RNA VIRAL DIVERSITY AND DYNAMICS ALONG THE ANTARCTIC PENINSULA

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

OCEANOGRAPHY

MAY 2015

By

Jaclyn A. Mueller

Dissertation Committee:
Grieg Steward, Chairperson
Alexander Culley
Matthew Church
Craig Smith
Guylaine Poisson, Outside member

Key words: marine RNA viruses, viral ecology, metagenomes, viromes, nucleic acid extraction, reverse transcription quantitative PCR (RT-qPCR), Antarctica

ACKNOWLEDGEMENTS

This research would not have been possible without the generosity and support of a number of people and organizations, for which I am very thankful. I would first like to acknowledge the Department of Oceanography, the Center for Microbial Oceanography: Research and Education (C-MORE), and the National Science Foundation (NSF) for supporting my research and providing such a fulfilling and enriching graduate career. I am forever indebted to the C-MORE ‘Ohana for their unwavering support and enthusiasm, while providing me the opportunity to experience outreach and education, enhance my leadership and professional development skills, and conduct my doctoral research. In particular, I thank Dr. David Karl for his insights, enthusiasm for oceanography, and continued support of my research. It has been a pleasure to work with so many bright, motivated, and inspiring scientists.

I am forever grateful for the patience, guidance, encouragement, and support of my advisor, Dr. Grieg Steward. His enthusiasm and creativity are contagious, and have made this experience quite an enjoyable one. Working with him has not only taught me to be more a perceptive and critical thinker, but also a better writer and scientist. I thank Dr. Alexander Culley for his mentorship over the years and for the inspiration to focus on marine RNA viruses. I am grateful for his patience, support, and countless insightful discussions. I thank Dr. Matthew Church for his continuous enthusiasm, perceptive discussions, and expertise in biogeochemical cycling. I thank Dr. Craig Smith for providing me the opportunity to join his lab group on a research cruise in the Larsen A Ice Shelf, Antarctica, and the ability to collect samples for an entire chapter of this dissertation. I thank Guylaine Poisson for her expertise in bioinformatics and collaboration with our earlier work on marine RNA viruses. It has been an honor to work with such a supportive and collaborative thesis committee.

As a member of the Steward Laboratory, I have had the opportunity to interact and work with a number of excellent and talented colleagues. I offer my gratitude to past and present members for their friendship, support, and assistance in the lab and field sampling, in particular Dr. Elisha Wood-Charlson, Dr. Olivia Nigro, Chris Schvarcz, Gordon Walker, and Kirena Clah. I would also like to thank the scientists, technicians, crew, and staff associated with the Palmer LTER program and the NBP1203 cruise on board the R/V *Nathaniel B. Palmer*. Their support, expert

advice, and assistance both at sea and onshore were instrumental to the success of instrument deployment and sample collection.

Most importantly, I would like to thank my family and friends for their continued encouragement, support, and love that greatly contributed to my sanity and happiness. My parents have always supported whichever direction or path I chose, believing I could accomplish whatever I put my mind to. Much love and thanks to my siblings, grandparents, aunts, uncles, and cousins for always being a source of inspiration. I thank the Culleys, Grabowskis, and Foremans for being such wonderful families and sharing such a special part of their lives with me. Finally, I am most grateful for the love and sacrifice of my partner and biggest supporter, Javier Miranda. I thank him for the laughter and happiness, positive energy, always putting food on the table, and embarking with me on our greatest adventure.

ABSTRACT

Marine viruses are capable of driving species succession during plankton blooms and greatly influencing biogeochemical cycling and food web dynamics through cell lysis. Most of our knowledge has come from studies in temperate and sub-tropical waters, but whether they play a major role in eukaryotic phytoplankton blooms in polar waters has not been investigated. RNA viruses in particular have been shown to contribute up to half of the total virioplankton and are known to predominantly infect eukaryotes. The goals of this dissertation were to provide the first description of RNA viruses in the Antarctic, determine their relative and absolute abundances, and investigate the diversity and dynamics of these viruses to better understand how they contribute to the ecology of eukaryotic plankton. A temporal analysis of RNA viral metagenomes collected throughout the summer bloom in the Western Antarctic Peninsula (WAP) revealed an abundant and diverse community of positive-sense, single-stranded (+ss) RNA viruses within the order *Picornavirales*. RNA viral metagenomes from the Eastern Antarctic Peninsula were analyzed to investigate spatial diversity, and revealed a similar dominance of +ssRNA viruses in the order *Picornavirales*, but a much lower diversity than the WAP metagenomes. Novel RNA virus genomes were assembled and primers were designed to target the RNA-dependent RNA polymerase gene to track individual phylotypes throughout the season in the WAP. In order to determine the absolute abundance of these RNA viruses, two methods were evaluated and optimized. First, a method for extraction of viral nucleic acids from plankton harvested on nanoporous anodic aluminum oxide filters was improved and validated. Second, a reverse transcription quantitative PCR (RT-qPCR) assay was developed to accurately quantify specific RNA viruses in nucleic acid extracts. Both methods were applied to provide the first estimates of abundance of specific RNA virus phylotypes in the ocean, revealing three

distinct patterns of abundance from four unique virus phylotypes. The results of this dissertation not only suggest the pervasiveness and ecological significance of RNA viruses in polar waters, but also present molecular techniques that should be of interest to virologists, oceanographers, and microbial ecologists.

TABLE OF CONTENTS

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 1

1.1 Introduction to marine viruses

Over the past few decades, the study of marine viruses has progressed quickly from the first recognition of their abundance in the oceans to a detailed appreciation of their vital importance and varied roles in marine ecosystems. We now know that viruses are the most abundant microbe in the ocean with concentrations of approximately tens of millions of viruses per milliliter of surface seawater (Wommack and Colwell 2000). Viruses also play an integral role in the evolution and community structures of their hosts and they influence marine biogeochemical cycling and food web dynamics (Suttle 2007). Ever since the first direct counts of marine viruses were performed over twenty years ago, it was assumed that the virioplankton is dominated by DNA-containing bacteriophages, because of patterns of abundance, predicted contact rates, and morphologies of the viruses in seawater (Steward et al 1992, Weinbauer 2004). More recent studies have shown that other types of viruses, including those containing single-stranded (ss) DNA (Angly et al 2006, Hopkins et al 2014, Labonte and Suttle 2013, Rosario and Breitbart 2011) or ss- or double-stranded (ds) RNA (Culley et al 2003, Culley et al 2006, Culley and Steward 2007, Steward et al 2013), may also be important contributors to marine plankton ecology.

Virtually all of the RNA viruses from marine systems described in the early literature were those that infect marine animals. Furthermore, until a decade ago all but one of the known marine bacteriophages and all viruses of marine phytoplankton had DNA genomes. Therefore RNA-containing viruses were not expected to have a numerically significant contribution to the total marine virioplankton (reviewed by Lang et al 2009). However, the isolation of many new RNA viruses from the marine environment that infect protists, including diatoms (Nagasaki et al 2004, Shirai et al 2008, Tomaru et al 2009, Tomaru et al 2012, Tomaru et al 2013), dinoflagellates (Tai et al 2003, Tomaru et al 2004), raphidophytes (Tai et al 2003), prasinophytes (Brussaard et al 2004), or thraustochytrids (Takao et al 2005), led to an interest in determining the overall diversity and relative abundance of RNA viruses in the ocean. In order to characterize the diversity of the RNA virus community, Culley et al (2003) developed degenerate primers to target the RNA-dependent RNA polymerase (RdRp) gene shared among the marine protistan

virus isolates within the order *Picornavirales* (marine picorna-like viruses). Subsequent metagenomic and targeted gene approaches have been used to describe the marine RNA virus communities in both temperate and tropical seawater (Culley et al 2006, Culley et al 2007, Culley and Steward 2007, Culley et al 2014, Gustavsen et al 2014). Of the identifiable sequences in the first RNA viral metagenomes prepared from marine samples, the overwhelming majority derived from members of a single order, the *Picornavirales*. Members of this order have positive-sense, single-stranded (+ss) RNA genomes. Only a very small percentage of viral sequences had hits matching dsRNA viruses in both tropical and temperate communities, 3% and 0.02% respectively (Culley et al 2006, Culley et al 2014). This was a surprising contrast to the enormous variety of positive-sense, negative-sense, and double-stranded RNA viruses found in plants and animals. Subsequent surveys targeting genes of marine picornavirads revealed enormous genomic diversity within the order *Picornavirales* that far exceeds that of cultured isolates. Many of these sequences appear to represent new virus families, demonstrating the value of cultivation-independent techniques for discovery of novel biodiversity. The identification of highly diverse RNA viruses that infect diverse eukaryotic hosts raises the question of how abundant these viruses may be.

Unlike DNA viruses, RNA viruses are small and cannot be reliably counted via standard methods of epifluorescence microscopy and flow cytometry (Brussaard et al 2000, Tomaru and Nagasaki 2007), thus the relative prevalence of RNA viruses in the environment could not be determined directly. A recent study, using an indirect approach to compare the total masses of viral RNA and DNA, suggested that RNA-containing viruses that infect eukaryotes may account for up to half or more of the virioplankton in tropical seawater (Steward et al 2013). However, this study was conducted in a single location and it is unknown whether this observation is typical of other regions of the ocean. In contrast to DNA viruses in the ocean, most of which infect prokaryotes (Fuhrman 1999), the marine RNA viruses almost exclusively infect eukaryotes (Lang et al 2009). Because eukaryotes are orders of magnitude lower in abundance than prokaryotes (Suttle 2007), the high abundance of RNA viruses was unexpected. These data might be explained by the much larger burst sizes of eukaryotic-infecting RNA viruses (Lang et al 2009), relative to bacteriophages (Wommack and Colwell 2000).

Viral infection and zooplankton grazing are the two major sources of plankton mortality, and these processes have different consequences for how carbon is routed through the food web (Fuhrman and Suttle 1993, Fuhrman 1999). Grazing leads to the incorporation of a portion of prey carbon into larger size classes, while viruses, by lysing cells, shunt the carbon into the dissolved and colloidal carbon pools (Wilhelm and Suttle 1999). Variation in the relative importance of these two pathways therefore directly affects global biogeochemical cycling. Since marine RNA viruses are predominantly lytic in nature (reviewed by Lang et al 2009) their higher-than-expected abundances suggest greater lysis and release of nutrients and carbon into the dissolved fraction from eukaryotic cells than was previously assumed. The specificity of viral infections means that an epidemic infection that spreads through a population can result in community succession, with a non-susceptible species filling the niche left open by the lysis of the first.

1.2 Viruses in the Antarctic

Although we now have many examples of the important ecological roles played by viruses, almost all of the studies have been conducted in temperate to sub-tropical waters. Whether viruses play a role in seasonal succession during the summer bloom in polar waters is not known. The Antarctic Peninsula undergoes a highly productive and seasonal cycle every year with the advancement and retreat of the sea ice, with a summer period of high productivity that supports a diverse array of fauna. The bimodal bloom dynamics in the Western Antarctic Peninsula (WAP) portray a succession from diatoms to flagellates through the season. This successional pattern is reminiscent of that observed in other natural and experimental systems, where bloom termination and phytoplankton community succession were clearly attributable to viral activity (Bratbak et al 1993, Jacquet et al 2002, Nagasaki et al 1994). At present, ecosystem models of the Antarctic food web include grazing, but not viral lysis (Ducklow et al 2006). Given the very different consequences of these two processes, a better understanding of the ecology of viruses around the Antarctic Peninsula may help to improve models of ecosystem dynamics in this region.

The few studies conducted in polar waters indicate that viruses are present in seawater and sea ice (Gowing et al 2004, Pearce and Wilson 2003). Their abundances (Marchant et al 2000, Pearce et al 2007), genome sizes (Steward et al 2000), and morphologies (Gowing et al 2002) are

similar to those seen elsewhere, but little is known about the ecology of these viruses. Additionally, since these studies largely focused on the dsDNA viruses, we know close to nothing about the majority of protistan-infecting RNA viruses. Because of the dominance of protistan phytoplankton in these annual blooms, and the fact that RNA viruses predominantly infect protists, it is likely that these viruses play an important role in the bloom dynamics of the Antarctic. This dissertation presents an investigation of RNA viruses in Antarctic waters and represents a significant contribution to our knowledge of marine RNA viruses and their influence on eukaryotic plankton.

1.3. Research objectives and structure

My overall objective for this dissertation was to utilize a suite of molecular techniques to provide the first description of RNA viruses in the Antarctic and contribute a more comprehensive understanding of the ecology of marine RNA viruses in polar waters. I investigated the temporal and spatial diversity and dynamics of the RNA viral communities along the Antarctic Peninsula to understand how these viruses contribute to the ecology of the eukaryotic plankton in the Antarctic. I also improved and validated two important methods including the collection and extraction of viral nucleic acids, and the quantification of specific RNA virus phylotypes. These two methods were then applied to obtain the first estimates of absolute abundance for RNA viruses in the environment.

This dissertation is organized in seven chapters. In **Chapter 1,** this chapter, I introduce the field of marine viral ecology and the importance of viruses in the Antarctic. In **Chapter 2**, I present an investigation of the RNA virus community along the Western Antarctic Peninsula in which I obtain the first estimates of relative RNA virus abundances in the Antarctic and determine the temporal variability of the community. RNA viral metagenomes were collected and analyzed on six occasions throughout the summer phytoplankton bloom at Palmer Station from November 2010 to March 2011 to determine the changes in community composition. The mass ratio of RNA and DNA in the viral fraction was used to determine the RNA virus abundances relative to the entire virioplankton throughout the season. Five near complete, novel RNA viral genomes were assembled and their relative abundances were also estimated based on the percentage of reads assembled to form a given contig.

In order to determine the spatial variability in diversity, I investigated the RNA virus communities on the Eastern Antarctic Peninsula and present the results in **Chapter 3**. As part of that study, RNA viral metagenomes were analyzed at four stations along a transect in the Larsen A Ice Shelf to compare viral community composition at each station and on either side of the Peninsula. The phytoplankton communities in the EAP appear to be dominated by large diatoms, which increase in abundance from the coastal margins to the shelf (personal communication, Maria Vernet). Therefore, I expected to see a shift in virus communities along the transect, and assemblages distinct from those found in the WAP (with small diatom and flagellate hosts).

Another of my main objectives was to quantify individual populations of uncultivated RNA viruses. Having a means to directly quantify populations would allow me to independently and more directly determine whether RNA viruses are as abundant as previous work suggested. This would also allow me to track the population dynamics of RNA viruses in seawater for the very first time. To accomplish this objective I had some technical hurdles to overcome. The first was to ensure I could extract nucleic acids from planktonic viruses with very high efficiency. Therefore, **Chapter 4** focuses on improving and validating a method for extraction of viral nucleic acids harvested from water samples by filtration onto nanoporous anodic aluminum oxide filters. These filters have been recommended for collection of viral nucleic acids (Steward and Culley 2010) due to their small pore size (0.02 μm) and high porosity with a disposable housing unit that allows for the quick and efficient capture of total microbial plankton, including viruses, with a high throughput capability. However, their small pore sizes could also lead to the trapping of nucleic acids, and the filter membrane material has been shown to irreversibly bind to nucleic acids in the presence of guanidinium buffers (a common reagent in popular spin-column extraction kits; Dames et al 2006, Gerdes et al 2001). For these reasons, I felt an assessment of their efficiency was worthwhile. Chapter 4 has been published in *Applied and Environmental Microbiology* (Mueller JA, Culley AI, Steward GF (2014). Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters. *Applied and Environental Microbiology* **80:** 3930-3942).

My second technical challenge was to develop a molecular assay that could accurately quantify specific phylotypes of RNA viruses present in nucleic acid extracts. Reverse-transcription quantitative PCR is currently the most common way to quantify specific RNA targets in a

sample, but it is highly sensitive to reaction conditions and must be optimized for each application. In **Chapter 5**, I describe my investigation of some of the variables that influence the performance of the RT-qPCR assay and provide insights into the proper use of this assay for quantifying RNA viruses. A number of variables were tested, including the source of the reverse transcriptase, concentrations of background RNA, target RNA, and primers, and the nature of the RT-primer. I also investigated how the choice of the material for the standard curve (linearized plasmids, amplicons, or RNA transcripts) influences the interpretation of the results.

I then applied the RT-qPCR method from Chapter 5 to investigate the dynamics of specific RNA virus phylotypes throughout the phytoplankton bloom at Palmer Station in **Chapter 6**. In this study, I tracked the abundances of individual assembled RNA viruses that appeared to be relatively abundant based on the metagenome assemblies from Chapter 2. Since RNA viruses are typically lytic and very host-specific, I expected their abundance dynamics to reflect that of their hosts, and provide some insight to host ecologies. Samples were collected from Palmer Station over the same time period as the metagenome samples, but with a higher temporal resolution (biweekly).

Finally, **Chapter 7** provides the overall conclusions and implications of this work, and outlines some potential directions and ideas for future research.

1.4 REFERENCES

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C *et al* (2006). The Marine Viromes of Four Oceanic Regions. *PLoS Biology* **4:** e368.

Bratbak G, Egge J, Heldal M (1993). Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series* **93:** 39-48.

