cytometry as previously described (Kimmance et al. 2011) and DNA extraction was carried out with the DNAeasy Tissue Kit (Qiagen). Reference water samples were also taken from the unamended fjord water at equivalent times

PCR amplification and pyrosequencing

PCR amplification of the v3 region of the 16S rRNA gene was accomplished using the primers 338F and 533R (Huse et al. 2008). The 156-bp fragments (after primer removal) were amplified using Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). PCR products were purified with the Qiaquick PCR cleanup kit (Qiagen, Valencia, CA). A second round of PCR (6 cycles) was performed on the individual PCR products to add 454 fusion primers and multiplex identifiers (MIDs) to each sequence (Hamady et al. 2008, see Table A1). Amplification products were pooled, purified using the QiaQuick PCR cleanup kit and prepared for pyrosequencing with 454 FLX chemistries according to manufacturer's protocols (454 Life Sciences, Branford, CT) at the UTK/ORNL Joint Institute for Biological Sciences (Oak Ridge, TN). Pyrosequencing flowgrams were converted to sequence reads and subjected to quality control filtering using software provided by 454 Life Sciences. Further quality control was done using the processing commands (i.e. trim.seqs) within the program Mothur v1.13 (Schloss et al. 2009). A library of 20,327 high quality bacterial sequences (excluding reads belonging to cyanobacteria and

chloroplasts) was used for subsequent analysis, about 2,030 per sub-library (or sample) on average.

Quantitative PCR validation

Quantitative PCR (qPCR) assays for the Roseobacter Clade-Associated (RCA) and the CHAB-I-5 subgroups (see Buchan et. al (2005) for descriptions of these groups) were applied to DNA from the sample sets following previously described methods (Buchan et al. 2009). In addition to the Roseobacter specific primers, general bacterial 16S rRNA primers (Suzuki et al. 2000) were also applied to the field samples. Each DNA sample was analyzed in duplicate reactions of at least two dilutions ranging from 0.5 to 0.05 X in molecular grade water. To each 25 µl qPCR mixture, 2.5 µl of diluted DNA was added and qPCR performed as described above. For each run, a standard curve was determined by analyzing a dilution series (10^1 to 10^5 gene copies per 25 µl reaction mixture) of the appropriate standard in duplicate. For each measurement, a standard deviation of a minimum of two duplicates was determined. Theoretical amplification efficiencies were calculated from the slope of the standard curve using the equation $10^{-1/slope}$ -1. Standard curves were determined as the correlation between the log of gene copy numbers and the cycle threshold (C_T). In all cases, correlation coefficients for standard curves assays were above 0.98. To provide the data in a relevant frame of reference, subgroup abundances were expressed as relative contributions of the total bacterial 16S rRNA gene pool.

Phylogenetic tree building and species level classification

Sequence alignment and neighbor-joining tree construction were performed within the Mothur software suite. Sequences were aligned using the Silva reference database (v102) (Pruesse et al. 2007) and relaxed neighbor joining trees were constructed with the Clearcut (Sheneman et al. 2006) implementation. Aligned sequences and neighbor joining trees were imported into the program ARB (www.arb-home.de) and clades were collapsed into groups and representative sequences submitted to the Classifier tool of the Ribosomal Database Project (RDP) for taxonomic assignment (Wang et al. 2007). Sequences belonging to individual bacterial families

or classes were subjected to an additional alignment, where they were exported from ARB and re-aligned using RDP's pyrosequencing aligner, which is based on the Infernal program (Nawrocki et al. 2009). Infernal uses a model-based approach to align 16S rDNA sequences using both base pair and RNA secondary structure consensuses. Because the aligner relies on secondary structure models hyper-variable regions found in the 16S rRNA gene that do not follow a known model are not aligned. To align these regions the sequences underwent a pair-wise alignment in ClustalW-MPI (Li 2003). Each sequence alignment of the separated bacterial families/classes was manual checked for errors and then re-imported individually into ARB where neighbor-joining trees were constructed. Clades within the neighbor-joining trees were collapsed into groups again and reference sequences were identified using RDP's SeqMatch tool and imported into ARB. The number of sequences associated with each phylotype in the family level trees for the individual samples were recorded and used as species abundance data in the statistical analyses. Sequences belonging to cyanobacteria and chloroplasts were excluded from all analyses. Also, only phylotypes containing greater than 15 sequences (0.07% of the complete library) were considered. Lastly, a Linnaean taxonomy table was created for the phylotypes that defines which genera, families, orders, etc. that each species belongs to. This table is referred to as an aggregation file and is used in the below statistical analyses.

OTU clustering and analysis within mothur

To generate OTU clusters, sequences were first aligned against the Silva database in Mothur and then clustered using the furthest neighbor algorithm. A matrix was constructed from the uncorrected pair-wise distances between sequences. OTU abundances using multiple distance cut-off values (0.05-0.02) were compared to the phylotype abundance data to assist in choosing a distance cut-off value for subsequent analysis. The goal was to identify a distance value that produced OTUs most similar to the phylotypes by comparing the number of sequences in each OTU with those in the corresponding phylotype. Based on these results (see below) a distance cut-off of 0.044 was chosen for all subsequent commands. Further analyses within Mothur included the generation of rarefaction and collectors curves, identification of sequences shared

across sub-libraries, and tests of community structure using the weighted and unweighted UniFrac algorithms (Lozupone and Knight 2005). In addition, sequences belonging to individual bacterial families identified within ARB (see above) were also subjected to these analyses, where they were re-aligned and re-clustered individually using the same distance cut-off value.

Statistical analysis

Species abundance data generated with ARB and mothur were imported into the software program PRIMER v6 (PRIMER-E Ltd, Plymouth Marine Laboratory, UK) along with the phyto- and bacterioplankton abundance data, as well as Unifrac distances. OTUs from mothur were divided into rare and abundant taxa using the number of species identified in ARB. There were 95 total phylotypes (species) recorded in ARB, which included all phylotypes with greater than 15 total reads, so the 95 most abundant OTUs were considered the abundant sequences and the remaining OTUs considered the rare. This method for separating the OTUs was preferred over setting a cutoff value (e.g. 0.07%) so as to maintain the same number of taxa when comparing biodiversity measures between the abundant OTUs and the phylotypes. Although, sequence compositions after separation using a set cutoff value were not much different in terms of the number of OTUs or number of sequences present (see results). Lastly, Spearman rank correlations (SPSS ,2-tailed, t-test) were performed using standardized (by total number of sequences in each sample) phylotype abundances and the phyto- and bacterioplankton abundance data.

Within PRIMER an extensive set of biodiversity indices were computed with the abundance data including taxonomic distinctness measures (making use of the above mentioned aggregation file). The taxonomic distinctness measures, Average Taxonomic Diversity (AvTD) and Variation in Taxonomic Distinctness (VarTD), were calculated using equal weights for all step lengths between taxonomic levels as well as using weights defined by taxon richness, as described by (Clarke and Warwick 1999). The same trends in the data were seen with both weighting methods and the data presented use weights based on taxon richness. AvTD and

VarTD were also tested for departure from expected distinctness using a univariate and bivarient analysis.

Abundance data was subjected to an array of pre-treatment options within PRIMER, including standardizations and power transformations. To characterize species abundance patterns a variety of resemblance matrices were constructed across samples (sub-libraries) and variables (phylotypes), including zero-adjusted Bray-Curtis similarities and correlation (Spearman rank and Pearson) similarities. These resemblance matrices as well as UniFrac distances (exported from mothur) were subjected to non-metric multi-dimensional scaling (MDS) ordination and agglomerative, hierarchical clustering using the group average linkage method. To measure the similarity amongst the multivariate data the above resemblance matrices were subjected to the RELATE routine, which calculates a non-parametric form of the Mantel test. Matrices were compared amongst themselves where appropriate and also to a seriation model.

Results

Phyto- and bacterioplankton abundances throughout the bloom are presented in Figure 4.1. As described by Kimmance et al (2011) the data can be divided into three characteristic temporal phases. Briefly, phase 1 was between days 2-7 and was characterized by an increase and subsequent decline of pico- and nanoeukaryotes (not *E. huxeleyi*) as well as a gradual decrease in bacterial and *Synechococcus* concentrations. Phase 2 was between days 7 and 13, and consisted of an exponential increase in both the bacterial and coccolithophorid populations. The bacterial community quickly collapsed by day 12. Other photosynthetic organisms also gradually increased during this time. Phase 3, days 13-17, are characterized by the collapse of the *E. huxeleyi* population and a concurrent exponential rise in the other photosynthetic populations. In general, the phosphorous replete and deplete treatments showed the same population trends, although differences are apparent, most notably in the overall abundances, which were considerably higher in the replete bags.

OTU and phylotype abundances

OTU abundances calculated with multiple distances were compared to the phylotype abundance data to assist in choosing the most accurate distance cutoff value for clustering. We calculated the number of 'misclustered' sequences, which is the difference in the number of sequences between an OTU and the corresponding phylotype. We found that individual OTUs most closely matched the phylotype abundances at different distances. In other words, a distance that accurately clustered one species was not necessarily the ideal distance for another species (see Supp. Table 4.2). The distance cutoff value used subsequent analysis produced the least amount of misclustered sequences, which was 0.044 or 95.6%. This distance was also found to be ideal when clustering individual groups at the class or family level separately (data not shown). At distances greater than 0.04 most of the OTUs contained more sequences than their corresponding phylotype (ARB) and a distance of 0.03 most contained far less than expected, leading to an over estimate of sample richness.

To assess the overall distribution of the sequence reads and OTUs amongst the deplete, replete, and fjord samples a venn diagram of shared sequences was constructed (Supp. Fig. 4.10). Approximately 84% of all sequences were shared amongst the samples and over 93% were shared between the replete and deplete treatments. Using the RDP classifier the distribution of the abundant bacterial lineages was graphed over time (Fig. 4.2A). The major bacterial classes identified were the α-proteobacteria, γ-proteobacteria, and Flavobacteria. Classification statistics are reported in Table 4.1 and show that the majority of the sequences were left unclassified beyond the family level. In particular, γ-proteobacteria were typically not classified beyond the class level. The majority of the α-proteobacteria were termed 'unclassified Rhodobactereace' (Fig. 4.2B), which showed the highest abundance in the beginning and at the end of the experiment. Flavobacteria appeared to peak towards day 11 and the γ-proteobacteria remained relatively constant throughout.

Sampling coverage of the 454 data is described through rarefaction and species accumulation curves for the fjord sample (Supp. Fig. 4.11). The fjord sample showed the highest richness and diversity out of all the samples (see below). All the other sub-libraries produced very similar curves though. The rarefaction curve (Supp. Fig. 4.11A) does not reach saturation,

but the collectors' curves based on the Shannon index (Supp. Fig. 4.11B) and the Chao 1 estimator (Supp. Fig. 4.11C) do. Furthest neighbor clustering at a distance of 0.044 produced a total of 718 OTUs for the entire library. On average, each sub-library contained 198 OTUs (SD 65). Singleton sequences represented 41.3% of the OTUs, doubletons 12.9%, and 26.6% had between 3 and 10 sequences each.

Phylotype analysis using a combination of ARB and the RDP database allowed for a robust, manual sequence classification. Phylotypes containing 15 or greater sequences were classified to the species or near species level, which contributed to almost 89% of the entire library (Table 4.1). The α-proteobacteria, comprised mainly of the family Rhodobacteriaceae (84%), showed the highest average number of sequences per phylotype, having a large distribution, where the maximum sequences per phylotype was twice as high as the second largest, the Flavobacteria, comprised mainly of the family Flavobacteriaceae (86%). The γ-proteobacteria had the least number of sequences classified and also had the lowest average number of sequences per phylotype. Half of the entire library's reads came from the 12 most abundant phylotypes and their distribution over time in the deplete and replete treatments is shown in Figure 4.3. The abundances of all the phylotypes are depicted by a bubble heat map (Fig. 4.4).

Comparisons of the phylotype abundances and QPCR abundances of specific *Roseobacter* subgroups are shown in Figure 4.5. Overall the data from the two methods followed the same temporal trends, although the 454 data tended to be considerably higher at the upper extremes for each of the subgroups. Spearman rank coefficients for the RCA data were not as significant in the deplete treatment (0.800, p=0.104) as in the replete treatment (0.943, p=0.005). For the CHAB data, the coefficients were significant for both the deplete (0.900, p=0.037) and replete (0.943, p=0.005) treatments.

Biodiversity

Bacterial diversity, richness, and evenness were calculated for the phylotype and OTU (total, abundant, and rare, distance of 95.6%) abundance data (Fig. 4.6). Only the replete data is shown,

for the deplete treatment showed the same overall trends and in essence acted as a duplicate sample in interpretation of the biodiversity data, although, temporal shifts in the deplete treatment's biodiversity indices were muted overall when compared to the replete treatment (data not shown). Biodiversity measures were also calculated using a distance of 97% and these as well showed the same trends as shown in Figure 4.6, except for obviously having higher absolute richness values (data not shown). Unsurprisingly, biodiversity of the abundant OTUs matched almost exactly (except for a few data points) to that of the phylotypes.

Richness of the bacterial community was driven by the rare species, decreasing at the beginning of the bloom and reaching a minimum around day 8 (loss of >50%) before beginning to rise by day 14. Richness of the abundant species showed little if any decrease, except for γ-proteobacteria richness, which showed a slight decrease at day 5. Roseobacter richness actually rose slightly at day 2, which is in contrast with the Flavobacteria and γ-proteobacteria, whose richness decreased. Roseobacter richness began to rise sooner than that of the γ-proteobacteria, rebounding on day 11 opposed to day 14. In contrast, evenness of the bacterial community was driven by the abundant species, decreasing at the beginning and reaching a minimum around day 8 before beginning to rise by day 11 and then evening out by day 14. Evenness of the rare species showed little if any decrease, remaining essential flat throughout the experiment. As with richness, Roseobacter evenness rose at day 2, which is again in contrast with the Flavobacteria and γ-proteobacteria, whose evenness decreased. Roseobacter evenness only rose in the abundant species. Roseobacter evenness decreased significantly at the beginning, reaching a minimum around day 8 before rising back to its initial levels by day 14. γ-proteobacteria evenness decreased slightly, with a minimum at day 8.

Diversity was calculated with the Shannon and Simpson indices. As expected, the Shannon index was influenced most by sample richness, where evenness drove the Simpson index. The fjord sample (day 1) was the most diverse, with γ-proteobacteria diversity being the highest of the groups, followed by that of the Flavobacteria and then the Roseobacter. Bacterial diversity reached a minimum between day 5 and 8, as did the diversity of the individual bacterial groups. A funnel plot depicting simulated VarTD (Supp. Fig. 4.12A) indicates lower than expected values for days 5 and 8, reflecting unevenness in the taxonomic hierarchy. Simulated

AvTD (Supp. Fig. 4.12B) indicates that most samples fall within expected ranges, except for day 8 of the deplete sample, which showed higher than expected values. A bivariant eclipse plot of the simulated VarTD and AvTD summarizes these funnel plots showing that the data agree well with the traditional diversity indices and species abundance plots, showing a taxonomic shift and a loss of biodiversity through day 8, and a subsequent return similar to pre-entrapment levels (Fig. 4.7).

Multivariate analysis

MDS ordination of the phylotype abundance data was used to reflect the similarity amongst the species. Initial attempts failed to produce ordinations that were interpretable. This was based on comparisons with phylotype abundance plots and the within species correlations. It was found that matrices constructed based on Bray-Curtis resemblances between variables (species), using even the most severe of transformations (Log), clustered the species based on their total abundance in the library (data not shown). To alleviate this, resemblance matrices were constructed using Spearman rank correlations instead. Additionally, initial MDS ordinations had high stress values (>0.3). Ordination stress was reduced by simplifying the data prior to analysis, here the phylotype abundances were divided into groups based on their general abundance patterns observed in the two treatments. Phylotype patterns showing a single dominant peak in abundance that was observed in both treatments (type 1) were separated from phylotypes displaying multiple peaks or from those behaving differently in the two treatments (type 2).

MDS ordination of Spearman rank correlation resemblances of type 1 phylotypes clearly separated the species into clusters corresponding to the day on which they peaked (Figure 4.8) and produced stress values much lower (<0.2) than ordinations based on the entire data set. The species clusters could be classified by their correlations with the phyto- and bacterioplankton abundance data as well as to the biodiversity measures. The type 1 species represented 52% of all the phylotypes and 80% of all the phylotype sequences (70% of the entire library).

MDS ordination plots of the type 2 phylotypes were difficult to interpret due to the presence of many different abundance patterns and exhibited high stress values (>0.3). Resemblance matrices created with certain coefficients (Bray-Curtis, Spearman rank, etc) offered

no clear clustering of species (data not shown). MDS ordination using a resemblance matrix based on the Pearson correlation coefficient did however cluster some species into several sub-groups that were supported by correlation values greater than 0.70 (data not shown). Although, the stress of the ordination was still relatively high (>0.25) and many species remained unclustered. The type 2 phylotypes represented 45% of all the phylotypes and 19% of all the phylotype sequences (17% of the entire library). Overall, the determination of species-specific responses amongst the type 2 phylotypes was hampered by the great variation in abundance patterns, the low average number of reads per phylotype, and a lack of biological replicates for the two treatments.

Transformation of the abundance data was also required prior to the construction of Bray-Curtis resemblance matrices between samples (sub-libraries). Both fourth root and log transforms produced interpretable ordinations with low stress values (<0.1). Plots based on the phytoplankton and phylotype abundances (Fig. 4.9) as well as plots based on the OTU abundances and UniFrac distances (data not shown) all followed a linear time sequence. The two treatments (replete/deplete) were very similar (>70%) across all days. Plots based on the 454 data showed an overall lower similarity between samples than did plots based on the phytoplankton abundances (Fig. 4.8). Also, plots based on the 454 data showed a higher similarity between days 11 and 14 than did plots based on the phytoplankton abundances (Fig. 4.8). The matching (non-parametric Mantel) coefficients (rho, ρ) between the plots and the seriation model matrix were all highly significant indicating a good match. Ordinations including both the replete and deplete treatments (Fig. 4.8) had higher significance values ($\rho > 0.6$, $p < 0.01$) to the seriation model than ordinations of the replete treatment alone. A good portion of the differences between the replete and deplete treatments could be explained by the type 2 phylotypes. Their removal from the data set caused the similarity values to increase between the treatments (>90% similarity) (data not shown). Also, the type 1 phylotype ordinations lost aspects of their linear sequence (ρ decreased from 0.726 to 0.688), where the different treatments became more similar to each other then to the corresponding days (data not shown).

Discussion

Understanding the functional relevance of bacterial diversity is a fundamental goal of marine microbial ecology. The tremendous genetic and metabolic diversity seen within natural populations challenges our ability to connect the presence of individual taxa within a community to an ecological niche or functional role(s). Patterns in bacterial community functions and interactions are observed at both high and low levels of taxonomic organization (Acinas et al. 2004; Fuhrman and Campbell 1998; Philippot et al. 2010). In the ocean, certain marine bacterial lineages (above genus level) are known to be associated with particular ecosystems or assigned general physiological roles, notable examples include Flavobacteria (Alonso et al. 2007; Gómez-Pereira et al. 2010; Kirchman 2002), Roseobacters (Buchan et al. 2005; Mayali et al. 2008; Newton et al. 2010), certain γ–proteobacteria (Cho and Giovannoni 2004; Cho et al. 2007), and the ε-proteobactereria (Campbell et al. 2006). Also reported for many abundant marine taxa is microdiversity, where distinct population distributions are occurring at or below species level classifications (García-Martínez and Rodríguez-Valera 2000; Rocap et al. 2002; Selje et al. 2004). Interpreting the ecological relevance of these patterns is often hampered by the complexity of niche differentiation in bacteria, where species often overlap in distributions and their presence or even abundance does not always translate to activity. Microbial ecologists are further restricted by a lack of physiological information on many marine species and also by the limits of traditional molecular tools used to measure biodiversity. But with recent technological advances in single cell analysis and in genomic, metagenomic, metatranscriptomic, and deep amplicon sequencing efforts, we are poised to significantly advance our knowledge of ecophysiological diversity within microbial populations.

Here we used pyrosequencing of the 16S rRNA gene to produce a high-resolution description of the bacterial community associated with a mesocosm induced algal bloom. Bacterial patterns at both high and low taxonomic levels were observed to coincide with the succession of the phytoplankton populations. Taxonomic classification demonstrated that the bacterial community was dominated by the classes α-proteobacteria, Flavobacteria, and γ-proteobacteria, which over time showed obvious patterns, including dramatic shifts in dominance. This generalized picture of bacterial succession is easily obtained through the use of

an automated classifier (see Fig. 4.2A), but hides a great deal of the system's biodiversity and reveals little about the potential physiological or mechanistic factors leading to the success of the individual taxa. Through the creation of a custom reference database we maximized the classification resolution of our reads allowing succession patterns to be observed within genera and at times amongst sub-species.

Microdiversity

The genus *Polaribacter* for example, of the family Flavobacteriaceae, is often found in high-latitude waters and contributed significantly to our total library. Only half of the *Polaribacter*-like reads were assigned to the genus level by the automated classifier, the others were termed 'unclassified Flavobacteriaceae'. We were able to resolve six relatively abundant and distinct *Polaribacter* phylotypes, many displaying a different temporal pattern; *Polaribacter*-1 peaking on day 2, *Polaribacter*-2 peaking on day 11, and *Polaribacter*-4 having peaks on day 2 and 14 (see Supp. Fig. 4.13). The RCA clade, of the *Roseobacter* lineage, offers an example of possible sub-species level niche development. With no appropriately described isolate available, the sequencing reads belonging to the RCA clade were all auto-assigned to 'unclassified Rhodobacteriaceae', but through manual annotation we were able to not only identify reads belonging to RCA but also distinct sub-populations whom differed in their abundance patterns by a single base pair (Supp. Fig. 4.14). More work needs to be done to confirm that micro-clades are in fact distinct organisms, ruling out any possibility of interoperon variability or sequencing artifacts. But as mentioned above, similar observations are common amongst marine bacteria, and in this study the sequencing depth was high, capable of revealing many such relationships from a single source. Diversity was found within other abundant taxa from the mesocosm as well, including 3 phylotypes belonging to the *Pelagibacter* genus, 2 MS024-3C/*Cellulophaga*-like phylotypes, 2 species from the genus *Vibrio*, 3 species from the genus *Lacinutrix*, and 4 closely related phylotypes within the Verrucomicrobia. Not all related phylotypes displayed noticeably different abundance patterns, many similar organisms in fact behaved identical during the experiment, further obscuring the relevance of diversity and function. Nevertheless, a proper

phylogenetic classification of the rRNA reads was a necessary first step to fully utilize the potential of the pyrosequencing dataset. A thorough comparison of the classified reads with currently available public databases allowed us to determine specifically which bacteria were abundant at each stage of the mesocosm and to begin connecting phylogeny and function within the population.

Most of the bacteria living in marine environments are not amendable to culture (although some are, e.g. Roseobacters). Instead researchers have developed alterative cultivation approaches (Connon and Giovannoni 2002) and utilized single cell genomics (Stepanauskas and Sieracki 2007; Woyke et al. 2009) to provide metabolic information on some of the most abundant taxa. Cultivation, characterization and genomic sequencing provides a wealth of information about marine bacteria that can offer insight into the means and roles bacteria utilize to succeed in the environment. In this study we connected physiological and phenotypic characteristics of specific bacterial populations with their abundance over time. Using the relative abundances of the type-1 phylotypes we were able to cluster the majority of taxa by the day in which they peaked. In theory this places the phylotypes in groups based on when they were the most 'competitive' during the experiment. By examining the phylogenetic relationships of the species in these individual clusters, we attempted to identify commonalities in physiology, morphology, ecology, gene content, etc., to tease out the functional role or ecological niche these organisms held during the bloom. It should be noted though, that in reality the individual phylotypes were likely most competitive in the time leading up to their peak, where their rate of change in relative abundance was highest. But due to the temporal resolution of the dataset, their peak day was the most informative in this regard.

Day 2 and 5

Initial nutrient additions of the experiment caused a rise in diatom and pico- and nanoeukaryotic algal abundances along with a dramatic shift in the bacterial population. By the second day of the experiment, both evenness and richness of the bacterial community had decreased substantially (Fig. 4.6). The changes were driven by losses of rare OTUs from the γ-proteobacteria and the

Flavobacteria, as well as the loss of a number of abundant Flavobacteria phylotypes. These shifts corresponded to small increases in the contribution of both rare and abundant α-proteobacteria (mainly Roseobacters) to overall biodiversity. An increase in the abundance and diversity of α-proteobacteria after nutrient enrichment is not surprising. Roseobacters have recently been termed generalists, thriving under nutrient rich conditions through the utilization of a variety of genes involved in carbon metabolism and energy generation (Newton et al. 2010) and as mentioned previously are often associated with marine mesocosm studies. Upon inspection of the phylotype data, we see that although the α-proteobacteria appear to out compete the Flavobacteria overall during the first few days of the experiment (based on the automatic classifications, see Fig 2A), there are in fact Flavobacteria phylotypes that do increase during this time (Fig. 4.3 & 7). One notable example would be the Flavobacteria phylotype *Polaribacter*-1, that increased in relative contribution from 3% at time zero to almost 14% by day two (replete treatment), making the species the single most dominate phylotype at the time. Additionally, although the contribution of γ-proteobacteria remains stable, there is in fact community succession occurring, where specific γ-proteobacteria phylotypes increase substantially, some of which were considered novel, where no identical sequences were found in the database searched. The success of the phylotypes on days 2 and 5 may be the result of several factors, such as the utilization of organic compounds produced by the dominating algae, inorganic nitrogen/phosphorous assimilation efficiencies, and/or perhaps the possession of a photosystem.

Looking closely at the phylogenetic relationships of the sequences peaking on days 2 and 5 reveals that many of the phylotypes have similarity with bacteria shown to be associated with certain phytoplankton. Phylotype DG150 sp. clusters with isolates obtained from algal cultures (dinoflagellates, GenBank DQ486490) and diatom blooms (AM945591), RCA-1 groups with the *Roseobacter* Clade Affiliated cluster (RCA), which was shown to be associated with algal cultures (dinoflagellates) (Mayali et al. 2008) and decaying blooms (Giebel et al. 2011), and the phylotypes MS024-1C-like (similar to NAC11-7) and SAR92-like who were both shown to be enriched during a diatom bloom in a high-nutrient low-chlorophyll zone (West et al. 2008). Phylotypes peaking on days 2 and 5 are also similar (>97% 16 rRNA sequence similarity) to available genome sequences from cultured isolates (RCA-HTCC2150, *Loktanella vestfoldensis*,

Polaribacter irgensii, SAR92-HTCC2207) and single amplified genomes (SAGs) (MS024-1C, MS056-3A) (Stepanauskas and Sieracki 2007). We searched the genome annotations of these similar 'organisms' for the presence of nitrate reductases and the available SAG marker (NasA) data. Only one out of the six genomes contains a known assimilatory nitrate reductase. Although, the phylogenetic distribution of prokaryotic nitrate reductases amongst many marine bacterial groups is seemingly random (Cai and Jiao 2008; Richardson et al. 2001), where the distribution of appropriate genes can vary amongst closely related species. For example, approximately half of the available genomes belonging to the *Roseobacter* lineage contain NAS-like (see (Richardson et al. 2001) for description of NAS genes) assimilatory nitrate reductases and no phylogenetic patterns are observed (Newton et al. 2010). In contrast though, very few of the currently available marine Flavobacteria genomes contain nitrate reductases, which may in part explain their overall drop in the early days of the mesocosm.

Another genetic commonality between the genomes of the early mesocosm phylotypes could be their possession of a photosystem. Four out of the six genomes contained either a chlorophyll- or a rhodopsin-type light harvesting system. These systems utilize photon energy to generate a proton-motive force that can drive phosphorylation and produce ATP. To what ends the bacteria use this energy is unclear though, but it is likely to be exploited for a variety of purposes, such as CO_2 fixation, motility, and the active transport of molecules (Fuhrman et al. 2008; Zubkov 2009). Phylotypes possessing such photosystems could potentially be able to better compete for the added inorganic nitrate and phosphate, possibly explaining their dominance early in the mesocosm when reduced forms of these nutrients are at lower concentrations. Even though many of the genomes searched did not contain known nitrate reductases, some did contain ATP dependent nitrate/nitrite-like transporters (see Genbank ZP_01742952 and ZP_01002040), uncharacterized molybdopterin-binding oxidoreductases (GenBank ZP_01741082, ZP_01004317 and ZP_01118141) and known nitrogen sensing regulators (GenBank ZP_01742679 and ZP_01003596). Nitrogen sensing regulators of this type (NtrX-like) have been shown to control the expression of genes involved in the assimilation of alternative nitrogen sources (Merrick and Edwards 1995) and interact with regulators of the *puf* operon, affecting the expression of photosynthesis genes (Gregor et al. 2007). Given this and the

enigma of photosystem genes in marine heterotrophic bacteria, in would seem prudent to consider that the assimilation of alternative forms of nitrogen, such as nitrate, may be assisted by light induced proton pumping.

Day 8

Continuing past day 5 of this study, the abundance of the nano- and picoeukaryotic algae declines, reaching pre-mesocosm levels by day 7. This decline is thought to be in part due to protistian grazers (see introduction). Coinciding with this assumed high grazing activity, between days five and eight of the experiment, the bacterial population undergoes another large community shift and begins to increase in abundance. Such biodiversity shifts are often reported in marine mesocosms and other containment experiments and often it is concluded that grazing pressure is acting as a strong selective force on the bacterial community (Lebaron et al. 1999; Riemann et al. 2000). Many of the phylotypes that peaked on days 2 and 5 had decreased substantially in their relative contribution by day 8, except for a few γ-proteobacteria phylotypes who's abundance remained steady (see Fig 8). Only phylotypes belonging to the Flavobacteria and γ-proteobacteria peak on day 8, all α-proteobacteria show a decrease in relative contribution at this time. The Flavobacteria phylotype MS024-3C-like, is the second most abundant sequence-type in the library (total sequence number), after the RCA-1 phylotype. The phylotype increased from a relative abundance of <2.0% on day five to almost 20% by day 8. MS024-3C-like is phylogenetically similar to the SAG, MS024-3C (Woyke et al. 2009). MS024-3C is thought to be highly abundant in North Atlantic coastal waters and has a streamlined genome, predicted to be about 2 Mbp. Similarly, the phylotype HTCC2188-like, who increased from 2 to 10% by day 8, clusters with a γ-proteobacterium isolate of the same name, whom is considered oligotrophic and has a cell size less than a micrometer in width and length (Cho and Giovannoni 2004). Less can be inferred about the other phylotypes from day 8, except that environmental sequences very similar to the 64-10-like phylotype have been shown to incorporate methanol (Neufeld et al. 2008), which is produced by phytoplankton and from decaying matter (Heikes et al. 2002). It is not surprising that Flavobacteria phylotypes would begin to dominate at this point in the bloom,

given their known ability to degrade high molecular mass compounds (Kirchman 2002) and because of the assumption that grazing pressure was recently high, producing a variety of particulate matter. This is further supported by the genome of MA024-3C, which was shown to have predicted proteins involved in cell-surface interactions, such as attaching to detritus (Woyke et al. 2009). Lastly, data from a similar mesocosm study showed a dramatic shift in the number of free bacteria versus particle associated bacteria during this time, where the abundance of non-attached bacteria dropped significantly (Løvdal et al. 2008). This may indicate a competitive advantage for bacteria capable of attaching to the increasing amount of decaying matter and also hints that grazing pressure may have been high on free-living cells.