Brussaard CP, Noordeloos AA, Sandaa RA, Heldal M, Bratbak G (2004). Discovery of a dsRNA virus infecting the marine photosynthetic protist Micromonas pusilla. *Virology* **319:** 280-291.

Brussaard CPD, Marie D, Bratbak G (2000). Flow cytometric detection of viruses. *Journal of Virological Methods* **85:** 175-182.

Culley AI, Lang AS, Suttle CA (2003). High diversity of unknown picorna-like viruses in the sea. *Nature* **424:** 1054-1057.

Culley AI, Lang AS, Suttle CA (2006). Metagenomic analysis of coastal RNA virus communities. *Science* **312:** 1795-1798.

Culley AI, Lang AS, Suttle CA (2007). The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. *Virology Journal* **4:** 69.

Culley AI, Steward GF (2007). New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Applied and Environmental Microbiology* **73:** 5937-5944.

Culley AI, Mueller JA, Belcaid M, Wood-Charlson EM, Poisson G, Steward GF (2014). The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *MBio* **5**.

Dames S, Bromley LK, Herrmann M, Elgort M, Erali M, Smith R *et al* (2006). A Single-Tube Nucleic Acid Extraction, Amplification, and Detection Method Using Aluminum Oxide. *The Journal of Molecular Diagnostics* **8:** 16-21.

Ducklow HW, Fraser W, Karl DM, Quetin LB, Ross RM, Smith RC *et al* (2006). Water-column processes in the West Antarctic Peninsula and the Ross Sea: Interannual variations and foodweb structure. *Deep Sea Research Part II: Topical Studies in Oceanography* **53:** 834-852.

Fuhrman JA, Suttle CA (1993). Viruses in Marine Planktonic Systems. *Oceanography* **6:** 51-63.

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

Gerdes JC, Marmaro JM, Roehl CA (2001). Nucleic acid archiving. Xtrana: US.

Gowing MM, Riggs BE, Garrison DL, Gibson AH, Jeffries MO (2002). Large viruses in Ross Sea late autumn pack ice habitats. *Marine Ecology Progress Series* **241:** 1-11.

Gowing MM, Garrison DL, Angela H. Gibson JMK, Jeffries MO, Fritsen CH (2004). Bacterial and viral abundance in Ross Sea summer pack ice communities. *Marine Ecology Progress Series* **279:** 3-12.

Gustavsen JA, Winget DM, Tian X, Suttle CA (2014). High temporal and spatial diversity in marine RNA viruses implies that they have an important role in mortality and structuring plankton communities. *Frontiers in microbiology* **5:** 703.

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

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

Labonte JM, Suttle CA (2013). Previously unknown and highly divergent ssDNA viruses populate the oceans. *Isme Journal* **7:** 2169-2177.

Lang AS, Rise ML, Culley AI, Steward GF (2009). RNA viruses in the sea. *FEMS Microbiology Reviews* **33:** 295-323.

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

Nagasaki K, Ando M, Itakura S, Imai I, Ishida Y (1994). Viral mortality in the final stage of Heterosigma akashiwo (Raphidophyceae) red tide. *Journal of Plankton Research* **16:** 1595-1599.

Nagasaki K, Tomaru Y, Katanozaka N, Shirai Y, Nishida K, Itakura S *et al* (2004). Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom Rhizosolenia setigera. *Applied and Environmental Microbiology* **70:** 704-711.

Pearce DA, Wilson WH (2003). Viruses in Antarctic ecosystems. *Antarctic Science* **15:** 319-331.

Pearce I, Davidson AT, Bell EM, Wright S (2007). Seasonal changes in the concentration and metabolic activity of bacteria and viruses at an Antarctic coastal site. *Aquatic Microbial Ecology* **47:** 11-23.

Rosario K, Breitbart M (2011). Exploring the viral world through metagenomics. *Current opinion in virology* **1:** 289-297.

Shirai Y, Tomaru Y, Takao Y, Suzuki H, Nagumo T, Nagasaki K (2008). Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom Chaetoceros tenuissimus Meunier. *Applied and Environmental Microbiology* **74:** 4022-4027.

Steward GF, Wikner J, Cochlan WP, Smith DC, Azam F (1992). Estimation of virus production in the sea: II. Field results. *Marine Microbial Food Webs* **6:** 79-90.

Steward GF, Montiel JL, Azam F (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnology and Oceanography* **45:** 1697-1706.

Steward GF, Culley AI (2010). Extraction and purification of nucleic acids from viruses. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of aquatic virology*. American Society of Limnology and Oceanography: Waco, TX. pp 154-165.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G (2013). Are we missing half of the viruses in the ocean? *Isme Journal* **7:** 672-679.

Suttle CA (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology* **5:** 801-812.

Tai V, Lawrence J, Lang A, Chan A, Culley A, Suttle C (2003). Characterization of HaRNAV, a single-stranded RNA virus causing lysis of Heterosigma akashiwo (Raphidophyceae). *Journal of Phycology* **39:** 343-352.

Takao Y, Nagasaki K, Mise K, Okuno T, Honda D (2005). Isolation and characterization of a novel single-stranded RNA Virus infectious to a marine fungoid protist, Schizochytrium sp. (Thraustochytriaceae, Labyrinthulea). *Applied and Environmental Microbiology* **71:** 4516-4522.

Tomaru Y, Katanozaka N, Nishida K, Shirai Y, Tarutani K, Yamaguchi M *et al* (2004). Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga Heterocapsa circularisquama. *Aquatic Microbial Ecology* **34:** 207-218.

Tomaru Y, Nagasaki K (2007). Flow Cytometric Detection and Enumeration of DNA and RNA Viruses Infecting Marine Eukaryotic Microalgae. *Journal of Oceanography* **63:** 215-221.

Tomaru Y, Takao Y, Suzuki H, Nagumo T, Nagasaki K (2009). Isolation and characterization of a single-stranded RNA virus infecting the bloom-forming diatom Chaetoceros socialis. *Applied and Environmental Microbiology* **75:** 2375-2381.

Tomaru Y, Toyoda K, Kimura K, Hata N, Yoshida M, Nagasaki K (2012). First evidence for the existence of pennate diatom viruses. *Isme Journal*.

Tomaru Y, Toyoda K, Kimura K, Takao Y, Sakurada K, Nakayama N *et al* (2013). Isolation and characterization of a single-stranded RNA virus that infects the marine planktonic diatomChaetocerossp. (SS08-C03). *Phycological Research* **61:** 27-36.

Weinbauer MG (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28:** 127-181.

Wilhelm SW, Suttle CA (1999). Viruses and Nutrient Cycles in the Sea. *BioScience* **49:** 781-788.

Wommack KE, Colwell RR (2000). Virioplankton: Viruses in Aquatic Ecosystems. *Microbiology and Molecular Biology Reviews* **64:** 69-114.

Chapter 2

RNA viruses as major contributors to Antarctic virioplankton

Jaclyn A Mueller[1]

Alexander I Culley[1, 2], Christopher R Schvarcz[1], Grieg F Steward[1]

[1]Department of Oceanography, Center for Microbial Oceanography: Research and Education, University of Hawai‘i at Mānoa, 1950 East-West Road, Honolulu HI 96822 USA

Present Address: [2]Département de biochimie de microbiologie et de bio-informatique, Université Laval, Quebec City, Quebec Canada.

2.1 ABSTRACT

Evidence has been accumulating that diverse RNA-containing viruses that infect eukaryotic plankton are abundant in seawater. The metagenomic data available so far suggest that these viruses are dominated by types having positive-sense, single-stranded RNA (+ssRNA) genomes that are phylogenetically related to known protist-infecting viruses within the order *Picornavirales*. Previous data have come from temperate and tropical waters, but we hypothesized that RNA viruses would also be present and abundant in polar marine waters, where intense seasonal blooms are dominated by eukaryotic phytoplankton, most notably diatoms. To test this, we harvested viruses from seawater samples from a coastal site near Palmer Station, Antarctica on six occasions spanning the rise and decline of the summer phytoplankton bloom (November to March). We then estimated the relative abundances of RNA and DNA viruses from the mass ratios of RNA and DNA in the purified viral samples and verified the composition of the operationally-defined viral RNA using shotgun metagenomics. Our data suggest that RNA viruses contributed from 8 to 65% of the total virioplankton in these polar waters depending on the sampling date. The lowest percentage occurred at the beginning of the study in early austral spring (November) and the highest in mid-summer (January). Our metagenomic analyses suggest that, as observed in tropical and temperate waters, RNA viruses in coastal Antarctic seawater have +ssRNA genomes most closely related to viruses in the order *Picornavirales*. Assembly of the metagenomic reads resulted in five novel, nearly complete genomes, three of which had features similar to diatom-infecting viruses. Our data are consistent with the hypothesis that RNA viruses influence diatom bloom dynamics in Antarctic waters.

2.2 INTRODUCTION

Viruses influence the structure of marine plankton communities and influence marine biogeochemical cycling and food web dynamics by selectively infecting and lysing cells (Fuhrman 1999, Suttle 2007). The focus of earlier literature on marine viruses was on double-stranded (ds) DNA-containing viruses that infect bacteria (Breitbart et al 2002, Steward et al 1992, Weinbauer 2004) or eukaryotic phytoplankton (Chen and Suttle 1995, Dunigan et al 2006, Van Etten et al 1991, Van Etten et al 2002). More recent work suggests that other types of viruses, namely those containing single-stranded (ss) DNA (Angly et al 2006, Hopkins et al 2014, Labonte and Suttle 2013, Rosario and Breitbart 2011) or ss- or dsRNA (Culley et al 2003, Culley et al 2006, Culley and Steward 2007, Steward et al 2013), may also be important contributors to marine plankton ecology.

In contrast to the dsDNA viruses, which appear to be numerically dominated by those infecting bacteria (Edwards & Rohwer 2005), marine RNA viruses appear to be dominated by those that infect eukaryotes (Lang et al 2009). Despite the much lower concentrations of eukaryotes relative to prokaryotes in the ocean, a recent study in coastal tropical waters provided evidence that these eukaryote-infecting RNA viruses were just as abundant as prokaryote-infecting DNA viruses (Steward et al 2013). Since there has been only this one study in one location so far, it is not clear whether the observed high abundance of RNA viruses is a general phenomenon.

We hypothesized that viruses, and RNA viruses in particular, would play an important role in the summer phytoplankton bloom in Antarctic waters. The typical bloom in the Western Antarctic Peninsula is initiated by diatoms, which reach highest concentrations in early to mid-summer. The diatom abundances then decline with a transition to dominance by photosynthetic flagellates, which can generate a second peak in chlorophyll (Moline et al 2004). One possible driver of the succession in phytoplankton is increasing mortality of diatoms when they reach high population densities. While many studies have investigated grazing as a source of phytoplankton mortality in the Antarctic (Burkill et al 1995, Calbet et al 2005, Ross et al 2008), very few studies have looked at the role of viruses (Evans and Brussaard 2012, Guixa-Boixereu et al 2002). Viruses could be a major contributor to succession by driving density-dependent infection and mass lysis of phytoplankton blooms (Bratbak et al 1993, Nagasaki et al 1994). The diatom viruses isolated

to date contain ssRNA or ssDNA and none are known yet that contain dsDNA (Tomaru and Nagasaki 2011), suggesting that viral contributions to the Antarctic phytoplankton bloom may be discerned among the RNA viruses.

In this paper, we present the first metagenomic analysis of RNA viral communities in coastal polar waters. We analyze the changes in community composition through the course of the summer phytoplankton bloom near Palmer Station, Antarctica and present five near-complete genomes assembled from the metagenomes.

2.3 METHODS

2.3.1 Field sample collection and processing

Seawater samples were collected from the Palmer Long-Term Ecological Research (PAL-LTER) Station B (64° 46.77'S, 64° 04.35'W) located south of Anvers Island on the Western Antarctic Peninsula on six occasions from November 13, 2010 to March 3, 2011 (Table 2.1). Seawater was pumped from 5 m depth with a submersible stainless steel pump (Monsoon XL DTW; Proactive Environmental Products, Bradenton, FL, USA) passing directly through an in-line polyethersulfone membrane filter capsule (Whatman 0.2 μm Polycap TC 36; GE Healthcare Bio-Sciences, Piscataway, NJ, USA) into acid-washed, polycarbonate carboys. The filtered seawater was transported to a cold room (4 °C) within 1 hr and stored there until it could be processed (within 2 hr). High-molecular-weight dissolved and colloidal organic matter (which included viruses) was concentrated to a final volume of about one liter via tangential flow ultrafiltration with 0.3 m^2 of Ultracel composite regenerated cellulose membranes (three stacked Pellicon 2 mini cassettes, 30 kDa; Millipore, Billerica, MA, USA). The retentates were concentrated further with centrifugal ultrafiltration devices (30 kDa, Centricon 70, Millipore) by centrifuging at 2,500 RPM for 20 minutes in a swinging bucket rotor (GH-3.8A; Beckman Coulter Inc., Brea, CA, USA). Each sample was concentrated on one device by repeated emptying of the filtrate and refilling of the sample reservoir between spins until the entire sample was concentrated. Samples were recovered by inverting the filter into the collection cup provided and centrifuging at 2,000 rpm for 3 min, then transferring the retentate (final volume = ~1.5 ml) to a 2 ml polypropylene tube (screw-cap with o-ring). Concentrates were flash-frozen in liquid N_2, and stored at -80°C. Upon arrival to our home laboratory, samples were thawed and viruses were purified through

sequential step and continuous CsCl buoyant density gradients (Lawrence and Steward 2010). Fractions of approximately 0.5 ml were collected from the final gradient from top to bottom using an Auto Densi-Flow (Labconco, Kansas City, MO, USA).

2.3.2 Buoyant density fraction analysis

The density of each fraction was measured using a positive-displacement micropipette and analytical balance (Lawrence and Steward 2010). After buffer exchange into SM (0.4 M NaCl, 0.02 M $MgSO_4$, 0.05 M Tris, pH 7.5) using centrifugal ultrafiltration units (0.5 ml Amicon Ultra 30 kDa; Millipore), nucleic acids were extracted from a portion of each fraction with the MasterPure Complete DNA and RNA Purification Kit (Epicentre, Madison, WI, USA), according to the manufacturer's instructions. Extracted fractions were split for subsequent DNA and RNA analyses. The total DNA in each fraction was measured by fluorometry using the Quant-iT DNA Assay Kit (Life Technologies, Carlsbad, CA, USA) in a multi-well plate fluorometer (2030 Multilabel Reader, PerkinElmer, Waltham, MA, USA). Samples for RNA analysis were treated with TURBO DNase (Life Technologies) to avoid non-specific signal and RNA was quantified by fluorometry using the Quant-iT RNA Assay Kit (Life Technologies) in a cuvette fluorometer (TD-700, Turner Designs, Sunnyvale, CA, USA).

2.3.3 Calculation of RNA and DNA virus abundance

To calculate the number of RNA and DNA viruses present in each sample, we used a mass apportioning approach, as previously described (Steward et al 2013). The mass of RNA was summed for the fractions in a narrow buoyant density range (1.39-1.51 g ml^{-1}) that seems to be typical of many marine RNA viruses (Culley et al 2014, Steward et al 2013). We summed the DNA content from a slightly broader buoyant density range (1.34-1.51 g ml^{-1}) that encompassed the main DNA peak in the virus density range. Based on the data in the Ninth Report of the International Committee on Taxonomy of Viruses and as reported previously (Steward et al 2013), we calculated an average RNA mass of 5.1×10^{-18} g per 9 kb ssRNA virus. We assumed that all unknown sequences from the viral RNA fraction (classified as "No hits" below) were derived from RNA viruses. We assumed an average DNA mass of 5.5×10^{-17} g per virion, which is equivalent to 50 kb of dsDNA, an average that was found in a wide range of environments (Steward et al 2000). After converting the RNA and DNA masses to viral abundances, we

calculated the percentage of total viruses that were RNA-containing. To get a sense of what the possible range of the percentage might be, we used more extreme assumptions as previously described (Steward et al 2013). For the high estimate, we assumed the average DNA content per dsDNA virus was two-fold greater (100 kb), as calculated for marine metagenomic data (Angly et al 2006). For the low estimate, we assumed that the average dsDNA viral genome was two-fold smaller (25 kb), and that the sequences from the putative viral RNA fraction with no significant hits in GenBank (47% of the total) were not viral.

2.3.4 Metagenome construction

For each sampling date, the viral fraction with the greatest amount of RNA (the peak fraction) was processed for metagenomic sequencing. The RNA in these peak fractions was amplified with a random-priming sequence-independent single-primer amplification (RP-SISPA) approach (Djikeng et al 2008, Steward and Culley 2010, Steward et al 2013), followed by a reconditioning reaction to both increase yield and minimize heteroduplex formation (Duhaime et al 2012). Discrepancies in the volumes and concentrations for the amplification reaction have appeared in the literature, so the complete reaction conditions are provided hereof. For random-priming cDNA synthesis, 5-10 µl of extracted viral RNA was added to a 0.2 ml PCR tube with SISPA primers: 20 pmol FR26RV-N (GCCGGAGCTCTGCAGATATCNNNNNN) and 0.1 pmol FR40RV-T (GCCGGAGCTCTGCAGATATC$(T)_{20}$), 10 nmol each dNTP, and a final volume of 13 µl. The reaction was heated to 65 °C for 5 min to disrupt RNA secondary structure, then placed on ice for at least 1 min. While still on ice, 1 × First Strand Buffer, 0.1 µmol DTT, 40 U RNase OUT, and 200 U Superscript III were added to the tube and mixed gently (20 µl final volume). The reactions were incubated sequentially at 25 °C for 10 min to keep random hexamers annealed and initiate transcription, 50 °C for 1 hr to allow efficient reverse transcription, and 94 °C for 3 min to denature the product and allow annealing of additional FR26RV-N primer to the first-strand cDNA product. Samples were then rapidly cooled to 4 °C. A complementary second DNA strand was then synthesized by adding 2.5 U Klenow Fragment 3′-5′ exo– (New England Biolabs Inc., Ipswich, MA, USA) and incubating at 37 °C for 1 hr. The reaction was terminated by incubating at 75 °C for 10 min.

Amplification of the cDNA was carried out in triplicate 50 µl reactions containing 1-5 µl of

template with a final concentration of 1 × Reaction Buffer, 0.2 mM dNTPs, 1 µM FR20RV (SISPA) primer, and 2.5 U Expand High Fidelity Plus Enzyme (Roche, Indianapolis, IN, USA). PCRs were incubated at 94 °C for 2 min, followed by 32 cycles of denaturation at 94 °C for 30 s, annealing at 65 °C for 30s, extension at 68 °C for 4 min, and a final extension at 68 °C for 7 min. A fraction of each reaction was loaded onto a 1% agarose gel containing 1 × SYBR Safe stain and 1 × TAE, and run at 100V for 40 min to ensure that amplification resulted in visible PCR products (typically a smear 250-1000 bp). Reconditioning reactions were then performed on 10-fold dilutions of amplified DNA in a fresh PCR mix (200 µl reactions with 20 µl original PCR mixture as template, and all other reagents in same concentrations as small-scale PCR). Thermal cycling conditions were the same as before, but with only three cycles of denaturation, annealing, and extension. All PCR products were pooled and concentrated in a centrifugal ultrafiltration unit (0.5 ml Amicon Ultra 100 kDa; Millipore), purified with QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA), and eluted with 150 µl TE which was warmed to 80 °C immediately prior to adding to the column. Samples were loaded onto a 0.5% agarose gel containing 1 × SYBR Safe stain and 1 × TAE, and run at 100 V for 60 min. Amplicons of 500-1000 bp were gel purified to obtain 100-500 ng per sample, and diluted to equal concentrations prior to sequencing.

Amplicons were sequenced with the 454 GS FLX+ System (Roche) at the ASGPB sequencing facility at the University of Hawai‘i at Mānoa. The yield of high-quality reads was low, so samples were resubmitted for additional sequencing with the 454 GS FLX System (Roche). For each metagenome, the sequences from both runs were pooled and passed through the same quality-control parameters to remove any low quality reads. SISPA primers were removed, and reads were quality filtered (≥ 21 average phred score, dereplication) using PRINSEQ v0.20.3 (Schmieder and Edwards 2011). Trimmed, high-quality sequences from each metagenome were assembled separately, as well as combined and treated as a single library, with the CLC Genomics Workbench version 5.0 (CLCBio, Cambridge, MA, USA). Assemblies were constructed with the default parameters using a non-global alignment.

2.3.5 Metagenomic and genomic analyses

Rarefaction curves were generated with quality filtered reads in Metavir (Roux et al 2011) using

a clusterization percentage of 75. Quality filtered reads and assembled contigs from individual metagenomes and the pooled libraries were analyzed with BLASTx (Altschul et al 1990) searches in the NCBI nr ("non-redundant") database. Hits with an *E* value $\leq 10^{-5}$ were considered significant. This output was analyzed using MEGAN v. 5.5.0 (Huson et al 2011), which assigns each contig, and all of the reads that comprise it, to the top hit of the blast search (top 5%). Krona charts were made from the MEGAN analyses using Krona (Ondov et al 2011). Open reading frames (ORFs) were identified using Geneious 6.1.8 and used to search the Conserved Domain Database (CDD) NCBI database (Marchler-Bauer et al 2011). Primers pairs were designed using Primer3 (Koressaar and Remm 2007, Untergasser et al 2012) to target regions of both the structural and non-structural proteins of about 800-1,000 bp, including the untranslated region between the two ORFs, to verify the consensus sequence of the assembled genomes.

In addition, primers were designed to target the 3' end of contig PAL128 in order to close the second ORF. Equal volumes of extracted viral RNA (2 µl) were pooled from each metagenome sample, and a 3 µl aliquot of the mixture was used for cDNA synthesis. RNA was converted to cDNA using only a poly-T tagged primer (FR40RV-T; (Djikeng et al 2008)) under the same reaction conditions as described above, but without FR26RV-N primer. The resulting product was PCR amplified with a forward primer designed from the assembled contig and reverse primer FR20RV under the same PCR conditions described above, but without a reconditioning step. The amplicon was concentrated and purified with a MinElute PCR Purification Kit (Qiagen) and run on a 0.5 % agarose gel in 1 × TAE stained with SYBR Safe. The single band was visualized with a Safe Imager 2.0 Blue-Light Transilluminator (Life Technologies), excised and purified with a MinElute Gel Extraction Kit (Qiagen), and verified by sequencing. For that purpose, the amplicons were cloned into the pSMART HC Kan vector (Lucigen) according to the manufacturer's instructions.

2.3.6 Identification and analysis of RdRp phylotypes

HMMER 3.0 (Eddy 1998) was used to identify RdRp-like sequences from quality-filtered assembled and unassembled reads using a profile hidden Markov model (profile HMM). A profile HMM was produced based on amino acid alignments of conserved RdRp regions 4 through 7 (Koonin et al 2008) of representative picornavirad isolates and amplified or assembled

environmental virus sequences (Culley et al 2006, Culley and Steward 2007, Culley et al 2014). Sequences fitting the criteria of the model were retrieved from the libraries translated in all six frames. RdRp sequences that matched a known RdRp gene by BLAST with an *E* value $\leq 10^{-5}$ were considered to be significant. Translated RdRp sequences that were more than 98% identical at the amino acid (aa) level were considered a single phylotype. Sequences were also clustered at a level of 75.7% aa identity to provide some sense of taxonomic diversity. This aa sequence identity threshold was chosen as an arbitrary, but probably conservative, approximation of species-level diversity based on the ICTV classifications of picornavirads. It represents the greatest distance between any two officially classified strains of picornavirads currently classified as the same species (Human Rhinovirus 2 and Human Rhinovirus 89). A maximum-likelihood tree of RdRp gene sequences was created with PHYML, WAG model, and 1,000 bootstraps (Guindon et al 2010) from protein sequences aligned with MAFFT (Katoh et al 2005). The sequences in this tree can be found in Table 2.S1.

2.3.7 Accession numbers

Metagenomic data referred to in this manuscript are available through the GenBank Sequence Read Archive (SRA). The raw reads are available under the accession numbers SRR1648122 (RNAV_PAL20101113), SRR1648127 (RNAV_PAL20101218), SRR1648128 (RNAV_PAL20110107), SRR1648129 (RNAV_PAL20110127), SRR1648130 (RNAV_PAL20110217), and SRR1648131 (RNAV_PAL20110303).

2.4 RESULTS

2.4.1 Buoyant density and relative abundance of RNA viruses

Nucleic acids in the density gradients displayed some variability both in distribution and in total mass, but both DNA and RNA were always observed in the fractions with buoyant densities in the range expected for marine viruses (1.34-1.51 g ml^{-1}). Gradient fraction 16, having an average density of 1.45 g ml^{-1}, was the peak RNA fraction in 5 of 6 samples (RNAV_PAL20101113 through RNAV_PAL20110217). In the final sample, RNAV_PAL20110303, peak RNA was in fraction 17, which had an average density of 1.47 g ml^{-1}. The estimated percentage of the total virus community comprised of RNA viruses varied from a low of 8% (range 2-15%) on

November 13, 2010 at the beginning of the season to a high of 65% (range 30-79%) on January 27, 2011 in mid-austral summer (Table 2.1) where the ranges reflect the application of more or less conservative assumptions as described in the Methods section. Excluding the lowest value at the first sampling, the contribution of RNA viruses varied only from 30-65% with an average of around 50% of the total viruses.

2.4.2 Metagenomic assembly analysis

The number of quality-controlled, de-replicated reads varied from 19,150 to 78,321 per sample for a total of 365,902 reads (Table 2.2). The majority of reads formed contigs (86-98%) in each separate library, and a total of 96% formed contigs when pooled together. The maximum contig length ranged from 3,996 to 8,187 bp in the individual library assemblies, and in the total reads assembly the longest contig was 8,117 bp. After sequencing the 3' end of PAL128, this contig was extended to 8,660 bp. The GC content of assembled contigs ranged from 43-45%, on average.

Rarefaction curves produced by Metavir (Roux et al 2011) with a threshold of 75% suggest that sample RNAV_PAL20101113 had the highest richness, while RNAV_PAL20101218 and RNAV_PAL20110303 had the lowest (Fig. 2.1). The number of sequences per sample varied, but normalizing the data to compare an equal number of sequences from each sample resulted in the same trends (data not shown).

Only 15% of the filtered reads were classified as viral sequences, and most (82.7%) had no similarity to a known sequence prior to assembly (Fig. 2.2). After assembly, nearly half of the total reads (47.4%) assembled to contigs classified as viral sequences based on top BLASTx hits. In contrast, assembly had little influence on the number of reads that were classified as cell-derived. This number increased from 2.1 to 2.2% (2.0% bacterial, 0.2% eukaryotic). Of the total classified sequences, 97.8% were classified as +ssRNA viruses and most of those were affiliated with viruses within the order *Picornavirales* (Table 2.3). Only 0.01% of the sequences were classified as dsDNA viruses and 0.04% classified as dsRNA viruses.

Taxonomic classification of filtered reads and assembled contigs from each individual library revealed some variability throughout the season (Table 2.3). All libraries contained contigs

similar to +ssRNA viruses from unclassified and uncultured members of the *Picornavirales* and members from the genus *Bacillarnavirus*. Libraries RNAV_PAL20101218, RNAV_PAL20110107, and RNAV_PAL20110217 were dominated by sequences similar to those found in viruses in the genus *Bacillarnaviruses* (>90%), and in libraries RNAV_PAL20110127 and RNAV_PAL20110303 roughly half were classified as *Bacillarnaviruses*. Library RNAV_PAL20101113 was the only library to contain contigs similar to members from the genus *Labyrnavirus*, and the *Dicistroviridae* family was only found in small percentages in libraries RNAV_PAL20101113, RNAV_PAL20110107, and RNAV_PAL20110127. Based on filtered reads and assembled contigs, library RNAV_PAL20101113 appears to contain the highest diversity of representatives within the *Picornavirales*. Overall, the taxonomic classification of assemblies for each library suggests a similar composition throughout the season, with a persistent dominance of +ssRNA viruses in the order *Picornavirales*.

2.4.3 Genome reconstruction

Reconstruction of novel RNA genomes from the metagenomic sequences of each individual library and from the pooled libraries revealed phylotypes with similar genomic organization, three of which appear to be most closely related to diatom-infecting viral isolates. The contigs (PAL128, PAL156, PAL438, PAL473, and PAL_E4) ranged in size from 7,047 to 8,660 bp with a GC content of 39.2 to 46.6% (Table 2.4). Of the pooled reads, 12.1% assembled to form PAL156, and from the individual libraries, 46% of reads from RNAV_PAL20110107 and 13% from RNAV_PAL20110127 mapped to PAL156. Mapping parameters were set as the default for CLC Genomics Workbench version 5.0. PAL_E4 was assembled with reads from library RNAV_PAL20110217, of which 60.7% formed the contig. Over half of the reads in each individual library mapped to PAL_E4, except for in RNAV_PAL20110107 (only 23%). PAL_E4 had 100% aa identity with another contig from the total reads assembly and one from assembly of library RNAV_PAL20110107, both of which only covered the length of open reading frame (ORF) 1 of the PAL_E4 genome. A portion of the genome with at least 98% aa identity was also assembled from the other individual libraries, and when pooled together, 53% of the total reads from all libraries mapped back to PAL_E4.

Each of the five assembled genomes contained two ORFs ranging in size from 1,872 to 5,574 bp (Table 2.3, Fig. 2.3). The first ORF encodes for non-structural proteins (i.e., RdRp) in all contigs, while the second encodes for structural proteins (i.e., capsid) in contigs PAL128, PAL156, PAL473, and PAL_E4. All of the genomes contained a region with significant BLASTx hits to +ssRNA RdRps in the first ORF. A helicase gene was identified in ORF1 for all genomes except for PAL473, but none contained an identifiable protease gene. Three to four structural proteins were identified in all of the genomes, except for PAL438, that were most similar to the capsid binding site of picornaviruses and the VP4 and capsid protein of dicistroviruses. The second ORF of PAL438 represents a polyprotein without any structural proteins similar to those found in the database, but a portion of this ORF did have a top hit to a helicase gene (*E* value: 4.2E-11). A polyadenylated [poly(A)] tail was detected for both PAL438 and PAL473.

2.4.4 RdRp phylogenetics

A maximum-likelihood tree of RdRp gene sequences shows the identified environmental sequences from this study relative to other known viruses within the order *Picornavirales* (Fig. 2.4). A search for RdRp sequences in the assemblies and unassembled reads for each metagenome, based on four of the seven conserved motifs (mpl-like region), returned 22 unique picornavirad-like sequences from RNAV_PAL20101113, 3 from RNAV_PAL20101218, 29 from RNAV_PAL20110107, 17 from RNAV_PAL20110127, 5 from RNAV_PAL20110217, and 8 from RNAV_PAL20110303. Clustering at a 75.7% aa identity threshold to approximate conservative species-level diversity resulted in 13 clusters in RNAV_PAL20101113, 2 in RNAV_PAL20101218, 2 in RNAV_PAL20110107, 3 in RNAV_PAL20110127, 4 in RNAV_PAL20110217, and 3 in RNAV_PAL20110303. Of those clusters, four were found in at least one other sample at 100% amino acid (aa) identity, and 7 were found at ≥ 94% aa identity.

Phylogenetic analysis of the translated RdRp genes from individual reads and assembled contigs revealed that most of the identified picornavirad-like sequences clustered around other environmental RdRp sequences and diatom-infecting virus isolates, and were only distantly related to other cultivated isolates (Fig. 2.4). These phylotypes were clustered based on a 98% aa identity, and the relative proportion of reads from each library that comprise that sequence or contig assembly are shown as pies or bars, respectively (Fig. 2.4). Many of the identified

phylotypes consist of sequences from library RNAV_PAL20101113, and most contigs were assembled with sequences from two or more libraries. None of the RdRp sequences clustered with other known viruses from the *Dicistroviridae*, *Secoviridae*, *Picornaviridae*, *Iflaviridae*, or *Labyrnaviridae*. The RdRp from contig PAL473 clustered most closely with sequences from two pennate diatom-infecting viruses (CcloRNAV01 and 02, bootstrap value 96). The RdRp from PAL156 clustered most closely with the RdRp of a centric diatom-infecting virus in the same group with one of the bacillarnaviruses (Csp03RNAV), while PAL128 and PAL438 only clustered with other environmental sequences from both temperate and tropical coastal seawater. The PAL_E4 RdRp formed a deeply branching cluster with other RdRp contigs from the total library assemblies that was not similar to any known isolates.

2.5 DISCUSSON

2.5.1 RNA virus abundance

Our estimates of the numerical contribution of RNA viruses to the virioplankton in the polar waters of the Western Antarctic Peninsula were similar to those previously reported for tropical waters (Steward et al 2013). The two study sites are geographically distant, have large differences in temperature and seasonality of light flux, and differ in the variable limiting primary production (nutrients vs. light). Despite these differences, both study sites are coastal habitats prone to blooms of eukaryotic plankton, including a large contribution from diatoms. It therefore remains unclear whether a high relative abundance of RNA viruses is a general phenomenon in the ocean or is restricted to such coastal habitats.

The evidence available so far suggests that marine planktonic viruses containing RNA infect eukaryotes, while those containing DNA are dominated by viruses infecting prokaryotes (Edwards and Rohwer 2005, Kristensen et al 2010, Lang et al 2009). Given the strong seasonality in the community composition in polar waters, with blooms of eukaryotic phytoplankton in the summer transitioning to bacterial heterotrophy through the long dark winter (Ducklow et al 2007, Ducklow et al 2012, Steinberg et al 2012), we hypothesized that there would be a seasonal cycle in the relative abundance of RNA viruses. Our data are consistent with this hypothesis, with the lowest relative abundance of RNA viruses occurring in the earliest spring sampling and the highest in mid-summer, but there is considerable variability among the

samples. Analysis of additional samples over multiple seasons and at least one complete annual cycle will be necessary to properly test this idea.