Another attribute similar amongst the phylotypes peaking on day 8 and perhaps the group of γ-proteobacteria that did not decrease after peaking on day 2 and 5, could be cell size. Although it is difficult to speculate the cell size for all the phylotypes, MA024-3C-like would probably be classified as picobacterioplankton, considering genome size is arguably a good proxy for cell size (Raes et al. 2007). It is well established that grazing pressure results in size selection amongst bacterioplankton, leaving extremely large (or microcolony forming) and small cells to proliferate (Pernthaler 2005). It has also been shown that individual protistian grazers can be better suited to consume either free-living or attached bacteria (Boenigk and Arndt 2002). The day 8 phylotypes may indeed of been very small bacteria, who have evolved to grow on particulate matter, and although they may be considered oligotrophic, have a competitive advantage in high productive environments that are under grazing pressure from certain types of protists. A general characteristic of oligotrophic bacteria is their ability to grow under low nutrient conditions, but this does not necessarily imply they will be unable to grow when nutrients are high (Cho and Giovannoni 2004). Given the dominance of these phylotypes and the fact that total bacterial abundance was on the rise at this time, these organisms may have been growing and not simply inactive cells who escaped predation.

Day 11

Continuing past day 8 and on to day 11, we see the bacterial community change again, where both bacterial abundance and biodiversity increased substantially, which coincides with assumed increases in copepod grazing (see introduction). The phylotypes mentioned above, including those that peaked on day 8 and the γ-proteobacteria that carried on from days 2 and 5, all decreased in relative contribution by day 11. They were replaced by peaks of mainly Flavobacteria phylotypes and very few α-proteobacteria, but did not include increases in γ-proteobacteria. The rise in biodiversity was the result of increases in the relative contributions of Flavobacteria phylotypes (Fig. 4.8, Group D and F), some of who were present at the start of the experiment, but who were most likely outcompeted after the initial nutrient additions (Fig. 4.8 group F). Many of these Flavobacteria phylotypes to the genus *Lacinutrix*. Phylotype *Lacinutrix*-1 was most similar to an isolate that has been shown to produce large amounts of exopolysacchrides, which are thought to influence the sequestration of metals (Nichols et al. 2005). Phylotype *Lacinutrix*-2 and *Lacinutrix*-3 were found to be similar to isolates shown to attach to copepods (Bowman and Nichols 2005) and macroalgae (Nedashkovskaya et al. 2008), respectively. Another Flavobacteria phylotype, MS056-2A-like was related to the genus *Ulvibacter*, where the type strain of the genus was isolated from the surface of the green alga *Ulva* (Nedashkovskaya et al. 2004). Particle attachment provides an obvious advantage during this time of the mesocosm and could be one reason for the success of these phylotypes. Furthermore, the production of exopolysaccharides to trap free or ligand bound metals would also prove beneficial (see introduction). The few α-proteobacteria phylotypes peaking on day 11 tended to contribute more in the deplete treatment than the replete treatment. These phylotypes *Ruegaria*-like and *Jannaschia*-like were similar to a methyl halide oxidizing bacterium (Schafer et al. 2005) and ethane oxidizing bacteria (Redmond et al. 2010) respectively. The presence of organisms capable of oxidizing simple hydrocarbons is not surprising, given the assumed level of decomposition occurring at this time in the mesocosm.

Day 14

By day 8 phytoplankton abundance is on the rise again, most notably is *E. huxleyi* cell concentrations, which increase quickly after day 7, reaching a peak around day 12. After day 12 the coccolithophorid bloom collapses. The phylotypes mentioned above, that peaked on day 11, all decreased in their relative contribution by day 14. These were replaced by two groups of phylotypes (based on MDS ordinations). The first, Group G (Fig. 4.8), consists mainly of Flavobacteria phylotypes that decreased in the first few days of the experiment only to begin increasing on day 11 and eventually peak by day 14. Many of the Group G phylotypes returned to contribution levels similar to those observed in the fjord sample and are thus highly correlated with Flavobacteria biodiversity. The single most common attribute amongst the Group G phylotypes is their high similarity to uncultured bacteria that are ubiquitous and highly abundant in North Atlantic waters (Alonso et al. 2007; Gómez-Pereira et al. 2010). The Flavobacteria phylotypes dominated Group G in terms of total contribution and this again may be attributed to particle attachment. A recent study suggested that certain Flavobacteria, many similar to these phylotypes, were significantly correlated with phytoplankton, and that they were found to be specifically enriched in the phycosphere compared to other bacterial lineages analyzed (Gómez-Pereira et al. 2010). Particle attachment has been a common trait linked to many of the Flavobacteria phylotypes in this study. Based on the overall community structure of this mesocosm bloom, it is unlikely that all the Flavobacteria phylotypes are attaching to the same types of particles. For example, the Group G Flavobacteria may be better adapted at forming close associations with actively growing nanoeukaryotes, where the Group E, D, and F phylotypes have evolved to attach more readily to various types of detritus or zooplankton. Associations with certain particle-types are likely a reflection of the bacterium's ability to utilize the specific carbon and nutrient forms emitted or contained within the particle.

In contrast to the Group G phylotypes, members of Group I generally showed a marked increase from day 11 to 14 and included many different phylotypes, including α-proteobacteria, γ-proteobacteria, and other lineages (Rhodospirillales, Verrumicrobia, etc.), but no Flavobacteria. Also, many Group I (see Group J) phylotypes did not contribute significantly (<0.5%) to the initial bacterial community. The coccolithophorid population crash after day 12

coincided with an exponential increase in coccolithovirus concentrations (Kimmance et al. 2011) indicating viral lysis as the likely cause. The most abundant phylotype on day 14 was by far the α-proteobacterial phylotype CHAB-1. CHAB-1 increased from a relative contribution of 3.5% on day 11 to 27% by day 14. The phylotype belongs to the uncultured *Roseobacter* cluster CHAB-I-5, whose members are very abundant in the ocean, particularly in coastal systems (Buchan et al. 2005). No information is available about the genomic content or physiology of this ubiquitous *Roseobacter* clade. Another α–proteobacteria phylotype (SAR116-like) showed a marked increase on day 14 and was similar to the SAR116 clade, an abundant group of marine bacteria (Mullins et al. 1995). Members of this clade have been shown to compete successfully for a variety of dissolved organic compounds (Mou et al. 2007) and a recent genome announcement indicates the presence of genes encoding proteorhodopsin, DMSP demethylase, and C_1 compound metabolism (Oh et al. 2010). One of the only β–proteobacterial sequences recovered in the library also peaked during this time (Methylophiales-like). This phylotype is similar to clones recovered from a DMSP degrading bacterial community associated with an *E. huxleyi* bloom in the North Sea (Zubkov et al. 2002) and to bacteria shown to be abundant in the German Bight (Sekar et al. 2004). Four phylotypes belonging to the phylum Verrucomicrobia were also found to peak on day 14. These phylotypes were closely related and grouped near the genera *Coraliomargarita* and *Puniceicoccus*. The Verrucomicrobia are common in of marine systems and are often found in eutrophic, sulfide-rich, and polar habitats (Schlesner et al. 2006). The role these organisms play in pelagic environments is not well understood, although they have been found in association with algae (Isolate DG1235, see GenBank genome project 36179). Recently, it has been demonstrated that members of the Verrucomicrobia were involved in the uptake or binding of lipids (R. Stepanauskas, personal communication. Need to confirm and ask his permission). The study added fluorescently labeled lipids to a coastal seawater incubation, flow cytometrically sorted single cells, and created SAGs. All of the SAGs belonged to members of the Verrucomicrobia. Lastly, the phylotype Microbacteriaceae-like, belonging to the Actinobacteria, also peaked on day 14. Interestingly, the Microbacteriaceae-like phylotype was the only abundant (contributing greater than 1% to the total library) species to peak early (day 2) and late (day 14) in the mesocosm. The phylotype was most similar to sequences

obtained from high-latitude waters, including freshwater lakes, estuaries, fjords, sea-ice, and surface waters of the Arctic Ocean. This clade of Actinobacteria may represent a group of euryhaline cold adapted bacteria.

Conclusions

By including environmental sequence data in our taxonomy classifications and generating phylogenetic trees, close to 90% of all sequence reads were identified to the 'species' level, revealing the temporal dynamics of the most abundant phylotypes present during an induced coccolithophorid bloom. This annotation unmasked straightforward but important questions about the microbial community, including, simply, which specific bacteria were present. The wealth of sequence information that has been deposited in public databases allowed us to construct a representation of the different possible species-level niches present in the bloom. This generalized picture of the bacterial community acts as a foundation for developing testable hypotheses on the roles and factors that lead to the success of individual microbial species. Examples of future investigations include unraveling the contribution of various photosystems to the physiology of the early peaking phylotypes and also the likelihood that certain phylotypes avoided grazing during days 5 to 8. Deep-amplicon sequencing alone cannot answer these questions though. Multiple molecular approaches will be required to accurately characterize changes in microbial populations over time and space as well as the utilization of microscopy. In this study almost 18% of the phylotypes described were novel sequences, not identical to anything in the public databases searched and many more phylotypes had no cultured relatives or available genomic sequence data. This is especially evident in the γ-proteobacteria, where approximately 25% of the phylotypes could not be classified to even the family level. Directed cultivation and single cell genome sequencing of particular ecotypes will provide useful information on these organisms and more allow more thorough interpretations of deep-amplicon data sets.

Often the goal of high-throughput sequencing endeavors is to maximize the length and the number of reads per sample. Increases in read depth and size will go a long way in helping us

to understand microbial biodiversity in natural environments. In this study, the community differences between the two nutrient regimes were largely due to the type-2 phylotypes, many who peaked on different days depending on the treatment, making a correlation analysis difficult. Removal of the type-2 phylotypes from the data analysis demonstrated that these organisms largely influenced the observed beta diversity between samples. The average relative abundance of these phylotypes was significantly lower than that of the type-1 phylotypes, further complicating the analysis. Higher sequencing depth would of helped to alleviate this. But our 'resolution' of the microbial community was also limited by a lack of biological replicates. This is evident with the type-1 phylotypes, many of which contributed significantly to the total library and were thus not lacking in read depth. These phylotypes had observable differences in peak intensity between the treatments, but without any measures of variance it was not possible to give them credence. Sequencing of the replicate treatments from the experiment would of provided much needed statistical support for the small community differences found between the deplete and replete samples. These apparently small changes within the community, including the presence of micro-clades and of rare organisms may have a substantial influence on sample heterogeneity, differentiation, and functional potential. Micro-environments were also likely contributing to differences in overall community structure and perhaps between treatments as well. Clearly the mesocosm's bacterial community was partitioning into free-living and particle attached forms, but there is no direct evidence of this. Size-fractionation prior to sequencing would of greatly assisted in describing the ecophysiology of the community. Thus, the biological patterns observed in this study emphasize the importance of allocating sequencing coverage in ways that will better support the hypotheses being tested, especially when these methods are semi-quantitative at best. Lastly, studies that utilize high-throughput sequencing to examine microbial community dynamics should incorporate more traditional methods such as microscopy (including FISH-based approaches), rate measurements (grazing and viral production), and physiochemical parameters into the experimental plans. Basic supporting metadata is essential for researchers to interpret the biological patterns found within the ever-increasing size of environmental sequencing libraries.

References

Acinas, S. G. and others 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature **430:** 551-554.

Alonso, C., F. Warnecke, R. Amann, and J. Pernthaler. 2007. High local and global diversity of Flavobacteria in marine plankton. Environ Microbiol **9:** 1253-1266.

Andersson, A. F., L. Riemann, and S. Bertilsson. 2010. Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. ISME J **4:** 171-181.

Boenigk, J., and H. Arndt. 2002. Bacterivory by heterotrophic flagellates: community structure and feeding strategies. Antonie Van Leeuwenhoek **81:** 465-480.

Bowman, J. P., and D. S. Nichols. 2005. Novel members of the family *Flavobacteriaceae* from Antarctic maritime habitats including *Subsaximicrobium wynnwilliamsii* gen. nov., sp nov., *Subsaximicrobium saxinquililnus* sp. nov., *Subsaxibacter broadyi* gen. nov., sp. nov., *Lacinutrix copepodicola* gen. nov., sp nov., and novel species of the genera *Bizionia*, *Gelidibacter* and *Gillisia*. Int J Syst Evol Micr **55:** 1471-1486.

Bratbak, G., J. K. Egge, and M. Heldal. 1993. Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. Mar Ecol-Prog Ser **93:** 39-48.

Bratbak, G., W. Wilson, and M. Heldal. 1996. Viral control of *Emiliania huxleyi* blooms? Journal of marine systems **9:** 75-81.

Brown, C. W., and J. A. Yoder. 1994. Coccolithophorid blooms in the global ccean. J Geophys Res **99:** 7467-7482.

Buchan, A., J. M. González, and M. A. Moran. 2005. Overview of the marine *Roseobacter* lineage. Appl Environ Microb **71:** 5665-5677.

Buchan, A., M. Hadden, and M. T. Suzuki. 2009. Development and application of quantitative-PCR tools for subgroups of the Roseobacter clade. Appl Environ Microbiol **75:** 7542-7547.

Cai, H. Y., and N. Z. Jiao. 2008. Diversity and abundance of nitrate assimilation genes in the northern South China sea. Microb Ecol **56:** 751-764.

Campbell, B. J., A. S. Engel, M. L. Porter, and K. Takai. 2006. The versatile epsilon-proteobacteria: key players in sulphidic habitats. Nat Rev Microbiol **4:** 458-468.

Castberg, T. and others 2001. Microbial population dynamics and diversity during a bloom of the marine coccolithophorid Emiliania huxleyi (Haptophyta). Mar Ecol-Prog Ser **221:** 39-46.

Chakravorty, S., D. Helb, M. Burday, N. Connell, and D. Alland. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods. **69:** 330-339.

Charlson, R., J. Lovelock, M. Andreae, and S. Warren. 1987. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. Nature **326:** 655-661.

Cho, J. C., and S. J. Giovannoni. 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria. Appl Environ Microb **70:** 432-440.

Cho, J. C. and others 2007. Polyphyletic photosynthetic reaction centre genes in oligotrophic marine *Gammaproteobacteria*. Environ Microbiol **9:** 1456-1463.

Clarke, K. R., and R. M. Warwick. 1999. The taxonomic distinctness measure of biodiversity: weighting of step lengths between hierarchical levels. Mar Ecol-Prog Ser **184:** 21-29.

Connon, S. A., and S. J. Giovannoni. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl Environ Microb **68:** 3878.

Egge, J. K., and D. L. Aksnes. 1992. Silicate as regulating nutrient in phytoplankton competition. Mar Ecol-Prog Ser **83:** 281-289.

Evans, C., S. D. Archer, S. Jacquet, and W. H. Wilson. 2003. Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a *Micromonas* spp. population. Aquat Microb Ecol **30:** 207-219.

Fuhrman, J. A., and L. Campbell. 1998. Microbial microdiversity. Nature **393:** 410-411.

Fuhrman, J. A., M. S. Schwalbach, and U. Stingl. 2008. Proteorhodopsins: an array of physiological roles? Nat Rev Microbiol **6:** 488-494.

García-Martínez, J., and F. Rodríguez-Valera. 2000. Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. Mol Ecol **9:** 935-948.

Giebel, H.-A. and others 2011. Distribution of Roseobacter RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. ISME J **5:** 8-19.

Gilbert, J. A. and others 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One **3:** e3042.

Gómez-Pereira, P. R., B. M. Fuchs, C. Alonso, M. J. Oliver, J. E. E. Van Beusekom, and R. Amann. 2010. Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. Isme Journal **4:** 472-487.

González, J. M. and others 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. Appl Environ Microb **66:** 4237-4246.

Gregor, J., T. Zeller, A. Balzer, K. Haberzettl, and G. Klug. 2007. Bacterial regulatory networks include direct contact of response regulator proteins: Interaction of RegA and NtrX in Rhodobacter capsulatus. J Mol Microb Biotech **13:** 126-139.

Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight. 2008. Error-correcting Barcoded Primers for Pyrosequencing Hundreds of Samples in Multiplex. Nature Methods **5:** 235-237.

Heikes, B. G. and others 2002. Atmospheric methanol budget and ocean implication. Global Biogeochem Cy **16:** -.

Heldal, M., K. M. Fagerbakke, P. Tuomi, and G. Bratbak. 1996. Abundant populations of iron and manganese sequestering bacteria in coastal water. Aquat Microb Ecol **11:** 127-133.

Huse, S. M., L. Dethlefsen, J. A. Huber, D. Mark Welch, D. A. Relman, and M. L. Sogin. 2008. Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. PLoS Genetics **4:** e1000255.

Hutchins, D. A., and K. W. Bruland. 1994. Grazer-Mediated Regeneration and Assimilation of Fe, Zn and Mn from Planktonic Prey. Mar Ecol-Prog Ser **110:** 259-269.

Jacobsen, A., J. K. Egge, and B. R. Heimdal. 1995. Effects of increased concentration of nitrate and phosphate during a springbloom experiment in mesocosm. J Exp Mar Biol Ecol **187:** 239-251.

Jacquet, S., M. Heldal, D. Iglesias-Rodriguez, A. Larsen, W. Wilson, and G. Bratbak. 2002. Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. Aquat Microb Ecol **27:** 111-124.

Kiene, R. P., and L. J. Linn. 2000. The fate of dissolved dimethylsulfoniopropionate (DMSP) in seawater: Tracer studies using S-35-DMSP. Geochim Cosmochim Ac **64:** 2797-2810.

Kiene, R. P., L. J. Linn, and J. A. Bruton. 2000. New and important roles for DMSP in marine microbial communities. J Sea Res **43:** 209-224.

Kimmance, S. A., M. J. Allen, J. Martínez-Martínez, A. Pagarete, and W. H. Wilson. 2011. Uncoupling of *Emiliania huxleyi* photosynthesis: virus infection versus nutrient stress. Unpublished.

Kirchman, D. L. 2002. The ecology of *Cytophaga-Flavobacteria* in aquatic environments. Fems Microbiol Ecol **39:** 91-100.

Larsen, A. and others 2001. Population dynamics and diversity of phytoplankton, bacteria and viruses in a seawater enclosure. Mar Ecol-Prog Ser **221:** 47-57.

---. 2004. Spring phytoplankton bloom dynamics in Norwegian coastal waters: Microbial community succession and diversity. Limnol Oceanogr **49:** 180-190.

Lebaron, P. and others 1999. Changes in bacterial community structure in seawater mesocosms differing in their nutrient status. Aquat Microb Ecol **19:** 255-267.

Li, K. B. 2003. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. Bioinformatics **19:** 1585-1586.

Liu, Z., T. Z. Desantis, G. L. Andersen, and R. Knight. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res **36:** e120.

Lomas, M. W., and P. M. Glibert. 2000. Comparisons of nitrate uptake, storage, and reduction in marine diatoms and flagellates. J Phycol **36:** 903-913.

Løvdal, T. and others 2008. Competition for inorganic and organic forms of nitrogen and phosphorous between phytoplankton and bacteria during an *Emiliania huxleyi* spring bloom. Biogeosciences **5:** 371-383.

Lozupone, C., and R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol **71:** 8228-8235.

Malin, G., S. Turner, P. Liss, P. Holligan, and D. Harbour. 1993. Dimethylsulfide and dimethylsulphoniopropionate in the Northeast Atlantic during the summer Coccolithophore bloom. Deep-Sea Res Pt I **40:** 1487-1508.

Malmstrom, R. R., R. P. Kiene, and D. L. Kirchman. 2004. Identification and enumeration of bacteria assimilating dimethylsulfoniopropionate (DMSP) in the North Atlantic and Gulf of Mexico. Limnol Oceanogr **49:** 597-606.

Matrai, P. A., and M. D. Keller. 1993. Dimethylsulfide in a large-scale coccolithophore bloom in the Gulf of Maine. Cont Shelf Res **13:** 831-843.

Mayali, X., P. J. S. Franks, and F. Azarn. 2008. Cultivation and ecosystem role of a marine Roseobacter clade-affiliated cluster bacterium. Appl Environ Microb **74:** 2595-2603.

Merrick, M. J., and R. A. Edwards. 1995. Nitrogen control in bacteria. Microbiol Rev **59:** 604-&.

Mitra, S., J. A. Gilbert, D. Field, and D. H. Huson. 2010. Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. ISME J **4:** 1236-1242.

Mou, X. Z., R. E. Hodson, and M. A. Moran. 2007. Bacterioplankton assemblages transforming dissolved organic compounds in coastal seawater. Environ Microbiol **9:** 2025-2037.

Mou, X. Z., M. A. Moran, R. Stepanauskas, J. M. González, and R. E. Hodson. 2005. Flow-cytometric cell sorting and subsequent molecular analyses for culture-independent identification of bacterioplankton involved in dimethylsulfoniopropionate transformations. Appl Environ Microb **71:** 1405-1416.

Muller, F. L. L., S. P. Jacquet, and W. H. Wilson. 2003. Biological factors regulating the chemical speciation of Cu, Zn, and Mn under different nutrient regimes in a marine mesocosm experiment. Limnol Oceanogr **48:** 2289-2302.

Muller, F. L. L., A. Larsen, C. A. Stedmon, and M. Sondergaard. 2005. Interactions between algal-bacterial populations and trace metals in fjord surface waters during a nutrient-stimulated summer bloom. Limnol Oceanogr **50:** 1855-1871.

Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni. 1995. Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnol Oceanogr **40:** 148-158.

Nawrocki, E. P., D. L. Kolbe, and S. R. Eddy. 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics **25:** 1335-1337.

Nedashkovskaya, O. I. and others 2004. *Ulvibacter litoralis* gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the green alga *Ulva fenestrata*. Int J Syst Evol Micr **54:** 119-123.

Nedashkovskaya, O. I., K. K. Kwon, S. H. Yang, H. S. Lee, K. H. Chung, and S. J. Kim. 2008. *Lacinutrix algicola* sp. nov. and *Lacinutrix mariniflava* sp. nov., two novel marine alga-associated bacteria and emended description of the genus *Lacinutrix*. Int J Syst Evol Micr **58:** 2694-2698.

Nejstgaard, J. C., I. Gismervik, and P. T. Solberg. 1997. Feeding and reproduction by *Calanus finmarchicus*, and microzooplankton grazing during mesocosm blooms of diatoms and the coccolithophore *Emiliania huxleyi*. Mar Ecol-Prog Ser **147:** 197-217.

Neufeld, J. D., Y. Chen, M. G. Dumont, and J. C. Murrell. 2008. Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. Environ Microbiol **10:** 1526-1535.

Newton, R. J. and others 2010. Genome characteristics of a generalist marine bacterial lineage. Isme Journal **4:** 784-798.

Nichols, C. M., J. P. Bowman, and J. Guezennec. 2005. Olleya marilimosa gen. nov., sp nov., an exopolysaccharide-producing marine bacterium from the family Flavobacteriaceae, isolated from the Southern Ocean. Int J Syst Evol Micr **55:** 1557-1561.

Oh, H. M. and others 2010. Complete genome sequence of "Candidatus Puniceispirillum marinum" IMCC1322, a representative of the SAR116 clade in the alphaproteobacteria. J Bacteriol **192:** 3240-3241.

Øvreås, L. and others 2003. Response of bacterial and viral communities to nutrient manipulations in seawater mesocosms. Aquat Microb Ecol **31:** 109-121.

Paulino, A. I., J. K. Egge, and A. Larsen. 2008. Effects of increased atmospheric CO2 on small and intermediate sized osmotrophs during a nutrient induced phytoplankton bloom. Biogeosciences **5:** 739-748.

Pernthaler, J. 2005. Predation on prokaryotes in the water column and its ecological implications. Nat Rev Microbiol **3:** 537-546.

Philippot, L. and others 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol **8:** 523-529.

Pinhassi, J. and others 2005. Dimethylsulfoniopropionate turnover is linked to the composition and dynamics of the bacterioplankton assemblage during a microcosm phytoplankton bloom. Appl Environ Microb **71:** 7650-7660.

Pruesse, E. and others 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35:** 7188-7196.

Raes, J., J. O. Korbel, M. J. Lercher, C. Von Mering, and P. Bork. 2007. Prediction of effective genome size in metagenomic samples. Genome Biol **8:** -.

Reay, D. S., D. B. Nedwell, J. Priddle, and J. C. Ellis-Evans. 1999. Temperature dependence of inorganic nitrogen uptake: Reduced affinity for nitrate at suboptimal temperatures in both algae and bacteria. Appl Environ Microb **65:** 2577-2584.

Redmond, M. C., D. L. Valentine, and A. L. Sessions. 2010. Identification of novel methane-, ethane-, and propane-oxidizing bacteria at marine hydrocarbon seeps by stable isotope probing. Appl Environ Microb **76:** 6412-6422.

Richardson, D. J., B. C. Berks, D. A. Russell, S. Spiro, and C. J. Taylor. 2001. Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. Cell Mol Life Sci **58:** 165-178.

Riemann, L., G. F. Steward, and F. Azam. 2000. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. Appl Environ Microb **66:** 578-587.

Rocap, G., D. L. Distel, J. B. Waterbury, and S. W. Chisholm. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. Appl Environ Microb **68:** 1180-1191.

Sato, M., S. Takeda, and K. Furuya. 2007. Iron regeneration and organic iron(III)-binding ligand production during in situ zooplankton grazing experiment. Mar Chem **106:** 471-488.

Schafer, H., I. R. Mcdonald, P. D. Nightingale, and J. C. Murrell. 2005. Evidence for the presence of a CmuA methyltransferase pathway in novel marine methyl halide-oxidizing bacteria. Environ Microbiol **7:** 839-852.

Schlesner, H., C. Jenkins, and J. Staley. 2006. The Phylum Verrucomicrobia: A Phylogenetically Heterogeneous Bacterial Group, p. 881-896. *In* M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stackebrandt [eds.], The Prokaryotes. Springer New York.

Schloss, P. D. and others 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75:** 7537-7541.

Sekar, R., B. M. Fuchs, R. Amann, and J. Pernthaler. 2004. Flow sorting of marine bacterioplankton after fluorescence in situ hybridization. Appl Environ Microb **70:** 6210-6219.

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

Sheneman, L., J. Evans, and J. A. Foster. 2006. Clearcut: a fast implementation of relaxed neighbor joining. Bioinformatics **22:** 2823-2824.

Stepanauskas, R., and M. E. Sieracki. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. P Natl Acad Sci USA **104:** 9052-9057.

Suzuki, M. T., L. T. Taylor, and E. F. Delong. 2000. Quantitative Analysis of Small-Subunit rRNA Genes in Mixed Microbial Populations via 5'-Nuclease Assays. Appl. Environ. Microbiol. **66:** 4605-4614.

Vila, M. and others 2004. Use of microautoradiography combined with fluorescence in situ hybridization to determine dimethylsulfoniopropionate incorporation by marine bacterioplankton taxa. Appl Environ Microb **70:** 4648-4657.

Vila-Costa, M., J. M. Rinta-Kanto, S. Sun, S. Sharma, R. Poretsky, and M. A. Moran. 2010. Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. ISME J.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol **73:** 5261-5267.

West, N. J., I. Obernosterer, O. Zemb, and P. Lebaron. 2008. Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. Environ Microbiol **10:** 738-756.

Westbroek, P. and others 1993. A model system approach to biological climate forcing. The example of *Emiliania huxleyi*. Global Planet Change **8:** 27-46.

Woyke, T. and others 2009. Assembling the marine metagenome, one cell at a time. PLoS One **4:** -.

Zubkov, M. V. 2009. Photoheterotrophy in marine prokaryotes. J Plankton Res **31:** 933-938.

Zubkov, M. V., B. M. Fuchs, S. D. Archer, R. P. Kiene, R. Amann, and P. H. Burkill. 2001. Linking the composition of bacterioplankton to rapid turnover of dissolved dimethylsulphoniopropionate in an algal bloom in the North Sea. Environ Microbiol **3:** 304-311.

---. 2002. Rapid turnover of dissolved DMS and DMSP by defined bacterioplankton communities in the stratified euphotic zone of the North Sea. Deep-Sea Res Pt Ii **49:** 3017-3038.

Appendix

Tables

Table 4.1. Comparison of classifications obtained from the RDP Classifier and creating a custom reference database with ARB.

		RDP Classifier			ARB Phylotypes		
	% of library	Order	Family	Genus	Phylotypes	% classified[a]	Sequences per phylotype[b]
α-proteobacteria	36.3	96.8	96.5	15.7	29	90.1	237, 604, 2917
Flavobacteria	37.3	100	98.5	17.3	38	92.1	185, 265, 1229
γ-proteobacteria	20.8	9.2	5.4	2.8	19	85.2	176, 196, 870
Total[c]	-	73.7	72.3	14.9	96	88.6	191, 358, 2917

a Percent of sequences classified within a particular class of bacteria.

b Average number of sequences per phylotype with standard deviation and maximum.

c Includes all bacterial classes

Supplemental Tables

Table 4.2. Number of misclustered sequences at a given distance cutoff between some of the most abundant phylotypes and OTUs

	Cutoff			
	0.045	**0.044**	**0.04**	**0.03**
RCA-1	164	160	-4	-10
MS024-3C-like	-1	-3	-1	-13
HTCC2188-like	-4	25	25	5
Polaribacter-1	-15	-18	-17	-30
CHAB-1	-36	-137	-136	-142
MS056-2A-like	-37	-38	-37	-55
Polaribacter-2	-10	-11	-10	-27
MS024-1F-like	-29	-30	-29	-32
Others	72	49	44	-259
Total	104	-3	-166	-563

Figures

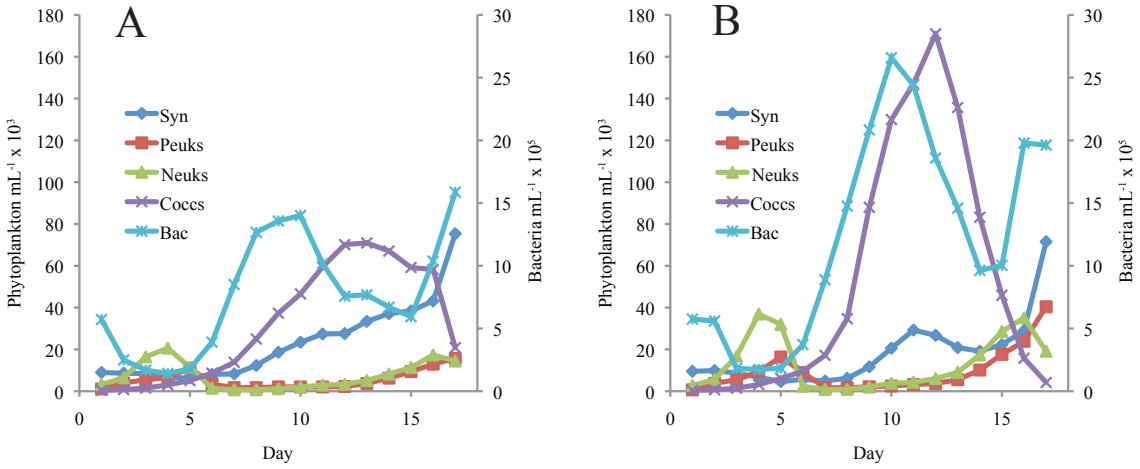


Figure 4.1. Phytoplankton and bacterioplankton abundance in the P-deplete (A) and P-replete (B) mesocosm treatments

The P (phosphorous)-deplete treatment contained a nitrogen to phosphorous ratio of 15:1 and the P-replete treatment had a N:P ratio of 75:1.