2.5.2 Diversity of RNA viruses

A phylogenetic comparison of RdRp genes found in the Antarctic metagenomes to other known RdRps revealed that most of these fall within a rather broad cluster of putative protistan-infecting viruses similar to the results from tropical waters (Culley and Steward 2007). Since viruses within the order *Picornavirales* tend to cluster by the hosts that they infect (Culley and Steward 2007) the environmental sequences most closely related to diatom-infecting viruses likely represent putative diatom-infecting clades. This is consistent with the predominance of diatoms in the early bloom and suggests that viral infection may contribute significantly to the seasonal phytoplankton dynamics in this region. Some of the RdRp phylotypes were more divergent and did not cluster with any other known viruses, yet appear to be persistent throughout the season and represent a significant abundance in the RNA virus community. A concerted effort to isolate more virus-protist systems from diverse ocean habitats may be helpful in placing novel viruses like these into an ecological context.

A greater diversity was found in the beginning of the season (RNAV_PAL20101113), as suggested by rarefaction curves and taxonomy, and may be explained by a more diverse community of hosts prior to the phytoplankton bloom. The diversity then appears to fluctuate minimally throughout the season, with a greater dominance of *Bacillarnavirus* groups in libraries, RNAV_PAL20101218, RNAV_PAL20110107, and RNAV_PAL20110217 in particular. Many of the RdRp phylotypes are represented in multiple metagenomes (four of which had 100% aa identity), which suggests a persistence of some virus types throughout the season. There may also be a succession of different phylotypes infecting similar hosts, as seen with previous virus-host bloom dynamics in mesocosm experiments (Schroeder et al 2003). However, the resolution of sampling and taxonomic classifications does not allow us to draw conclusions about whether variations in community composition we observed represent meaningful ecological shifts or simply reflect sampling variability.

2.5.3 Assembled genomes

The five nearly complete genomes assembled from the libraries represent a significant contribution to the number of marine picornavirad genome sequences available. However, without any host data, we can only assess potential ecological significance through phylogenetic analyses. The genomic configuration and RdRp phylogeny of PAL128, PAL156, and PAL473 are most similar to the viruses within the genus *Bacillarnavirus*, a diatom-infecting group of viruses, again suggesting the importance of these viruses in influencing the dynamics of their dominant diatom hosts. PAL438 and PAL_E4 also contain dicistronic genome structures, but contain RdRps that are highly divergent from any known picornavirads. Although the second ORF of PAL438 does not contain any genes similar to the structural genes in the NCBI database, and the syntenous regions of the genome were slightly smaller than the others, this contig was verified with two sets of primers overlapping both ORFs. Thus, this may be a very divergent type of virus with unique structural proteins since there are currently no examples of this genome configuration within the order *Picornavirales*. None of the genomes contained sequences with significant hits to any proteases, a similar finding to that reported by Culley et al (2014) where only two of the six genomes contained sequences with hits to a protease. However, since this gene is required for +ssRNA viral replication and an ORF exists in the location where a protease is normally encoded, we presume these viruses contain a highly divergent protease.

The number of reads that recruit to a given contig can provide some insight to the relative abundance of the assembled contig and the composition of the RNA viral community at a given time. These data suggest that in January (the dates of collection for libraries RNAV_PAL20110107 and RNAV_PAL20110127), PAL156 may represent nearly half of the total RNA viruses. PAL_E4 on the other hand, may represent nearly half of the total RNA viruses during all other sampling periods of the season. The high similarity of PAL_E4 to other contigs assembled from all of the individual library assemblies suggests that this phylotype was likely present throughout the season. The other contigs, PAL128, PAL438 and PAL473, recruited less than 1% of the reads from each of the metagenomes, suggesting that they do not represent a very large proportion of the viruses at any given sample date throughout the season. These estimates based on recruitment are subject to biases from metagenomic library

construction through cDNA synthesis or PCR amplification (Culley et al 2014) and would need to be confirmed through reverse-transcription quantitative-PCR (RT-qPCR).

2.5.4 RNA viral metagenomics

An overwhelming majority of the sequences in our metagenomes prepared from Antarctic marine RNA viruses matched +ssRNA virus groups identified within the order *Picornavirales*, a finding very similar to those of studies in temperate (Culley et al 2006) and tropical (Culley et al 2014, Steward et al 2013) systems. A comparison of the metagenomes in this study with the Kaneohe Bay RNA viral metagenomes (Culley et al 2014) revealed similarities among the communities. In both cases, the sequences were either dominated by those most similar to viruses or had "no hits" in the database, with less than 4% of sequences matching cellular organisms, suggesting that our multi-step purification procedure was highly selective for viruses. Of the viral sequences, the majority matched +ssRNA viruses within the order *Picornavirales*, and very few matched dsRNA viruses (0.04% at Palmer Station and 1.3% in Kaneohe Bay). Differences among the metagenomes were apparent within the ssRNA virus classifications (Fig. 2.5). The Kaneohe Bay samples had a much larger representation of *Dicistroviridae* and *Marnaviridae*, as might be expected in a coastal habitat with a high abundance of copepods and more terrestrial input and runoff. The Palmer Station metagenome had a higher percentage of *Bacillarnavirus*-like sequences, representing over 90% of sequences in three of the six samples (Table 2.3). Overall, fewer RdRp sequences were retrieved from the Antarctic samples, so it was difficult to provide a significant estimate of diversity using standard metrics.

2.6 CONCLUSIONS

We have shown the relative abundance of RNA viruses to be equal to or greater than DNA viruses at multiple times, first in tropical (Steward et al 2013) and now polar waters, using two different methods of virus concentration (iron chloride flocculation in the previous study vs TFF in the present study). When considered on a broad taxonomic level, the composition of metagenomic libraries of RNA viruses varied little over the course of the Austral summer in the coastal waters of Palmer Station, Antarctica. Though some phylotypes may dominate certain periods of the bloom, neither the depth nor resolution of sampling is sufficient to conclude that there were significant, systematic changes through the course of the season.

The dominance of +ssRNA picronavirad-like sequences in marine RNA virus metagenomes appears to be a general phenomenon, having been observed in all metagenomes from temperate, tropical, and now polar waters. This suggests that picornavirads are common in the marine environment, but the possibility remains that methodological biases in library preparation could account for their pronounced dominance.

2.7 ACKNOWLEDGEMENTS

We thank and the Palmer LTER group for assistance with sampling and logistics. In particular, Alice Alpert, Jennifer Brum, Kaycee Coleman, Carolina Funkey, Mike Garzio, and Edgar Woznica provided invaluable assistance with collecting samples under often challenging conditions. We also thank Mahdi Belcaid and Guylaine Poisson at the Bioinformatics Laboratory at the University of Hawai‘i at Mānoa for the use of their computer cluster for sequence assembly and BLAST searches. This work was supported by NSF grants to GFS and AIC (ANT 09-44851) and by NSF funding to the Center for Microbial Oceanography: Research and Education (EF 04-24599).

Table 2.1. Relative abundance of RNA viruses in seawater samples collected at Station B throughout the 2010-2011 season. The calculated percent contributions of RNA viruses are relative to the total. Ranges are extreme low and high estimates as described in the Methods section.

			Virus genome copies ($\times 10^9$)		% RNA Viruses	
Sample name[a]	**Collection date**	**Volume (L)**	**RNA**	**DNA**	**Mean**	**Range**
RNAV_PAL20101113	13 November 2010	150	1.3	15	8	2-15
RNAV_PAL20101218	18 December 2010	50	3.7	2.1	64	29-78
RNAV_PAL20110107	7 January 2011	45	5.5	7.2	43	15-60
RNAV_PAL20110127	27 January 2011	50	5.9	3.2	65	30-79
RNAV_PAL20110217	17 February 2011	45	3.0	6.8	30	9-47
RNAV_PAL20110303	3 March 2011	45	3.8	4.3	47	17-64

[a]Sample name indicates the sample name given to the metagenome in the GenBank Sequence Read Archive (SRA).

Table 2.2. Assembly data of the six viral metagenomes collected throughout the season. Listed are the number and length (nucleotides, nt) of initial quality controlled reads and of contigs after assembly, percent GC content, and the percentage of reads that were assembled to form a contig.

			Length (nt)				G+C	Assembled
Sample date (YYYYMMDD)		**Count**	**Mean**	**Min**	**Max**	**Total nt**	**(%)**	**(%)**
20101113	Reads	57,065	429	100	1,016	24,481,035	43	86
	Contigs	1,020	765	105	7,740	779,973		
20101218	Reads	74,580	412	100	708	30,749,857	45	90
	Contigs	248	663	153	3,996	164,395		
20110107	Reads	76,980	431	100	744	33,201,674	45	98
	Contigs	500	675	71	7,843	337,434		
20110127	Reads	78,321	416	100	857	32,551,951	43	96
	Contigs	430	697	70	5,752	299,824		
20110217	Reads	19,150	428	100	910	8,196,870	43	94
	Contigs	316	723	193	8,187	228,313		
20110303	Reads	59,806	415	100	1033	24,843,067	45	95
	Contigs	138	668	100	4,943	92,132		
Total	Reads	365,902	417	100	1 033	152,560,846	43	96
	Contigs	1 794	698	88	8,117	1,253,203		

Table 2.3. Distribution of hits to viral sequences for the individual sample libraries or a library representing all sequences combined (Total). In each case, BLAST analyses were done on assembled sequences. The combined library was assembled separately after combining the sequences. Values shown are percentages of viral hits. Assembled contigs and singletons were assigned based on comparison to the NCBI database using the BLASTx algorithm and an e-value of $\leq 10^{-5}$. Sequences were assigned to the Unclassified ssRNA or *Picornavirales* if they matched ssRNA viruses or members of the order *Picornavirales* that could not be further classified. Uncultured *Picornavirales* includes environmental sequences of marine picorna-like viruses JP-A or JP-B.

				Sample date							
				YYYY	**2010**		**2011**				
Group	**Order**	**Family**	**Genus**	**MMDD**	**1113**	**1218**	**0107**	**0127**	**0217**	**0303**	**Total**
ssRNA viruses	*Picornavirales*	*Dicistroviridae*			1.01	0	0.04	0.04	0	0	0.30
		Marnaviridae			0	0	0	0	0	0	0.14
		Picornaviridae			0	0	0	0	0	0	0.04
		Unassigned	*Labyrnavirus*		0.41	0	0	0	0	0	0.14
			Bacillarnavirus		13.8	93.9	97.5	52.3	90.9	55.4	43.8
		Unclassified			47.7	3.96	0.15	41.3	0.60	43.4	10.2
		Uncultured			22.9	0.39	1.42	0.49	2.83	0.83	43.2
	Unclassified				14.1	1.75	0.78	5.75	5.65	0.35	2.06
dsRNA viruses					0	0	0.01	0.05	0.02	0	0.04
dsDNA viruses					0	0	0	0	0	0	0.01
All other viruses					0.11	0.06	0.08	0.07	0.01	0.01	0.15

Table 2.4. Summary of assembled genomes. Listed are the estimated minimum genome size, percent GC content, average coverage, percent of reads mapped from the pooled metagenomes, sizes of open reading frames (ORFs), untranslated regions (UTRs), and intergenic regions (IGR), and whether or not a poly(A) tail was identified. The * indicates numbers relative to only library RNAV_PAL20110217, as this genome was assembled with only those reads.

Genome	Size (bp)	% GC	Average coverage of reads	% Reads mapped	ORF 1 size (bp)	ORF 2 size (bp)	5' UTR	IGR	3' UTR	% UTR	Poly (A) tail
PAL128	8,660	42.2	81	0.4	5,574	2,397	195	67	427	8.0	ND
PAL156	7,897	41.9	2,258	12.1	4,764	2,640	69	315	109	6.2	ND
PAL438	7,109	39.2	29	0.1	4,623	1,872	5	223	324	7.8	Y
PAL473	6,360	42.7	25	0.1	3,153	2,700	104	200	203	8.0	Y
PAL_E4	8,187	46.6	604*	52.9	4,980	2,748	65	133	261	5.6	ND

Table 2.S1. List of sequences and their accession numbers that were used to make the RdRp alignment and tree in Figure 4.

Virus Abbreviation	Full Name	Accession Number(s)	Family	Genus
ABPV	Acute bee paralysis virus	NC_002548	Dicistroviridae	Aparavirus
AEV	Avian encephalomyelitis virus	NC_003990	Picornaviridae	Tremovirus
AglaRNAV	Asterionellopsis glacialis RNA virus	NC_024489.1	Picornavirales	Bacillarnavirus
AiV	Aichi virus	NC_001918	Picornaviridae	Kobuvirus
PAL_A_A278J	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_A_A4PXB	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_A_ASLE9	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_A_B83H6	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_A_BHI98	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_A_BHRWM	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_A_BK0WV	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_A_BXE1V	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122	Unassigned	Unassigned
PAL_C_A0RCR	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648128	Unassigned	Unassigned
PAL_C_BDXWP	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648128	Unassigned	Unassigned
PAL_D_A6VSM	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648129	Unassigned	Unassigned
PAL_E_B03M2	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648130	Unassigned	Unassigned
PAL_F_AZ1QD	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648131	Unassigned	Unassigned
PAL_F_CE2C5	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648131	Unassigned	Unassigned
ASV	Avian sapelovirus	NC_006553	Picornaviridae	Sapelovirus
AuRNAV	Aurantiochytrium single-stranded RNA virus	NC_007522	Unassigned	Labyrnavirus
B	2003 RdRp assembled from Straight of Georgia, BC metagenome	AAQ21147	Unassigned	Unassigned
BBWV	Broad bean wilt virus 1	NC_005289	Secoviridae	Fabavirus
C	2003 RdRp assembled from Straight of Georgia, BC metagenome	AAQ21158	Unassigned	Unassigned
CcloRNAV01	Cylindrotheca closterium RNA Virus 01	NA	Unassigned	Unassigned
CcloRNAV02	Cylindrotheca closterium RNA Virus 02	NA	Unassigned	Unassigned
CPMV	Cowpea mosaic virus	NC_003549	Secoviridae	Comovirus
CRLV	Cherry rasp leaf virus	NC_006271	Secoviridae	Cheravirus
CrPV	Cricket paralysis virus	NC_003924	Dicistroviridae	Cripavirus
CsfrRNAV	Chaetoceros scoialis f. radians RNA Virus	NC_012212	Unassigned	Bacillarnavirus
Csp03RNAV	Chaetoceros sp. SS08-C03 RNA virus	AB63940	Unassigned	Bacillarnavirus
CtenRNAV	Chaetoceros tenuissimus RNA Virus	AB375474	Unassigned	Bacillarnavirus
D	2003 RdRp assembled from Straight of Georgia, BC metagenome	AAQ21156	Unassigned	Unassigned
DCV	Drosophila C virus	NC_001834	Dicistroviridae	Cripavirus
DHBV	Duck hepatitis A virus 1	NC_008250	Picornaviridae	Avihepatovirus
EMCV	Encephalomyocarditis virus	NC_001479	Picornaviridae	Cardiovirus
ERBV	Equine rhinitis B virus 1	NC_003983	Picornaviridae	Erbovirus
FMDV	Foot-and-mouth disease virus A	NC_011450	Picornaviridae	Aphthovirus
HaRNAV	Heterosigma akashiwo RNA virus	NC_005281	Marnaviridae	Marnavirus

Table 2.S1. (continued) List of sequences and their accession numbers that were used to make the RdRp alignment and tree in Figure 4.

Virus Abbreviation	Full Name	Accession Number(s)	Family	Genus
HAV	Hepatitis A virus	NC_001489	Picornaviridae	Hepatovirus
HPeV	Human parechovirus	NC_001897	Picornaviridae	Parechovirus
HRV	Human rhinovirus A	NC_001617	Picornaviridae	Enterovirus
IFV	Infectious flacherie virus	NC_003781	Iflaviridae	Iflavirus
JP-A	Marine RNA virus JP-A	NC_009757	Unassigned	Unassigned
JP-B	Marine RNA virus JP-B	NC_009758	Unassigned	Unassigned
JP11	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP16	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP162	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP1743	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP20	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP32	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP5	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP6	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP62	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
JP9	2006 RdRp assembled from Straight of Georgia, BC metagenome	DX420985-DX421142	Unassigned	Unassigned
KB1	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29304	Unassigned	Unassigned
KB10	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29313	Unassigned	Unassigned
KB11	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29314	Unassigned	Unassigned
KB12	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29315	Unassigned	Unassigned
KB13	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29316	Unassigned	Unassigned
KB14	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29317	Unassigned	Unassigned
KB15	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29318	Unassigned	Unassigned
KB16	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29319	Unassigned	Unassigned
KB18	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29321	Unassigned	Unassigned
KB19	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29322	Unassigned	Unassigned
KB2	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29305	Unassigned	Unassigned
KB20	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29323	Unassigned	Unassigned
KB2009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_1363	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned

Table 2.S1. (continued) List of sequences and their accession numbers that were used to make the RdRp alignment and tree in Figure 4.