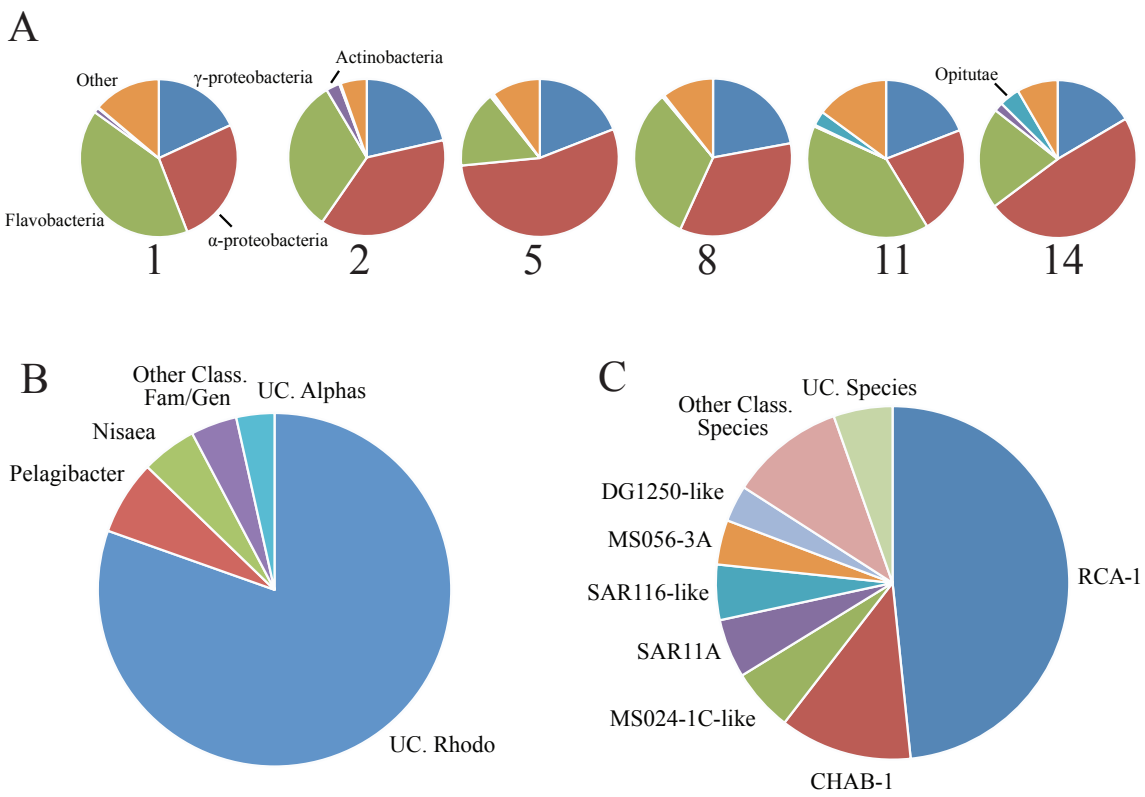


Figure 4.2. Taxonomic classification distributions of the bacterial community

Class level distributions (RDP Classifier) of bacteria from the replete treatment graphed over time (days) (A). Distribution of α–proteobacteria genra and species from the entire library determined using the RDP Classifier (B) or through manual annotation using ARB (C). Class.=Classified, UC=Unclassified.

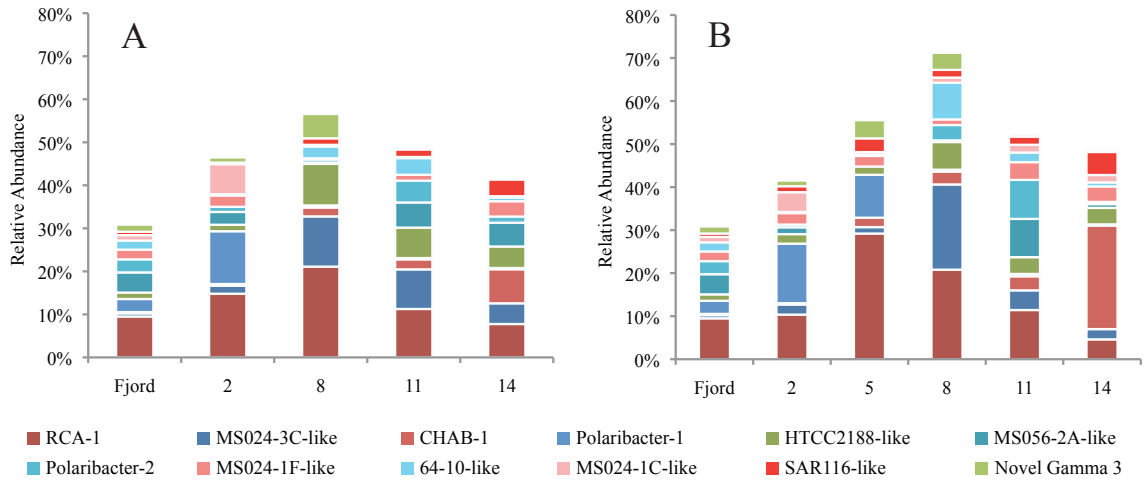


Figure 4.3. Distribution of 12 of the most abundant phylotypes over time

Distribution of 12 of the most abundant phylotypes over time in the deplete (A) and replete (B) treatments. Color shades depict the class the individual phylotypes belong to; α-proteobacteria (red), Flavobacteria (blue), and γ–proteobacteria (green).

Figure 4.4. Bubble heat map of phylotype abundances over time

Bubble heat map of phylotype abundances over time in the deplete and replete treatments. Bubble colors represent the classes α–proteobacteria (red), Flavobacteria (blue), γ-proteobacteria, and others (orange). Bubble sizes depict the relative contribution of that phylotype to the total bacterial reads in a given sample, ranging from >10% contribution (largest bubble), 5 to <10%, 1 to <5%, 0.1 to < 1%, >0 to < 0.1%, and the smallest at 0%. Color shading indicates rank abundance of the phylotype, where darker shades depict higher sample abundance.

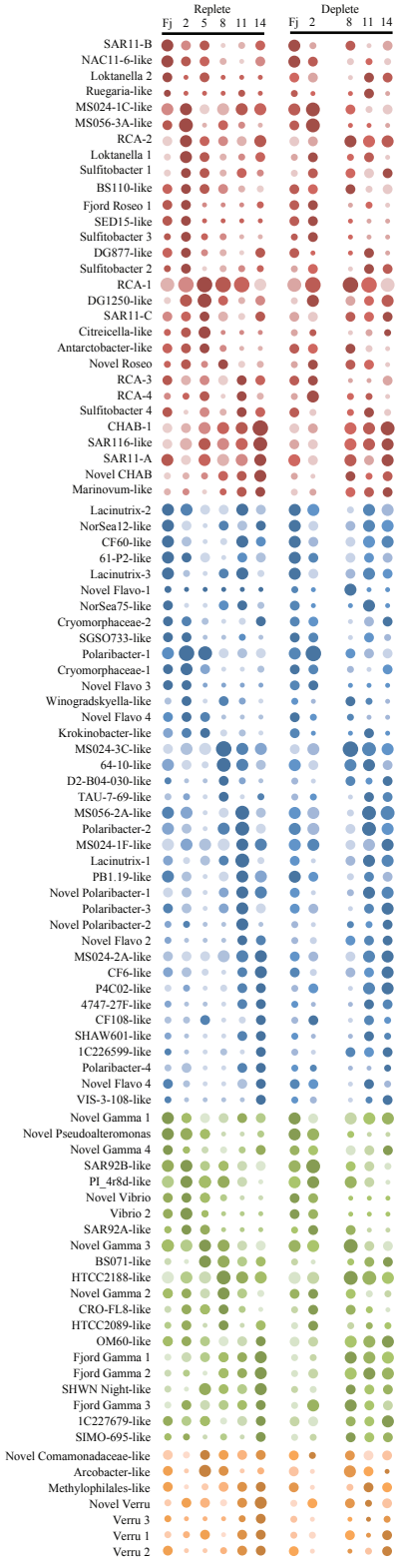

Replete
Deplete
Fj 2 5 8 11 14
Fj 2 8 11 14
SAR11-B
NAC11-6-like
Loktanella 2
Ruegaria-like
MS024-1C-like
MS056-3A-like
RCA-2
Loktanella 1
Sulfitobacter 1
BS110-like
Fjord Roseo 1
SED15-like
Sulfitobacter 3
DG877-like
Sulfitobacter 2
RCA-1
DG1250-like
SAR11-C
Citreicella-like
Antarctobacter-like
Novel Roseo
RCA-3
RCA-4
Sulfitobacter 4
CHAB-1
SAR116-like
SAR11-A
Novel CHAB
Marinovum-like
Lacinutrix-2
NorSea12-like
CF60-like
61-P2-like
Lacinutrix-3
Novel Flavo-1
NorSea75-like
Cryomorphaceae-2
SGSO733-like
Polaribacter-1
Cryomorphaceae-1
Novel Flavo 3
Winogradskyella-like
Novel Flavo 4
Krokinobacter-like
MS024-3C-like
64-10-like
D2-B04-030-like
TAU-7-69-like
MS056-2A-like
Polaribacter-2
MS024-1F-like
Lacinutrix-1
PB1.19-like
Novel Polaribacter-1
Polaribacter-3
Novel Polaribacter-2
Novel Flavo 2
MS024-2A-like
CF6-like
P4C02-like
4747-27F-like
CF108-like
SHAW601-like
1C226599-like
Polaribacter-4
Novel Flavo 4
VIS-3-108-like
Novel Gamma 1
Novel Pseudoalteromonas
Novel Gamma 4
SAR92B-like
PI_4r8d-like
Novel Vibrio
Vibrio 2
SAR92A-like
Novel Gamma 3
BS071-like
HTCC2188-like
Novel Gamma 2
CRO-FL8-like
HTCC2089-like
OM60-like
Fjord Gamma 1
Fjord Gamma 2
SHWN Night-like
Fjord Gamma 3
1C227679-like
SIMO-695-like
Novel Comamonadaceae-like
Arcobacter-like
Methylophilales-like
Novel Verru
Verru 3
Verru 1
Verru 2

Figure 4.4

Figure 4.5. Relative abundances of the *Roseobacter* subgroups RCA and CHAB determined using 454 pyrosequencing and QPCR.

Relative abundance (to total bacteria) over time in the deplete treatment (A), correlation between 454 and QPCR data in the deplete treatment (B), relative abundance over time in the replete treatment (C), and the correlation between 454 and QPCR data in the replete treatment (D). Blue lines are CHAB and red lines are RCA. Squares are QPCR data and circles are 454 data.

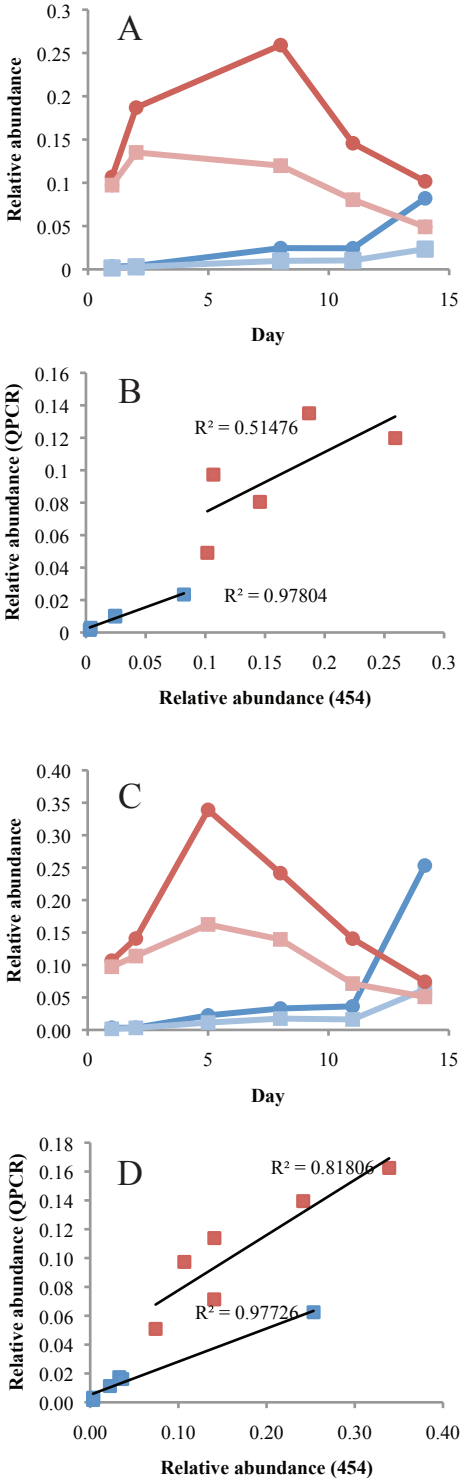
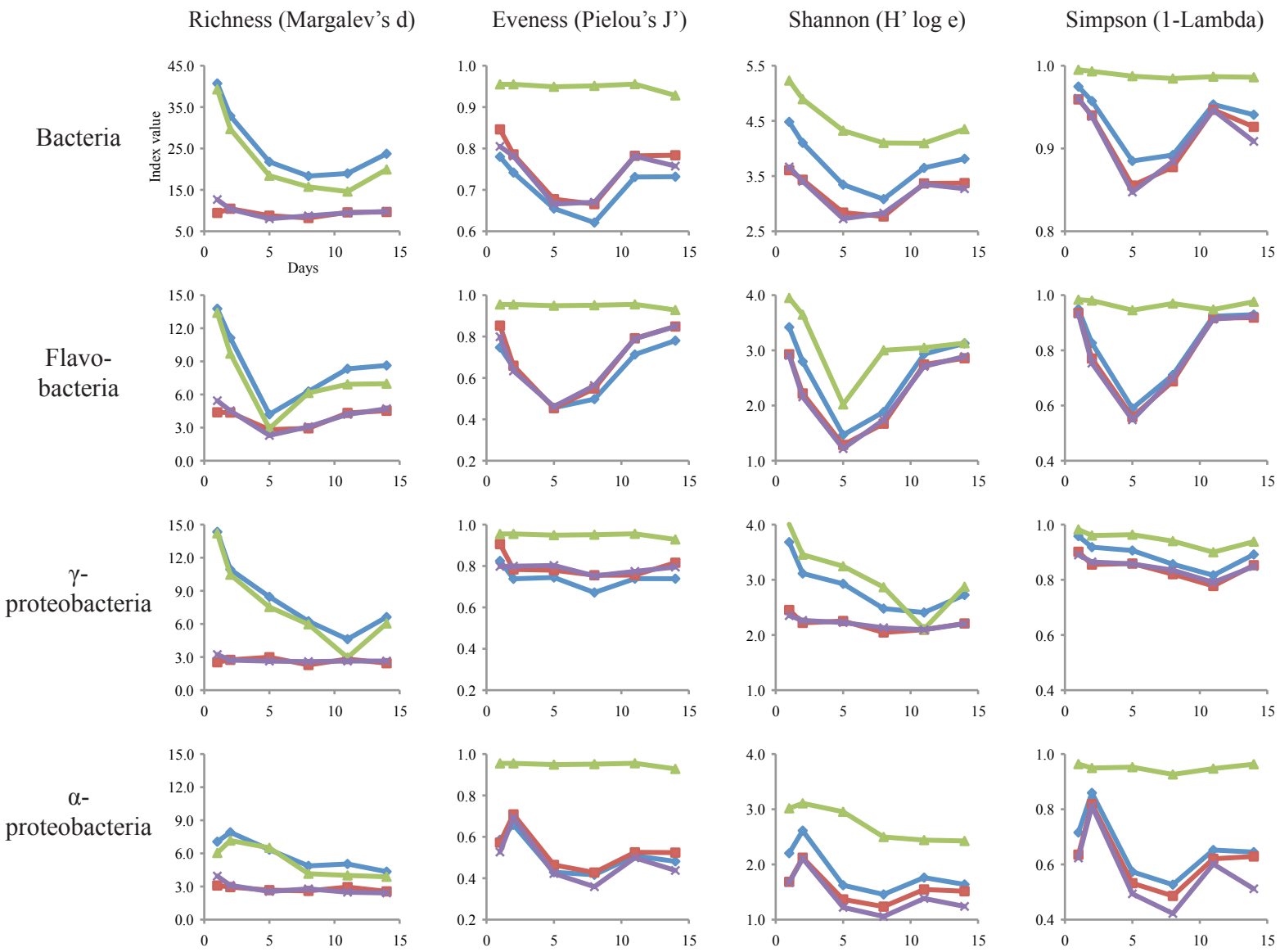

A
Relative abundance
Day
B
Relative abundance (QPCR)
R² = 0.51476
R² = 0.97804
Relative abundance (454)
C
Relative abundance
Day
D
Relative abundance (QPCR)
R² = 0.81806
R² = 0.97726
Relative abundance (454)

Figure 4.5

Figure 4.6. Biodiversity measures of the bacterial community over time

Biodiversity measures of the bacterial community over time from the replete treatment. Data plotted consists of phylotype data (purple line) from ARB and OTU data from mothur. OTU data is separated into abundant (red line), rare (green line), and total (blue line). Y-axis is the coefficient index value and the X-axis is days


Richness (Margalev's d)
Eveness (Pielou's J')
Shannon (H' log e)
Simpson (1-Lambda)
Bacteria
Flavo-bacteria
γ-proteobacteria
α-proteobacteria
Index value
Days

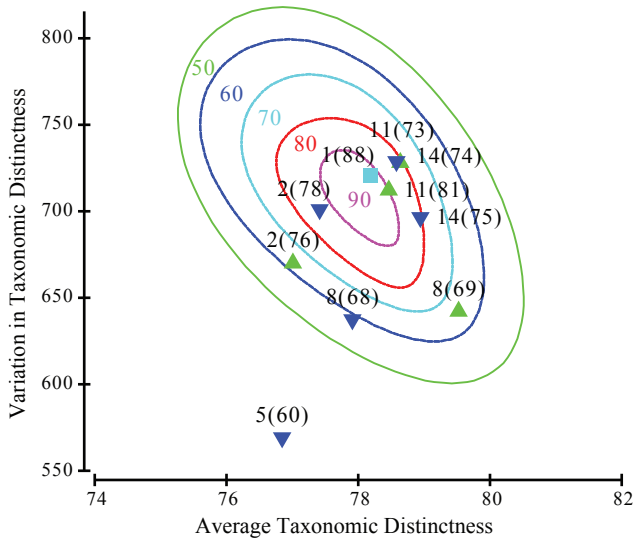
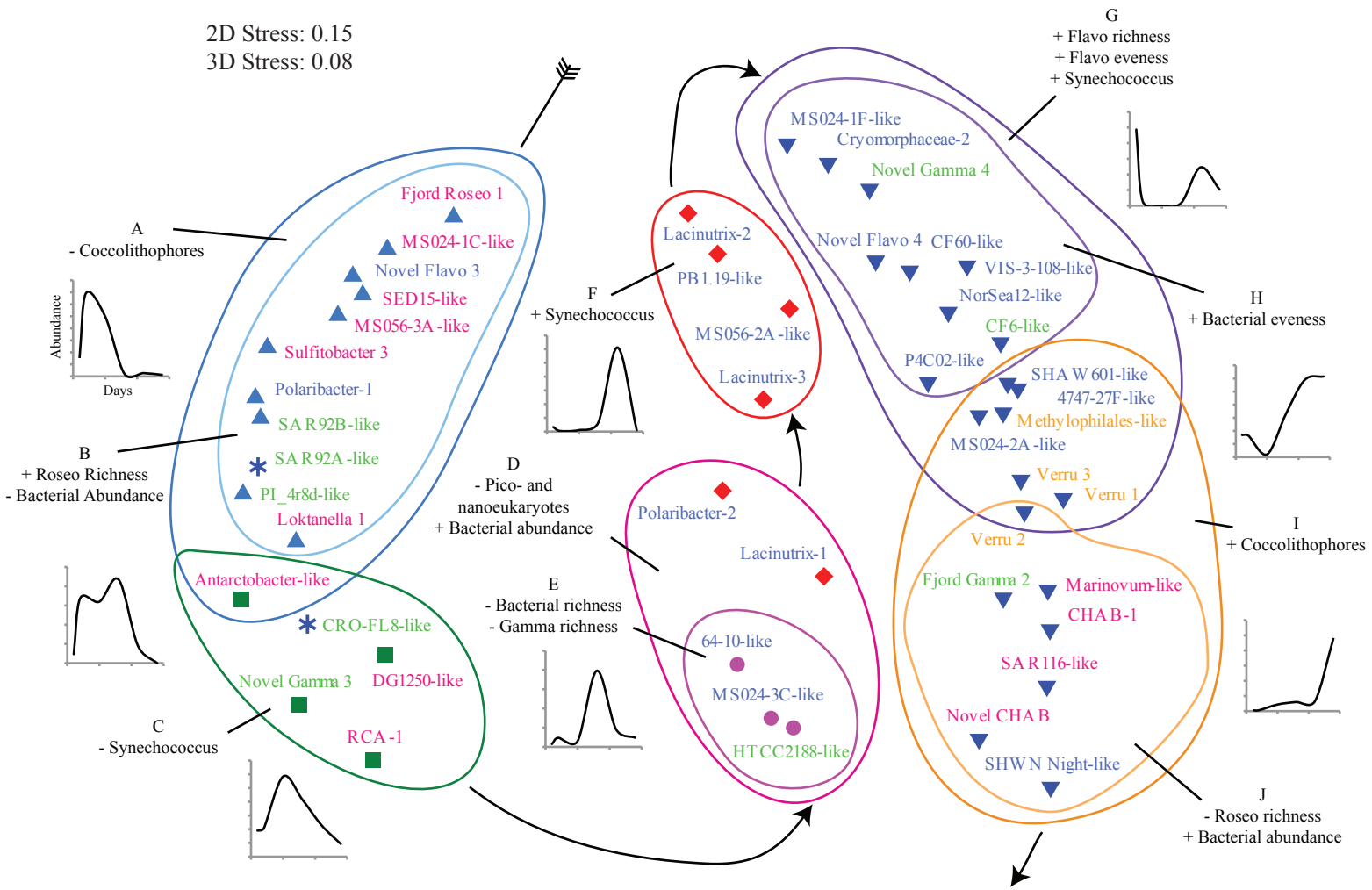


Figure 4.7. Bivariant eclipse plot of the simulated average taxonomic diversity and the simulated variation in taxonomic distinctness

Green triangles represent the deplete treatment and the blue triangles represent the replete treatment. The day is indicated above each data point along with the number of phylotypes present. Contours separate percent probability regions of the simulated values.

Figure 4.8. MDS ordination of Spearman rank correlation resemblances of the type 1 phylotypes

Marker points indicate when the individual species peaked; day 2 (upright blue triangle), day 2-5 (blue asterisk), day 5 (green square), day 8 (purple circle), day 11 (red diamond), and day 14 (inverted blue triangle). Font colors of the individual species indicate membership within the Flavobacteria (blue), α–proteobacteria (pink) or γ–proteobacteria (green). Example plots of corresponding phylotype abundance patterns are shown for each day. Groups (A thru J) encompass species showing significant correlations with the phytoplankton and biodiversity measures and are also referred to in the text. Arrow indicates a time trajectory.


2D Stress: 0.15
3D Stress: 0.08
A
- Coccolithophores
Abundance
Days
B
+ Roseo Richness
- Bacterial Abundance
C
- Synechococcus
D
- Pico- and
nanoeukaryotes
+ Bacterial abundance
E
- Bacterial richness
- Gamma richness
F
+ Synechococcus
G
+ Flavo richness
+ Flavo eveness
+ Synechococcus
H
+ Bacterial eveness
I
+ Coccolithophores
J
- Roseo richness
+ Bacterial abundance
Fjord Roseo 1
MS024-1C-like
Novel Flavo 3
SED15-like
MS056-3A-like
Sulfitobacter 3
Polaribacter-1
SAR92B-like
SAR92A-like
PI_4r8d-like
Loktanella 1
Antarctobacter-like
CRO-FL8-like
Novel Gamma 3
DG1250-like
RCA-1
Lacinutrix-2
PB1.19-like
MS056-2A-like
Lacinutrix-3
Polaribacter-2
Lacinutrix-1
64-10-like
MS024-3C-like
HTCC2188-like
MS024-1F-like
Cryomorphaceae-2
Novel Gamma 4
Novel Flavo 4
CF60-like
VIS-3-108-like
NorSea12-like
CF6-like
P4C02-like
SHAW601-like
4747-27F-like
Methylophilales-like
MS024-2A-like
Verru 3
Verru 1
Verru 2
Fjord Gamma 2
Marinovum-like
CHAB-1
SAR116-like
Novel CHAB
SHWN Night-like

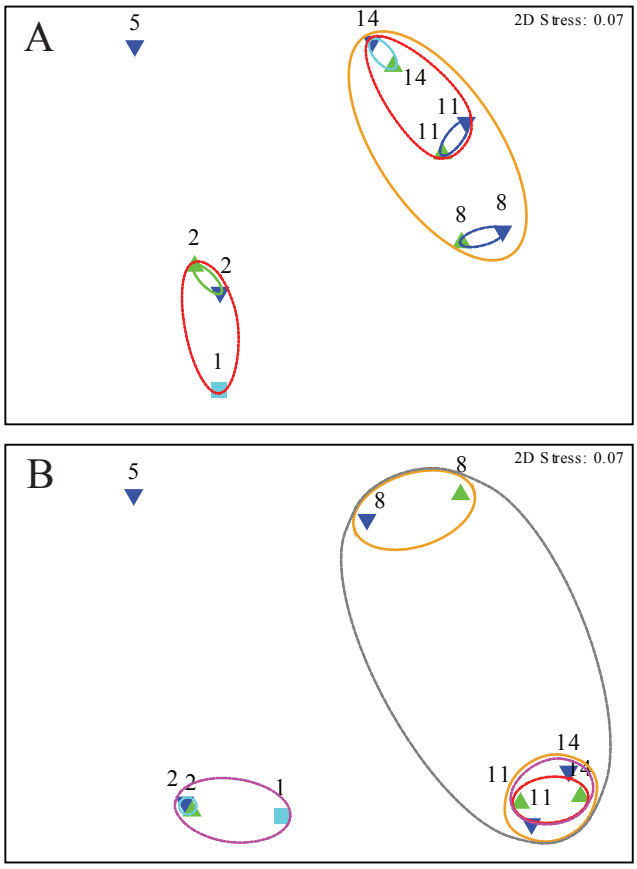


Figure 4.9. MDS ordinations of Bray Curtis resemblance matrices between samples

(A) Phytoplankton abundances (B) Phylotype abundances. Green triangles represent the deplete treatment and the blue triangles represent the replete treatment. The day is indicated above each marker point. Contours encircle samples with correlations of 0.70 (gray), 0.75 (orange), 0.80 (purple), 0.85 (red), 0.90 (light blue), 0.95 (blue), >0.95 (green).

Supplemental Figures

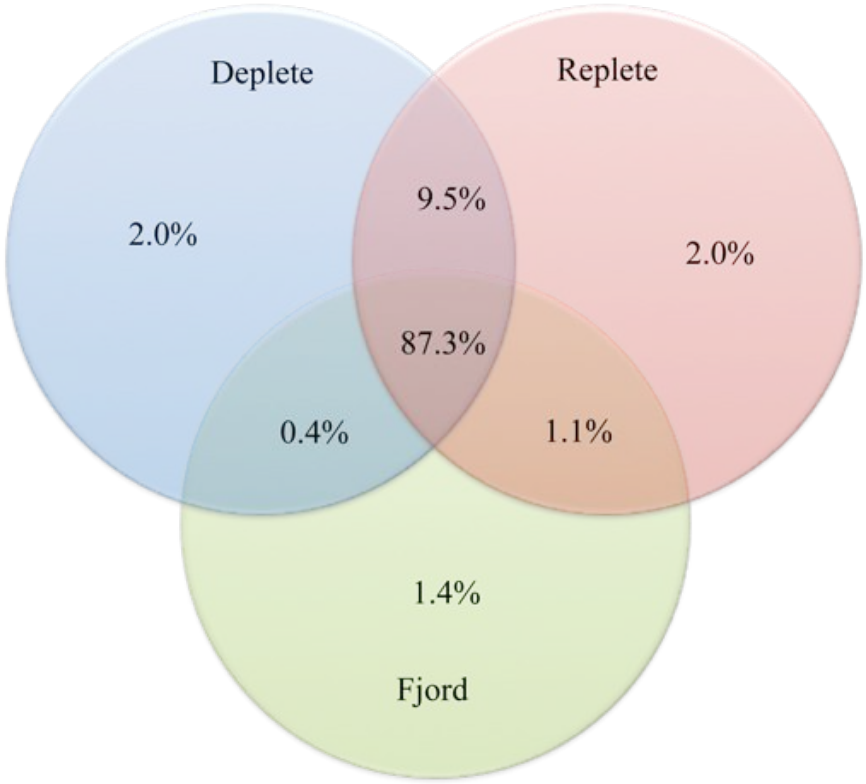


Figure 4.10. Venn diagram showing the percent of shared sequences and shared species

Venn diagram showing the percent of shared sequences between the fjord and the two treatments.

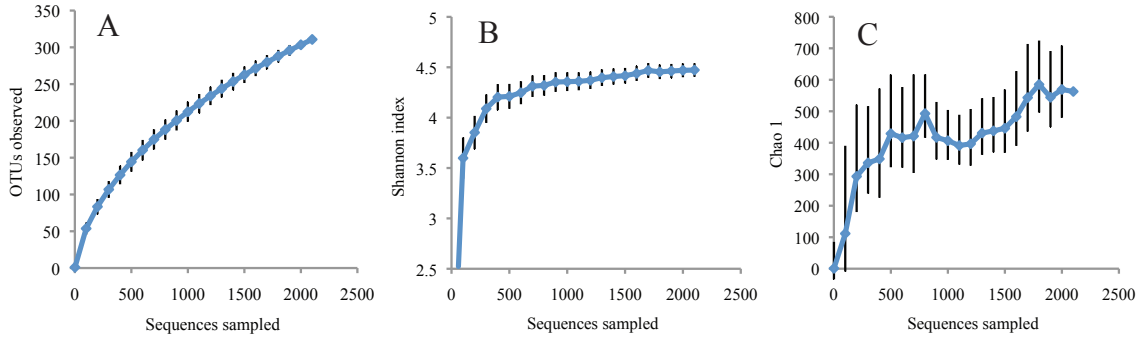


Figure 4.11. Rarefaction and species accumulation curves of the fjord sample

(A) Rarefaction (B) Shannon index H' (C) Richness estimator Chao 1

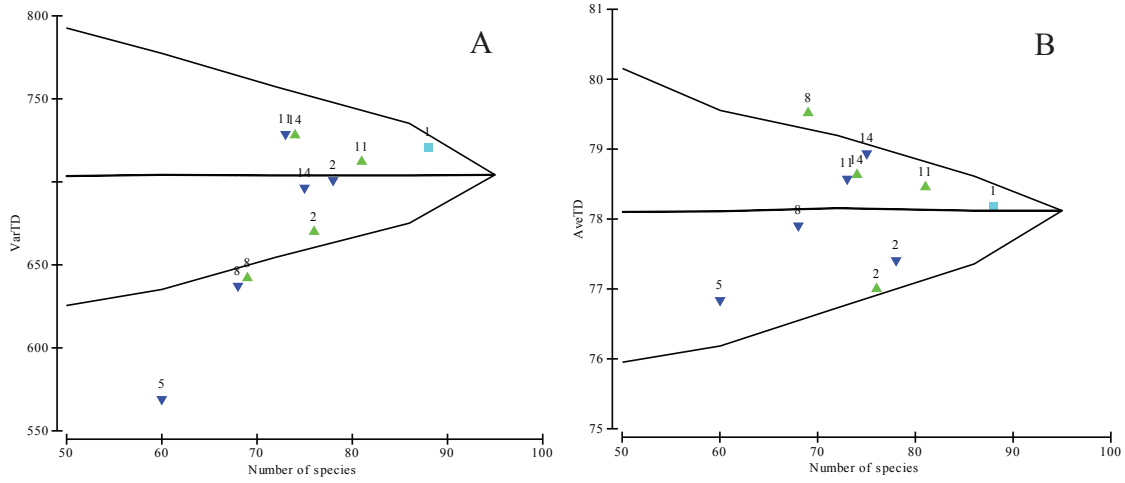
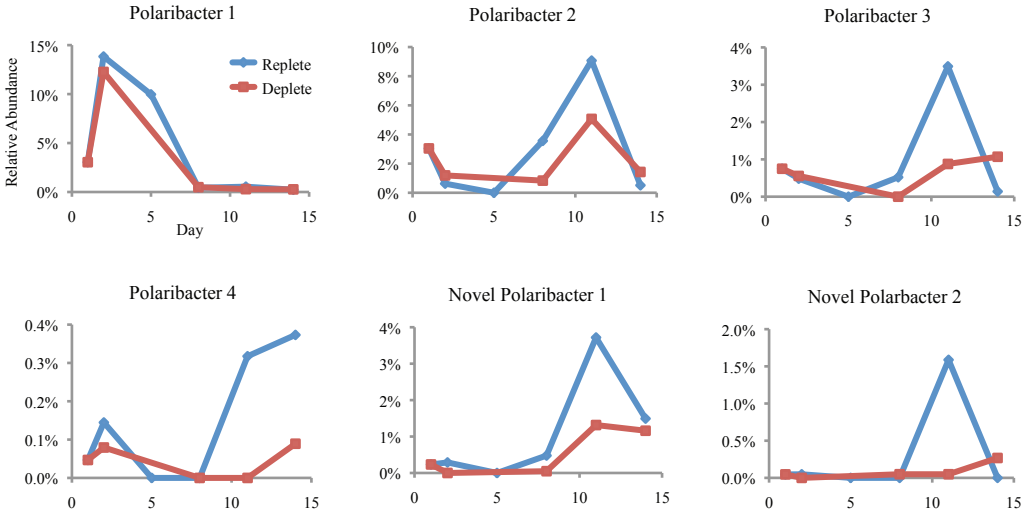


Figure 4.12. Biodiversity measures based on relatedness of species

(A) Simulated VarTD (B) Simulated AvTD. Green triangles represent the deplete treatment and the blue triangles represent the replete treatment. The day is indicated above each data point. Tick lines indicate limits within which 95% of the simulated values lie and the thin line indicates the mean from the master list (aggregation file).

A

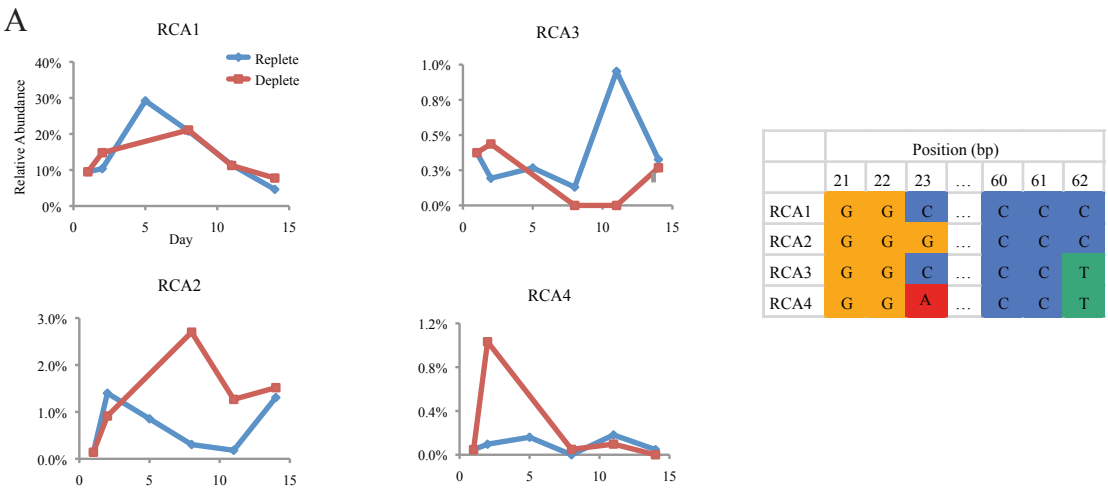


B

	Position (bp)										
	58	…	98	…	101	102	…	107	…	111	112
Polar 1	T	…	A	…	G	C	…	T	…	G	C
Polar 2	G	…	C	…	C	G	…	A	…	C	G
Polar 3	G	…	A	…	C	G	…	T	…	C	G
Polar 4	G	…	A	…	G	C	…	T	…	G	C
N. Polar 1	T	…	A	…	A	G	…	T	…	C	T
N. Polar 2	G	…	T	…	G	C	…	T	…	G	C

Figure 4.13. Differentiation of *Polaribacter*-like phylotypes

Differentiation of *Polaribacter*-like phylotypes demonstrated by relative abundance curves (A) and sequence nucleotide differences (B).



	Position (bp)						
	21	22	23	…	60	61	62
RCA1	G	G	C	…	C	C	C
RCA2	G	G	G	…	C	C	C
RCA3	G	G	C	…	C	C	T
RCA4	G	G	A	…	C	C	T

Figure 4.14. Differentiation of RCA-like phylotypes

Differentiation of RCA-like phylotypes demonstrated by relative abundance curves (A) and sequence nucleotide differences (B).

CHAPTER 5 -
ISOLATION, GENOMIC COMPOSITION, AND ECOLOGY OF ROSEOPHAGE

A version of this chapter will be published by Charles R. Budinoff, Matthew E. Jones, Steven W. Wilhelm and Alison Buchan

CRB coordinated the experimental design, performed the experiments, analyzed and interpreted the data, and drafted the manuscript. MEJ helped with isolating the bacteriophage. SWW participated in the design and coordination and helped to draft the manuscript. AB participated in the design and coordination of the study, aided in the interpretation of the data, and helped to draft the manuscript.

Abstract

Bacteriophage are an important biological component of marine systems, acting as significant agents of bacterial mortality. Lytic infection results in the transfer of organic matter into the dissolved pool, shunting fixed carbon away from higher trophic levels. Characterizing the diversity of marine bacteriophage, including their life cycle strategies and genomic content, is essential for understanding not only their impact on bacterial populations, but on overall ecosystem function. Here we describe the isolation and genomic characterization of bacteriophage infecting the *Roseobacter* lineage, a dominant group of marine heterotrophic bacteria. Roseophage were isolated from two distinct coastal environments, the Chesapeake Bay, a temperate estuary on the eastern coast of the United States and from Raunefjorden, a fjord on the western coast of Norway. Using three distinct *Roseobacter* groups as hosts who originated from these habitats, we isolated a total of seven phage: five Podoviridae-, one Siphoviridae-, and one Myoviridae-like phage. Transmission electron microscopy revealed two Podoviridae morphotypes, those with a small icosahedral capsid measuring 50 nm in diameter (ϕ1005) and another with a large 80 nm capsid size (ϕ1032, ϕ 2047A,B, &C). The Siphoviridae like phage (ϕ 2051) had a 65 nm capsid and a long (~350 nm) non-contractile tail. Plaque assays of ϕ2047A and ϕ 2047C presented bull's-eye-like plaques, which contrasted with the other phage who had clear plaques. Cross infectivity assays for all phage demonstrated very high host

selectivity below the species level, as determined by sequencing of the 16S-23S ITS region and by host genotyping using BOX-PCR. Interestingly, two of the Podoviridae phage (one of each morphotype) isolated from the Chesapeake Bay showed unique infectivity patterns amongst their closely related hosts, where one was capable of infecting two strain types within the species, but the other was only able to infect one strain-type. The genomes of the four phage isolated from Raunefjorden (ϕ2047A,B, & C and ϕ2051) were sequenced and revealed that ϕ2047A and ϕ2047C were nearly identical to each other, except for having distinct putative replication modules, containing an integrase gene and being most similar to known temperate phage, as determined through whole proteome tree building. Phage ϕ2047B was N4-like, similar to other recently isolated Roseophage, and ϕ2051 was a novel Siphoviridae, showing little homology to known lytic phage, but instead to prophage found with the genomes of Rhizobiales bacteria. The genomic details of the sequenced phage are presented, revealing a typical module nature, where groups of genes, such as tail fibers or replication machinery, and even individual genes (deoxycytidine deaminase, thymidylate synthases) appear to undergo horizontal transfer. Furthermore, using a comparison analysis of available genomes for the N4-like phage, we explore, in depth, the relationships and characteristics of the core genes for this group of phylogenetically distinct viruses. Lastly, we performed a thorough search of available metagenomic data to determine the ecology of our isolated phage and of the other known roseophage. Interestingly, the ecology of putative roseophage was similar to that of Roseobacters, often being detected in highly productive environments such as coastal zones, algal blooms, and polar environments, including oceanic waters and marine-derived lakes. Surprisingly, some of the isolated phage were very abundant in their respective habitats from which they were isolated. Using the genomic analysis of the phage, their physiology, and environmental distribution, we discuss the potential ecological life strategies of roseophage.

Introduction

Marine heterotrophic bacteria contribute significantly to the total biomass in the world's oceans and are an extremely metabolically active group of organisms. Bacteria are considered major

secondary producers, converting greater than 25% of the carbon fixed by phytoplankton into cell biomass (Robinson 2008). Although bacterial growth rates are high, their overall numbers remain stable, implying that their production is balanced by their mortality. Physical removal processes, such as sinking, do not typically apply to small organisms like bacteria (Raven 1998) nor do extended periods of starvation (Morita 1997). Instead, the processes of viral lysis and protistan grazing are currently considered the most important sources of microbial mortality in aquatic systems (Thingstad 2000). Recent reviews of the literature have shown a wide range of bacterial mortality rates via both viral and grazer mediated mechanisms and it appears that viral infection has a large impact on bacterial abundance, lysing approximately 20% of their biomass per day (Jürgens and Massana 2008; Pernthaler 2005). Lysis of bacterio- and phytoplankton will shunt energy and nutrients into dissolved pools, preventing their immediate transfer to higher trophic levels. Moreover, viruses are thought to play a role in structuring microbial communities (Pernthaler 2005). Bacteriophage have varying degrees of host specificity, where an individual phage is limited in which types of bacteria it is capable of infecting. By preying on specific taxa, viruses directly alter bacterial population structure and consequentially change community function. Additionally, viruses can introduce genetic traits through horizontal gene transfer, altering the fitness of specific population members, thus again impacting community function. And lastly, it has been recently suggested that temperate phage, including prophage, may alter the physiology of marine bacteria through the regulation of gene expression (Paul 2008), creating a scenario where bacterial activity may be in part be under the control of chronic viral infection by pseudolysogenic and lysogenic bacteriophage.

Given the central role that viruses play in aquatic systems it is important to garner a better understanding of how their diversity and life cycles can shape food webs. Recently, high-throughput sequencing technologies have facilitated metagenomic studies of natural viral communities (Breitbart et al. 2002; Culley et al. 2006; Edwards and Rohwer 2005). These studies revealed that many of the sequence reads obtained had no recognizable homology to the records in public databases and that there was limited overlap between different marine samples. The authors concluded viral diversity to be extremely high in the global ocean and that it is poorly represented in the current databases. The usefulness of metagenomic studies in

understanding the diversity and function of microbial populations is presently rooted in the availability of well-characterized genes and proteins present in sequence depositories. Information on the role and source of individual proteins is based on decades of research using cultured isolates. The isolation, characterization, and genomic sequencing of virus strains will assist in interpreting the wealth of data obtained from viral metagenomes. Furthermore, cultivation allows the development of model host-virus systems providing a framework to study how the environment and microbial physiological diversity shape one another. Part of our motivation for this study was to obtain genomic data that could be utilized in the design of quantitative molecular probes to monitor specific types of bacteriophage in natural microbial communities. Having a molecular toolbox to measure abundance and diversity of individual phage-types is an important step in understanding phage infectivity patterns in marine systems.

As of 2008, only a few published reports detailing the isolation and genetic structure of lytic bacteriophage from non-pathogenic marine heterotrophic bacteria were available (Lohr et al. 2005; Männistö et al. 1999; Rohwer et al. 2000). Marine bacteriophages infecting cyanobacteria and bacterial fish pathogens have been the focus of many genomic studies for the past decade (Hess 2008; Paul and Sullivan 2005), especially those infecting picocyanobacteria, for which dozens of genomes have been recently sequenced (Henn et al. 2010; Sullivan et al. 2010). Committing resources to studies of cyanophage is understandable given their host's abundance in the ocean and their significant contribution to primary production. Less attention though has been devoted to bacteriophage infecting non-pathogenic marine heterotrophic bacteria. This is surprising given the relative ease at which certain marine bacteria can be grown on agar plates, making them amendable to traditional bacteriophage isolation techniques. Another goal of our study was to help fill this void, through the isolation of bacteriophage that infect ecologically important marine heterotrophic bacteria.

Bacteria of the *Roseobacter* lineage are abundant marine heterotrophs, comprising greater than 15% of the bacterioplankton in certain marine systems and mediate several key biogeochemical processes, including transformations of organic and inorganic sulfur compounds, oxidation of carbon monoxide, and degradation of vascular plant material. (see (Buchan et al. 2005a) for a review of the *Roseobacter* lineage). Roseobacters are broadly distributed across

diverse marine environments, yet concentrations are often highest near phytoplankton blooms and in coastal systems. Most important to this study, unlike other numerically dominant marine bacteria, roseobacters are readily isolated into pure culture. These characteristics make them model organisms for studying bacteriophage ecology. Here we describe the isolation and characterization of roseobacters and "roseophage" from two distinct coastal environments. In the past three years other researchers have been interested in roseophage as well (e.g. Angly et al. 2009; Chan 2010; Huang et al. 2011; Zhao et al. 2009b), providing us subjects to perform a thorough genomic characterization of known lytic roseophage.

Materials and Methods

Sampling

Samples were collected from the Chesapeake Bay, USA (July 2007) and the Raunefjorden, Norway (June 2008). The Chesapeake Bay samples originated from site #2 South (38 28.00 N 76 23.16 W). This area of the Bay typically develops hypoxia in the summer months and was so at the time of sampling. Samples came from the surface (2 meters), the oxycline (~15 meters), and the hypolimnion (~25 meters). The samples from Raunefjorden were part of an international mesocosm experiment (see chapter 4 of this dissertation) and were of surface samples taken throughout the experiment. Seawater samples were used in direct plating of bacteria (see isolation below) and for making viral concentrates. Viral concentrates were made using a Labscale TFF System (Millipore) equipped with a Pellicon XL 50 Ultrafiltration Cassette (catalog #PXC030C50, Millipore). Seawater was first filtered through either a GF/F (Chesapeake Bay samples) or a GF/D filter (Raunefjorden samples) followed by passage through a 0.22 μm polycarbonate (Chesapeake Bay samples) or polyethersulfone (Raunefjorden samples) membrane. The pre-filtered seawater was then added to the TFF System and was concentrated according to the manufacturer's operation manual. Seawater samples were concentrated ~10X. These viral concentrates were then stored at 4°C.

Isolation of roseobacters

Bacterial isolation from the Chesapeake Bay samples has been previously described (Zhao et al. 2009a). Briefly, heterotrophic bacteria were isolated from the Bay by direct plating onto artificial seawater agar plates supplemented with dimethylsufoniopropionate (DMSP) and nutrients (ammonia, phosphate, etc.). Agar plates used on the Raunefjorden samples were made using 0.22 µm filtered fjord water instead of an artificial basal media. The filtered fjord water was mixed with sterile molten agar (Noble Agar, 0.8% final concentration) and the nutrients. Plates were incubated at 20°C. After 48 to 72 hours visible colonies were re-streaked onto original isolation media. Colonies were subsequently purified on a 2% w/v artificial seawater (ASW) media containing 0.25% yeast and 0.4% tryptone as previously described (see chapter 3 of this dissertation). After purification, the organism was frozen at -80 °C in 25% v/v glycerol. Strains were named with the prefix 'ACB1' or 'ACB2' if they were isolated from the Chesapeake Bay or from Raunefjorden, respectively. Unless otherwise noted, all further growth experiments with the bacteria and virus isolates were performed in 2% ASW and at 20°C. The 2% ASW media, minus vitamins, ammonia and phosphate, was used in subsequent purification steps, being referred to as MSB buffer.

Molecular analysis of roseobacters

Genomic DNA extraction from the bacterial isolates and PCR of their 16S rRNA gene and 16S-23S internal transcribed spacer region (ITS) was carried out as described previously (see chapter 3 of this dissertation). The 16S rRNA genes from the isolates were individually aligned to the Ribosomal Database Project (RDP) database (Cole et al. 2009) using Clustal X (Larkin et al. 2007) (append sequences option). The ITS sequences were aligned amongst themselves in Clustal X. The 16S rRNA genes from the isolates and type strains of the Roseobacter lineage from RDP, as well as the ITS sequences from the isolates, were tested for a model of evolution with JModelTest (Posada 2008) (default settings), and then used as input files for the program PhyML (Guindon et al. 2010), where maximum likelihood bootstrap trees were created. PhyML was run in parallel 'mode' on Newton high performance computing cluster at the University of

Tennessee. Trees were visualized with ARB (Ludwig et al. 2004). Genomic fingerprinting was done using repetitive element PCR, or BOX-PCR, with the primer BOXA1R as originally described (Versalovic et al. 1994). Individual BOX-PCR reactions were carried out in 25 µL volumes and contained 20 ng of genomic template, 1 µM of primer, 0.25 µL of FailSafe™ Enzyme Mix (EPICENTRE® Biotechnologies) and 1.5 µL of FailSafe™ PCR 2X PreMix (Type B). PreMix Type B was found to give the most optimal banding patterns (number and separation of bands), for the strains studied. Thermal cycler conditions were as follows: initialization for 3 min at 94°C; 30 cycles of denaturation at 94°C for 1 min, annealing at 53°C for 1 min, and elongation at 68°C for 8 min; final elongation at 68°C for 10 min; final hold at 4°C. PCR products were electrophorised through a 1.5% agarose gel for 3 hours at 100V, stained with ethdium bromide, and imaged.

Isolation, purification, and host specificity of roseophage

Roseophage were isolated using standard bacteriophage enrichment techniques. Strains used for isolation include; Rhodobacteraceae sp. ACB1005, Rhodobacteraceae sp. ACB1032 (sub-strain of ACB1005), *Sulfitobacter* sp. ACB2047, *Loktanella* sp. ACB2051 (all isolated in this study) and Rhodobacteraceae sp. SIO67 was kindly provided by F. Rower. For enrichment, liquid ASW was inoculated with a fresh bacterial colony and incubated shaking at 100 rpm over night or until reaching an OD_{540} between 0.20-0.30. The cell culture was then mixed 2:1:2 with viral concentrate (from the same environment that the host was isolated from, except for SIO67 where Chesapeake Bay water was used), fresh ASW, and incubated for 48 hours. Enrichments were spun at 5000 x *g* for 10 minutes and then filtered through a 0.22 µm syringe filter (cellulosic). Clarified enrichments were then checked for the presence of plaque forming units. Purification of plaques and preparation of phage stocks were based on standard methods (Sambrook and Russell 2001). Isolated phage were plaque-purified a total of six times. This number of purification passes was determined based on the distinct plaque morphologies of ϕ2047A and ϕ2047B, where these phage were co-occurring until the 5^{th} successive plaque assay. Once purified, concentrated lysates were made by gently washing soft agar from 10 completely lysed plates of

each phage strain using 7 mL of MSB buffer, followed by the above clarification procedure. Lysates were stored at 4°C. Cross-infectivity was first screened by direct plating of lysates onto bacterial lawns and then positive samples were further examined through a plaque assay. Strains used for cross infectivity included the above host strains, as well other *Roseobacter* strains; *Sulfitobacter* sp. EE-36, *Ruegeria pomeroyi*, *Sulfitobacter pontiacus*, *Ruegeria* sp. TM1040, *Sulfitobacter* sp. NAS-14-1, *Phaeobacter* sp. Y4I, and Rhodobacteraceae sp. GAI21, as well as the phage strains EE36ϕ1 and DSS3ϕ1 (Zhao et al. 2009b) (kindly provided by F. Cheng)

Electron microscopy of roseophage

Electron microscopy grids (Cat. #FCF200-Cu, Electron Microscopy Sciences, Hatfield, PA) were glow discharged and then used for negative staining of phage within 12 hours. Grids were placed on to a drop of concentrated phage lysate for 1 min, removed, and wicked with Whatman paper. Grids were then placed on to a drop of 0.5% urynal acetate for 0.5 min, wicked, and then placed onto a drop of molecular grade water and then immediately wicked dry. Grids were visualized on a Hitachi H-800 transmission electron microscope.

Phage DNA purification

For virus strain ϕ2047B, concentrated phage lysate was purified using density gradient centrifugation as outlined previously (Lawrence and Steward 2010). Briefly, an Optiprep 4-step gradient (20, 25, 30, 35%) was poured into 12 mL ultracentrifuge tubes (Cat# 7030, Seton Scientific, Los Gatos, CA) using the underlaying approach and allowed to blend for 2 hours. Lysate (500 µL) was added to the gradient and spun at 38,000 RPMs for 8 hours at 15°C using an SW41ti rotor (Beckman Coulter, Inc., Brea, CA) and a L8-70M ultracentrifuge (Beckman Coulter, Inc.). Three bands were viable; a yellowish one at the bottom of the lysate layer, and two more very close to each other, one of which was white and the other, slightly lower, with a hint of blue. The blue tinted band was extracted from the tube using the direct unloading approach. Optiprep was removed from the purified phage particles by rinsing 3 times with MSB buffer using Amicon centrifugal ultrafiltration devices (Cat #UFC903024, Millipore, Billerica,

MA) according to accompanying manual. Special care was made to do a final rinse of the membrane using the purified phage solution, as is necessary to ensure good recovery of virus particles (J. Lawrence, personal communication). Fluorescent microscopy was performed at each stage of the purification to measure particle loss as previously described (see Chapter 2). DNA was extracted from the Optiprep/Amicon purified phage particles as described in Chapter 1. For ϕ2047A, ϕ 2047C, and ϕ 2051A phage particles for DNA extraction were obtained using standard methods (Sambrook and Russell 2001), where nucleases and chloroform were added during lysate preparation to remove host DNA. Briefly, soft agar from each lysed plate was rinsed with MSB buffer containing 50 µg/mL of both DNase and RNase, and 1.5% chloroform for 15 minutes at room temperature, on a rocker. This was followed by a final incubation of the lysate at 37°C for 30 minutes. Lysate were clarified as described above and then subjected to polyethylene glycol (PEG 8000) precipitation. The PEG pellets were resuspended in MSB and DNA extraction proceeded as described above. All phage DNA preparations were confirmed for purity using a restriction enzyme approach, where the DNA was cut with the enzymes HindIII and BamHI, and run on an agrose gel (data not shown).

Phage DNA sequencing, assembly, and annotation

Phage DNA from strains 2047A, 2047B, 2047C, and 2051A were submitted to the Broad Institute (Cambridge, MA) and sequenced under the Gordon and Betty Moore Foundation's (Palo Alto, CA) Marine Phage, Virus and Virome Sequencing Project. The Broad Institute supplied the sequencing data as standard flowgram files (.sff) upon request. These files were assembled in Lasergene's SeqMan Pro (DNASTAR, Inc., Madison, WI) using default settings. Assemblies resulted in the generation of a single large contig for each phage strain, which had sequencing coverage averaging 3X to 30X. A limited number of smaller contigs having low coverage (<2X) were present in some assemblies and were considered host/virus contamination. The single large contigs were annotated using default settings at the RAST Annotation Server (Aziz et al. 2008) and tRNAscan-SE Search Server (Schattner et al. 2005).

Whole genome analysis

The phage genomes were analyzed using many client and server-based applications. Peptides from all phage genomes were used as blastp queries (default parameters) to NCBI's Non-redundant protein sequences database to manually curate possible gene functions and to identify nearest phage or prophage relatives. Searches were performed against sequences available as of April 2011. GenBank genome files of related phage were used as-is and not reannotated. Identified relatives were used with the CoreGenes extension CGUG server (Mahadevan et al. 2009) to identify gene homologues (BLAST score threshold = 85). Genes identified by CGUG were considered 'core' if they appeared in all relatives of a given phage family (e.g. N4-like phage), 'extended core' if they appeared in more than one subfamily (e.g. N4 roseophage and N4 pseudomonas-like phage). If genes were identified in all members of a sub-family then they were considered specific (e.g. roseophage specific) or semi-specific (e.g. in two or more N4-like roseophage). Related phage genomes were aligned in Mauve 3.0 (Darling et al. 2004) using the progressive global aligner (default settings) to visualize nucleotide similarity and gene synteny across the genomes. The alignments were subjected to a similarity comparison using Sean Eddy's ALISTAT module (implemented at the National Centre for Biological Sciences Integrated Web Server, http://caps.ncbs.res.in) to obtain average percent pair wise identities. Here the collinear blocks identified by Mauve were analyzed individually and then combined to obtain values that were representative of ~ 85% of all nucleotides in each genome. Whole genome 'phylogenetic' comparisons of the roseophage strains were performed with the CVTree server (Xu and Hao 2009), using all available K-mer lengths. Here, the whole proteomes of the roseophage strains as well as the available phage proteomes in GenBank were used to construct distance trees based on a compositional vector approach using the relatedness of their oligopeptide content.

Single gene analysis

Select roseophage phage core genes were subjected to a phylogenetic analysis. Only the back half of the virion-encapsidated RNA polymerase (vRNAP) gene was used in these analysis.

Translated genes were aligned using Clustal X (default parameters, manual post-check), tested for a model of evolution with ProtTest (Abascal et al. 2005), and then used as input files for PhyML or FastTree (Price et al. 2010), where maximum likelihood trees were created. Select gene trees were then used in the program pplacer (Matsen et al. 2010), where partial-gene, environmental sequences greater than 100 amino acids (see below) were incorporated into the 'reference tree'. Select core gene alignments were also subjected to similarity and identity comparisons using the program MatGAT (Campanella et al. 2003), for the creation of a BLOSUM62 similarity matrix (default parameters), and the ALISTAT module. The BLOSUM62 similarity matrices were imported into the program PRIMERv6 (Clark and Gorley 2006) where MDS plots were generated. The terminase gene, the only phage ORF that is capable of being aligned and compared from practically all know tailed bacteriophage (including roseophage), was aligned using PSI-Coffee (Kemena and Notredame 2009) (default parameters), used as input for tree creation by FastTree 2.1 (Price et al. 2010) (-gamma option), and then visualized with iTOL (Letunic and Bork 2007)

Comparison to environmental samples

The roseophage genomes obtained in this study were used as BLAST queries against environmental sequences found in the databases of CAMERA (Sun et al. 2011), NCBI, and MG-RAST (Meyer et al. 2008). Samples found within MG-RAST were uploaded into CAMERA for analysis. Searches were performed against sequences available as of April 2011 and used all default parameters found at CAMERA and NCBI, except the number of returned hits and the e-value cut-off thresholds of CAMERA searches, which were set rather high (i.e. 40,000 results, e-value of 10^{-5}) to obtain a comprehensive dataset. The results were later parsed at set cut-off thresholds depending on the types of comparisons examined. A variety of searches were performed using individual select core peptides (tBLASTn, blastp) as well as whole genomes (tblastx). Subject matches to select core genes were aligned and phylogenetically assessed as described above. To reduce the number of environmental sequences appearing on the phylogenetic trees a 'de-replication' step was performed by creating consensus sequences of

overlapping reads from the protein alignments. Here, subject sequences were formed into contigs if they had at least 50% of their residues overlapping at >98% similarity with another sequence. The BLAST results from the whole genome searches were parsed into recruitment plots to visualize sample coverage. Metagenomic samples demonstrating high recruitment to roseophage genomes were assembled using CAMERA's 454 Read Assembly to form contigs. These assemblies were annotated using MG-RAST and were subjected to various BLAST searches as described above. These metagenomic samples were also processed by CAMERA to predict rRNA sequences using an implementation of hmmer 3.0. Predicted 16S rRNA sequences were then submitted to RDP's Classifier (Wang et al. 2007) to determine gross phylogeny and read contribution of the bacterial communities present in the samples.

Results and Discussion

Phage morphology

The *Roseobacter* strains utilized in this study made excellent hosts for the isolation of marine bacteriophage. Single-host enrichments with seawater or viral concentrates almost always produced plaque-forming units on agar overlay plates regardless of the *Roseobacter* strain chosen. Culture conditions affecting the success of an enrichment were not investigated thoroughly, but higher incubation temperatures (>25°C) were not as effective, likely a result of increased host growth rates and less time spent in exponential phase (data not shown). Using just four *Roseobacter*-like species as hosts, we isolated 7 morphologically or genetically distinct phage, including at least 4 unique podoviruses, a siphovirus and a putative myovirus. Transmission electron microscopy of phage particles is shown in Figure 5.1 and 5.2. Phage ϕSIO67-Myo resembles Myoviridae-like phage, having a long contractile tail (~150 nm) with a capsid about 85 nm in diameter (Fig. 5.1). Phage ϕ2047A, ϕ2047B, ϕ2047C, and ϕ1032 were indiscernible from each other, resembling large Podoviridae-like phage with capsids about 80 nm in diameter. Phage ϕ1005 also resembles Podoviridae-like phage, but its capsid is smaller, around 50 nm in diameter. Lastly, phage ϕ2051 has a Siphoviridae-like morphology, evidenced by a long non-contractile tail (~350 nm) with a capsid about 65 nm in diameter. The majority of

phage produced clear plaques ranging in size from ~0.5 mm to ~2 mm, except ϕ2047A and ϕ2047C, who's plaques were characterized by a 'bull's-eye' morphology having distinct rings (Fig. 5.3A). Such plaque morphology is often indicative of a phage with a temperate or lysis-inhibited life-style (Abedon 2008), which was the case for these two phage, as determined through genomic sequencing (see below). Temperate phage are a common in marine environments and often observed in bacteriophage isolations (Moebus 1983; Ohki and Fujita 1996; Paul et al. 1998).

Cross infectivity

Cross infectivity assays demonstrated that the isolated phage were highly host specific, infecting only the original isolation host or a sub-species of that host. Phage ϕ1005, ϕ1032, ϕ2051 and their hosts offer excellent examples of roseophage host specificity. Bacterial isolates ACB1005 and ACB1032 represent novel type strains from a group of highly similar isolates from the Chesapeake Bay, referred herein as clade 1005. The strains of clade 1005 share >99% 16S rRNA gene similarity amongst each other and based on their ITS sequences, can be divided into three groups (Fig. 5.3A). Strains in group I and III are characterized by a 125 bp insert in the ITS region that is absent from group II. BOX-PCR fingerprints (Fig. 5.3B) of strains ACB1083 and ACB1088 were more similar to each other than to ACB1032. Strains ACB1005 and ACB1023 had distinct fingerprinting patterns that were quite different from that of the other strains. Group I strains of clade 1005 are infected by both ϕ1005 and ϕ1032, but in contrast, strains in group II of clade 1005 are only infected by ϕ1032 and group III (strain ACB1023) is infected by neither phage (Fig. 5.3) Lastly, strain ACB1028 showed a significantly lower infection rate by ϕ1032 compared to ACB1032. A recent survey of *Roseobacter* diversity within the Chesapeake Bay using the Rhodobacterales GTA major capsid protein gene (g5) as a marker (Zhao et al. 2009a) showed that members of clade 1005 were highly abundant in certain hypoxic regions within the Bay. Furthermore, initial studies using specific ITS primers have shown that both Group I and II of clade 1005 are detectable in the Bay (data not shown), providing a potential opportunity to study the role viruses play in maintaining microdiversity within a natural bacterial population.

The bacterial isolate *Loktanella* sp. ACB2051 represents the type strain from a group of highly similar isolates from Raunefjorden. The strains of clade 2051 share 100% 16S gene and ITS region similarity with each other. Although, the strains can be differentiated based on their BOX-PCR fingerprints (Supp. Fig. 5.12). Phage ϕ2051 was only capable of cross infection with *Loktanella* strain ACB2058, who's fingerprinting pattern was practically identical to the original isolation host, ACB2051 (Supp. Fig. 5.12).