Virus Abbreviation	Full Name	Accession Number(s)	Family	Genus
KB2009_con_16	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_165	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_1978	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_202	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_2048	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_224	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_243	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_2732	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_2873	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_361	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_40	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_411	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_45	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_505	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_68	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_708	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con_90	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con15	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con28	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con74	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_con88	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_sing_1178	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_sing_16692	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_sing_21981	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_sing_33009	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2009_sing_41282	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2010	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010_con_161	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010_con_191	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2010_con_26	2009 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2010_con_391	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010_con_59	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010_con_61	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000815	Unassigned	Unassigned
KB2010_con_780	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010_con_883	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010_sing_GNV001E03G6KQO	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned

Table 2.S1. (continued) List of sequences and their accession numbers that were used to make the RdRp alignment and tree in Figure 4.

Virus Abbreviation	Full Name	Accession Number(s)	Family	Genus
KB2010_sing_GNV 001E03G9ND2	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB2010_sing_GNV 001E03GSDFI	2010 RdRp assembled from Kaneohe Bay, HI	CAM_SMPL_000824	Unassigned	Unassigned
KB21	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29324	Unassigned	Unassigned
KB22	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29325	Unassigned	Unassigned
KB23	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29326	Unassigned	Unassigned
KB3	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29306	Unassigned	Unassigned
KB4	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29307	Unassigned	Unassigned
KB5	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29308	Unassigned	Unassigned
KB6	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29309	Unassigned	Unassigned
KB8	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29311	Unassigned	Unassigned
KB9	2007 RdRp amplicon from Kaneohe Bay, HI	ABV29312	Unassigned	Unassigned
MB1	2007 RdRp amplicon from Monterey Bay, CA	ABV29327	Unassigned	Unassigned
PAL104	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL107	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL128	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL156	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL_E4/PAL19	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL250	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL307	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL379	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL396	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL414	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL438	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL473	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL609	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL666	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL667	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PAL92	2010 RdRp assembled from Palmer Station, Antarctica	SRR1648122, SRR1648128-31	Unassigned	Unassigned
PTV	Porcine teschovirus 1	NC_003985	Picornaviridae	Teschovirus
PV	Human enterovirus C	NC_002058	Picornaviridae	Enterovirus
PYFV	Parsnip yellow fleck virus	NC_003628	Secoviridae	Sequivirus
RsRNAV	Rhizoselenia setigera RNA Virus	AB243297	Unassigned	Bacillarnavirus
RTSV	Rice tungro spherical virus	NC_001632	Secoviridae	Waikavirus
Sc1Fr13012617R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC620977	Unassigned	Unassigned

Table 2.S1. (continued) List of sequences and their accession numbers that were used to make the RdRp alignment and tree in Figure 4.

Virus Abbreviation	Full Name	Accession Number(s)	Family	Genus
Sc1Fr13012734R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC620983	Unassigned	Unassigned
Sc1Fr13020144R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC620986	Unassigned	Unassigned
Sc1Fr18012906R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC620996	Unassigned	Unassigned
Sc1Fr18012916R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC620998	Unassigned	Unassigned
Sc1Fr18012919R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC621001	Unassigned	Unassigned
Sc1Fr18012925R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC621003	Unassigned	Unassigned
Sc1Fr18020211R	RdRpsequence amplified with MplSc1 primers from 1.49 g/ml density fraction of 2009 virome	KC621009	Unassigned	Unassigned
Sc2Fr13020102R	RdRpsequence amplified with MplSc2 primers from 1.38 g/ml density fraction of 2009 virome	KC621013	Unassigned	Unassigned
Sc2Fr13020109R	RdRpsequence amplified with MplSc2 primers from 1.38 g/ml density fraction of 2009 virome	KC621020	Unassigned	Unassigned
Sc2Fr13020110R	RdRpsequence amplified with MplSc2 primers from 1.38 g/ml density fraction of 2009 virome	KC621021	Unassigned	Unassigned
Sc2Fr13020116R	RdRpsequence amplified with MplSc2 primers from 1.38 g/ml density fraction of 2009 virome	KC621024	Unassigned	Unassigned
SDV	Satsuma dwarf virus	NC_003785	Secoviridae	Sadwavirus
SVV	Seneca Valley virus	NC_011349	Picornaviridae	Senecavirus
TspRNAV01	Thalassiosira sp. RNA virus	NA	Unassigned	Unassigned
ToRSV	Tobacco ringspot virus 1	NC_005097	Secoviridae	Nepovirus
ToTV	Tomato torrado virus	NC_009013	Secoviridae	Torradovirus
TSV	Taura syndrome virus	NC_003005	Dicistroviridae	Aparavirus

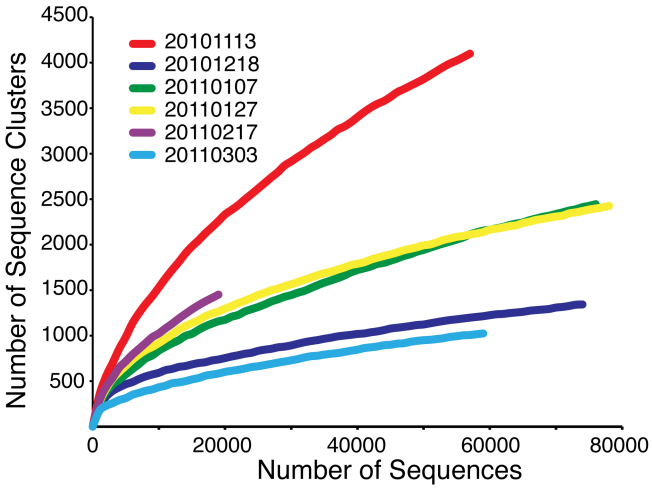


Figure 2.1. Rarefaction curves of all six metagenomes collected throughout the season at Station B as determined by Metavir (Roux et al 2011). Libraries are designated by the color indicated in the legend that corresponds to each collection date (YYYYMMDD). In this analysis, sequence reads from each library were plotted as a function of the number of unique clusters (threshold = 75%) identified.

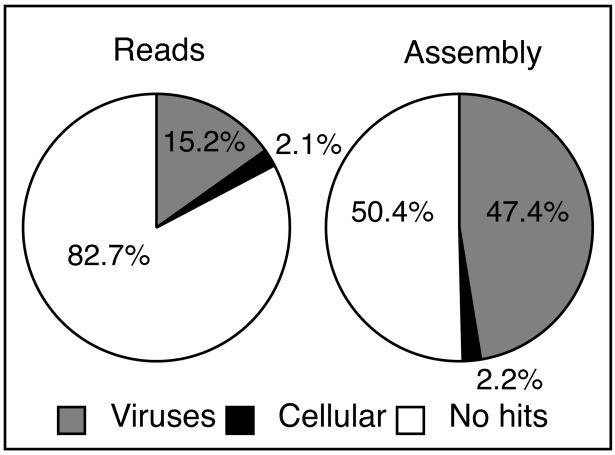


Figure 2.2. Taxonomic classification of reads and assembly from all six RNA virus metagenomes combined. Domain-level classification of sequence hits for quality filtered reads (left) and assembled contigs and singletons (right). Individual sequence reads (Reads) or assembled contigs plus singletons (Assembly) were assigned based on comparison to the NCBI nr database using the BLASTx algorithm and a threshold *E* value of $\leq 10^{-5}$.

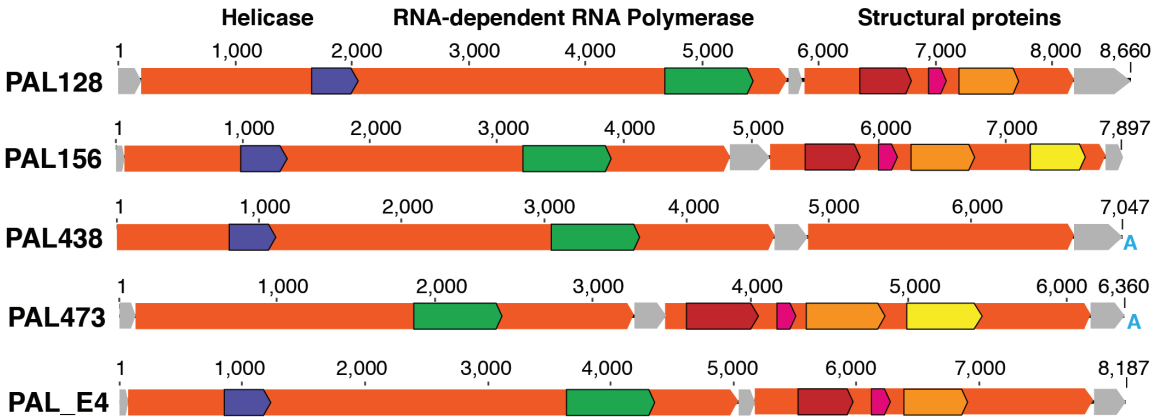


Figure 2.3. Genome maps of contigs PAL128, PAL156, PAL438, PAL473, and PAL_E4. The first four genomes were assembled from the pooled reads of all six metagenomes, and the last was assembled from reads in library RNAV_PAL20110217. The conserved gene regions were identified by the CDD NCBI BLAST searches. The helicase genes are shown in purple, the RNA-dependent RNA polymerase genes in green, and the structural genes in red, pink, light orange and yellow. Untranslated regions are shown in light grey and the ORFs are shown in orange. The light blue A indicates where a polyadenylated tail was identified.

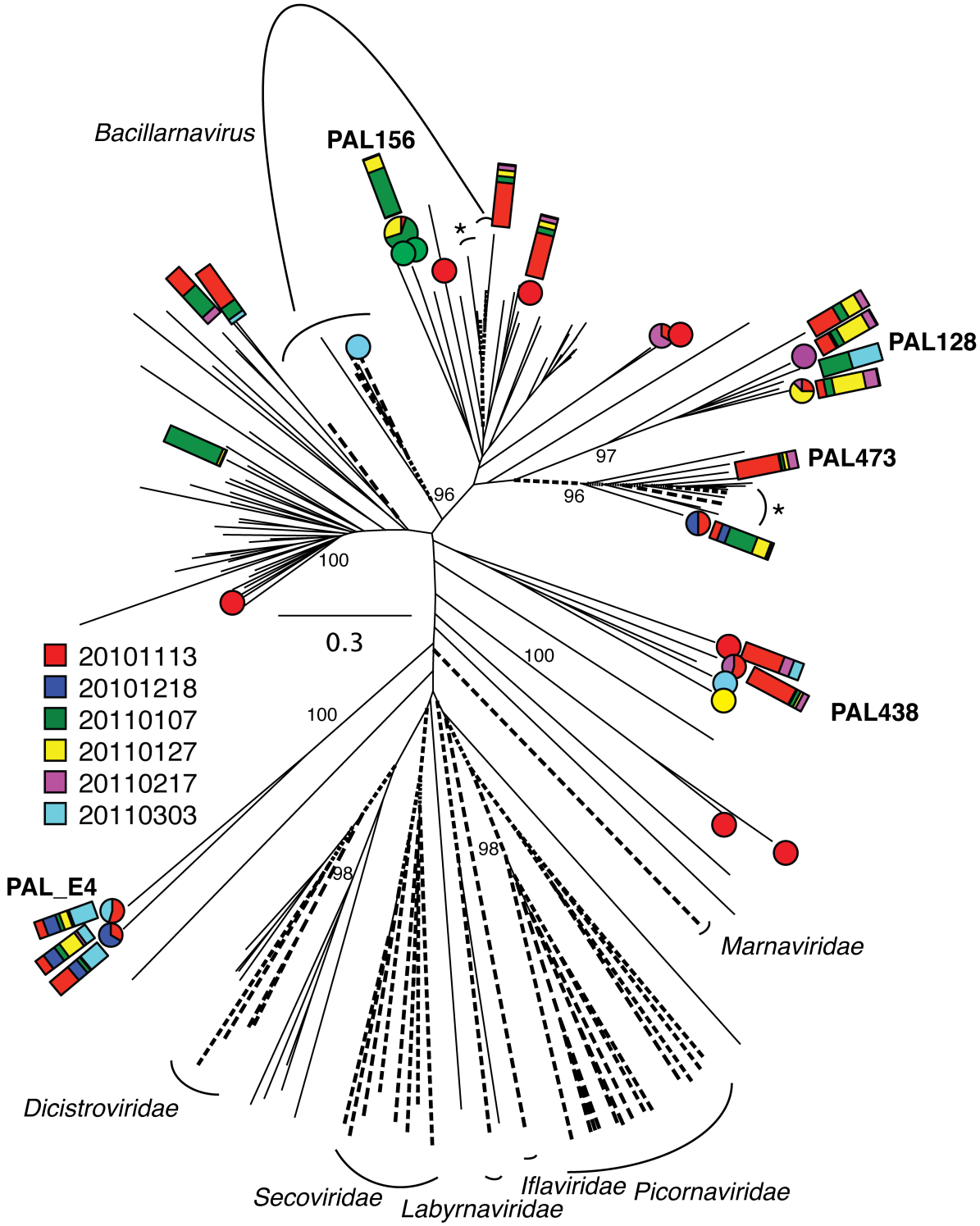


Figure 2.4. RdRp phylotypes. An unrooted phylogenetic analysis (maximum likelihood; model WAG; 1000 bootstraps) of the picorna-like region of the RdRp gene showing the position of environmental sequences from this study (marked with pies and bars) and previous studies (Culley and Steward 2003: Culley et al., 2006, 2007, 2014), relative to known viruses in the order *Picornavirales*. The branches of cultured isolates are shown with a dashed line, and the

pies (reads) and bars (contigs) represent phylotype groups (98% aa identity) from this study and indicate the relative number of reads from each sample represented in that sequence or contig, respectively. Libraries are designated by the color indicated in the legend that corresponds to each collection date (YYYYMMDD). The asterisk (*) indicates unpublished diatom-infecting viruses isolated from Kaneohe Bay. PAL128, PAL156, PAL438, PAL473, and PAL_E4 indicated on the tree represent the five nearly complete genomes assembled in this study. The scale bar is equal to 0.3 substitutions per site. The specific sequences from the alignment are listed in Table 2.S1.

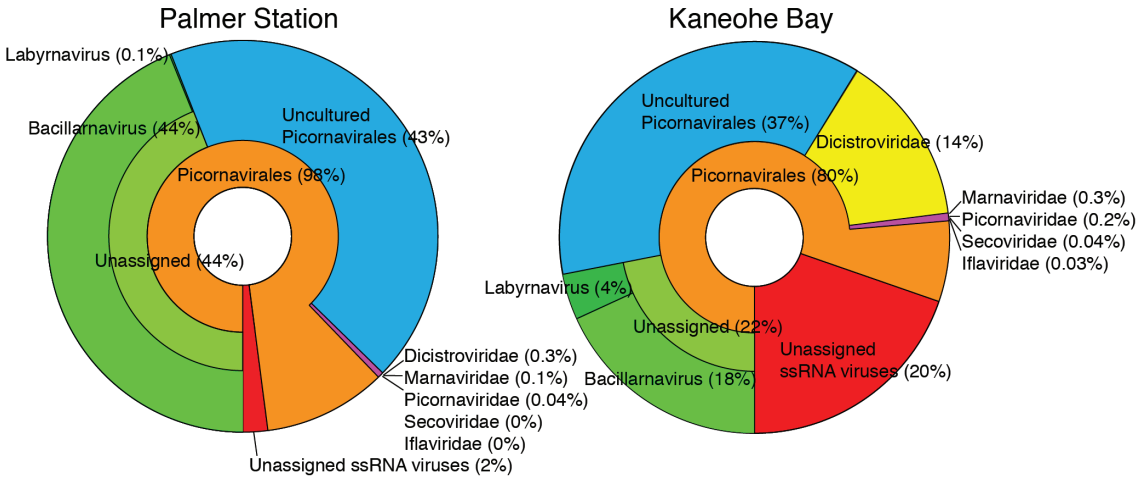


Figure 2.5. Taxonomic classification of assemblies from Palmer Station metagenomes compared with Kaneohe Bay metagenomes (Culley et al 2014). Percentages are relative to the total ssRNA virus hits for the pooled library assembly from Palmer Station and for the combined data from Kaneohe Bay 2009 and 2010 metagenomes. Krona charts were made with Krona (Ondov et al 2011).

2.8 REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of molecular biology* **215:** 403-410.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C *et al* (2006). The Marine Viromes of Four Oceanic Regions. *PLoS Biology* **4:** e368.

Bratbak G, Egge J, Heldal M (1993). Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series* **93:** 39-48.

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D *et al* (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences* **99:** 14250-14255.

Burkill PH, Edwards ES, Sleight MA (1995). Microzooplankton and their role in controlling phytoplankton growth in the marginal ice zone of the Bellingshausen Sea. *Deep Sea Research Part II: Topical Studies in Oceanography* **42:** 1277-1290.

Calbet A, Alcaraz M, Atienza D, Broglio E, Vaqué D (2005). Zooplankton biomass distribution patterns along the western Antarctic Peninsula (December 2002). *Journal of Plankton Research* **27:** 1195-1203.

Chen F, Suttle C (1995). Amplification of DNA polymerase gene fragments from viruses infecting microalgae. *Applied and Environmental Microbiology* **61:** 1274-1278.

Culley AI, Lang AS, Suttle CA (2003). High diversity of unknown picorna-like viruses in the sea. *Nature* **424:** 1054-1057.

Culley AI, Lang AS, Suttle CA (2006). Metagenomic analysis of coastal RNA virus communities. *Science* **312:** 1795-1798.