High host specificity amongst our phage isolates is in agreement with previous studies of roseophage, which virtually all reported no cross-infectivity (except for one SIO1-phage strain see (Angly et al. 2009)). Amongst cyanophage infecting marine picocyanobacteria (the most studied marine bacteriophage-type) it appears that the cyanomyoviruses have a much broader host range than cyanopodoviruses or cyanosiphoviruses (Sullivan et al. 2003; Wang and Chen 2008). The putative roseomyovirus isolated in this study did not cross-infect any of the tested strains, although no strains phylogenetically similar to host SIO67 were available. In general, past infectivity studies of marine heterotrophic bacteria have shown large variabilities in phage host range, where both narrow and broad susceptibilities have been found for multiple bacterial groups (Comeau et al. 2006; Holmfeldt et al. 2007; Kellogg et al. 1995; Moebus 1983; Shivu et al. 2007; Wichels et al. 1998; Wichels et al. 2002). Deciphering infectivity patterns is difficult for a few reasons: (1) the large diversity of marine hosts and phage present in the environment (2) the limited number of comprehensive studies incorporating the genomic diversity of both host and phage (3) host susceptibility has often been measured using the direct plating assay or 'spot test', without any follow-up confirmation with a dilution to extinction plaque assay. The direct plating method in our experience, although quite easy to perform, often results in false positives, where lawn clearing was not the result of a true infection, but was due to lysis from without, where the bacteria are lysed by a high-multiplicity of virion adsorption instead of from within, through the production of phage (Abedon 2011). Furthermore, cross-infectivity is rarely examined in detail so as to categorize the host-range type (adsorptive, penetrative, productive, etc.) as outlined by (Hyman and Abedon 2010). Lastly, most bacteriophage isolation protocols employ a single-host enrichment step (as utilized in this study) that may be selecting for the most virulent phage present in the sample, which in turn may also be phage with very narrow host

ranges (Jensen et al. 1998). It is quite likely that the culture conditions used for enrichment are creating a bias in what roseophage-types have been isolated thus far and future isolation efforts should use innovative or untraditional enrichment approaches to further our understanding of phage diversity and host-range. The development of large-scale, hypothesis driven isolation experiments coupled to viral genomic and metagenomic sequencing pipelines can help shed light on the significance of bacteriophage cross-infectivity within marine heterotrophic bacteria.

Genomics of N4-like roseophage

Phage ϕ2047B is an N4-like bacteriophage, highly similar, but genetically distinct to other recently described roseophage (Chan 2010; Zhao et al. 2009b). N4 phage are a unique group of bacteriophage characterized by: (1) the use of three different DNA-dependent RNA polymerases during growth, including a virion-encapsidated RNA polymerase (vRNAP) that is injected into the host cell during infection and is responsible for the transciption of the early genes (2) the use of single-stranded DNA binding proteins as transcriptional activators (3) a genome with 3' single-stranded ends and (4) a lysis-inhibited lytic cycle, where complete virions are constructed but cell membrane integrity is maintained (Kazmierczak and Rothman-Denes 2006). First isolated on *Escherichia coli* K-12 over 40 years ago, N4-like phage have recently been isolated from *Roseobacter* spp. (Zhao et al. 2009b), *Pseudomonas* spp. (Ceyssens et al. 2010), other Enterobacteriaceae (Born et al. 2011), as well being identified as prophages in the genomes of several bacterial strains belonging to the genus *Moraxella* (see GenBank accession numbers AERK01000018 and AERH01000030).

DNA alignments with publically available N4-like roseophage genomes (see Supp. Fig. 5.13) showed that ϕ2047B shared 43.9, 44.4, and 39.6% nucleotide similarity with ϕDSS3P2, ϕEE36P1, and ϕRvP1 respectively. In comparison, ϕEE36P1 shared 84.1 and 54.8% nucleotide similarity with ϕDSS3P2 and ϕRvP1 respectively. Similarity analysis using a number of single conserved markers (see Supp. Figs. 3 and 5) as well as whole proteome tree building (Fig. 5.5) further supported the distinctiveness of ϕ2047B. Regions of low nucleotide similarity amongst the N4 roseophage genomes were found within the early, middle, and late encoding ORFs (see

Supp. Fig. 5.13). Variation in sequence homology in the early genes is restricted to small ORFs with no known gene function, except for the presence of a putative endonuclease in ϕEE36P1 (which is not present in the other N4-like roseophage), and what appears to be out of place structural genes (Fig. 5.6). Recently, (Chan 2010) performed a mass spectrometric analysis of a N4 roseophage virion and demonstrated the presence of certain early genes within the virion, which are likely a putative structural or head-decorating module based on homology searches. The small ORFs in the early genes are likely not protein encoding genes, but rather represent a set of vRNAP promoter recognition sites that form stable hairpin structures, as has been determined for ϕN4 (Kazmierczak and Rothman-Denes 2006). Low similarity regions were also apparent in genes involved in host metabolism and interaction. One such region contained genes encoding for deoxycytidine deaminases and thymidylate synthases. These genes are located in relatively the same positions across the N4 genomes (Fig. 5.6), but are represented by analogous proteins (Table 5.1). Most N4 phage contain a deoxycytidine deaminase including all of the N4 roseophage. Although, ϕ2047B and ϕN4 contain a deoxycytidine triphophate deaminase instead of a deoxycytidine monophophate deaminase found in the other N4 phage. Additionally, phage ϕ2047B contains a ThyX host-like thymidylate synthase where the other N4 roseophage and ϕN4 have a distantly related Thy-like protein, similar to predicted proteins in a limited number of other phage. The late or structural genes show high homology overall except for regions within putative tail fiber regions, as well as the large ORF encoding for the vRNAP. Low sequence homology is only observed in the 5' half of the vRNAP gene, with the back half of the gene, the actual location of the conserved RNA polymerase motifs, exhibiting relatively high similarity. The vRNAP protein, although a single ~3500 amino acid polypeptide, is thought to be involved in other phage functions besides transcribing early genes, including injection of the genome into the host and a role in genome replication (Kazmierczak and Rothman-Denes 2006) and references therein). Interestingly, the vRNAP protein contains no cysteine residues, which is thought to aid its passage through the periplasmic space (Ceyssens et al. 2010). BLAST detected conserved domains in the vRNAP of ϕ2047B and ϕLUZ7 at exactly 500 bp from the C-terminus of the peptides. These regions displayed motifs having E-values of 8.90e-03 to an N-acetylmuramoyl-L-alanine amidase and 2.88e-04 to a membrane fusion protein for ϕ2047B and

ϕLUZ7 respectively, which may also function to gain passage through the host's membranes. Similar conserved regions were not found in the vRNAPs of the other N4-like phage. The genetic mosaicism of tailed bacteriophage is well known and is often observed amongst groups of genes, with individual genes, and even within certain protein domains (Hendrix 2003). It is thus not surprising to see such recombination modules in N4-like phage.

Based on protein homology comparisons 22 core genes were identified amongst the N4-phage, including the capsid, terminase, and DNA polymerase. Core extended genes numbered 12 (see methods for description), including the lysis inhibition genes RIIA and RIIB (Supp. Table 5.2), which were found in all the genomes except ϕEcP1. N4-like roseophage-specific and –extended core genes numbered 12 and 24 respectively (Fig. 5.6), many of which had no known functions, except for three host-like ORFs encoding for a thioredoxin, a ribonucleotide reductase, and a putative lysin. N4 roseophage and ϕN4 all have a transcriptional DNA binding protein as well as a MazG-like nucleoside triphosphate pyrophosphohydrolase. T4-like cyanophage also harbor a MazG-like protein and it is thought to regulate transcription and possibly extend host cell survival during infection (Sullivan et al. 2010) and references therein). A similar role for MazG in N4 roseophage seems plausible, given the lysis inhibited growth cycle of N4-like phage. These two genes are not present in the *Pseudomonas* N4s, although ϕLIT1 and ϕLUZ7 have conserved peptides directly upstream of the RNAP1 and RNAP2 subunits that could serve as DNA binding proteins for the vRNAP. Two other ORFs from the N4 phage show potential analogous protein modules, EE36 gp55/N4 gp46 and N4 gp51/LIT1 gp72 (Figure 5.6 and Supp. Table 5.2). The function of these proteins is not known, although N4 gp51/LIT1 gp72 have been shown to be present in the virion. The *Moraxella* N4-prophage have nearly complete N4-like genomes, being similar in size to the other N4 (~70 kb) and contain all 22 core genes. These 22 core genes were practically the only ORFs shared though, where the prophage were missing many of the core-extended genes, including the two lysis inhibition genes. These prophage were more similar to the *Pseudomonas* N4-like phage than to ϕN4 or ϕ2047B. In contrast to the *Pseudomonas* N4-like phage, the prophage contained an integrase, a thymidylate synthase and a ribonucleotide reductase. As with the N4-like roseophage, the prophage ribonucleotide reductase was host-like, similar to those found in other related gamma-proteobacteria. Perhaps

coincidently, the *Pseudomonas* N4-phage are thought to contain an entire tail fiber module (see Fig. 5.6) that shares homology with temperate phage within Pseudomonads (Ceyssens et al. 2010).

The N4-core genes were further analyzed to provide insight into gene conservation and aid in the design of molecular tools that could target the individual N4-subgroups. Based on average amino acid similarities the two most conserved genes were the terminase (60%) and the capsid (59%), followed by the DNA polymerase and the portal protein, both having 53% average similarity (see Supp. Fig. 5.15). Other genes showing high conservation include N4 gp42, 43, and 44 in the replication module and the late encoding gene N4 gp69. Genes demonstrating low conservation were the lysis inhibition gene RIIA and multiple genes in the structural module. Multi-dimensional scaling of protein similarity matrices provided a novel means to determine gene relationships between the different N4-phage types, especially those proteins that have no known function. MDS ordinations of the individual core peptides provided the following general observations (see Supp. Fig. 5.16): (1) In general, the N4-like roseophage grouped together, with ϕ2047B as an outlier, as did the N4-like *Pseudomonas* phage. This was in contrast with the Enterobacteriaceae phage ϕEcP1 and ϕN4, who rarely grouped close to one another. (2) In general, ϕN4 grouped closer to the roseophage than did any of the other phage strains, most notably with the portal peptide, indicating a potential difficulty in creating roseophage specific probes using these genes (3) Ordinations of N4 gp52 and N4 gp57 provided a wide distribution of the phage strains, demonstrating no clear grouping patterns. (4) The RNAP2, N4 gp25, and helicase genes were the only peptides that demonstrated clear phage strain grouping by the taxonomic family of their respective hosts (i.e. Rhodobacteriaceae, Enterobacteriaceae, and Pseudomonadaceae), indicating their possible use in designing N4 group-specific probes. A second-stage MDS plot is shown in Figure 5.10 with the most notable observation being the tight grouping of proteins involved in the manipulation of the genomic DNA (translation, transcription, etc.) such as the RNA polymerases, DNA polymerase, single-stranded DNA binding protein, helicase, and terminase. N4 gp53 also groups with these proteins, but it has no known function. The evolutionary forces behind such a correlation are elusive, given that the genes are located throughout a bipartite genome and are likely affected by modular gene transfer

events. The mechanisms of N4 DNA replication and head packaging are not fully elucidated, including the role of the vRNAP and the unusual 3' genome extensions. A phylogenetic analysis of the terminase gene indicated that the N4-like phage grouped with other phage having no known DNA packaging strategy (Supp. Fig. 5.14). More work understanding transcription and replication mechanisms within N4 may provide insight into gene evolution and selective pressures on phage lytic growth. Lastly, the second stage MDS also showed strong correlations between N4 gp43, the capsid, and N4 gp67. N4 gp52, N4 gp57, and N4 gp69 were outliers on the ordination.

Genomics of ϕ2051A

Phage ϕ2051A is a novel lytic-member of the Siphoviridae and is most similar to prophage found in the genomes of *Sinorhizobium meliloti* AK83 (NC_015590) and *Agrobacterium* sp. ATCC31749 (NZ_AECL01000049), both of the bacterial family Rhizobiaceae. Phage ϕ2051 shared high protein and nucleotide similarity with these prophage at multiple loci, including structural and replication genes (Fig. 5.8, Supp. Fig. 5.17, Supp. Table 5.3). Although, ϕ2051 contained no integrase gene. Single gene analysis and whole proteome tree building showed ϕ2051 to be weakly similar to known lytic Siphoviridae, but showed little to no nucleotide similarity with these phage and showed protein homology with a few loci, where only one or two signatures phage genes were shared between strains (Supp. Fig. 5.18, Supp. Table 5.3). For example, the DNA polymerase gene of ϕ73 and ϕBcepGomr have high protein similarity to ϕ2051A, but the capsid and terminase do not. In contrast, ϕBcepNazgul has homologous terminase and capsid genes to ϕ2051A, but a non-homologous DNA polymerase. The roseophage ϕRDJL is also a Siphoviridae, but overall shares little homology to ϕ 2051A, with the notable exception being a ribonucleotide reductase, which for all roseophage, is host-like in origin. These phage comparisons illustrate not only the mosaic or modular nature of bacteriophage (as mentioned above), but also the lack of a suite of core genes amongst many related Siphoviridae phage. Recently, as part of the Gordon and Betty Moore Foundation MMI Marine Phage, Virus, and Virome Sequencing Initiative, multiple cyanosiphoviruses have been

sequenced, and preliminary analysis demonstrates a similar picture of genome diversity, including not only gene content, but also genome size (J. Lennon and F. Chen, personal communication). As discussed below, such scenarios make it challenging to detect and quantify certain individual phage types in environmental metagenomic datasets and through the use of molecular probes. Similar to other roseophage, ϕ2051A contained a deoxycytidine deaminase (dCMP-type) and a thymidylate synthase (ThyX-like). Lastly, a phylogenetic analysis of the terminase gene indicated that the ϕ2051A grouped with phage having a λ-like DNA packaging strategy (Supp. Fig. 5.14), which is confirmed through the presence of conserved GpA domains within the terminase gene. Further genomic comparisons with ϕ2051A are limited at this time due to a lack of highly similar relatives available in public databases and indicate the need for further isolation and characterization of roseosiphoviruses, which are likely to be quite diverse.

Genomics of ϕ2047A/C

Phage ϕ2047A and C are nearly identical in DNA sequence content, except for ~3000 bp region that shares no similarity (Fig. 5.9). Phage ϕ2047A/C are temperate, contain an integration module and based on single gene analysis and whole proteome tree building group with known temperate phage and prophage (Fig. 5.5 and Supp. Fig. 5.19) The nearest relatives to ϕ2047A are ϕEPV2, an uncultured phage genome assembled from a viral metagenome of a lab-scale enhanced biological phosphorous removal reactor (Skennerton et al. 2011) and prophage within the Rhizobiaceae. Although ϕEPV2 is uncultured, the genome was assembled from the viral size fraction and based on BLAST analysis many of the encoded host-like proteins are similar to bacteria or prophage of the Rhizobiaceae and Rhodobacteraceae. Given this, and the fact that most of the dominate α-proteobacteria detected in the reactor were similar to *Rhodospirillum* and *Rhizobiales* (García Martin et al. 2006), it likely that ϕEPV2 was actively infecting. Phage ϕ2047A also shared homology with the temperate phage *Myxococcus* ϕMx8 and *Pseudomonas* ϕF116, but not at the nucleotide level and only for a few select genes (Supp. Table 5.4). Phage ϕ2047A and its closest relatives (ϕEPV2, *S. medicae* prophage, ϕMx8, ϕF116, and ϕBcep22) do not contain identifiable DNA polymerases and may rely largely on the host's machinery for

genome replication. Although, ϕ2047A/C (and ϕEPV2) do contain a gene module in which certain ORFs share homology with proteins known to be involved in replication, such as a putative helicase and DnaD- and DnaA-like replication proteins. Similarly, the phage ϕMx8 and ϕBcep22 contain DnaC-like genes. Surprisingly, the replication module is the only significant difference between ϕ2047A and C, supporting the reliance of these phage on host DNA replication machinery and implying that a dynamic portion of the genome is involved in 'assisting' genome replication, perhaps facilitating its efficiency and subsequently increasing the rate at which virions are created. In this vein, ϕ2047A/C contain at least four ORFs encoding for putative lysin- or lysozyme-like proteins (Fig. 5.9), which is in contrast to other known roseophage who only have one identifiable lysin-like protein. Although, certain tail fiber proteins can also contain lysozyme-like or peptidoglycan-binding domains to facilitate DNA injection through the bacterial cell wall (Rossmann et al. 2004) and this is likely the case for two of these ORFs (see below) that are conserved between ϕ2047A/C and ϕEPV2. Nevertheless, such redundancy in lysin and lysozyme proteins suggests a functional mechanism to ensure timely infection and release of virus progeny. Preliminary results in our lab support the above genomic evidence, where ϕ2047A has an extremely short latent period and a very high burst size compared to that of ϕ2047B (N. Ankrah, personal communication). These genomic features and life cycle traits point to a distinct ecophysiology of ϕ2047A compared to other roseophage and may have implications for niche development in natural systems (see below).

In contrast to other known roseophage, the genomes of ϕ2047A and C do not contain a thymidylate synthase or a ribonucleotide reductase, nor do they have a deoxycytidine deaminase, which is found in both roseosipohviruses and the N4-like phage (but not ϕSIO1). Although, ϕEPV2 does have a ribonucleotide reductase within its 'replication' module. Furthermore, all the temperate relatives of ϕ2047A (ϕEPV2, ϕMx8, ϕF116, ϕBcep22, and ϕBcepC6B) contain at least one identifiable DNA methylase, which is not present in ϕ2047A/C. It is important to note that ϕ2047A and C have numerous conserved hypothetical proteins with unknown function. The significance behind the absence of these DNA metabolism proteins is not known, but may be indicative of individual hosts or ϕ2047A/C's life cycle, which appears to be either rapid replication and lysis or entry into a lysogenic state. Both ϕ2047A/C and ϕEPV2 have a DNA

BreC-like integrase and an XRE-like transcriptional regulator (a putative immunity repressor, see below) based on BLAST analysis, but the proteins of each phage share no homology. Rather, the integrase of ϕ2047A/C and the XRE transcription regulator show homology to proteins found within ϕF116 and ϕBcepC6B respectively (Supp. Table 5.4). This is further evidence of single gene mosaicism where related phage show high nucleotide and protein homology for many signature genes (capsid, terminase, etc.), but then also show little homology between other functionally analogous genes, instead sharing homology with genes from more distant relatives. Interestingly, ϕ2047A/C contained an HNH-like endonuclease that was most similar to an endonuclease in roseophage ϕEE36P1. Phage ϕ2047A also contained an additional endonuclease (RusA-like) in the replication module. The location of these two endonucleases may indicate the high mobility of their associated gene modules (Fig. 5.9). Lastly, a phylogenetic analysis of the terminase gene indicated that the ϕ2047A/C grouped with phage having a P22-like DNA packaging strategy (Supp. Fig. 5.14).

Some of the most intriguing ORFs within the genomes of ϕ2047A and C are a group of possible temperate phage specific DNA injection related proteins (Fig. 5.9). BLAST homology results with other temperate Podoviridae support this hypothetical role. For example, ϕ2047A gp5 shows low homology to a protein (PPLUZ24_gp60, YP_001671933) from *Pseudomonas* phage ϕLUZ24 that was determined to be associated with the virion based on mass spectrometry of the phage particle (Ceyssens et al. 2008). Another ORF, gp1 (or Mx8p54-like), shows weak homology at its C-terminus to a putative 'phage injection protein' belonging to *Enterobacteria* phage phiEco32 (see YP_001671769). Interestingly, three of the predicted proteins within the module have high homology to ORFs within ϕMx8 that have no known function; Mx8p54 (NP_203468), Mx8p52 (NP_203466), and Mx8p51 (NP_203465). These three genes are also found together within prophage from many diverse proteobacterial genomes (Supp. Fig. 5.19) and within numerous environmental datasets (see below). Also, the ϕMx8 protein Mx8p56, although not present in ϕ2047A/C, is in close proximity to the putative injection module and is similar to proteins in certain *E. coli* strains thought to be involved in DNA injection (see YP_002403625). Lastly, ϕ2047A/C have another putative structural gene that is exclusively shared with ϕMx8 and certain other temperate/prophage (Supp. Table 5.4), ORF Mx8p57-like,

which in the genomes of ϕ2047A/C and ϕEPV2 is located between two putative lysozyme/tailfiber-like proteins (Fig. 5.9). To our knowledge, it is unusual to find such high homology amongst multiple structural proteins from a diverse range of phage and host-types, not including of course the capsid, portal and terminase genes. The high conservation of these proteins and the fact that they are seemingly exclusive to temperate phage and prophage encourages further discussion as to their role in lysogeny and superinfection.

One well-known reason for temperate phage to enter into a lysogenic state (or lytic phage into a lysis-inhibited state) is a high ratio of infecting phage to hosts, also known as a high multiplicity of infection or superinfection. In bacteriophage lambda the signal of superinfection is mediated through a high dosage of regulator genes transcribed from the multiple genomes present within the host cell after superinfection (Kourilsky 1973; Kourilsky 1974) causing lysogeny. Alternatively, in bacteriophage T4 the signal indicating an excess of phage leading to lysis inhibition appears to be derived from a molecule injected into the host (not a protein created within the host), likely a virion packaged protein, a peptide generated from cleavage of a phage structural component, or the DNA itself (Paddison et al. 1998 and references therein). Based on the genomic information presented above, it appears that certain temperate Podoviridae phage and prophage share numerous conserved proteins that are involved in disrupting the host's membrane and injecting the viral DNA, which could play a role in signaling superinfection (similar to phage T4) and determining the switch to lysogeny. In phage lambda after lysogeny occurs, subsequent infections from identical phage are prevented through immunity control. Here, the prophage prevents the infecting phage from entering into a lytic replication cycle and from integrating into the host genome through the expression of the CI repressor protein. Immunity in ϕ2047A is likely accomplished through the use of a repressor as well. Phage ϕ2047A contains an XRE-like transcriptional regulator that belongs to a large family of DNA binding helix-turn helix proteins, which includes the lambda CI repressor. The putative immunity repressor from ϕ2047A is most similar to proteins found in a Siphoviridae-like prophage in *Rhodobacter sphaeroides* (YP_351703), the Siphoviridae phage Bacillus ϕ105 (NP_690787), and the Podovirdae phage Burkholderia ϕBcepC6B (YP_024964), which groups closely with ϕ2047A in whole proteome trees (Fig. 5.5 and Supp. Table 5.4). Another type of immunity

control is seen in the Podoviridae temperate phage *Xyella* ϕXfas53 (who groups with ϕ2047A in whole proteome trees, but shares practically no protein homology) and bacteriophage P22 that are termed superinfection exclusion proteins (Summer et al. 2010) and references therein). These proteins appear to be associated with the bacterial membrane and are thought to prevent entry of superinfecting phage DNA. No evidence of these proteins was seen in ϕ2047A. Overall, there appears to be multiple mechanisms amongst temperate Podoviridae to not only detect superinfection and invoke lysogeny and but also in immunity control.

Ecology of roseophage

BLAST searches against environmental samples using select roseophage 'core' genes and genomes returned hits from many diverse environments. To help determine which environmental samples actually contained roseophage or roseophage related viruses we examined the subject reads using multiple methods; (1) plotting the distribution of returned reads with respect to their E-value score to determine cross-loci conservation and the suitability of a particular gene(s) for environmental searches, (2) reciprocal BLAST searches of select returned reads, (3) insertion of subject peptides into reference trees using maximum likelihood and parsimony methods (see methods section), and (4) generating recruitment plots to whole roseophage genomes using multiple E-value thresholds. We contend that such a stringent analysis of environmental BLAST searches is necessary to accurately estimate the global distribution and ecology of specific bacteriophage types. We base this on multiple lines of reasoning. First, the degree of conservation between phage 'core' genes amongst closely related isolates or subjects found in environmental samples varies dramatically and thus the number of returned environmental hits from a BLAST search is highly dependent not only on the chosen gene but also on the chosen e-Value threshold. This can be illustrated using N4-like phage as an example, which are unique amongst known roseophage by having a large set of core genes. Figure 5.10 shows that even for the most conserved N4-like genes the distribution of returned reads varies with respect to the e-Value score. In other words, an e-Value threshold that determines if a subject read is or is not an N4-like phage will be different for practically each gene query. To accurately determine

threshold values reciprocal BLAST analysis and tree building should be utilized. Furthermore, the genome presence, conservation, and evolutionary relationships of N4-like signature phage proteins (capsid, terminase, DNA polymerase) are consistent amongst the N4-phage. But this is not the case for the other roseophage. This is illustrated through single gene (peptide) phylogenetic analysis of ϕ2047A and ϕ2051A, where the phage relatives for the individual proteins differ (Supp. Fig. 6 and 5.8). The DNA polymerase of ϕSIO1 offers another example, where the gene is found in practically all T7-like phage isolates and often turns up in environmental samples, but homologues of the capsid and terminase are only found in a few T7-like phage and are far less abundant in environmental databases (data not shown). Thus, the application of only one gene marker in estimating the distribution of roseophage related viruses within environmental samples becomes problematic and the use of multiple signature markers during BLAST searches was required for these phage. Lastly, the size of the subject read greatly influences the returned e-Value, where in general, small subject peptides with 50 or less amino acids did not return values greater than 10^{-15}, even if they were identical to that region of the query (data not shown). Regardless of the homology score, small peptides may recruit only to the highly conserved domains or active sites of the particular protein, which can be found in many types of organisms, such as the host-like phage genes (i.e ribonucleotide reductase). Because of these issues it was not possible to ascertain the number of roseophage-like hits from samples sequenced with FLX technology using single marker BLAST searches. Instead, recruitment plots using the roseophage genomes were constructed to determine the likely presence of roseophage-like viruses in those samples. In this case, a sample was said to contain roseophage-like viruses if reads were recruited to multiple loci on the genome at relatively low scores (10^{-5} to 10^{-15}).

N4-like phage

N4-like phage were found in 17 distinct marine samples (see Fig. 5.11), ten of which were coastal environments, four were open ocean environments or beyond a shelf break, two were marine-derived ponds/lakes, and one was the surface of a marine animal (shark skin). N4-like phage were also found in non-marine sources, including a wallaby's gut and activated sludge. Protein alignments and phylogenetic analysis of the N4-like environmental hits showed they

were quite diverse. For example, 22% of the N4-like DNA polymerase reads (>100 amino acids) were unique (116 out of 524). A phylogenetic tree of the DNA polymerase gene containing the reference N4 phage and the environmental reads (Supp. Fig. 5.20) shows that overall, the reads do not cluster based on habitat, although some do. For example, there are a large number of sequences from an Antarctic tundra pool that are quite similar to a group of sequences from the Ross Sea. Only three distinct reads grouped within the N4-roseophage clade, two from a Galapagos Island mangrove sample and one from an Antarctic tundra pond. Many of the reads were most similar to N4-phage infecting the γ-proteobacteria. Similar results of high diversity, a lack of habitat clustering, and similarity to γ-proteobacterial N4-like phage was also observed with phylogenetic trees of the terminase and capsid genes. Lastly, generation of recruitment plots from short-read metagenomic data sets indicated two samples that were likely to contain N4-like phage, a solar saltern and a tilapia farm (data not shown). In general, N4-like phage were predominantly found in highly productive environments and were not constrained by latitude, salinity, or depth. These trends are discussed in more detail in the conclusions section below.

Many of the N4-like metagenomic 454 reads came from an Antarctic tundra pond (aka an organic lake, see CAMERA sample CAM_SMPL_SRA022182) and were part of a larger library of Antarctic sequences that were created from size-fractionated water samples (0.1-0.8, 0.8-3.0, 3.0-200 μm) and sequenced with 454 Titanium™ chemistry (see CAMERA project CAM_PROJ_AntarcticaAquatic). The majority of the reads were highly similar to each other and formed a distinct N4-like clade (Supp. Fig. 5.20), but there were also dozens of singleton and doubleton sequences that were unique, indicating the presence of multiple distinct N4-like phage in the samples. Because these N4-like reads were from non-viral metagenomes and the fact that N4-like prophages are now known, we questioned whether these sequences were actually free virions. To answer this we examined the libraries more closely, comparing not only the number of N4-like hits from the three size fractions available, but also by quantifying 16S rDNA reads and by assembling the libraries. Assemblies of metagenomic libraries could potentially produce contigs containing host sequence data if the viruses were indeed prophage. But N4-like contigs from these assemblies were only representative of the most conserved regions of the N4-like genomes and had little overlap with known variable regions or genes with low conservation

(Supp. Fig. 5.21). Although, this is not surprising based on the number of distinct signature genes found and is further evidence of high N4-like phage diversity in this environment. The ends of the contigs (adjacent to conserved genes) were subjected to BLAST analysis and indicated that some of the variable regions were most similar to various peptides from marine γ-proteobacteria (data not shown), which is again consistent with the single gene analysis. Quantifying the number of N4-like hits in these libraries indicated that the smallest size fraction contained ~5 fold more hits (per GB) than the larger size fractions (Supp. Fig. 5.22A). This is in contrast to 16S rDNA reads, which increased ~2 fold in the higher size fractions. N4-like read number was only 30% less than 16S read number in the smallest size fractions, leaving the possibility that a certain bacterial taxa containing a N4-like prophage may be over represented in <0.8 μm size class. To investigate this we classified the 16S reads to the taxonomic level of class, where 90% of the 16S reads were divided amongst γ-proteobacteria, α-proteobacteria, and Flavobacteria. The γ-proteobacteria were the most abundant bacterial class in the smallest size fraction, but were almost 2 fold less abundant than the N4-like reads (Supp. Fig. 5.22B). Given that the vast majority of N4-like reads from these samples were likely from phage infecting γ-proteobacteria, based on BLAST results, phylogenic trees, and assemblies, it seems implausible that they are representative of prophage and are likely from virion DNA. The capsid of N4-like phage is one of the largest of the Podoviridae, around 80 nm, and thus some virions likely would have been caught on the filters used to create the above libraries.

ϕ2047A and C-like phage

Phage ϕ2047A/C signature genes used for environmental BLAST searches of included the capsid, terminase, and a putative structural injection protein (Mx8p52-like). Phage ϕ2047A/C does not contain a recognizable DNA polymerase. The putative injection protein was selected based on its level of conservation, where it is found in all close relatives of ϕ2047A/C (see Supp. Fig. 5.19) and BLAST searches with the peptide revealed it to be abundant in environmental databases. The Mx8p52-like peptide was found in diverse environments, including marine, freshwater, and soils (Supp. Table 5.5). Many of the samples were from viral metagenomes, indicating its presence in free virions and not prophage. Interestingly, the gene is often found in

aphotic environments including deep marine waters and groundwater. For example, the Mx8p52-like gene was found in a viral metagenomic depth profile at the San Pedro Ocean Time Series off the coast of California (see CAMERA sample CAM_SMPL_001014), where it increased 10 fold from the upper euphotic mixed layer to the aphotic zone and then doubled from the aphotic zone to the bottom waters of the channel. The environmental distribution of the Mx8p52-like gene and the fact that increased lysogenic phage production is often associated with conditions characterized by low host abundance and growth ((Paul 2008) and references therein), such as deep sea environments (Weinbauer et al. 2003) and potentially groundwater (Ghosh et al. 2009), provides further evidence that the putative injection proteins of ϕ2047A/C and related phage are indicative of a temperate lifestyle and may play an important role in controlling the lytic/lysogenic decision in certain natural systems. Intriguingly though, these putative injection proteins were also found in highly productive environments, including algal blooms.

BLAST searches detected ϕ2047A related phage in ten distinct marine environments (see Fig. 5.11). Because ϕ2047A lacks a large set of core genes like the N4-like phage, environments deemed to contain ϕ2047A-like phage were required to have at least two signature genes present in the sample to be considered, which was a requirement for the other roseophage BLAST searches as well (ϕ2051A, ϕSIO1, and ϕRDJL). Many of the returned subject reads similar to ϕ2047A were from polar or sub-polar environments and temperate estuaries. ϕ2047A-like sequences were also found in a hypersaline lagoon, a marine cold seep, and in a freshwater lake (Yellowstone Lake). Three of the samples containing ϕ2047A-like sequences were viral metagenomes, indicating these phage types do exist as free virions. Unlike ϕ2047B and ϕ2051A, ϕ2047A/C was detected in a viral metagenome from its original isolation source, Raunefjorden, Norway (see chapter 4 of this dissertation). This metagenome was created from samples taken during the same mesocosm experiment in which the host and phage were isolated (see methods). A recruitment plot of the sample to the ϕ2047A genome shows coverage spanning greater than 70% of the genome (Supp. Fig. 5.23). Phylogenetic analysis of retrieved signature peptides from the sample show at least three distinct ϕ2047A-like phage present in the sample, with one type being almost identical to ϕ2047A based on nucleotide similarities (>85%, data not shown). Such good coverage and high similarity of the sample to ϕ2047A indicates that

this roseophage-type was present during the mesocosm experiment and offers a unique opportunity to study phage temporal patterns. Recent ancillary projects have focused on characterizing the phyto- and bacterioplankton dynamics during the mesocosm, which will provide a foundational backdrop to apply molecular probes designed for the genome of ϕ2047A that can monitor roseophage life cycle dynamics during the algal bloom.

ϕ2051A- and ϕRDJL-like phage

Phage related to the roseosiphoviruses were less abundant in environmental samples than the other roseophage-types. Both roseoshipoviruses were detected in marine-derived Antarctic lakes, which was the only environment where all of the roseophage (or roseophage-like) were detected. Signature genes grouping closely with ϕ2051A were detected only in the Antarctic lakes. In contrast, signature genes grouping phylogenticaly close with ϕRDJL were found not only in Antarctic lakes, but also in a coastal embankment and a hypersaline lagoon. Phage ϕRDJL was isolated from *Roseobacter denitrificans*, which is one of the only Roseobacters capable of anaerobic growth (Huang et al. 2011). Although, no correlation with said host physiology and environmental distribution of the phage was seen. The high diversity of marine Siphoviridae isolates observed to date (see above) may contribute to the low overall abundance of individual phage types in environmental samples.

ϕSIO1-like

Phage related to ϕSIO1 were found in many environments and sequences highly similar to ϕSIO1 were quite numerous, making this phage the most abundant known roseophage in marine metagenomic samples. The ϕSIO1 signature genes were found in 12 distinct environments, with another 10 environments containing reads more distantly related to the phage. Similar to sequences related to the other roseophage, coastal environments dominated the BLAST results. But in contrast, ϕSIO1 related phage were in more samples near the equator and were not detected in non marine-sources such as hot springs or freshwater lakes. Retrieved reads grouping strongly with all three ϕSIO1 signature genes were found in three distinct coastal environments,

including Raunefjorden. Viral metagenomic samples from the west coast of North America, the original isolation site of ϕSIO1, contained the highest number of ϕSIO1-like reads and the highest number of roseophage reads from a single sample. An explanation of ϕSIO1's higher overall abundance in available metagenomic samples compared to other roseophage may in part be due to the life cycle of the phage. In contrast to the other roseophage, the genome of ϕSIO1 does not show evidence of a temperate, prophage, or lysis inhibition type life style. Instead, ϕSIO1 is the only known roseophage to encode a PhoH-like protein that was originally described in *E. coli* as a cytoplasmic protein that was induced by phosphate starvation, but appears to be quite cosmopolitan amongst marine bacteriophage, including all known cyanomyoviruses and many other marine viruses (Goldsmith et al. 2011; Sullivan et al. 2010). The function of PhoH homologs in bacteriophage are not clear, although in bacteria some have been linked to phospholipid metabolism and RNA modification and others to fatty acid beta-oxidation (Kazakov et al. 2003). The PhoH-like protein in ϕSIO1 shows the highest homology to proteins in bacteria belonging to the Rhizobiales (*Methylobacterium* spp.) and not to other bacteriophage or prophage.

Conclusions

Isolation of roseophage from multiple seawater samples and host strains, using standard plaque assays techniques was quite successful: demonstrating the utility of the *Roseobacter* lineage for model studies of marine bacteriophage. Members of the *Roseobacter* lineage are typically thought to be generalists (Buchan et al. 2005b; Newton et al. 2010), growing well on a variety of carbon substrates and at high nutrient concentrations, which is undoubtedly the primary reason for their amenability to culture. Such excess nutrient conditions are typically not the circumstances that many Roseobacters experience in the natural environment. Our reliance on traditional culturing conditions certainly influenced the types of roseophage that were isolated in this study, possibly selecting for phage that were highly virulent and whom had very narrow host ranges. Future marine bacteriophage isolation attempts must be more creative and attempt to mimic the natural conditions of the system/host being studied.

The genomes of ϕ2047A/C and ϕ2047B suggest that timing of the lysis event is critical to the success of certain roseophage. The two phage use contrasting strategies to control the release of phage progeny; a lysis-inhibited life cycle, prolonging an inevitable liberation of virions (ϕ2047B) and a quick lytic cycle with a lysogenic option (ϕ2047A). These distinct infection modes may be related to the ecology Roseobacters in general, who dominate coastal systems, which are highly dynamic and productive environments experiencing sudden variations in nutrient supply, salinity, and temperature. These rapid changes can quickly alter the growth and abundance of bacterial populations, in turn leading to increases or decreases in the supply of phage hosts. By controlling lysis in fundamentally different ways, these two roseophage are optimized to how bacterial growth and abundance might change in the future; choose a long-term solution to low host abundance, the lysogenic state, where lytic replication is halted entirely or instead maintain host integrity for the short-term, a lysis-inhibited state, keeping intact virions within the cell, hoping for a quick return in host abundance. Often successful viral infection is conceptualized as an 'arms race' between host and predator, but another aspect of natural viral communities that receives little attention is two-predator one-prey competition, where distinct phage-types with contrasting life strategies are relying on similar prey. Lastly, temperate phage and prophage are generally thought to have lifestyles that are beneficial in environments where overall bacterial growth and abundance is low, but it may also be true that certain temperate phage in highly productive environments take advantage of rapid and short-term changes in host availability.

Roseophage genomic structure and content is consistent with that of previously described marine bacteriophage. The mosaic nature of groups of genes and individual genes is a hallmark of tailed bacteriophage; and roseophage are no exception. Examples of module gene(s) in roseophage are numerous, but one striking example is the putative replication module in ϕ2047A and C, which despite the two genomes being practically identical at the nucleotide level, they share no homology in this group of genes. Like many marine bacteriophage, roseophage also contain a variety of auxiliary metabolic genes (AMGs) encoding critical phage functions. Currently, DNA metabolism or nucleotide synthesis genes are the most prevalent AMGs in

roseophage, although ϕSIO1 does contain genes that are possibly involved in phosphorus metabolism. It will be interesting to see the genomic content of the roseomyoviruses, perhaps they too will be geared towards improving and prolonging host fitness and physiology like cyanomyoviruses. As additional roseophage are isolated and sequenced, more AMGs are likely to be found and may broaden the role that phage play in *Roseobacter* fitness during infection.

Expectedly, the ecology of known roseophage is similar to that of its host, being found predominantly in highly productive environments. Somewhat unexpectedly, bacteriophage related to known roseophage are also often found in similar environments (Fig. 5.11), eluding to a specific ecology for these types of phages. Salinity and temperature were not determining factors for the presence of these phage-types; rather nutrient levels seemed to be the primary driver. Non-saline and hypersaline eutrophic environments contained phage-types related to roseophage, such as sewage related environments, man-made enclosed water bodies (fish farms, salterns) and hypersaline lagoons. Such an observation is not entirely unexpected, as phage are not impacted by temperature and salinity to the degree that a vegetative bacterial cell is. Rather the physiological growth rate and trophic strategy of the host are more important. Testing these observations through the quantification of specific phage-types is not straightforward, though. Like other groups of bacteriophage, roseophage do not have a signature gene marker for designing molecular probes. Based on the genomes available to date, the diversity of roseophage is high; three distinct Podoviridae and two Siphoviriade. Roseomyoviruses also exist, bringing the total number of roseophage types to six. Further complicating matters is the temperate nature of at least one group of roseophage, so that the abundance of free virions can be differentiated from prophage in molecular studies. Initial measures of roseophage population diversity and dynamics may be better suited as focused studies on specific habitats, such as environments where multiple known roseophage types have been observed and their temporal patterns can be investigated (i.e. Raunefjorden). Little is known about short-term temporal variability (hours to days) of viral abundance in general, with reports from different systems often giving conflicting results (Breitbart 2012). Even less is known about short-term variability in the diversity and abundance of particular viral types. Lastly, studies of viral dynamics typically encompass measures of total viral abundance or of particular host types (i.e. cyanophage). Another approach

would be to measure the diversity and activity of particular genomic phage groups and begin connecting function and gene content of viruses beyond the confines of specific hosts. For example, the N4-like phage are a genetically distinct group of viruses that are found in a multitude of environments and dominate in some metagenomic libraries. Understanding the ecology of phage with similar gene content will give us a better understanding of the environmental and ultimately the evolutionary forces driving viral diversity.

References

Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21:** 2104.

Abedon, S. T. [ed.]. 2008. Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses. Cambridge Univ Press.

---. 2011. Lysis from without. Bacteriophage **1:** 46.

Angly, F. and others 2009. Genomic analysis of multiple Roseophage SIO1 strains. Environmental Microbiology **11:** 2863-2873.

Aziz, R. K. and others 2008. The RAST Server: rapid annotations using subsystems technology. BMC genomics **9:** 75.

Born, Y., L. Fieseler, J. Marazzi, R. Lurz, B. Duffy, and M. J. Loessner. 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.

Breitbart, M. 2012. Marine Viruses: Truth or Dare. Annual Review of Marine Science **4:** Online First.

Breitbart, M. and others 2002. Genomic analysis of uncultured marine viral communities. Proceedings of the National Academy of Sciences of the United States of America **99:** 14250.

Buchan, A., J. M. Gonzalez, and M. A. Moran. 2005a. Overview of the marine *Roseobacter* lineage. Appl Environ Microbiol **71:** 5665-5677.

Buchan, A., J. M. González, and M. A. Moran. 2005b. Overview of the marine *Roseobacter* lineage. Appl Environ Microb **71:** 5665-5677.

Campanella, J. J., L. Bitincka, and J. Smalley. 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC bioinformatics **4:** 29.

Ceyssens, P. J. and others 2010. Molecular and physiological analysis of three *Pseudomonas aeruginosa* phages belonging to the "N4-like viruses". Virology **405:** 26-30.

---. 2008. The intron-containing genome of the lytic *Pseudomonas* phage LUZ24 resembles the temperate phage PaP3. Virology **377:** 233-238.

Chan, J. 2010. Bacteriophages of marine Roseobacter. University of Warwick.

Choi, K. H., J. McPartland, I. Kaganman, V. D. Bowman, L.B. Rothman-Denes, M. G. Rossman. 2008. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J Mol Biol **378:** 726-736

Clark, K., and R. Gorley. 2006. PRIMER v6: User Manual/Tutorial. PRIMER-E.

Cole, J. and others 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic acids research **37:** D141.

Comeau, A. M., A. M. Chan, and C. A. Suttle. 2006. Genetic richness of vibriophages isolated in a coastal environment. Environmental Microbiology **8:** 1164-1176.

Culley, A. I., A. S. Lang, and C. A. Suttle. 2006. Metagenomic analysis of coastal RNA virus communities. Science **312:** 1795.

Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna. 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Research **14:** 1394.

Edwards, R. A., and F. Rohwer. 2005. Viral metagenomics. Nat Rev Microbiol **3:** 504-510.

García Martin, H. and others 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotechnol **24:** 1263-1269.

Ghosh, D., K. Roy, K. E. Williamson, S. Srinivasiah, K. E. Wommack, and M. Radosevich. 2009. Acyl-Homoserine Lactones Can Induce Virus Production in Lysogenic Bacteria: an Alternative Paradigm for Prophage Induction. Appl Environ Microb **75:** 7142-7152.

Goldsmith, D. B. and others 2011. Pho Regulon Genes in Phage: Development of phoH as a Novel Signature Gene for Assessing Marine Phage Diversity. Appl. Environ. Microbiol.**:** AEM.05531-05511.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology **59:** 307.

Hendrix, R. W. 2003. Bacteriophage genomics. Curr Opin Microbiol **6:** 506-511.

Henn, M., M. B. Sullivan, N. Strange-Thomann, M. S. Osburne, A. M. Berlin, and L. Kelly. 2010. Analysis of highthroughput sequencing and annotation strategies for phage genomes. PLoS One **5:** e9083.

Hess, W. R. 2008. Comparative genomics of marine cyanobacteria and their phages, p. 89-116. *In* A. Herrero and E. Flores [eds.], The cyanobacteria: molecular biology, genomics, and evolution. Caister Academic Press.

Holmfeldt, K., M. Middelboe, O. Nybroe, and L. Riemann. 2007. Large Variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobactetium* hosts. Appl Environ Microb **73:** 6730-6739.

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

Hyman, P., and S. T. Abedon. 2010. Chapter 7 - Bacteriophage Host Range and Bacterial Resistance, p. 217-248. *In* I. L. Allen, S. Sima and M. G. Geoffrey [eds.], Advances in Applied Microbiology. Academic Press.

Jensen, E. C. and others 1998. Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. Appl Environ Microb **64:** 575-580.

Jürgens, K., and R. Massana. 2008. Protistan Grazing on Marine Bacterioplankton, p. 383-441. *In* D. L. Kirchman [ed.], Microbial Ecology of the Oceans (Second Edition).

Kazakov, A. E., O. Vassieva, M. S. Gelfand, A. Osterman, and R. Overbeek. 2003. Bioinformatics classification and functional analysis of PhoH homologs. In Silico Biol **3:** 3-15.

Kazmierczak, K., and L. Rothman-Denes. 2006. Bacteriophage N4. *In* R. Calendar [ed.], The Bacteriophages. Oxford University Press.

Kellogg, C. A., J. B. Rose, S. C. Jiang, J. M. Thurmond, and J. H. Paul. 1995. Genetic Diversity of Related Vibriophages Isolated from Marine Environments around Florida and Hawaii, USA. Mar Ecol-Prog Ser **120:** 89-98.

Kemena, C., and C. Notredame. 2009. Upcoming challenges for multiple sequence alignment methods in the high-throughput era. Bioinformatics **25:** 2455.

Kourilsky, P. 1973. Lysogenization by bacteriophage lambda. I. Multiple infection and the lysogenic response. Mol Gen Genet **122:** 183-195.

---. 1974. Lysogenization by bacteriophage lambda. II. Identification of genes involved in the multiplicity dependent processes. Biochimie **56:** 1511-1516.

Larkin, M. and others 2007. ClustalW and ClustalX version 2. Bioinformatics **23:** 2947-2948.

Lawrence, J. E., and G. F. Steward. 2010. Purification of viruses by centrifugation. Manual of Aquatic Viral Ecology. ASLO**:** 166-181.

Letunic, I., and P. Bork. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics (Oxford, England) **23:** 127.

Lohr, J. E., F. Chen, and R. T. Hill. 2005. Genomic analysis of bacteriophage {Phi} JL001: insights into its interaction with a sponge-associated alpha-proteobacterium. Appl Environ Microb **71:** 1598.

Ludwig, W. and others 2004. ARB: a software environment for sequence data. Nucleic acids research **32:** 1363-1371.

Mahadevan, P., J. F. King, and D. Seto. 2009. CGUG: in silico proteome and genome parsing tool for the determination of" core" and unique genes in the analysis of genomes up to ca. 1.9 Mb. BMC Research Notes **2:** 168.

Männistö, R., H. Kivelä, L. Paulin, D. Bamford, and J. Bamford. 1999. The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. Virology **262:** 355.

Matsen, F. A., R. B. Kodner, and E. Armbrust. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC bioinformatics **11:** 538.

Meyer, F. and others 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics **9:** 386.

Moebus, K. 1983. Lytic and Inhibition Responses to Bacteriophages among Marine-Bacteria, with Special Reference to the Origin of Phage-Host Systems. Helgolander Meeresun **36:** 375-391.

Morita, R. Y. 1997. Bacteria in oligotrophic environments : starvation-survival lifestyle. Chapman & Hall.

Newton, R. J. and others 2010. Genome characteristics of a generalist marine bacterial lineage. Isme J **4:** 784-798.

Ohki, K., and Y. Fujita. 1996. Occurrence of a temperate cyanophage lysogenizing the marine cyanophyte Phormidium persicinum. J Phycol **32:** 365-370.

Paddison, P. and others 1998. The roles of the bacteriophage T4 r genes in lysis inhibition and fine-structure genetics: A new perspective. Genetics **148:** 1539-1550.

Paul, J. H. 2008. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? The ISME Journal **2:** 579-589.

Paul, J. H., S. C. Jiang, and C. A. Kellogg. 1998. Characterization of marine temperate phage-host systems isolated from Mamala Bay, Oahu, Hawaii. Appl Environ Microb **64:** 535-542.

Paul, J. H., and M. B. Sullivan. 2005. Marine phage genomics: what have we learned? Current opinion in biotechnology **16:** 299-307.

Pernthaler, J. 2005. Predation on prokaryotes in the water column and its ecological implications. Nat Rev Microbiol **3:** 537-546.

Posada, D. 2008. jModelTest: phylogenetic model averaging. Molecular biology and evolution **25:** 1253.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One **5**.

Raven, J. 1998. The Twelfth Tansley Lecture, Small is Beautiful: The Picophytoplankton. Functional Ecology**:** 503-513.

Robinson, C. 2008. Heterotrophic Bacterial Respiration, p. 299-334. *In* D. L. Kirchman [ed.], Microbial Ecology of the Oceans (Second Edition).

Rohwer, F. and others 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. Limnol Oceanogr **45:** 408-418.

Rossmann, M. G., V. V. Mesyanzhinov, F. Arisaka, and P. G. Leiman. 2004. The bacteriophage T4 DNA injection machine. Curr Opin Struct Biol **14:** 171-180.

Sambrook, J., and D. W. Russell. 2001. Molecular cloning: a laboratory manual, 3 ed. Cold Spring Harbor Laboratory Press.

Schattner, P., A. N. Brooks, and T. M. Lowe. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic acids research **33:** W686.

Shivu, M. M., B. C. Rajeeva, S. K. Girisha, I. Karunasagar, G. Krohne, and I. Karunasagar. 2007. Molecular characterization of Vibrio harveyi bacteriophages isolated from aquaculture environments along the coast of India. Environmental Microbiology **9:** 322-331.

Skennerton, C. T. and others 2011. Phage Encoded H-NS: A Potential Achilles Heel in the Bacterial Defence System. PLoS One **6**.

Sullivan, M. B. and others 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4 like myoviruses from diverse hosts and environments. Environmental Microbiology.

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

Summer, E. J. and others 2010. Genomic and Biological Analysis of Phage Xfas53 and Related Prophages of *Xylella fastidiosa*. J Bacteriol **192:** 179-190.

Sun, S. and others 2011. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. Nucleic acids research **39:** D546.

Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr **45:** 1320-1328.

Versalovic, J., M. Schneider, F. De Bruijn, and J. R. Lupski. 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods in Molecular and Cellular Biology **5:** 25-40.

Wang, K., and F. Chen. 2008. Prevalence of highly host-specific cyanophages in the estuarine environment. Environmental Microbiology **10:** 300-312.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microb **73:** 5261.

Weinbauer, M. G., I. Brettar, and M. G. Hofle. 2003. Lysogeny and virus-induced mortality of bacterioplankton in surface, deep, and anoxic marine waters. Limnol Oceanogr **48:** 1457-1465.

Wichels, A., S. S. Biel, H. R. Gelderblom, T. Brinkhoff, G. Muyzer, and C. Schutt. 1998. Bacteriophage diversity in the North Sea. Appl Environ Microb **64:** 4128-4133.

Wichels, A., G. Gerdts, and C. Schutt. 2002. *Pseudoalteromonas* spp. phages, a significant group of marine bacteriophages in the North Sea. Aquat Microb Ecol **27:** 233-239.

Xu, Z., and B. Hao. 2009. CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. Nucleic acids research **37:** W174.

Zhao, Y. and others 2009a. Gene transfer agent (GTA) genes reveal diverse and dynamic Roseobacter and Rhodobacter populations in the Chesapeake Bay. The ISME Journal **3:** 364.

Zhao, Y., K. Wang, N. Jiao, and F. Chen. 2009b. Genome sequences of two novel phages infecting marine roseobacters. Environmental Microbiology **11:** 2055-2064.

Appendix

Tables

Table 5.1. Common proteins found amongst lytic roseophage

	Podoviridae, N4-like				Siphoviridae		Podoviridae, temperate-like		Podoviridae, T7-like
	2047B	EE36P1	DSS3P2	RvP1	2051A	RDJL	2047A	2047C	SIO1
DNA Polymerase	Type I	Type I	Type I	Type I	Type II	Type I	-	-	Type I
RNA Polymerase[2]	T3/T7-like	T3/T7-like	T3/T7-like	T3/T7-like	-	-	-	-	-
DNA Helicase	UrvD, RecD	UrvD, RecD	UrvD, RecD	UrvD, RecD	DEXDc	HELICc	-	-	GP4d, DnaB[1]
DNA Primase	-	-	-	-	Primpol, PriCT 2	Primpol, PriCT 2	-	-	TOPPRIM, DnaG[1]
Ribonuleotide/side reductase	Class II	Class II	Class II	Class II	Class II	Class II	-	-	Class II
dCTP/dCMP deaminase	dCTP	dCMP	dCMP	dCMP	dCMP	dCMP	-	-	-
Thymidylate synthase	ThyX	Thy1	Thy1	Thy1	ThyX	ThyA	-	-	ThyX
DNA methylase	-	-	-	-	N-4/N-6	-	-	-	-
Integrase	-	-	-	-	-	-	P4-like integrase, tyrosine type	P4-like integrase, tyrosine type	-
Terminase	Type 6[3]	Type 6	Type 6	Type 6	Type GpA, λ-like	Type 3, P22-like	Type 3, P22-like	Type 3, P22-like	Type 6
Thioredoxin	TRX-like	TRX-like	TRX-like	TRX-like	-	-	-	-	GRX-like

Protein or enzyme type notation was determined by BLAST analysis

No recognizable homo- or analog (-)

Phage in red were isolated in this study

[1]Primase and helicase domains are found on the same protein

[2]N4-like bacteriophage also contain a virion-encapsulated RNA polymerase

[3]Type 6 terminases are unclassified in terms of head packaging strategy

Supplemental Tables

Table 5.2. Presence of the core-extended genes in N4-like bacteriophage

Gene/ORF	Potential function	N4	2047B	EE36P1	DSS3P2	RvP1	LUZ7	LIT1	EcP1
RIIA	Lysis inhibition	+	+	+	+	+	+	+	
RIIB	Lysis inhibition	+	+	+	+	+	+	+	
dCD	dCMP/dCTP deaminase	+	+	+	+	+	+	+	
N4 gp14	unknown (conserved hypothetical)	+	+	+	+	+		+	
N4 gp2	DNA Binding protein (RNAP)	+	+	+	+	+			
N4 gp36	NT pyrophosphohydrolase	+	+	+	+	+			
N4 gp22	HNH endonuclease	+		+	+	+		+	
EE36P1 gp55*	unknown (hypothetical)		+	+	+	+			+
gp46	unknown (hypothetical)	+					+	+	
N4 gp51	structural protein	+	+	+	+	+			
LUZ7 gp74	structural protein						+	+	+
Thy	Thymidylate synthase	+	+	+	+	+			+

Table 5.3. Presence of ϕ 2051 signature genes in other related phage and prophage

Single plus symbol indicates protein homology as determined with CoreGenes. Double plus symbols indicates nucleotide similarity as determined from Mauve genome alignments.

	S. meliloti prophage	**ϕRDJL**	**ϕSAI**	**ϕ73**	**ϕBcepGomr**	**ϕBcepNazgul**
DNA Polymerase	++			+	+	
Terminase	++		+			+
Capsid	++	+	++			+
prohead protease	++		++			+
Portal	++	+	++			+
Helicase	++		+	++		+
Primase	++	+	+	+		+
Tail tape measure protein	++	+	+	+	+	+
Tail assembly protein	+	+	+	+		+
DNA Ligase			+			
Exonuclease	+	+				
ribonucleotide reductase	++	++	+			
deoxycytidylate deaminase	++	+		+		

Table 5.4. Presence of ϕ2047A/C signature genes in other related phage and prophage

Single plus symbol indicates protein homology as determined with CoreGenes. Double plus symbols indicates nucleotide similarity as determined from Mauve genome alignments

	ϕEPV2	**S. medicae prophage**	**ϕBcepC6B**	**ϕMx8**	**ϕF116**	**ϕBcep22**
Terminase	++	++	+	+	+	+
Capsid	++	++	+			
Integrase					+	
Structural protein gp10	++	++			+	
Tail fiber Mx8p54-like	++	++		+		
Tail fiber Mx8p52-like	++	++		+		
Tail fiber Mx8p51-like	++	++		+		
Structural Mx8p57-like	++			+	+	+
XRE-like transcriptional regulator			+			
Unknown function gp43					+	
Unknown function gp48					+	

Table 5.5. Environmental samples that contained the putative injection protein Mx8p52-like

Data is from tblastn searches and only returned subject reads with E-values >10^{-10} were considered. Only samples with >25 hits/mega-read are shown. The red superscript 'v' indicates the samples were from a viral metagenome (<0.2 µm size fraction). The column 'Max E-value' is the maximum E-value returned amongst all returned reads.

Habitat	Sample	Hits/mega-read	Max E-value
Antarctica	Ace Lake	52	7.07E-45
	Organic Lake	32	7.27E-39
Arctic/Sub-arctic	Raunefjorden, Norway [v]	132	6.73E-75
	North Sea [v]	68	1.06E-30
	Subarctic Pacific [v]	26	7.31E-21
Atlantic	Gulf of Maine [v]	51	8.37E-17
	Equatorial deep water [v]	110	2.53E-24
	Chesapeake Bay [v]	102	1.24E-33
	BATS, deep water	200	4.39E-26
Pacific	Eel River methane seep [v]	60	1.50E-32
	Hydrocarbon seep field [v]	41	1.10E-27
	Southern California Bight [v]	25	6.23E-29
Freshwater	Groundwater, Florida [v]	54	1.43E-19
	Groundwater, Rifle Colorado [v]	154	2.45E-29
Other	Yellowstone hot springs	29	1.10E-37
	Minnesota farm soil	43	6.79E-42

Figures

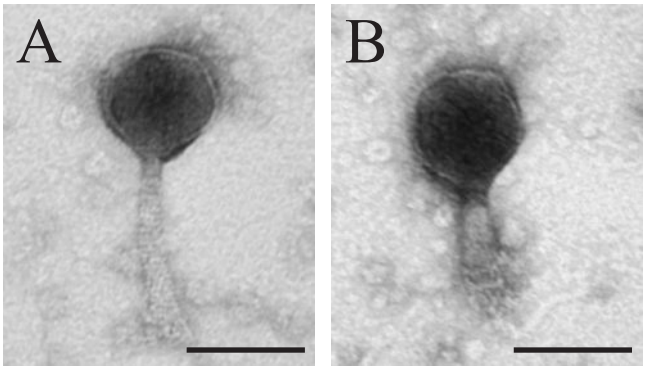


Figure 5.1. Transmission electron micrographs of roseophage ϕSIO67-Myo isolated in this study

(A) tail extended (B) tail contracted. Scale bars are 100 nm.

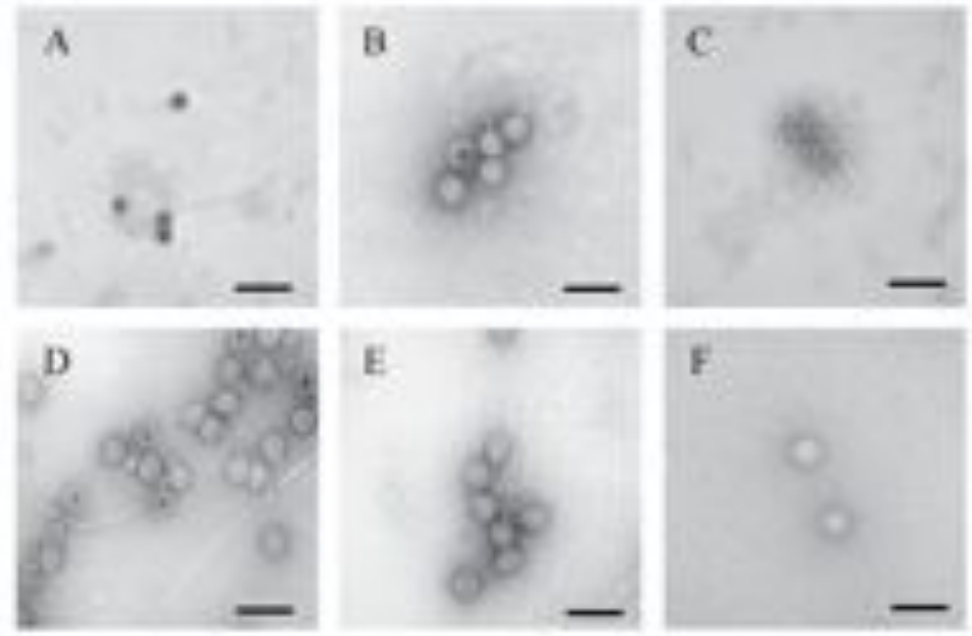


Figure 5.2. Transmission electron micrographs of roseophage isolated in this study

(A) ϕ1005 (B) ϕ1032 (C) ϕ2051 (D) ϕ2047A (E) ϕ2047B (F) ϕ2047C. Scale bars are 200 nm.

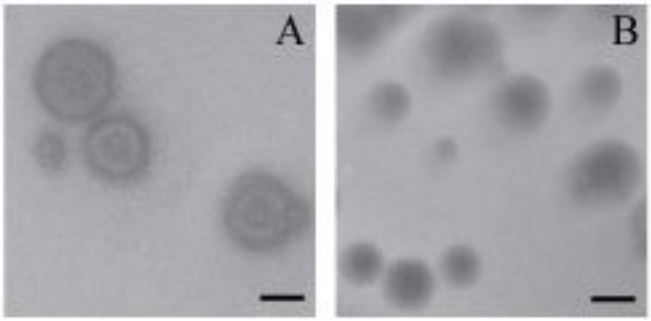


Figure 5.3. Digital photographs of plaques

Digital photographs of plaques belonging to ϕ2047A (A) and ϕ2047B (B). Scale bars are 1 mm. Photographs were converted to black and white and enhanced by modifying the brightness levels of the images' histograms.

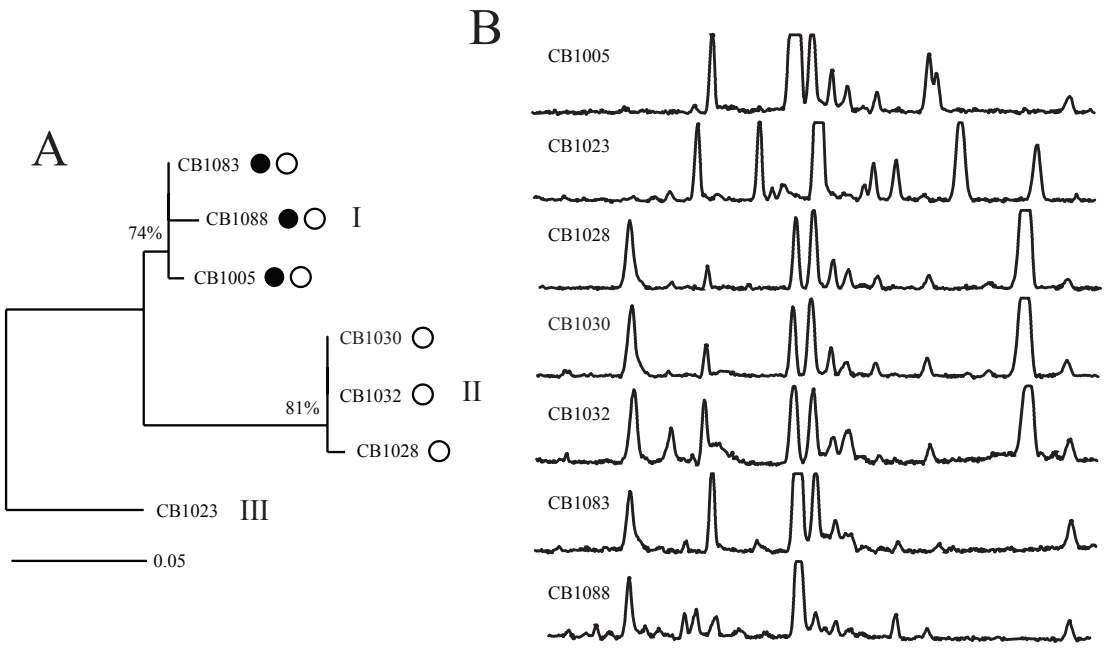


Figure 5.4. Strain differentiation of clade 1005

(A) Maximum likelihood tree inferred from the 16S-23S inter transcribed spacer region. Bootstrap values at nodes are from 100 replications. Closed circles next to strain names indicate infection by phage ϕ1005 and the open circles represent infection by ϕ1032 (B) BOX-PCR fingerprint histograms of the strains.