Culley AI, Steward GF (2007). New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Applied and Environmental Microbiology* **73:** 5937-5944.

Culley AI, Mueller JA, Belcaid M, Wood-Charlson EM, Poisson G, Steward GF (2014). The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *MBio* **5**.

Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N *et al* (2008). Viral genome sequencing by random priming methods. *BMC Genomics* **9**.

Ducklow HW, Baker K, Martinson DG, Quetin LB, Ross RM, Smith RC *et al* (2007). Marine pelagic ecosystems: the West Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362:** 67-94.

Ducklow HW, Schofield O, Vernet M, Stammerjohn S, Erickson M (2012). Multiscale control of bacterial production by phytoplankton dynamics and sea ice along the western Antarctic Peninsula: A regional and decadal investigation. *Journal of Marine Systems* **98–99:** 26-39.

Duhaime MB, Deng L, Poulos BT, Sullivan MB (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environmental Microbiology* **14:** 2526-2537.

Dunigan DD, Fitzgerald LA, Van Etten JL (2006). Phycodnaviruses: a peek at genetic diversity. *Virus Research* **117:** 119-132.

Eddy SR (1998). Profile hidden Markov models. *Bioinformatics* **14:** 755-763.

Edwards RA, Rohwer F (2005). Viral metagenomics. *Nature Reviews Microbiology* **3:** 504-510.

Evans C, Brussaard CP (2012). Regional variation in lytic and lysogenic viral infection in the Southern Ocean and its contribution to biogeochemical cycling. *Applied and Environmental Microbiology* **78:** 6741-6748.

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

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307-321.

Guixa-Boixereu N, Vaqué D, Gasol JM, Sánchez-Cámara J, Pedrós-Alió C (2002). Viral distribution and activity in Antarctic waters. *Deep Sea Research Part II: Topical Studies in Oceanography* **49:** 827-845.

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

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome research* **21:** 1552-1560.

Katoh K, Kuma K-i, Toh H, Miyata T (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33:** 511-518.

Koonin EV, Wolf YI, Nagasaki K, Dolja VV (2008). The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nature Reviews Microbiology* **6:** 925-939.

Koressaar T, Remm M (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23:** 1289-1291.

Kristensen DM, Mushegian AR, Dolja VV, Koonin EV (2010). New dimensions of the virus world discovered through metagenomics. *Trends in Microbiology* **18:** 11-19.

Labonte JM, Suttle CA (2013). Previously unknown and highly divergent ssDNA viruses populate the oceans. *Isme Journal* **7:** 2169-2177.

Lang AS, Rise ML, Culley AI, Steward GF (2009). RNA viruses in the sea. *FEMS Microbiology Reviews* **33:** 295-323.

Lawrence JE, Steward GF (2010). Purification of viruses by centrifugation. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. Aquatic Sciences Limnology and Oceanography: Waco, TX. pp 166-181.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C *et al* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* **39:** D225-D229.

Moline MA, Claustre H, Frazer TK, Schofield O, Vernet M (2004). Alteration of the food web along the Antarctic Peninsula in response to a regional warming trend. *Global Change Biology* **10:** 1973-1980.

Nagasaki K, Ando M, Itakura S, Imai I, Ishida Y (1994). Viral mortality in the final stage of Heterosigma akashiwo (Raphidophyceae) red tide. *Journal of Plankton Research* **16:** 1595-1599.

Ondov BD, Bergman NH, Phillippy AM (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12:** 385.

Rosario K, Breitbart M (2011). Exploring the viral world through metagenomics. *Current opinion in virology* **1:** 289-297.

Ross RM, Quetin LB, Martinson DG, Iannuzzi RA, Stammerjohn SE, Smith RC (2008). Palmer LTER: Patterns of distribution of five dominant zooplankton species in the epipelagic zone west of the Antarctic Peninsula, 1993–2004. *Deep Sea Research Part II: Topical Studies in Oceanography* **55:** 2086-2105.

Roux S, Faubladier M, Mahul A, Paulhe N, Bernard A, Debroas D *et al* (2011). Metavir: a web server dedicated to virome analysis. *Bioinformatics* **27:** 3074-3075.

Schmieder R, Edwards R (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27:** 863-864.

Schroeder DC, Oke J, Hall M, Malin G, Wilson WH (2003). Virus succession observed during an *Emiliania huxleyi* bloom. *Applied and Environmental Microbiology* **69:** 2484-2490.

Steinberg D, Martinson D, Costa D (2012). Two decades of pelagic ecology of the Western Antarctic Peninsula. *Oceanography* **25:** 56-67.

Steward GF, Wikner J, Cochlan WP, Smith DC, Azam F (1992). Estimation of virus production in the sea: II. Field results. *Marine Microbial Food Webs* **6:** 79-90.

Steward GF, Montiel JL, Azam F (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnology and Oceanography* **45:** 1697-1706.

Steward GF, Culley AI (2010). Extraction and purification of nucleic acids from viruses. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of aquatic virology*. American Society of Limnology and Oceanography: Waco, TX. pp 154-165.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G (2013). Are we missing half of the viruses in the ocean? *Isme Journal* **7:** 672-679.

Suttle CA (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology* **5:** 801-812.

Tomaru Y, Nagasaki K (2011). Diatom Viruses. In: Seckbach J, Kociolek P (eds). *The Diatom World*. Springer Netherlands. pp 211-225.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M *et al* (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research* **40:** e115.

Van Etten JL, Lane LC, Meints RH (1991). Viruses and viruslike particles of eukaryotic algae. *Microbiological reviews* **55:** 586-620.

Van Etten JL, Graves MV, Müller DG, Boland W, Delaroque N (2002). Phycodnaviridae - large DNA algal viruses. *Archives of Virology* **147:** 1479-1516.

Weinbauer MG (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28:** 127-181.

Chapter 3

Picornavirales dominate RNA viral metagenomes along the Larsen A Ice Shelf, Antarctica

Jaclyn A. Mueller

Grieg F. Steward

Center for Microbial Oceanography: Research and Education, Department of Oceanography, University of Hawai‘i at Mānoa, 1950 East-West Road, Honolulu, Hawai‘i 96822

3.1 ABSTRACT

Marine RNA viruses appear to predominately infect eukaryotes and have been shown to represent diverse communities that reach abundances equivalent to DNA containing bacteriophage. Metagenomic studies suggest that these viruses are dominated by positive-sense, single-stranded RNA (+ssRNA) genomes that are phylogenetically related to protist-infecting viruses in the order *Picornavirales*. These trends have been shown in tropical and temperate waters, and more recently coastal polar waters along the Western Antarctic Peninsula (WAP). We hypothesized that diatom-infecting viruses would similarly dominate the RNA virus communities on the eastern side of the peninsula, yet be distinct from those found in the WAP, as the diatom host communities are different. In this study, we investigated the RNA virus community along a transect in the Larsen A Ice Shelf system on the Eastern Antarctic Peninsula. Our metagenomic analyses confirm our previous findings that most polar RNA viruses have +ssRNA genomes most closely related to viruses in the order *Picornavirales*. However, these communities appear much less diverse than those found in the WAP, based on the identified RNA dependent RNA polymerase (RdRp) genes. Assembly of the metagenomic reads resulted in one novel, nearly complete genome with genomic features and RdRp phylogeny similar to diatom-infecting viruses. Our data are consistent with the hypothesis that RNA viruses influence diatom dynamics in Antarctic waters.

3.2 INTRODUCTION

Marine viruses play an integral role in the evolution of life in the sea, influence plankton community structure, and contribute to marine biogeochemical cycling and food web dynamics (Suttle 2007). Most of the double-stranded (ds) DNA-containing viruses in the ocean appear to be bacteriophages (Edwards and Rohwer 2005, Weinbauer 2004), but other types of viruses that infect eukaryotic plankton, including those containing single-stranded (ss) DNA (Angly et al 2006, Hopkins et al 2014, Labonte and Suttle 2013, Rosario and Breitbart 2011) or ss- or dsRNA (Culley et al 2003, Culley et al 2006, Culley and Steward 2007, Steward et al 2013) may also be important contributors to marine plankton ecology. In particular, RNA viruses have been shown to contribute equally in abundance compared with the DNA containing viruses in both coastal tropical and polar seawater samples on multiple occasions (Mueller et al *Submitted,* Steward et al 2013).

The one previous study of RNA viruses in a polar marine habitat was conducted in coastal waters near Palmer Station in the Western Antarctic Peninsula (WAP) over the course of a summer phytoplankton bloom. In that study, metagenomes prepared from RNA in the viral fraction appeared to predominantly derive from positive-sense single-stranded RNA (+ssRNA) viruses most closely related to those in the order *Picornavirales* (Mueller et al *Submitted*), a finding similar to other marine RNA viral metagenome studies (Culley et al 2003, Culley et al 2006, Culley and Steward 2007, Culley et al 2014, Steward et al 2013). However, the Palmer Station study only represented one location in a coastal environment in polar seawater.

In order to assess spatial variability in the RNA virus community, we sampled seawater along a transect in the Larsen A region on the Eastern Antarctic Peninsula (EAP). The Larsen A region is an area of particular interest to oceanographers and polar ecologists, because warming temperatures have resulted in ice shelf collapses that have drastically changed the surrounding ecosystem (Bertolin and Schloss 2009, Domack et al 2005). This loss of ice shelf allowed sunlight to penetrate the surface oceans, increasing primary production and, consequently, the flux of organic material to the sea floor. A few studies have investigated the effects of this ice shelf collapse on the benthic communities and found an increase in chlorophyll and diatom valve abundance in the top 2-3 cm, with negligible concentrations below, suggesting deposition only

after the ice shelves collapsed (Sañé et al 2011a, Sañé et al 2011b). However, no studies to date have looked at the microbial composition or activity in the water column since the collapse of these ice shelves, or investigated the viruses present in this region.

The ecosystem dynamics of the EAP are distinct from those in the WAP, including cooler seawater temperatures and higher sea ice cover in the EAP (Ó Cofaigh et al 2014) and are likely more typical of conditions generally found on the Antarctic shelf. In contrast to the WAP, sea ice retreat appears to play a minor role in stratification of the Larsen embayments, where a relatively shallow mixed layer is persistent even in open water conditions. While the rates of net primary production (NPP) and temporal patterns of bloom dynamics are similar on both sides of the peninsula, the spatial dynamics of production appear asymmetrical. NPP on the eastern side of the peninsula is low along the coast and reaches a maximum 100-150 km offshore, whereas on the western side maximum productivity is typically found along the coast (Cape et al 2014). The phytoplankton communities in the EAP appear to be dominated by large diatoms, which increase in abundance from the coastal margins to the shelf (personal communication, Maria Vernet). Therefore, we expected to see a shift in virus communities along the transect, and assemblages distinct from those found in the WAP (with small diatom and flagellate hosts). This study provides a spatial comparison of the RNA viral community that complements the study in the WAP (Mueller et al *Submitted*), and allows us to make further connections between host and virus dynamics.

3.3 METHODS

3.3.1 Sample collection and processing

Seawater samples were collected on board the *R/V Nathaniel B. Palmer* at Stations G-K along the Larsen A Ice Shelf from 21 March to 6 April 2012 during the NBP1203 cruise (Fig. 3.1). Volumes of ~40 L were collected at 5 m with a Niskin rosette. Seawater samples were 0.2 μm filtered (Sterivex, Millipore) to remove cells, then concentrated via iron chloride flocculation (John et al., 2011). The viral concentrates were then purified in two sequential CsCl gradients (step and continuous; Lawrence and Steward 2010). Fractions of approximately 0.5 ml were collected from top to bottom of the final gradient with an Auto Densi-Flow (Labconco). The density of each fraction was measured using a micropipette and analytical balance. Portions of

those fractions in the range that encompasses the densities of most viruses (1.39-1.51 g ml^{-1}) were pooled and buffer exchanged into SM (0.4 M NaCl, 0.02 M $MgSO_4$, 0.05 M Tris, pH 7.5) with centrifugal ultrafiltration devices (0.5 ml Amicon Ultra 30 kDa; Millipore). Nucleic acids were extracted with the MasterPure Complete DNA and RNA Purification Kit (Epicentre), according to the manufacturer's instructions.

3.3.2 Metagenomic library preparation

After DNase treatment (TURBO DNase, Life Technologies) of the pooled viral fractions, cDNA was synthesized by a random priming sequence-independent single-primer amplification (RP-SISPA) reaction (Culley et al 2010, Djikeng et al 2008), and then amplified via PCR with a reconditioning step, as recommended in Duhaime et al. (2012). For random-priming cDNA synthesis, 5 µl of extracted viral RNA was added to a 0.2 ml PCR tube with 20 pmol FR26RV-N (GCCGGAGCTCTGCAGATATCNNNNNN), 0.1 pmol FR40RV-T (GCCGGAGCTCTGCAGATATC$(T)_{20}$), 10 nmol each dNTP, and a final volume of 10 µl. The reaction was heated to 65 °C for 5 min to disrupt RNA secondary structure, then placed on ice for at least 1 min. While still on ice, 1 × RT Buffer, 100 µmol $MgCl_2$, 0.2 µmol DTT, 40 U RNase OUT, and 200 U Superscript III reverse transcriptase were added to the tube and mixed gently (20 µl final volume). The reactions were incubated sequentially at 25 °C for 10 min to keep random hexamers annealed, 50 °C for 1 hr for reverse transcription, 94 °C for 3 min to denature the product and allow annealing of additional FR26RV-N primer, and then rapidly cooled to 4 °C. A complementary second strand was then synthesized at 37 °C for 1 hr with the addition of 2.5 U Klenow Fragment 3′-5′ exo– (New England Biolabs), and the reaction terminated at 75 °C for 10 min.

Amplification of the cDNA was carried out in 5 replicates of 50 µl reactions containing 2 µl template with a final concentration of 1 × Reaction Buffer, 0.2 mM dNTPs, 1 µM FR20RV primer, and 2.5 U Expand High Fidelity Plus Enzyme (Roche). PCRs were incubated at 94 °C for 2 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 65 °C for 30 s, extension at 68°C for 4 min, and a final extension at 68°C for 7 min. A fraction of each reaction was loaded onto a 1% agarose gel containing 1 × SYBR Safe stain (Life Technologies) and 1 × TAE, and run at 100V for 40 min to ensure that amplification resulted in visible PCR products

(typically a smear 250-1000 bp). Reconditioning reactions were then performed on 10-fold dilutions of amplified DNA in a fresh PCR reaction mix in 200 µl reactions with 20 µl original PCR mixture as template, and all other reagents in the same concentrations as small-scale PCR. Thermal cycling conditions were the same as before, but with only 3 cycles of denaturation, annealing, and extension. All PCR products were pooled and concentrated in an Amicon Ultra 0.5 ml 100 kDa centrifugal column (Millipore), purified with QIAquick PCR Purification Kit (Qiagen), and eluted with 40 µl TE heated to 80 °C. Samples were loaded onto a 0.5 % agarose gel containing 1 × SYBR Safe stain and 1 × TAE, and run at 100V for 50 min. Amplicons of 500-1200 bp were gel purified with MinElute Gel Purification Kit (Qiagen). The samples were then prepared with the Nextera kit for sequencing via the MiSeq PE150 (Illumina) platform by the Georgia Genomics Facility.

3.3.3 Metagenomic and genomic analysis

For each metagenome, the sequences were passed through the same quality-control parameters with the CLC Genomics Workbench version 7.5 (CLCBio). Sequences were imported as Illumina paired-end reads, the FR20RV primer removed, and reads were checked for quality using a PHRED score of 20 and a minimum length of 50 bp. Paired reads were merged and a final data set containing merged reads and ORFans was checked again for QC with a minimum length of 100 bp. Trimmed, high-quality sequences from each metagenome were assembled separately, as well as combined and treated as a single library. Assemblies were constructed with the default parameters using a non-global alignment.

Rarefaction curves were generated with quality filtered reads greater than 200 bp in Metavir (Roux et al 2011) using a clusterization percentage of 75 and subsampled to 50,000 sequences. Assembled contigs from individual metagenomes and the pooled libraries were analyzed with BLASTx (Altschul et al 1990) searches in the NCBI nr ("non-redundant") database. Hits with an *E* value $\leq 10^{-5}$ were considered significant. We used this conservative, but frequently used, cutoff value to further reduce the likelihood of misclassification of a sequence. This output was analyzed using MEGAN v5.6.3 (Huson et al 2011), which assigns each contig to the top hit of the blast search (top 5%). Open reading frames (ORFs) were identified using Geneious 6.1.8 and used to search the Conserved Domain Database (CDD) NCBI database (Marchler-Bauer et al

2011).

3.3.4 Identification and analysis of RdRp phylotypes

HMMER (Eddy 1998) was used to identify RdRp-like sequences from assembled contigs using a hidden Markov model. The Markov model was produced based on amino acid alignments of conserved RdRp regions four through seven (Koonin et al 2008) of representatives from within the order *Picornavirales* and amplified or assembled environmental virus sequences (Culley et al 2006, Culley and Steward 2007, Culley et al 2014, Mueller et al *Submitted*). Sequences fitting the criteria of the model were retrieved from the libraries translated in all six frames. RdRp sequences that matched a known RdRp gene by BLAST with an *E* value $\leq 10^{-5}$ were considered to be significant. A maximum-likelihood tree of RdRp gene sequences was created with PHYML, WAG model, and 1,000 bootstraps (Guindon et al 2010) from protein sequences aligned with MAFFT (Katoh et al 2005). The sequences in this tree can be found in supplementary material from Mueller et al (*Submitted*), with an addition of the assembled contigs from this study.