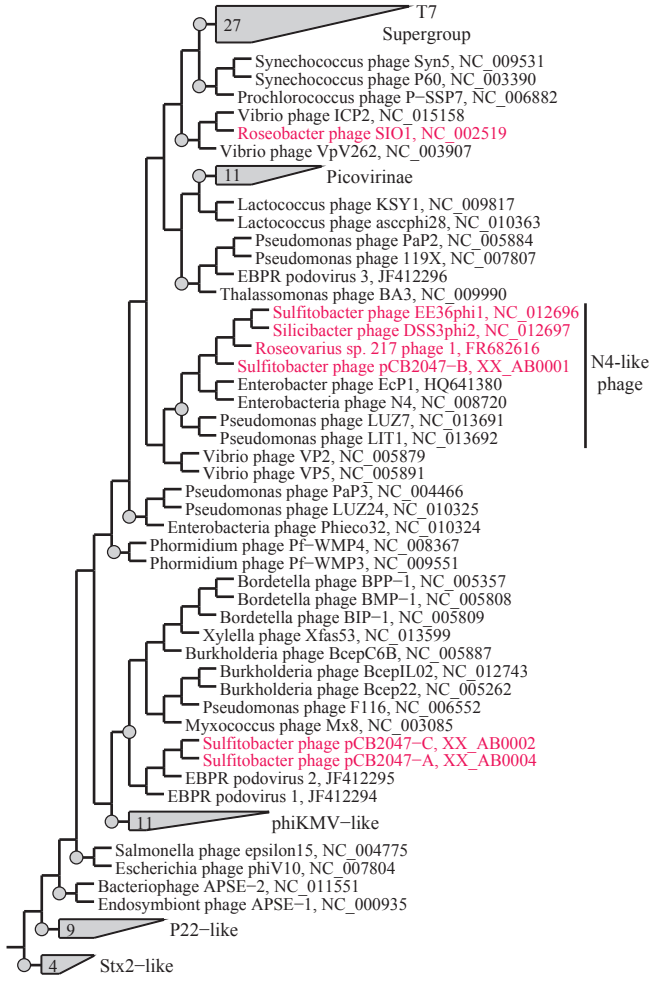
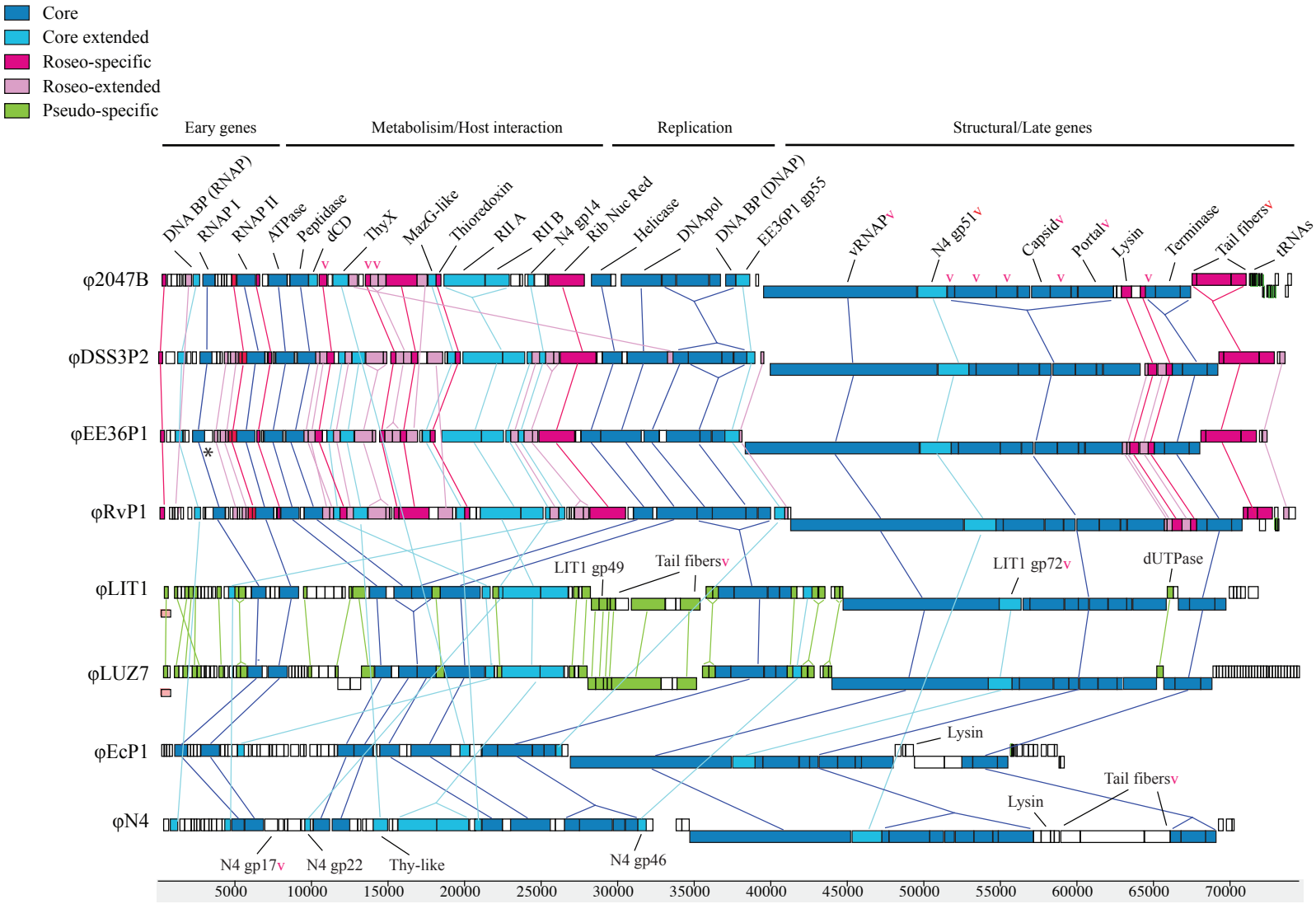


Figure 5.5. Neighbor joining tree of a whole proteome distance matrix of isolated Podoviridae

Neighbor joining tree of a whole proteome distance matrix of isolated Podoviridae created by CVTree. The distance matrix was created using a kmer value of 5. Circles represent nodes that are present in matrices created using a kmer of 4 and 5. Roseophage stains are indicated by a pink font color. Enterobacteria phage lambda (NC_001416) was used as an outgroup. Branch lengths have been normalized to aid visualization and do not represent true distances.

Figure 5.6. Genome comparison of N4-like bacteriophage

Open reading frames for each strain are shown as small colored blocks, which are positioned above or below a centerline depending on gene orientation (upper blocks are predicted to be transcribed left to right). ORF blocks are colored based on the total number of genomes containing the ORF as detailed in the text. Core (blue), core extended (light blue), Roseo-specific (pink) Roseo-extended (light pink), and Pseudo-specific (green) genes are noted. ORF showing no homology are white. The colored lines connecting the blocks indicate an ORF homology match, as determined by Core Genes. Lowercase letters 'v' (red font) indicate the peptide is known to be in the virion based on data from Chan (2010) and Choi et al. (2008). Scale bar is in base pairs.


Core
Core extended
Roseo-specific
Roseo-extended
Pseudo-specific
Eary genes
Metabolisim/Host interaction
Replication
Structural/Late genes
DNA BP (RNAP)
RNAP I
RNAP II
ATPase
Peptidase
dCD
ThyX
MazG-like
Thioredoxin
RII A
RII B
N4 gp14
Rib Nuc Red
Helicase
DNApol
DNA BP (DNAP)
EE36P1 gp55
vRNAPv
N4 gp51v
Capsidv
Portalv
Lysin
Terminase
Tail fibersv
tRNAs
φ2047B
φDSS3P2
φEE36P1
φRvP1
φLIT1
φLUZ7
φEcP1
φN4
LIT1 gp49
Tail fibersv
LIT1 gp72v
dUTPase
Lysin
Tail fibersv
Lysin
N4 gp17v
N4 gp22
Thy-like
N4 gp46
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
55000
60000
65000
70000

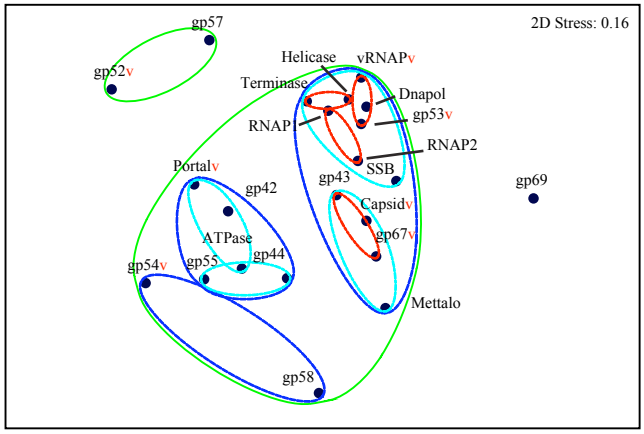


Figure 5.7. Second-stage MDS ordination of BLOSUM62 similarity matrices derived from N4-like core genes

Contour lines indicate the degree of correlation between similarity matrices; red (0.9), light blue (0.8), blue (0.7), and green (0.6). Protein gene product number (e.g. gp 43) follows naming convention of ϕN4.

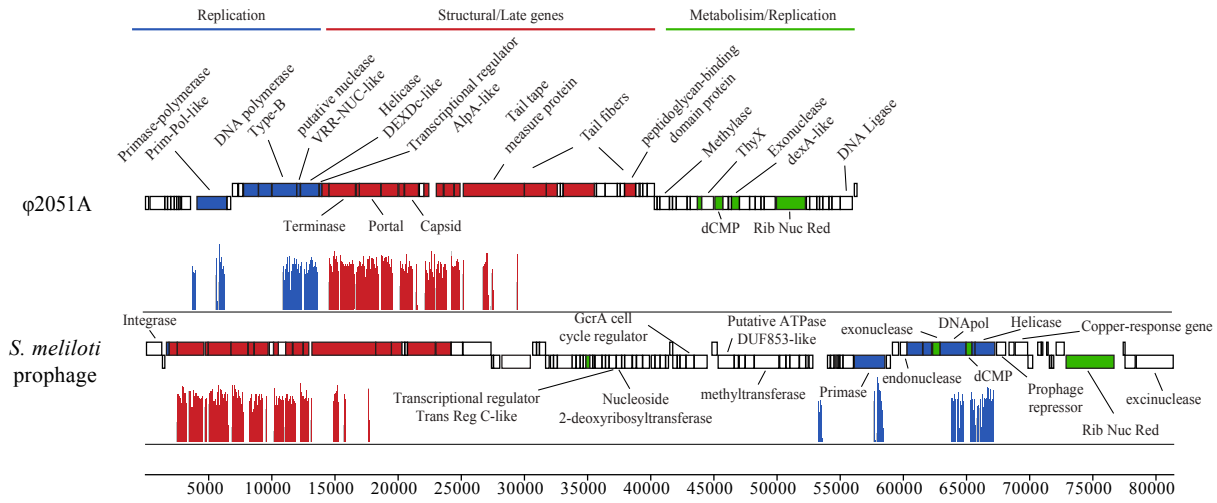


Figure 5.8. Genome comparison of ϕ2051A and the *S. meliloti* prophage

The prophage is found from locus tag Sinme_1353 to Sinme_1452 in the genome NC_015590. The figure is a modification of the alignment view from the program Mauve. Open reading frames for each strain are shown as small blocks, which are positioned above or below a centerline depending on gene orientation (upper blocks are predicted to be transcribed left to right). Similarity profiles of each genome sequence are shown below the blocks and the height of these profiles corresponds to the average level of conservation across the genomes. The colored blocks indicate an ORF homology match, as determined by Core Genes. ORF showing no protein homology are white. Scale bar is in base pairs.

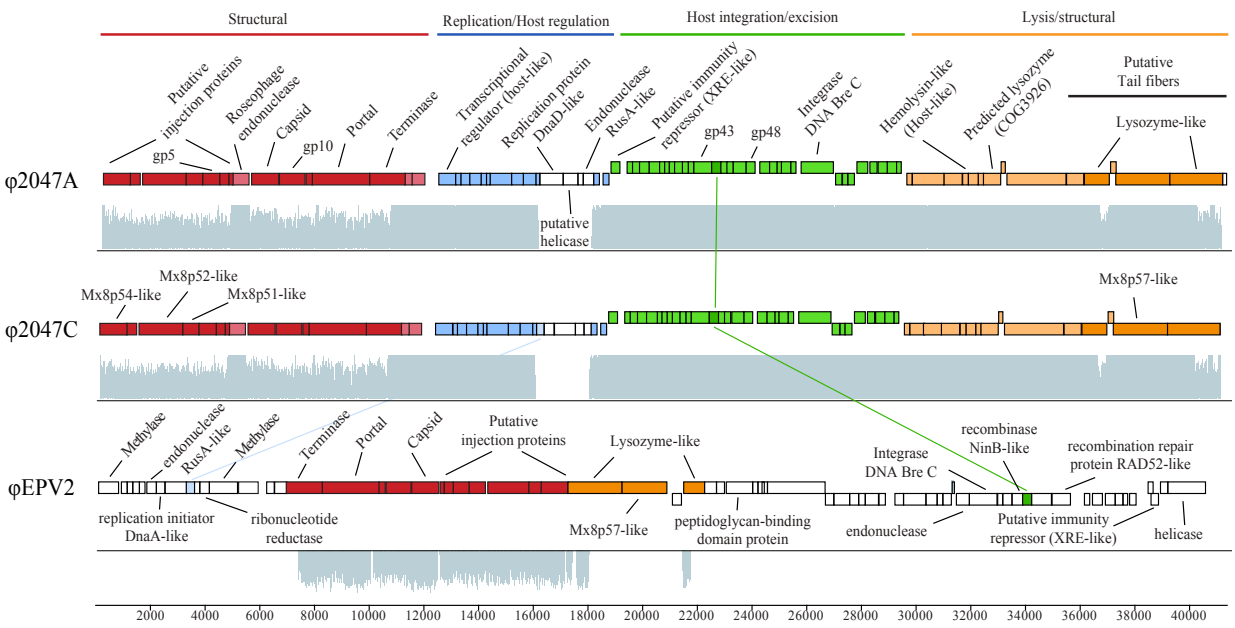


Figure 5.9. Genome comparison of ϕ2047A, ϕ2047C and ϕEPV2

The figure is a modification of the alignment view from the program Mauve. Open reading frames for each strain are shown as small blocks, which are positioned above or below a centerline depending on gene orientation (upper blocks are predicted to be transcribed left to right). Similarity profiles of each genome sequence are shown below the blocks and the height of these profiles corresponds to the average level of conservation across the genomes. The colored blocks indicate an ORF homology match, as determined by Core Genes with darker shades indicating homology across all three phage and lighter shades indicating homology between just two phage. ORF showing no homology are white. Scale bar is in base pairs.

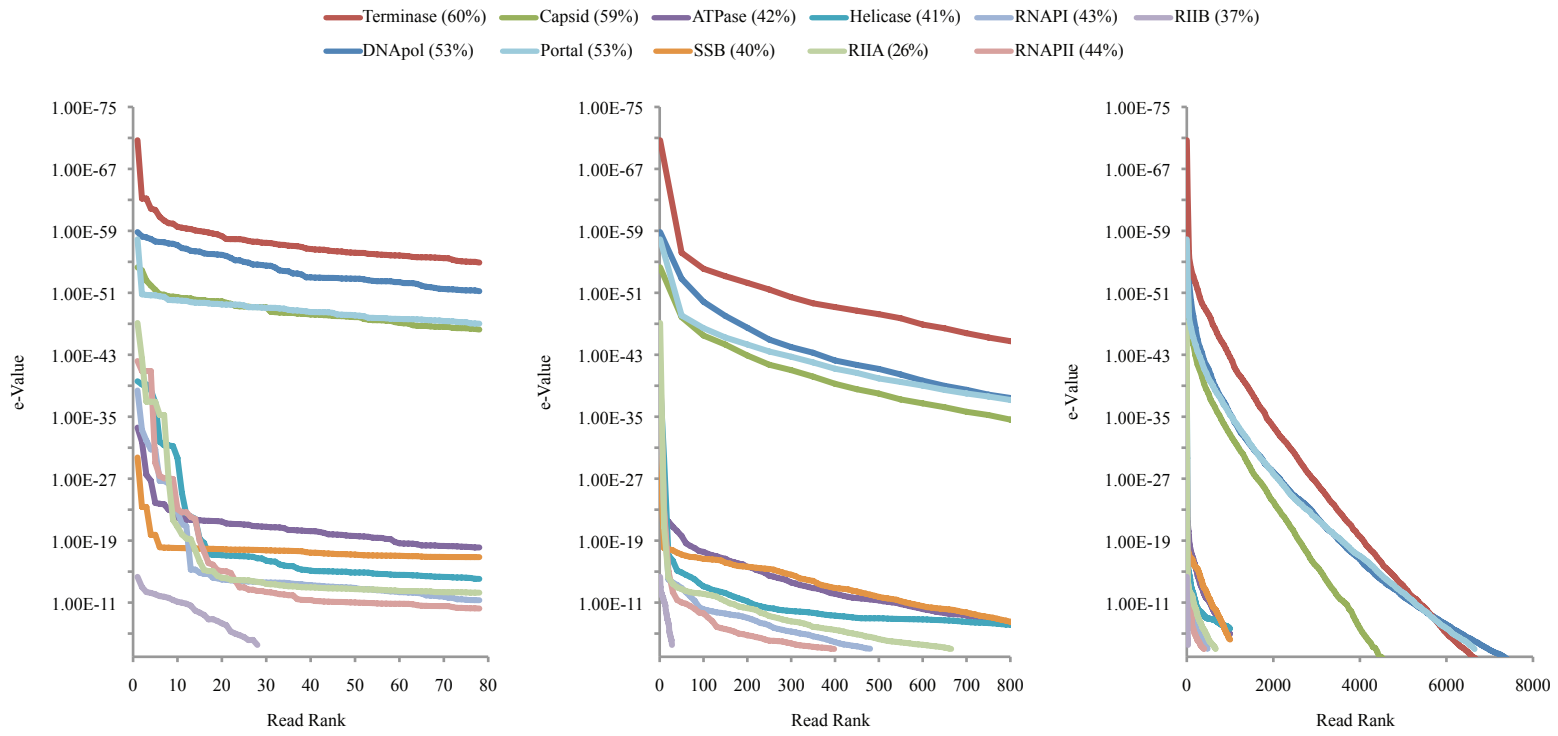
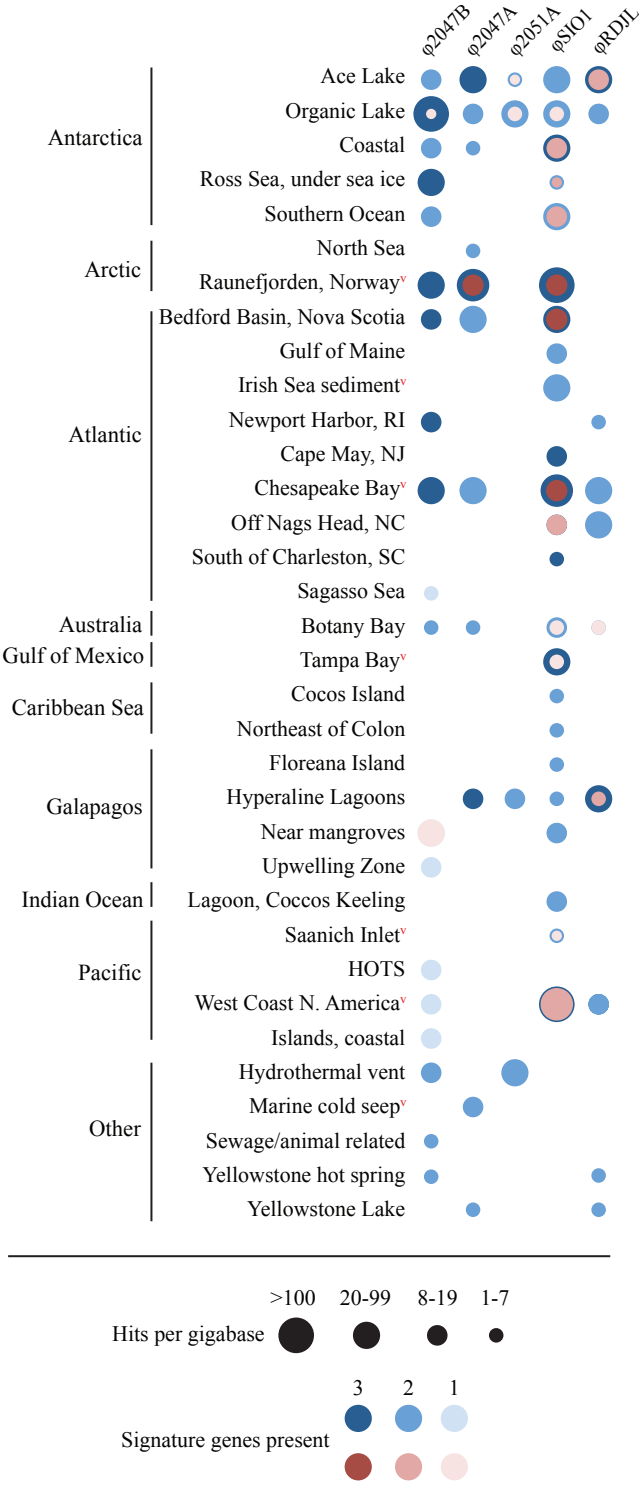


Figure 5.10. Rank similarity curves of select N4 bacteriophage core genes to environmental sequences

The y-axis is the BLAST e-value of reads obtained from a tblastn search with the CAMERA reference database that contained all metagenomic reads obtained by 454 sequencing. The average amino similarity is shown after the core gene name (as displayed in Supp. Fig. 4) The data is plotted at three x-axis scales; (A) 0-80 reads, (B) 0-800 reads, and (C) 0-8000 reads.

Figure 5.11. Distribution of roseophage and related viruses in environmental samples

Data is from tblastn searches of environmental databases (see methods) using the roseophage signature genes capsid, terminase, and DNA polymerase, except for ϕ2047A, which does not contain a DNA polymerase, so a conserved putative injection protein was used instead (Mx8p52-like). Only returned subject reads with E-values $>10^{-25}$ and lengths >100 amino acids were considered. Sizes of bubbles represent the abundance of hits returned per gigabase; >100, 20-99, 8-19, or 1-7. Color of bubbles indicates the presence of similar phage types (blue, i.e N4-like) and the amount of those reads that are roseophage (red, i.e. N4 roseophage), as determined by phylogenetic tree building. The shades of the two colors represent the number of signature genes detected in an individual sample, i.e. darkest shade indicates all three signature genes were found. Except for the N4-like phage, only environments with two or greater signature genes are shown for each phage (see text for details). The red superscript 'v' indicates that all or part of the samples from that environment were from a viral metagenome (<0.2 μm size fraction). As an example, >100 ϕSI01 related reads were found from samples off the west coast of North America, all three signature genes were present, almost all of the reads grouped strongly with ϕSI01, two signature roseophage genes were present, and some or all of the samples were from a viral metagenome.


φ2047B
φ2047A
φ2051A
φSIO1
φRDJL
Antarctica
Ace Lake
Organic Lake
Coastal
Ross Sea, under sea ice
Southern Ocean
Arctic
North Sea
Raunefjorden, Norwayv
Atlantic
Bedford Basin, Nova Scotia
Gulf of Maine
Irish Sea sediment v
Newport Harbor, RI
Cape May, NJ
Chesapeake Bayv
Off Nags Head, NC
South of Charleston, SC
Sagasso Sea
Australia
Botany Bay
Gulf of Mexico
Tampa Bayv
Caribbean Sea
Cocos Island
Northeast of Colon
Galapagos
Floreana Island
Hyperaline Lagoons
Near mangroves
Upwelling Zone
Indian Ocean
Lagoon, Coccos Keeling
Pacific
Saanich Inletv
HOTS
West Coast N. Americav
Islands, coastal
Other
Hydrothermal vent
Marine cold seepv
Sewage/animal related
Yellowstone hot spring
Yellowstone Lake
>100
20-99
8-19
1-7
Hits per gigabase
3
2
1
Signature genes present

Supplemental Figures

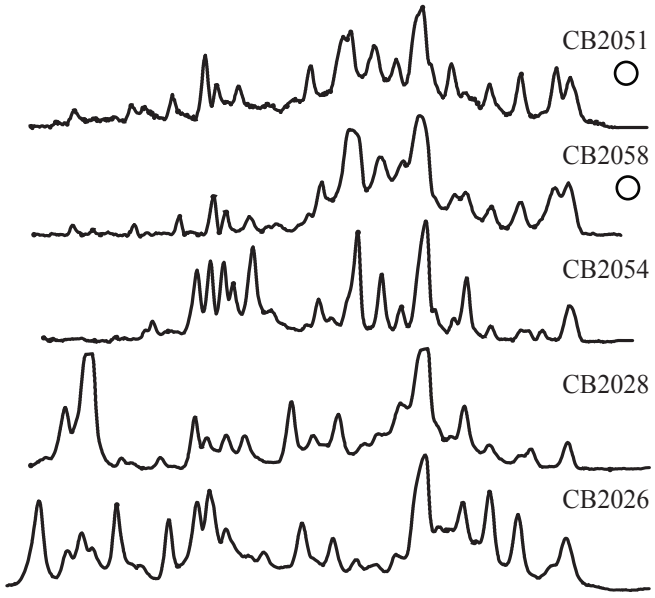


Figure 5.12. Strain differentiation of clade 2051

BOX-PCR fingerprint histograms of the strains. Open circles represent strains infected by ϕ2051.

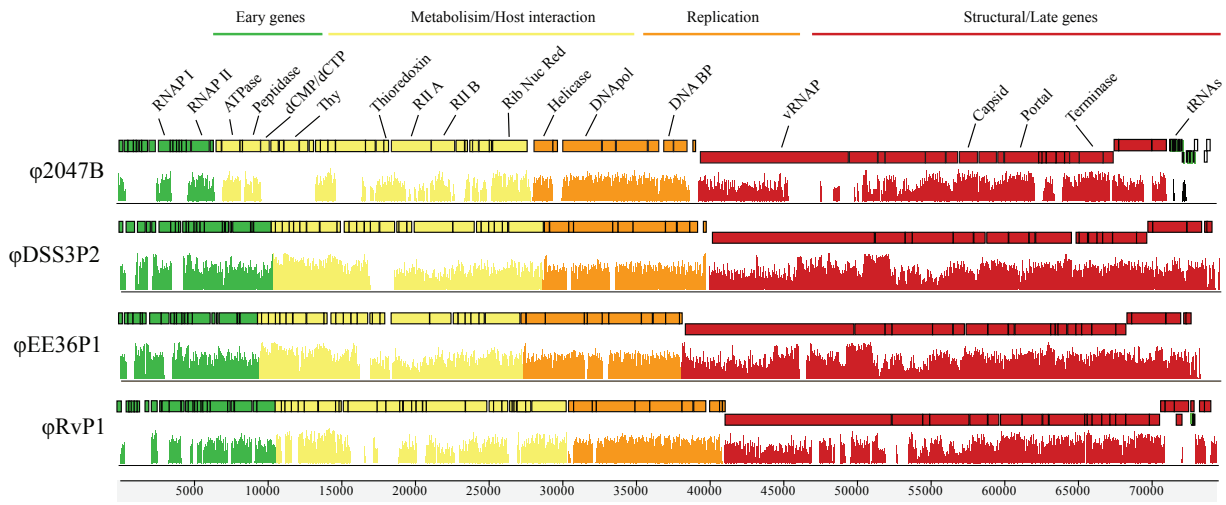


Figure 5.13. Genome-wide nucleotide similarity amongst N4-like roseophage

The figure is a modification of the alignment view from the program Mauve. Open reading frames for each strain are shown as small colored blocks, which are positioned above or below a centerline depending on gene orientation (upper blocks are predicted to be transcribed left to right). Similarity profiles of each genome sequence are shown below the colored blocks and the height of these profiles corresponds to the average level of conservation across the genomes.

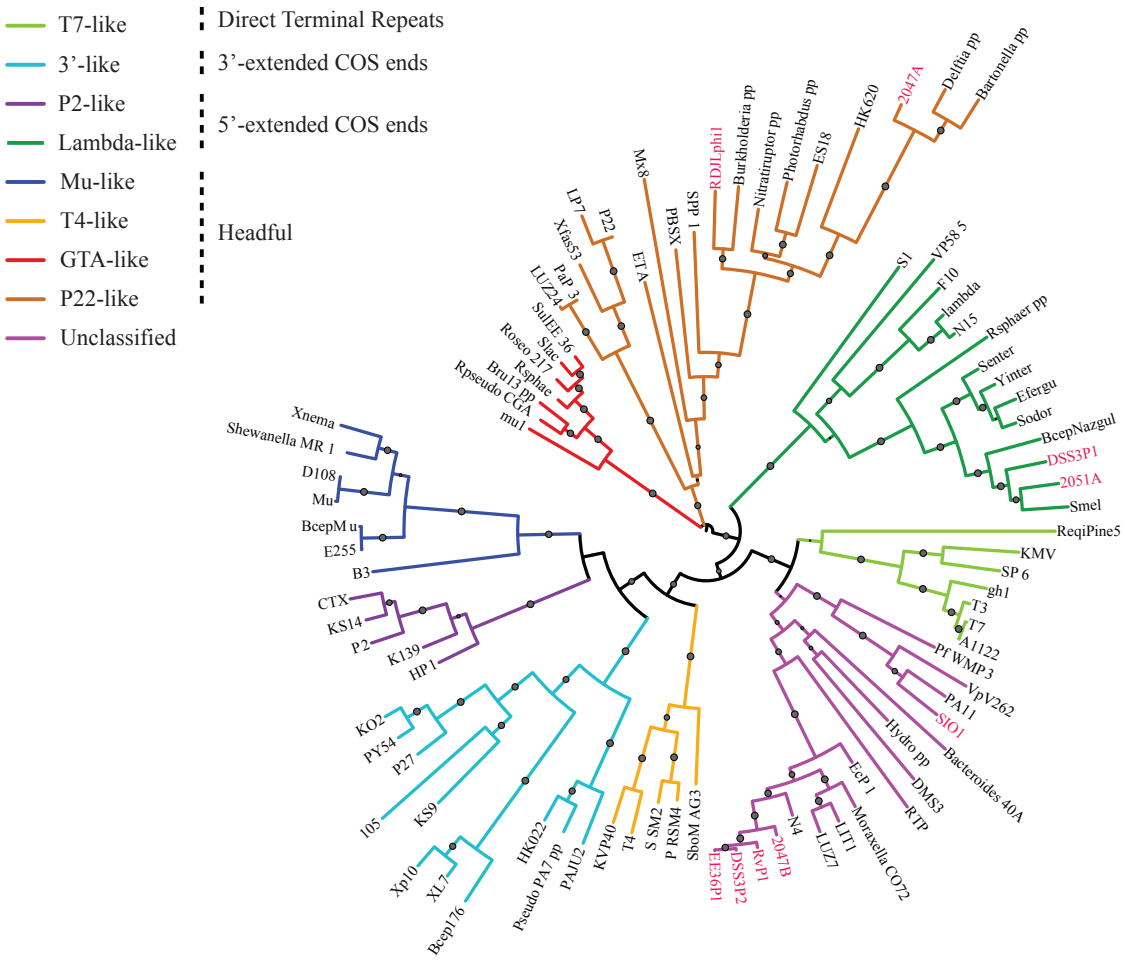


Figure 5.14. Maximum likelihood tree of bacteriophage terminase proteins

Tree shows select bacteriophage representing known DNA packaging strategies. Leaf identifiers depicted with a red colored font are roseophage. Solid circles at branch nodes indicate local support values calculated by FastTree that were >0.80. The tree is un-rooted.