3.3.5 Comparison to other marine RNA virus metagenomes

In order to investigate similarity between the two regions of the Antarctic Peninsula, reads from the EAP metagenomes were mapped to the assembled genomes from the WAP (PAL128, PAL156, PAL438, PAL473, and PAL_E4). Additionally, reads from the WAP samples were mapped to the LAR_K3 contig. A comparison of the Larsen A, Palmer Station, and Kāneʻohe Bay RNA virus metagenomes was conducted using a taxonomic classification of each of the assemblies as determined by BLASTx (Altschul et al 1990) searches in the NCBI nr ("non-redundant") database. Hits with an *E* value $\leq 10^{-5}$ were considered significant. This output was analyzed using MEGAN v5.6.3 (Huson et al 2011), which assigned each contig to the top hit of the blast search, and then displayed as Krona charts using Krona (Ondov et al 2011).

3.4 RESULTS

3.4.1 Metagenomic assembly analysis

The number of quality control filtered reads ranged from 2.6 to 3.0 million sequences for each library (Table 3.1). The majority of reads (92-95%) formed contigs in each separate library, and a total of 89% formed contigs when pooled together. The maximum length of sequences assembled to form a contig ranged from 5,386 to 6,476 bp and the GC content of assembled contigs ranged from 42-49%. Rarefaction curves produced by Metavir (Roux et al 2011) with a threshold of 75% suggest that library G, located nearest the coast, had the highest richness, while H and J had the lowest (Fig. 3.2). Station K was about 140 km east of Station G and the metagenome from K appears to contain the second highest diversity. A low percentage of sequences had no hits (0.4-14.9%), while the majority of sequences were classified as cellular (36-90%; Fig. 3.3A). The number of reads classified as viral varied for each metagenome, with a range from 9 to 58% in the individual libraries and 34% in the pooled library. Of these virus hits, the majority (72-100%) matched ssRNA viruses and none matched dsRNA viruses (Fig. 3.3B). In G, H, and J, the majority of sequences matched those from the genus *Bacillarnavirus* (60-83%).

3.4.2 Genome reconstruction

One novel RNA viral genome, LAR_K3, was reconstructed from metagenome K with a genomic structure and RdRp phylogeny that appears to be most closely related to two diatom-infecting isolates. These two isolates are proposed members of the genus *Bacillarnavirus*, Csp03RNAV that infects *Chaetoceros* sp. strain SS08-C03, and the other infects a putative *Thalassiosira* sp. isolated from Kāneʻohe Bay, Hawaii. The LAR_K3 genome was 6,476 bp long. A total of 32.8% of reads from library K mapped to this contig resulting in an average coverage of 24,900 bp per position. When the reads from all libraries were analyzed, 28.5% mapped to LAR_K3. Two open reading frames (ORFs) were identified in this assembled genome with a genomic average GC content of 40% (Fig. 3.4). The genome contained a region with significant hits to +ssRNA RdRps in the first ORF, but no helicase or protease genes were identified. Four structural proteins were identified that were most similar to the capsid binding site of picornavirads and the VP4 and capsid protein of dicistroviruses. Sequences with high similarity (>95% aa identity for a

565 aa fragment) to the capsid protein of LAR_K3 were found in libraries G, H, and J and were classified as *Bacillarnavirus* sequences based on BLASTx.

3.4.3 RdRp phylogenetics

A maximum-likelihood tree of RdRp gene sequences shows the identified environmental sequences from this study relative to other known viruses within the order *Picornavirales* (Fig. 3.5). A search for RdRp sequences based on four of the seven conserved motifs (mpl-like region) returned two picornavirad-like sequences from G, two from H, one from J, one from K, and one from the pooled library. Of the two sequences that were found in G and H, one from each had 100% amino acid (aa) identity to each other, the sequence from K, and the pooled library and is represented in the RdRp tree as one sequence. The other sequences from G and H had 98.6% aa identity to each other and the sequence from J. All identified RdRps clustered in a small group most similar to the genus *Bacillarnavirus*, including Csp03RNAV that infects a diatom *Chaetoceros* sp. strain SS08-C03 and an isolate that infects a putative *Thalassiosira* sp. isolated from Kāneʻohe Bay, Hawaii.

3.4.4 Marine RNA virus metagenome comparison

A comparison of the pooled metagenome from the Larsen A region with other marine RNA viral metagenomes revealed a similar dominance of ssRNA viruses, but with a much larger fraction of unknown ssRNA virus sequences (Fig. 3.6). Although the pooled assembly does not contain as high a percentage of *Bacillarnavirus*-like sequences as Palmer Station or Kāneʻohe Bay, a similar dominance of such sequences was seen in stations G, H, and J when analyzed individually. A recruitment analysis of the assembled genomes from Palmer Station with the total reads from Larsen A revealed little to no recruitment for all genomes except for PAL156. This genome recruited 0.02% of the reads, but the coverage was poor and sporadic. Recruitment of the Palmer Station reads to LAR_K3 resulted in only 0.0003% of the total reads, again with patchy coverage.

3.5 DISCUSSION

This study corroborates our previous findings of a dominance of +ssRNA viruses most similar to those from the order *Picornavirales* in the polar waters around the Antarctic Peninsula. While

the majority of the virus hits in each metagenome from the EAP consisted of ssRNA viruses, there were a large percentage of hits to cellular organisms when looking at the domain-level classification. We are not sure why there were so many cell-derived sequences in these libraries, since the method for constructing the metagenomes was similar to previous studies that resulted in predominately virus-like sequences (Culley et al 2014, Mueller et al *Submitted*). This may reflect a lower signal to noise ratio, because the total yield of viral nucleic acids from the EAP samples was much lower than in the other studies. The high sea ice cover in the study area may explain the lower yields at the time of sampling, which was toward the end of the summer season with little biomass in the upper water column. Despite the large background in cell-derived sequences, we were able to identify many RNA virus sequences and even assemble a nearly complete genome.

Station G was nearest to the coast and appears to contain the highest diversity, based on rarefaction curves and taxonomy, however the low percentage of viral sequences in this sample makes it difficult to assess whether these data represent meaningful ecological shifts in the viral community. While there appears to have been a dominance of viruses similar to the *Bacillarnaviruses*, a diatom-infecting group of viruses, at stations G, H, and J, station K seems to be dominated by viruses that cannot be classified further than ssRNA viruses. However, a large proportion (95%) of those sequences from library K were assembled to form contig LAR_K3, a virus that we determined likely infects a diatom based on RdRp phylogeny and genome structure. Thus, the majority of those sequences in library K should also be classified as putative diatom-infecting viruses in the order *Picornavirales*.

The only near-complete genome assembled in these datasets, LAR_K3, appears to be a significant component of the viral community in the Larsen A region at the time of sampling. Not only was the recruitment to this genome extraordinarily high in library K, but sequences similar to the structural ORF of this genome were present in all of the individual libraries along the transect. The genomic configuration and RdRp phylogeny is most similar to the viruses within the genus *Bacillarnavirus*, again suggesting the importance of these viruses in influencing the dynamics of their dominant diatom hosts. The contig did not contain any sequence with significant hits to any proteases, a similar finding to that reported by Culley et al (2014) and Mueller et al (*Submitted*) . However, since this gene is required for +ssRNA viral replication and

an ORF exists in the location where a protease is normally encoded, we presume this virus also contains a highly divergent protease.

There was surprisingly very little diversity among the identified RdRp sequences recovered from the EAP and the sequences all clustered closest to a member of the *Bacillarnavirus* genus and another diatom-infecting isolate. There were too few sequences to compare the diversity from these samples with other marine RNA viral metagenomes using standard metrics, but the glaring scarcity in diversity is apparent in the RdRp phylogenetic tree. While these results suggest a low diversity of host community during this time, we unfortunately do not have sufficient host data to determine their diversity. Nonetheless, similar to our findings in the WAP, it appears that RNA viruses that infect diatoms may dominate the community in the Larsen A region.

In the comparison of the pooled metagenome from the Larsen A to other marine RNA viral metagenomes, a majority of the sequences classified as ssRNA viruses would have been classified as *Bacillarnaviruses* if the Csp03RNAV, or Chaetoceros sp. RNA virus 2 was classified as a *Bacillarnavirus* in the database as suggested by Tomaru et al (2013). Taking this into consideration, 43% of sequences would have been classified as *Bacillarnaviruses* and 30% as uncultured Picornavirales, leaving only 24% classified as ssRNA viruses instead of 71%. These changes would obviously impact the Palmer Station and Kāneʻohe Bay results as well, providing a higher taxonomic classification of sequences classified as ssRNA viruses, but the impact appears greater for the Larsen A metagenome.

Nevertheless, these findings suggest that the RNA virus communities on the WAP and EAP represent distinct virus assemblages during the time periods sampled, with a similar dominance of diatom-infecting viruses. Given the difference in season and year of the samples, one might expect these results from probably distinct communities of protistan hosts. More spatial and temporal studies in the southern ocean are needed to fully understand the amount of viral diversity and variability in the communities present in these polar ecosystems.

3.6 ACKNOWLEDGEMENTS

We thank the scientists and crewmembers onboard the *R/V Nathaniel B. Palmer* (NBP1203). In particular, we thank Craig Smith and his lab group for inviting us to join the LARISSA

expedition, and Chung Hwang, Maria Vernet, and Mattias Cape for their collaboration and thoughtful discussions throughout this study. We thank Elisha Wood-Charlson for use of the CLC Genomics Workbench and Mahdi Belcaid and Guylaine Poisson at the Bioinformatics Laboratory at the University of Hawai‘i at Mānoa for the use of their computer cluster for BLAST searches. Funding for this project was provided by NSF grants to Grieg Steward (ANT 09-44851) and Craig Smith (OPP- 0732711), and by the Center for Microbial Oceanography: Research and Education (EF 04-24599).

Table 3.1. Assembly data of the four metagenomes and pooled total reads assembly. Listed are the number and length (nucleotides, nt) of initial quality controlled reads and of contigs after assembly, percent GC content, and the percentage of reads that were assembled to form a contig.

				Length (nt)				G+C	Assembled
Sample	**Date**		**Count**	**Mean**	**Min**	**Max**	**Total nt**	**(%)**	**(%)**
G	28 Mar 2012	Reads	2,632,931	162.5	100	300	427,844,432	42	92
		Contigs	4,898	543	254	5,386	2,663,478		
H	31 Mar 2012	Reads	3,248,985	154	100	300	501,618,311	46	95
		Contigs	380	591	276	5,386	224,829		
J	6 Apr 2012	Reads	3,282,899	160	100	300	526,302,310	49	92
		Contigs	372	601	300	5,386	223,624		
K	21 Mar 2012	Reads	3,031,561	162	100	300	489,867,060	47	95
		Contigs	526	629	254	6,476	331,251		
Total		Reads	12,196,376	160	100	300	1,945,632,113	43	89
		Contigs	2,452	749	496	5,386	1,838,010		

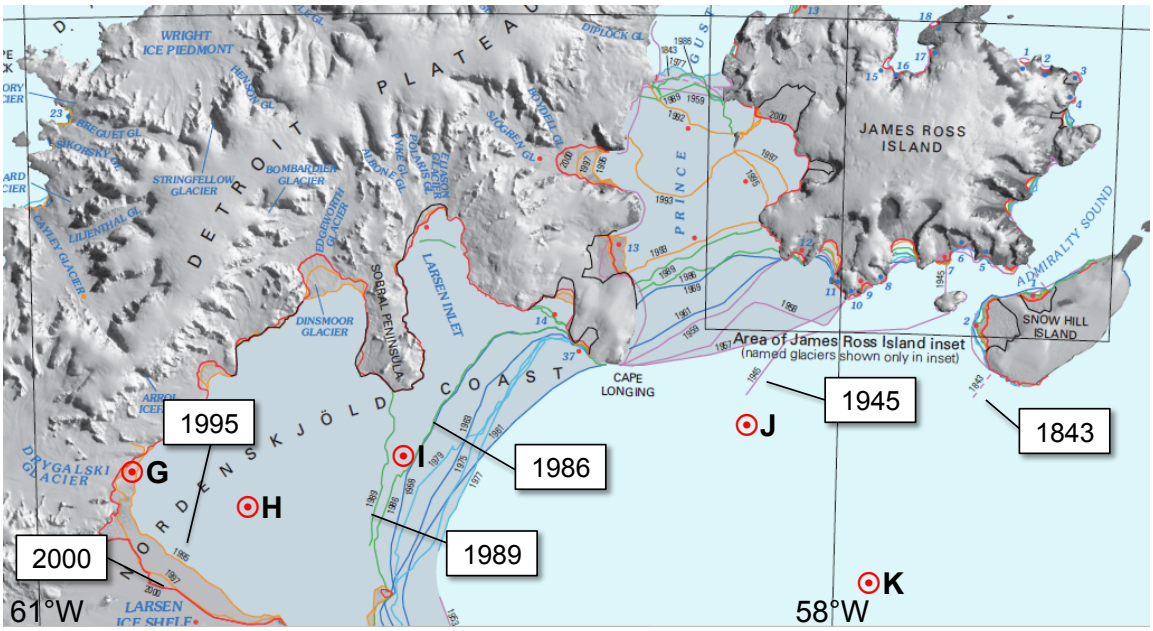


Figure 3.1. Map of the five main benthic stations from the LARISSA 2012 cruise track (G, H, I, J, K). The years and locations indicated represent the timing of the ice shelf retreat. Modified from Ferrigno et al (2006).

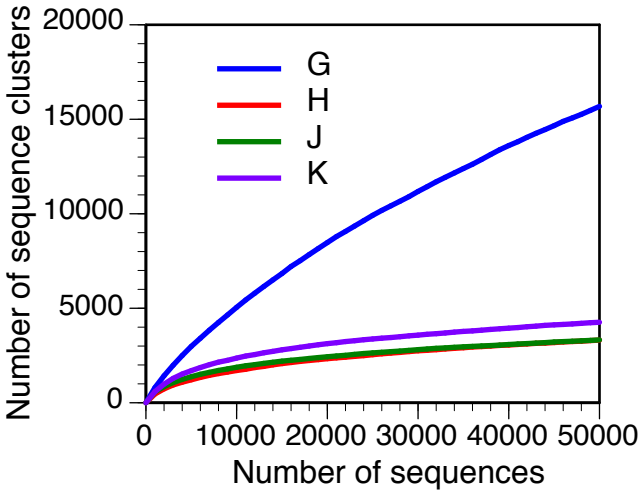


Figure 3.2. Rarefaction curves of libraries G, H, J, and K as determined by Metavir. In this analysis, sequence reads were subsampled to 50,000 sequences from each library and plotted as a function of the number of unique clusters (threshold = 75%) identified.

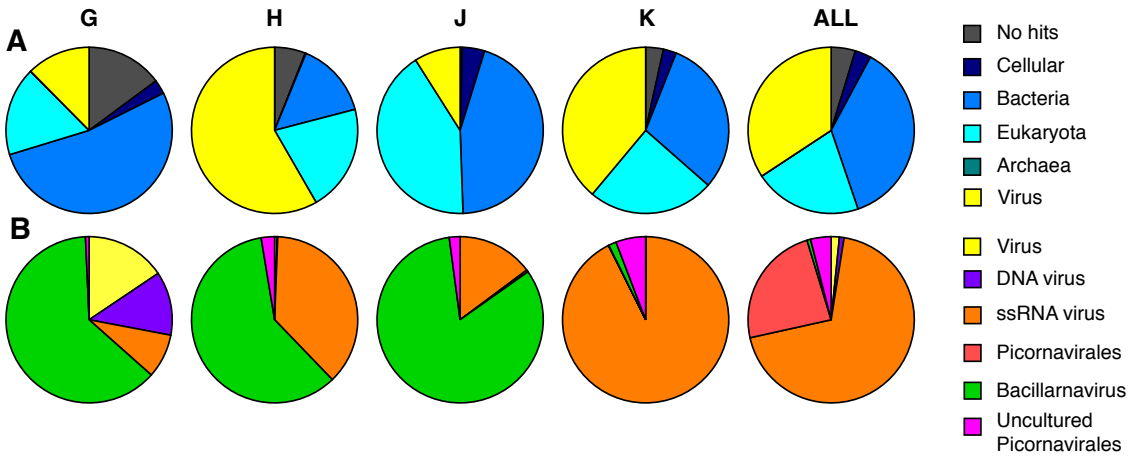


Figure 3.3. Taxonomic classification of assemblies from libraries G, H, J, K, and pooled reads. Panel A shows the domain-level classification of contigs and Panel B shows further classification of hits classified as viral. Those classified as "virus," "DNA virus," or "ssRNA virus" in Panel B could not be classified any further. The *Bacillarnavirus* genus is found within the order *Picornavirales* and the Uncultured *Picornavirales* include environmental sequences of marine picorna-like viruses JP-A or JP-B.