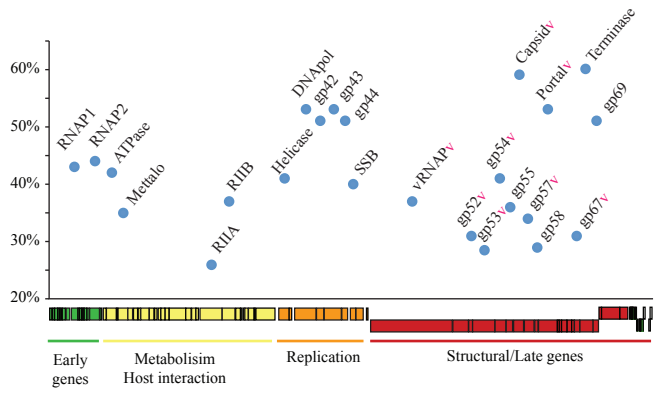


Figure 5.15. Average amino acid similarity amongst individual N4-like core genes

Average amino acid similarity (y-axis) amongst individual N4-like core genes plotted against genome position (x-axis) (ϕ2047B). Lowercase letters 'v' (red font) indicate the peptide is known to be in the virion. Protein gene product number (e.g. gp 43) follows naming convention of ϕN4.

Figure 5.16. MDS ordinations of BLOSUM62 similarity matrices derived from N4-like core genes

MDS ordinations of BLOSUM62 similarity matrices derived from N4-like core genes. Gene/protein abbreviation and average percent amino acid similarity is listed above each plot. Lowercase letters ‘v’ (red font) indicate the encoded peptide has previously been found in the virion.


RNAP1 - 43%
RNAP2 - 44%
ATPase - 42%
Mettalo - 35%
Helicase - 41%
DNApol - 53%
N4 gp42 - 51%
N4 gp43 - 53%
N4 gp44 - 51%
SSB - 40%
vRNAPv - 37%
N4 gp52v - 31%
N4 gp53v - 29%
N4 gp54v - 41%
N4 gp55 - 36%
Capsidv - 59%
N4 gp57v - 34%
N4 gp58 - 34%
Portalv - 53%
N4 gp67v - 31%
Terminase - 60%
N4 gp69 - 51%
Species
N4
EE36P1
RvP1
LUZ7
EcP1
DSS3P2
2047B
LIT1

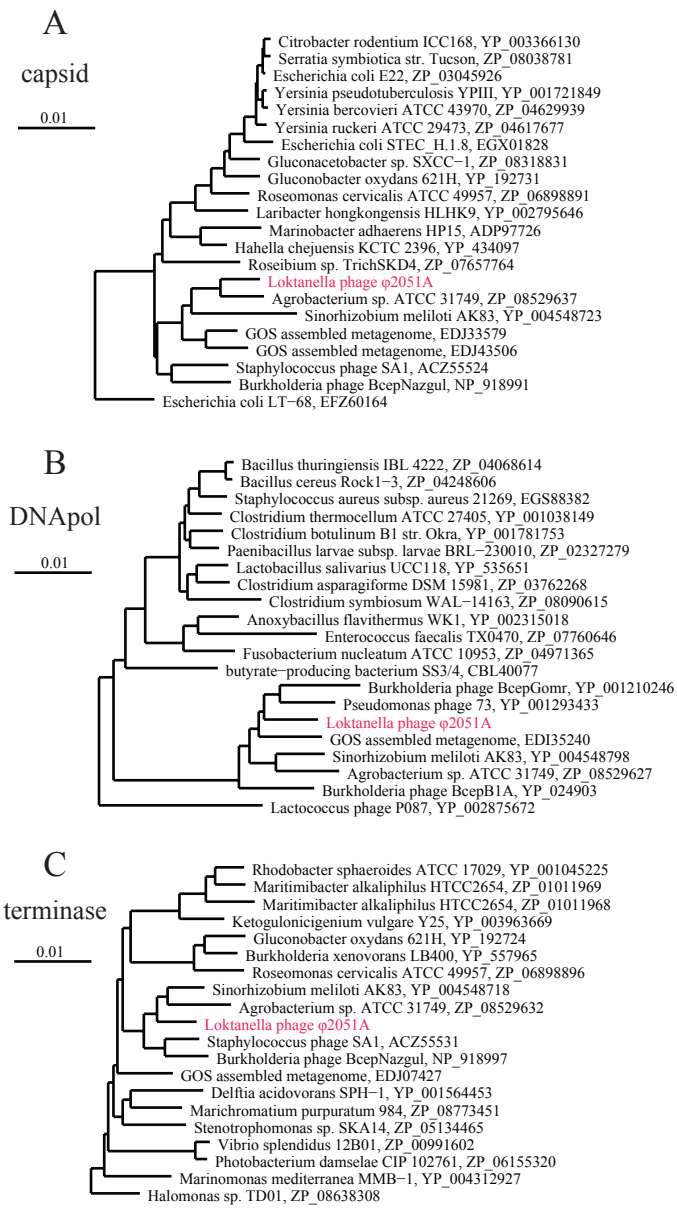


Figure 5.17. Maximum likelihood of ϕ2051 signature peptides

Maximum likelihood tree created with FastTree of the (A) capsid (B) DNA polymerase and (C) terminase peptides from ϕ2051 and its nearest relatives.

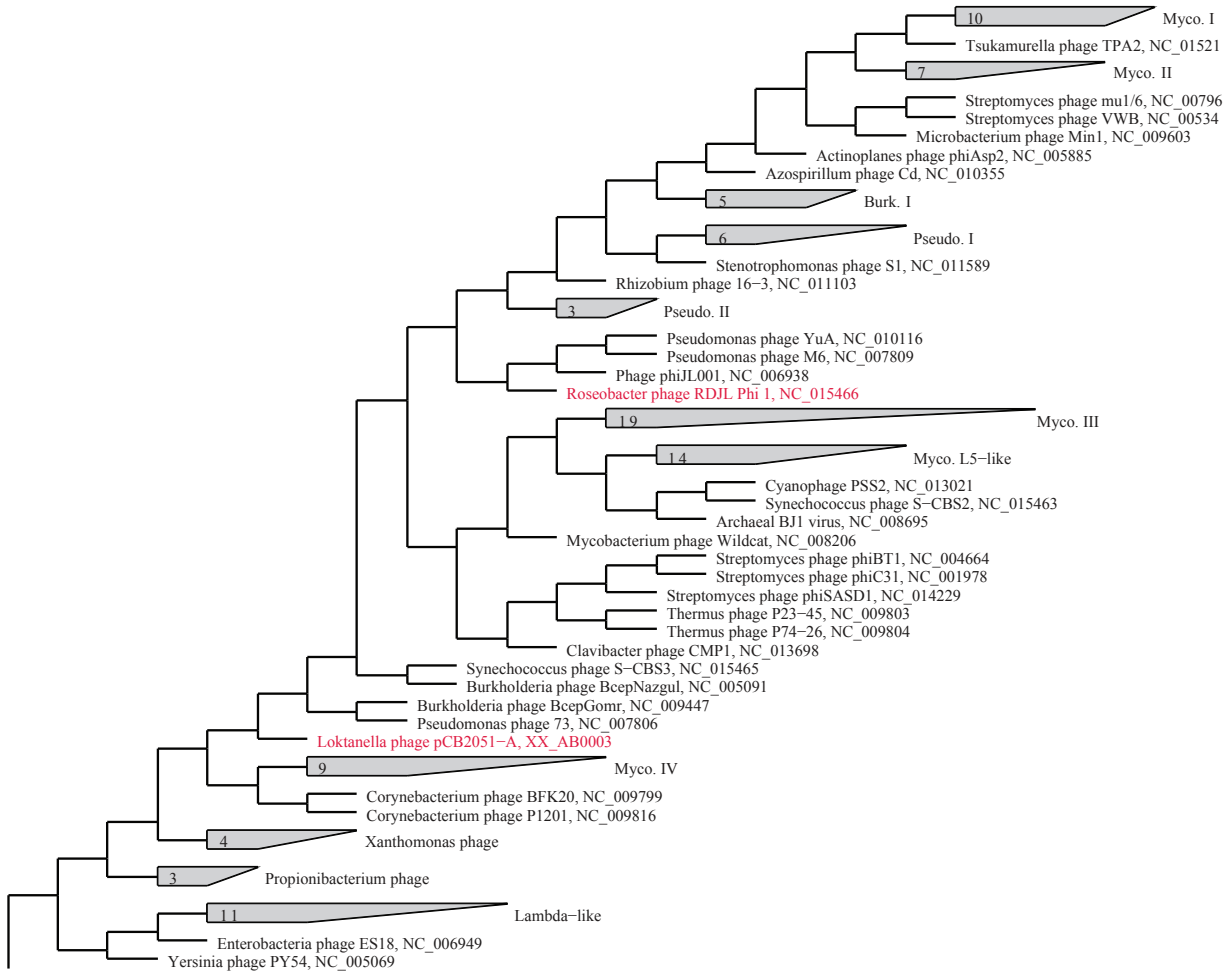


Figure 5.18. Neighbor joining tree of a whole proteome distance matrix of isolated Siphoviridae

Neighbor joining tree of a whole proteome distance matrix of isolated Siphoviridae created by CVTree. The distance matrix was created using a kmer value of 5. The tree presented is a portion of a larger tree containing all Siphoviridae. Roseophage stains are indicated by a pink font color. Enterobacteria phage T4 (NC_000866) was used as an outgroup. Branch lengths have been normalized to aid visualization and do not represent true distances.

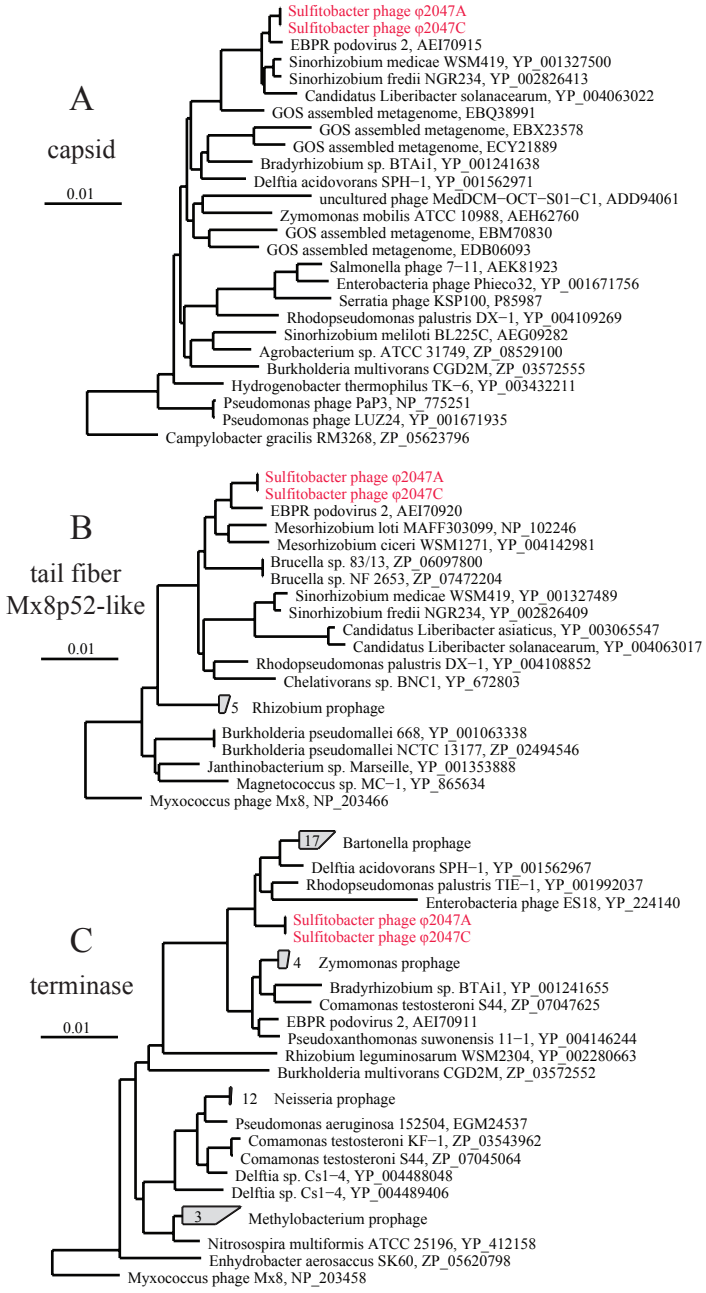


Figure 5.19. Maximum likelihood of ϕ2047A/C signature peptides

Maximum likelihood tree created with FastTree of the (A) capsid (B) DNA polymerase and (C) terminase peptides from ϕ2047A/C and their nearest relatives.

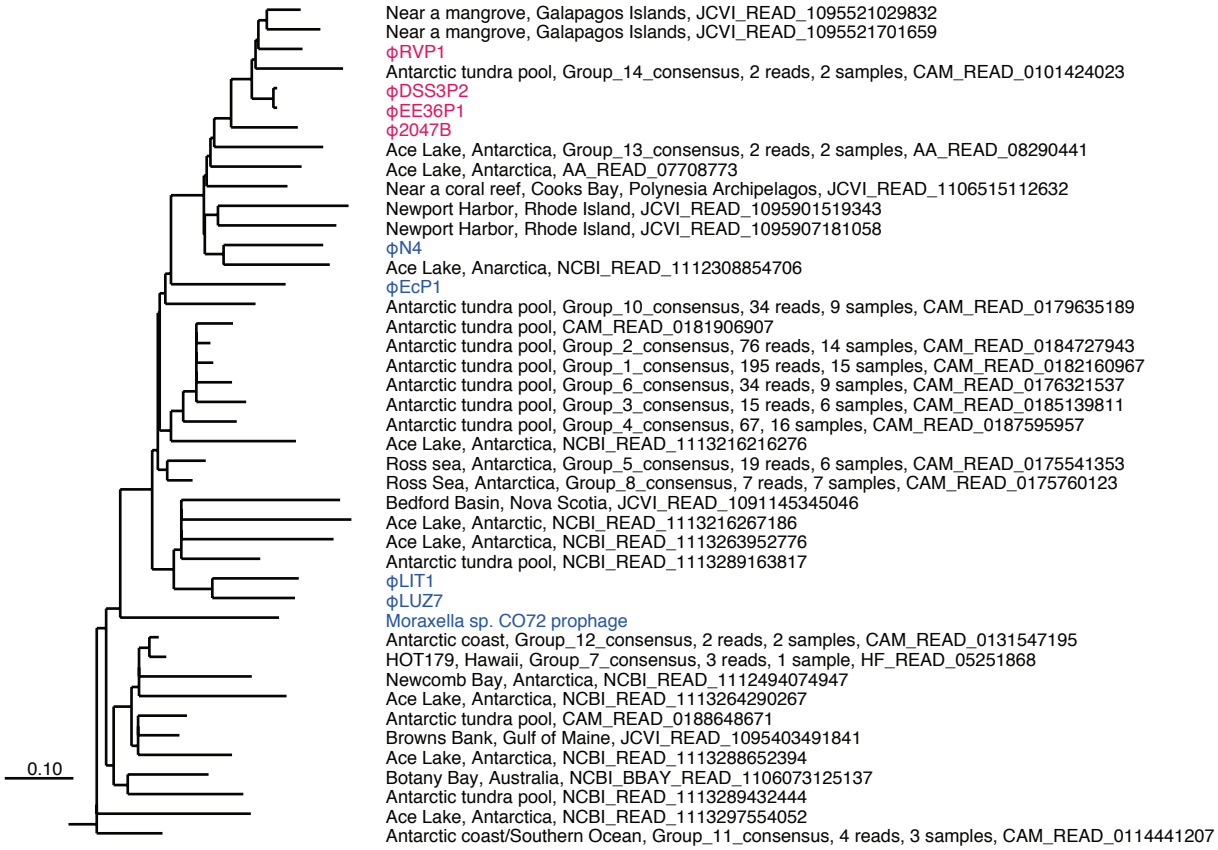


Figure 5.20. Phylogenetic tree of N4-like DNA polymerase sequences

The tree was constructed by inserting environmental sequences into a reference tree of full-length sequences. Information on individual reads is habitat, consensus group (if applicable), and read accession number (or representative accession number). Outgroup is *Thermus thermophilus* phage P74-26 (YP_001467981). Roseophage isolates are colored pink and other reference N4-like sequences are colored blue.

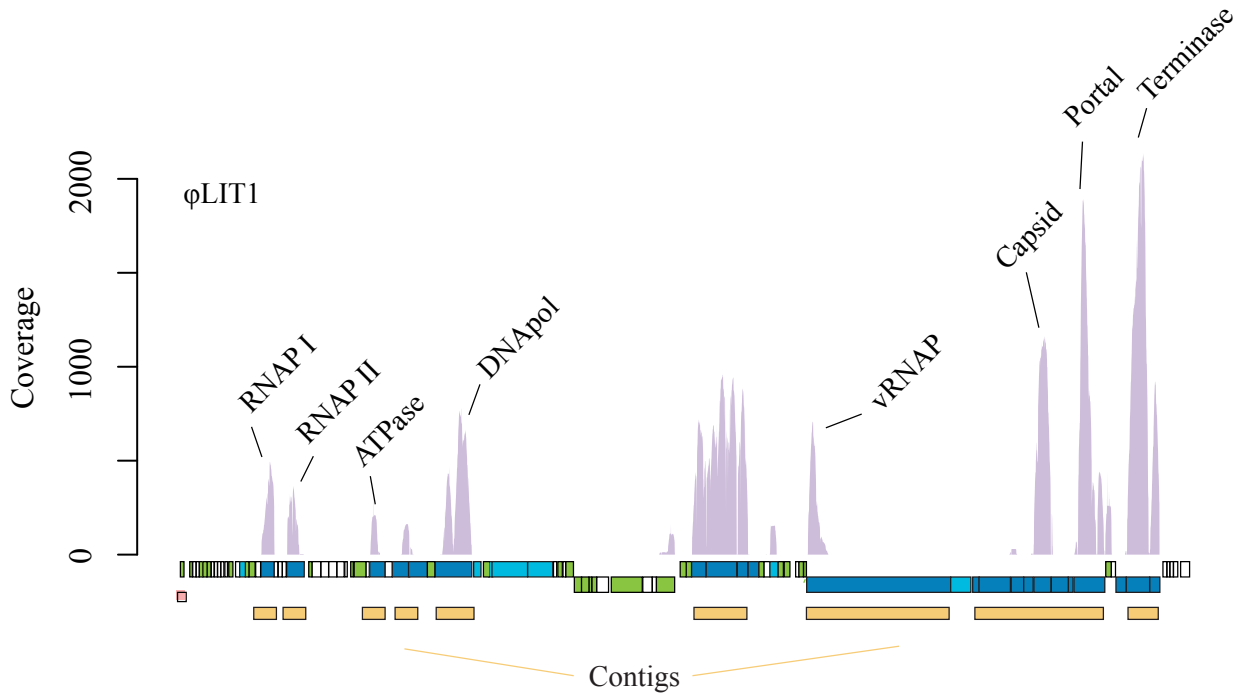


Figure 5.21. Recruitment plot of metagenomic reads from the Antarctic tundra pond samples to the genome of ϕLIT1

Recruitment plot of metagenomic reads (tblastn) from the Antarctic tundra pond samples to the genome of ϕLIT1. The y-axis is coverage and the x-axis is genome position overlaid with the gene map. The orange blocks represent the position of assembled contigs (>100 bp) from the same metagenomic samples.

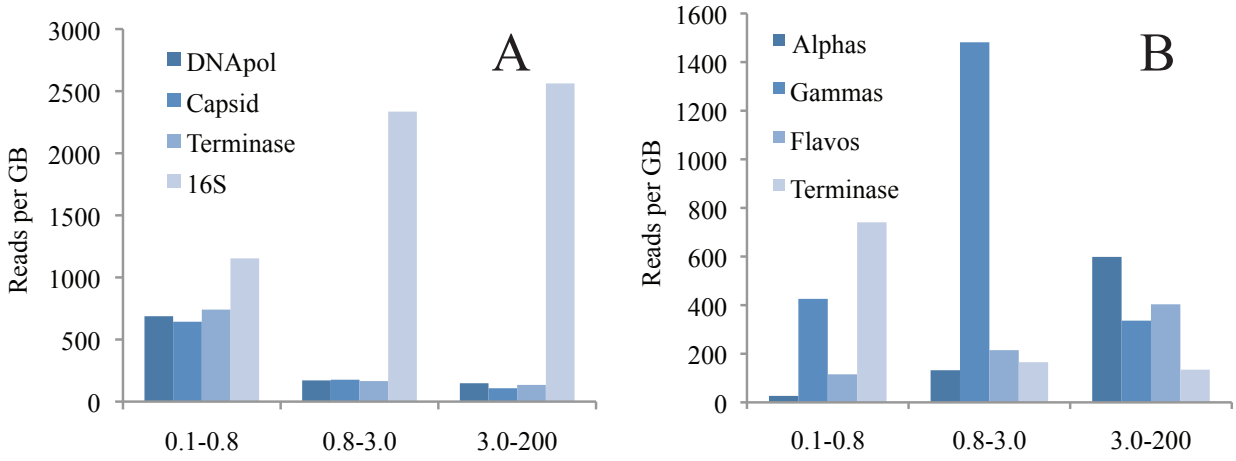


Figure 5.22. Comparison of the number of N4-like phage signature genes to the number of 16S rDNA genes from Antarctic tundra samples

Comparison of the number of N4-like phage signature genes to the number of 16S rDNA genes from Antarctic tundra samples representing the three metagenomic size fractions. (A) Three phage genes and all 16S rDNA reads. (B) The terminase gene and the 16S rDNA genes divided into taxonomic class.

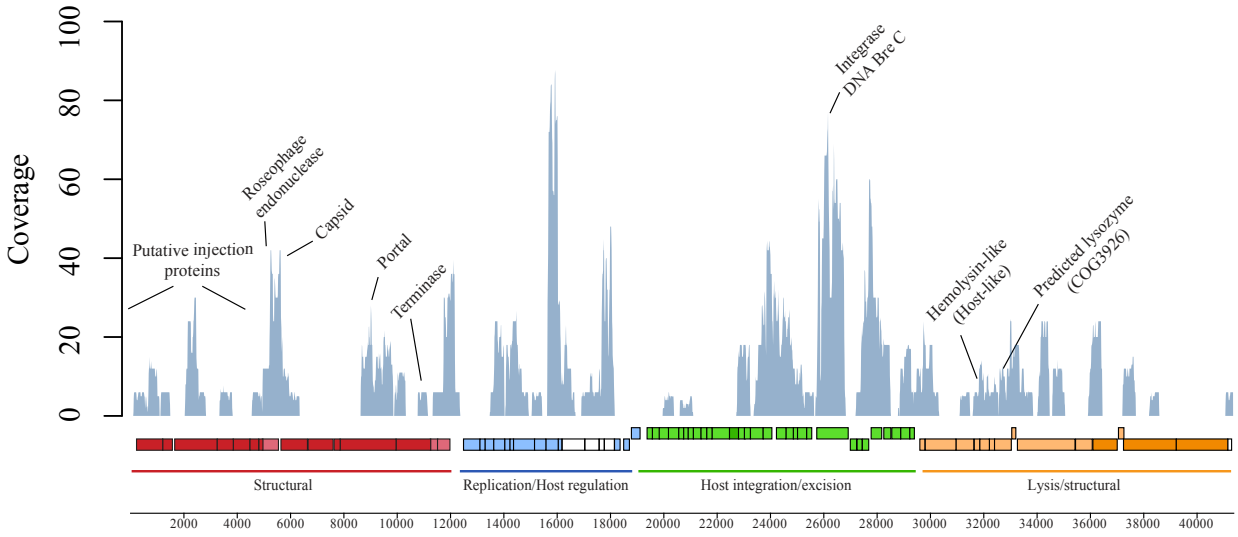


Figure 5.23. Recruitment plot (tblastn) of reads from the Raunefjorden viral metagenome to the genome of ϕ2047A

Recruitment plot (tblastn) of reads from the Raunefjorden viral metagenome to the genome of ϕ2047A. Threshold e-Value was 10^{-10}. The y-axis is coverage and the x-axis is genome position overlaid with the gene map.

CHAPTER 6 -
CONCLUSION

The roles that aquatic microbes play in the transfer of carbon and nutrients to higher trophic levels has been of great interest to me since I was an undergraduate. Microorganisms are metabolically diverse and contribute greatly to total biomass in the ocean, essentially controlling the flow of energy and elements (Falkowski et al. 2008). Roseobacters and roseophage are an excellent group of model organisms for studying these processes. For my doctorate, I wanted to investigate the influence that roseophage have on food web structure and nutrient cycling. Measuring the activity of specific microbial groups or metabolic functions is a continuing and challenging goal of microbial ecologists. These measurements are essential to accurately model the biogeochemical cycling of elements within the ocean and the atmosphere, ultimately leading to predictions on how changes in climate will alter ecosystems. In my naivety, I confidently assumed that by the time I graduated I would have measured the loss of *Roseobacter* production to viral lysis. But as the chapters of this dissertation reveal, the idea of such a measurement was unrealistic at the time. We first needed to understand the diversity and temporal dynamics of individual *Roseobacter* 'species' and obtain a general idea of what a roseophage actually was and develop tools to detect them in natural systems. These challenges are more than enough for a single dissertation and in the end provide a solid foundation for future graduate students to build upon.

Microscopy is a basic and dependable tool of a microbiologist. We rely on the technique for much of our work and must appreciate the method. Upon arriving at the University of Tennessee I was somewhat skeptical that viruses could actual be enumerated using a compound light microscope. But I was proven otherwise and became reliant on the method for parts of my dissertation. By developing a procedure using an alternative Anodisc™ membrane I become versed in the intricacies of enumerating viral particles and the importance of the statistical measures involved. Although not a large part of chapter one, the nucleic acid staining step was found to be the source of much variation between the two membranes, where backstaining often resulted in different counts than prestaining the sample. Method comparisons require a statistical analysis and for microscopy counts this includes not only replicate slides, but also considerations of sample volume, density of particles per field, and the number of fields (Chae et al. 2008). The

above fundamental properties are often not known to a novice microscopist (including myself), who is focused on secondary data analysis and not data acquisition.

Maintaining and contributing to culture collections is an essential mechanism to preserve the diversity of microorganisms from our ever-changing planet (Komagata 1997; Smith 2003). Over the last 30 years the concept of the microbial loop has shifted the classic view (Fenchel 2008) of marine plankton community structure to a more complex system, dominated by heterotrophic prokaryotes and pico- and nanoeukaryotes. This has resulted in an explosion of new species characterizations over the past ten years, with some laboratories devoting resources exclusive to the endeavor. Thus, it is basically uneconomical for groups focused on the actual ecology of marine bacteria to describe an isolate relevant to their work. In the words of my advisor, these projects become a "moving target", where new species are being described at such a rate that a year's worth of work can become irrelevant overnight. Journals publishing new species descriptions have recently attempted to curb the influx of submitted manuscripts by requiring a greater abundance of arguably non-ecologically relevant biochemical tests (Tindall et al. 2010). Nevertheless, chapter two was accepted for publication, and *Marivita roseacus* is a novel species. For the chapter, I tried to focus not only on the phenotypic and physiological properties of the organism, but also describe what we know about its ecology. The next generation of microbiologists will eventually be the ones responsible for deciding what defines a new species and I am thankful to have experienced the current process.

High-throughput sequencing continues to transform microbial ecology. In the last ten years the challenge of obtaining enough sequences for a thorough analysis of a natural community has been substituted by one of properly interpreting the obtained sequence data. Deep-amplicon datasets are no exception, although compared to metagenomic sequences the level of complexity is far less. My previous experience with sequence alignment and phylogenetic tree building proved beneficial to chapter four and allowed us to drill down into the bacterial community dynamics of the Bergen mesocosm to a level not previously reported for these types of data. Such an analysis revealed a fine resolution of the dataset and permitted us to follow the succession patterns of almost 100 distinct bacterial types. Our ability to interpret the data was not only possible by our methods of analysis, but on the tremendous amount of

publically available sequences and ancillary data collected over the past decade. Although the initial learning curve of processing the sequences was steep, I was able to perform secondary biodiversity and correlation analyses as well. Having experience with deep-amplicon sequences and the available computational methods to analyze these datasets will be invaluable to my career as a microbial ecologist.

Viruses were something of an enigma to me four years ago and I am amazed at how much I have learned thus far and even more so by how much more awaits to be discovered. It is difficult to make any generalizations about bacteriophage, for the mosaic nature of their genomes seemingly provides an endless combination of traits. What can be said though, is that groups of phage do have distinct life cycle traits and genomic content, which ultimately leads to their ecology. A future challenge of marine virology will be identifying and characterizing these phage groups and unraveling the evolutionary forces that created them. Metagenomics will undoubtedly contribute to these endeavors, but there is no substitute for phage isolation and characterization for connecting genome content to physiology and for determining host type and host selectivity.

Researchers focusing on specific lineages of bacteria (i.e Roseobacters) and viruses (i.e roseophage) have produced thorough descriptions of their physiological capabilities and global distributions, allowing for interpretations of their ecology and importance to an ecosystem. Writing the last two chapters of this dissertation provided the realization that there may be benefits to studying organisms as 'ecogroups'; bacteria and viruses that have a similar role or distribution in the environment, but may be phylogenetically distinct or infect phylogenetically distinct hosts. Examples of such would be the phylotypes that peaked early in the Bergen mesocosm experiment or the N4-like bacteriophage. What factors lead to the success of the *Polaribacter* phylotype peaking on day 2 of the mesocom? How did this phylotype compete so successful? Are the other phylotypes on day 2 using similar mechanisms to succeed? Do N4-like phage infecting α–proteobacteria and γ-proteobacteria dominate under similar ecological conditions? Investigating organisms that have a similar specific ecology, regardless of a phylogenetic connection or identical habitat-type, may help us to understand the functional mechanisms that allow them to succeed in their environment. Such alternative approaches for

studying marine microbes could advance our understanding of the abiotic and biotic factors that lead to the success of particular species and the consequences that climate change will have on these groups as a whole and individually.

References

Chae, G. T., J. Stimson, M. B. Emelko, D. W. Blowes, C. J. Ptacek, and M. M. Mesquita. 2008. Statistical assessment of the accuracy and precision of bacteria- and virus-sized microsphere enumerations by epifluorescence microscopy. Water Res **42:** 1431-1440.

Falkowski, P. G., T. Fenchel, and E. F. Delong. 2008. The microbial engines that drive Earth's biogeochemical cycles. Science **320:** 1034-1039.

Fenchel, T. 2008. The microbial loop-25 years later. J Exp Mar Biol Ecol **366:** 99-103.

Komagata, K. 1997. Microbial diversity and the role of culture collections.

Smith, D. 2003. Culture collections over the world. International Microbiology **6:** 95-100.

Tindall, B. J., R. Rossello-Mora, H. J. Busse, W. Ludwig, and P. Kampfer. 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Micr **60:** 249-266.

VITA

Charles Ryan Budinoff received an Associates of Science degree from the Community College of Aurora (Aurora, Colorado) in 1998, a Bachelors of Science in Microbiology from Arizona State University (Tempe, Arizona) in 2002 under advisor Ferran Garcia-Pichel, a Masters of Science in Marine Science from the University of Georgia (Athens, Georgia) in 2005 under advisor James T. Hollibaugh, performed an internship at the Environmental Protection Agency (Athens, Georgia) from 2005-2007 under supervisor Marirosa Molina, and began his doctoral work at the University of Tennessee (Knoxville, Tennessee) in 2007 under advisor Alison Buchan.