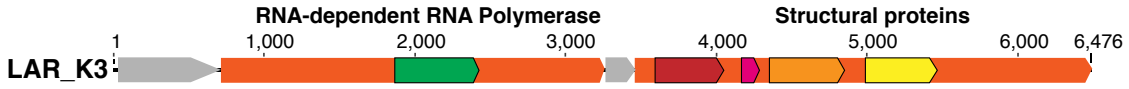


Figure 3.4. Genome map of contig LAR_K3. The identified RNA dependent RNA polymerase gene is shown in green, and the structural genes in red, pink, light orange and yellow. Untranslated regions are shown in light grey and the open reading frames are shown in orange.

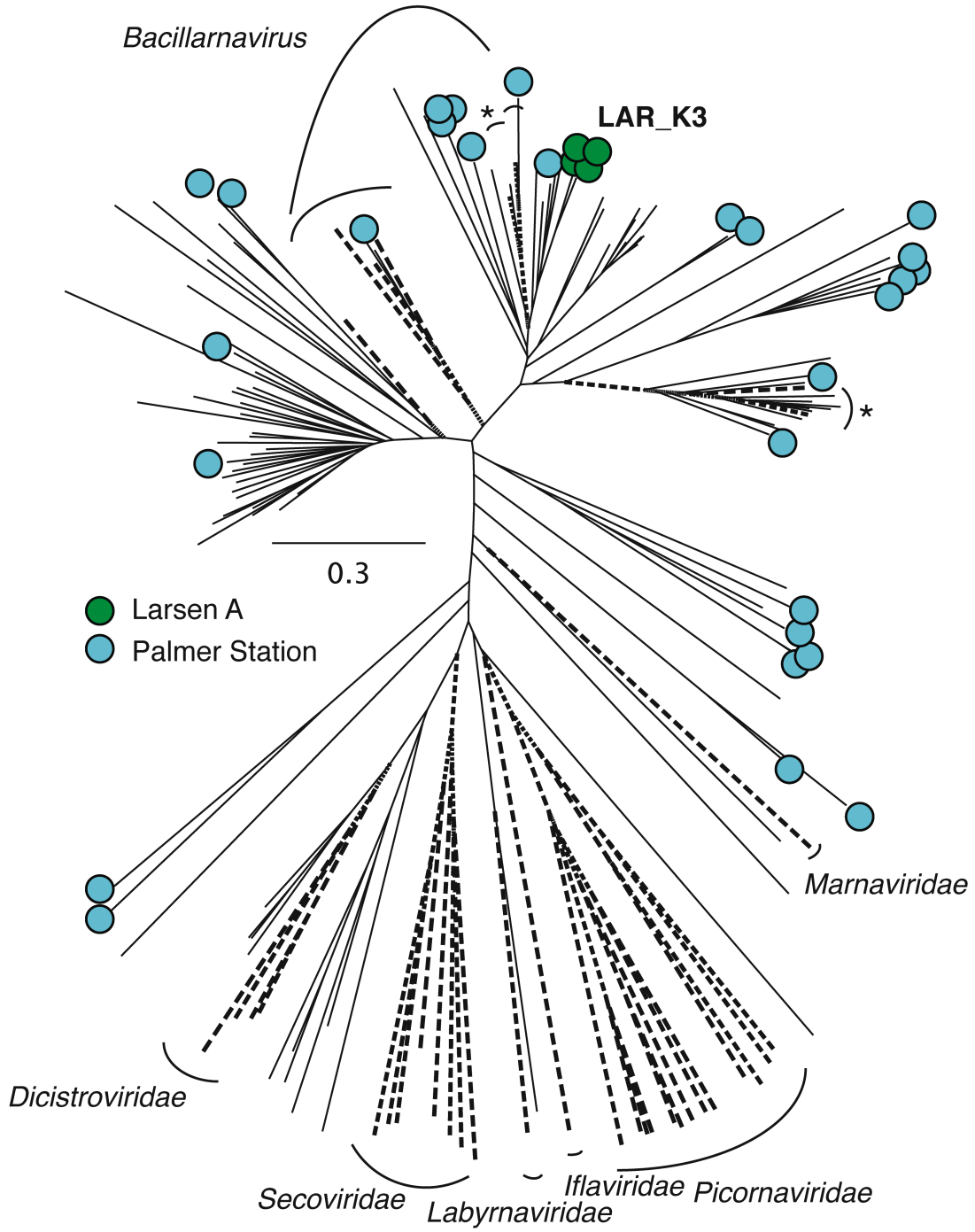


Figure 3.5. RdRp phylotypes. An unrooted phylogenetic analysis (maximum likelihood; model WAG; 1,000 bootstraps) of the picorna-like region of the RdRp gene showing the position of environmental sequences from this study (green circles), Palmer Station (blue circles (Mueller et al *Submitted*), and previous studies (solid lines (Culley and Steward 2007, Culley et al 2003, 2007, 2014), relative to known viruses in the order *Picornavirales*. The branches of cultured isolates are shown with a dashed line, the asterisk (*) indicate two unpublished diatom-infecting

viruses isolated from Kāneʻohe Bay. LAR_K3 indicated on the tree represents the genome assembled in this study. The scale bar represents 0.3 substitutions per site.

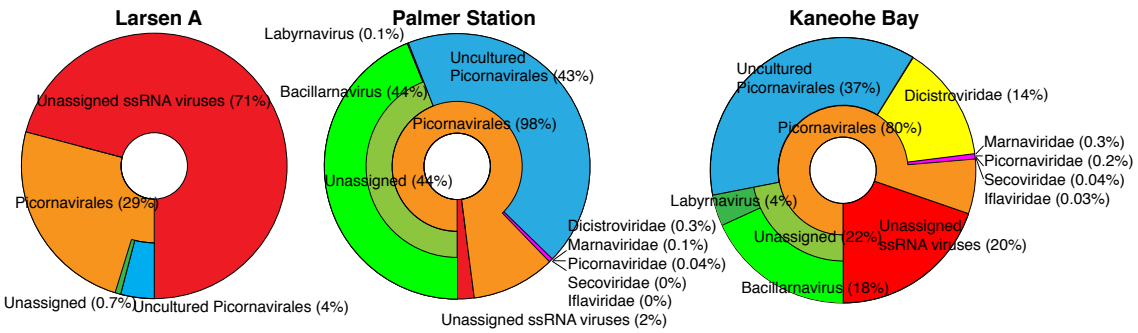


Figure 3.6. Taxonomic classification of assemblies from Larsen A metagenomes compared with Palmer Station (Mueller et al *Submitted*) and Kāneʻohe Bay metagenomes (Culley et al 2014). Percentages are relative to the total ssRNA virus hits for the pooled library assembly from Larsen A or Palmer Station and for the combined data from Kāneʻohe Bay 2009 and 2010 metagenomes. Krona charts were created using Krona (Ondov et al 2011).

3.7 REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C *et al* (2006). The Marine Viromes of Four Oceanic Regions. *PLoS Biology* **4:** e368.

Bertolin ML, Schloss IR (2009). Phytoplankton production after the collapse of the Larsen A Ice Shelf, Antarctica. *Polar Biology* **32:** 1435-1446.

Cape MR, Vernet M, Kahru M, Spreen G (2014). Polynya dynamics drive primary production in the Larsen A and B embayments following ice shelf collapse. *Journal of Geophysical Research: Oceans* **119:** 572-594.

Culley AI, Lang AS, Suttle CA (2003). High diversity of unknown picorna-like viruses in the sea. *Nature* **424:** 1054-1057.

Culley AI, Lang AS, Suttle CA (2006). Metagenomic analysis of coastal RNA virus communities. *Science* **312:** 1795-1798.

Culley AI, Steward GF (2007). New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Applied and Environmental Microbiology* **73:** 5937-5944.

Culley AI, Suttle CA, Steward GF (2010). Characterization of the diversity of marine RNA viruses. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. Aquatic Sciences Limnology and Oceanography: Waco, TX. pp 193-201.

Culley AI, Mueller JA, Belcaid M, Wood-Charlson EM, Poisson G, Steward GF (2014). The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *MBio* **5**.

Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N *et al* (2008). Viral genome sequencing by random priming methods. *BMC Genomics* **9**.

Domack E, Duran D, Leventer A, Ishman S, Doane S, McCallum S *et al* (2005). Stability of the Larsen B ice shelf on the Antarctic Peninsula during the Holocene epoch. *Nature* **436:** 681-685.

Duhaime MB, Deng L, Poulos BT, Sullivan MB (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environmental Microbiology* **14:** 2526-2537.

Eddy SR (1998). Profile hidden Markov models. *Bioinformatics* **14:** 755-763.

Edwards RA, Rohwer F (2005). Viral metagenomics. *Nature Reviews Microbiology* **3:** 504-510.

Ferrigno JG, Cook AJ, Foley KM, Williams RSJ, Swithinbank C, Fox AJ *et al* (2006). Coastal change and glaciological map of the Trinity Peninsula area and
South Shetland Islands, Antarctica: 1843–2001. *U. S. Geological Survey Geologic Investigations Map Series Map I-2600-A*. p 32.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307-321.

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

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Research* **21:** 1552-1560.

Katoh K, Kuma K-i, Toh H, Miyata T (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33:** 511-518.

Koonin EV, Wolf YI, Nagasaki K, Dolja VV (2008). The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nature Reviews Microbiology* **6:** 925-939.

Labonte JM, Suttle CA (2013). Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME Journal* **7:** 2169-2177.

Lawrence JE, Steward GF (2010). Purification of viruses by centrifugation. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. Aquatic Sciences Limnology and Oceanography: Waco, TX. pp 166-181.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C *et al* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* **39:** D225-D229.

Mueller JA, Culley AI, Schvarcz CR, Steward GF (*Submitted*). RNA viruses as major contributors to Antarctic virioplankton.

Ó Cofaigh C, Davies BJ, Livingstone SJ, Smith JA, Johnson JS, Hocking EP *et al* (2014). Reconstruction of ice-sheet changes in the Antarctic Peninsula since the Last Glacial Maximum. *Quaternary Science Reviews* **100:** 87-110.

Ondov BD, Bergman NH, Phillippy AM (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12:** 385.

Rosario K, Breitbart M (2011). Exploring the viral world through metagenomics. *Current Opinion in Virology* **1:** 289-297.

Roux S, Faubladier M, Mahul A, Paulhe N, Bernard A, Debroas D *et al* (2011). Metavir: a web server dedicated to virome analysis. *BMC Bioinformatics* **27:** 3074-3075.

Sañé E, Isla E, Grémare A, Gutt J, Vétion G, DeMaster DJ (2011a). Pigments in sediments beneath recently collapsed ice shelves: The case of Larsen A and B shelves, Antarctic Peninsula. *Journal of Sea Research* **65:** 94-102.

Sañé E, Isla E, Pruski AM, Bárcena MA, Vétion G, DeMaster D (2011b). Diatom valve distribution and sedimentary fatty acid composition in Larsen Bay, Eastern Antarctica Peninsula. *Continental Shelf Research* **31:** 1161-1168.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G (2013). Are we missing half of the viruses in the ocean? *ISME Journal* **7:** 672-679.

Suttle CA (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology* **5:** 801-812.

Tomaru Y, Toyoda K, Kimura K, Takao Y, Sakurada K, Nakayama N *et al* (2013). Isolation and characterization of a single-stranded RNA virus that infects the marine planktonic diatomChaetocerossp. (SS08-C03). *Phycological Research* **61:** 27-36.

Weinbauer MG (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28:** 127-181.

Chapter 4

Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters

Jaclyn A. Mueller[1]

Alexander I. Culley[1,2], Grieg F. Steward[1]

[1]Center for Microbial Oceanography: Research and Education, Department of Oceanography, University of Hawai‘i at Mānoa, 1950 East-West Road, Honolulu, Hawai‘i 96822

[2]Present address: Département de biochimie de microbiologie et de bio-informatique, Université Laval, 2420, rue de la Terrasse, Québec, Québec G1V 0A6

Published by American Society for Microbiology

Applied and Environmental Microbiology, 80(13):3930-3942.

Reprinted with permission

4.1 ABSTRACT

Anodic aluminum oxide (AAO) filters have high porosity and can be manufactured with a pore size that is small enough to quantitatively capture viruses. These properties make the filters potentially useful for harvesting total microbial communities from water samples for molecular analyses, but their performance for nucleic acid extraction has not been systematically or quantitatively evaluated. In this study, we characterized the flux of water through commercially produced nanoporous (0.02 μm) AAO filters (Anotop, Whatman), and used isolates (a virus, a bacterium, and a protist) and natural seawater samples to test variables that we expected would influence the efficiency with which nucleic acids are recovered from the filters. Extraction chemistry had a significant effect on DNA yield, and back flushing the filters during extraction was found to improve yields of high molecular weight DNA. Using the back-flush protocol, the mass of DNA recovered from microorganisms collected on AAO filters was ≥ 100% of that extracted from pellets of cells and viruses, and 94±9% of that obtained by direct extraction of a liquid bacterial culture. The latter is a minimum estimate of the relative recovery of microbial DNA, since liquid cultures include dissolved nucleic acids that are retained inefficiently by the filter. In conclusion, we demonstrate that nucleic acids can be extracted from microorganisms on AAO filters with similar efficiency to that achievable by direct extraction of microbes in suspension or in pellets. These filters are therefore a convenient means by which to harvest total microbial communities from multiple aqueous samples in parallel for subsequent molecular analyses.

4.2 INTRODUCTION

Direct filtration is one of the most common methods for harvesting microorganisms from natural water samples for molecular analyses (Fuhrman et al 1988, Pichard et al 1993, Somerville et al 1989), because it is technically simple and efficiently concentrates microorganisms onto a small surface area for subsequent nucleic acid extraction. Most direct filtration protocols focus on the capture of cells (bacteria, archaea and eukaryotes) using filters with pore sizes ≥ 0.2 μm (Valentin et al 2005). Since most viruses in aquatic environments are < 0.2 μm in diameter (Brum et al 2013), they are not quantitatively represented in the nucleic acids recovered by these protocols. Collection of viruses for molecular analyses is more commonly achieved using tangential flow ultrafiltration (Wommack et al 2010). This method can also be used to harvest viruses and cells simultaneously (Suzuki et al 2004), but the method is time intensive and requires relatively expensive equipment, making it unsuitable for high-throughput sampling. Other methods for harvesting cells or viruses from water include adsorption/elution or flocculation (Percival et al 2004), but many of these methods have highly variable recovery efficiencies depending on the microorganism, and have been typically used for water quality monitoring rather than comprehensive community analysis. An iron flocculation procedure was recently shown to concentrate viruses in natural communities with high efficiency (John et al 2011), but specifically targeted only the viral fraction.

A simple, direct filtration procedure that effectively captures all of the microorganisms, including viruses, from a water sample would be useful for many applications (e.g., studies of microbial ecology, monitoring of recreational and drinking water quality, or testing for pathogens in the process streams of food, biotechnology, and biopharmaceutical industries). Some of the membrane materials typically used for molecular analysis of microbial communities (e.g., polyvinyldifluoride, polycarbonate) can be manufactured with pore sizes small enough to retain most or all viruses, but have such low flux that they are less practical for routine collections. The best option, at present, for simultaneous capture of viruses and cellular microorganisms appears to be nanoporous anodic aluminum oxide (AAO) filters. These filters have high porosity with uniform pore size (Jani et al 2013), and are commercially manufactured with pores as small as 0.02 μm, which is close to the lower limit for known viruses (Steward et al 2013b). They are also available in disposable housings with luer-lock fittings, making them

convenient for field collections by in-line filtration. Many samples can be processed in parallel by this method using multi-channel peristaltic pumps.

Two of us reported previously on the use of AAO filters for molecular analyses of aquatic viruses (Culley and Steward 2007b, Culley et al 2009b). Extraction of the nucleic acids was accomplished using a commercial kit (MasterPure Complete DNA and RNA Purification Kit, Epicentre) that is based on a salting-out extraction protocol (Miller et al 1988). The salting-out procedure employs a buffer containing an ionic detergent (SDS) to solubilize tissues, cells, or viruses. Following the lysis step (usually at 37-65°C with proteinase K), SDS-protein complexes are precipitated by the addition of salt and pelleted by centrifugation. Nucleic acids in the supernatant are then precipitated with alcohol. Many variations on the salting out procedure have been published (Aljanabi and Martinez 1997, Ferrara et al 2006, Grimberg et al 1989, Laitinen et al 1994, Yu and Mohn 1999).

We chose the salting-out extraction procedure, rather than one of the popular guanidinium-based extraction methods, because guanidinium salts have been reported to cause irreversible binding of DNA to AAO (Dames et al 2006, Gerdes et al 2001). We later described a modification of our filter-extraction protocol (Steward and Culley 2010) in which extraction buffer is back flushed through the membrane rather than being pushed through in the same direction as the sample. This modification was adopted out of concern that the exceptionally small pore size of the filters may trap high molecular weight nucleic acids. Although there were reasonable theoretical bases for our choices about the extraction buffer composition and the direction in which it is applied to the filter, there has been no empirical evidence that these choices actually affect the yield of nucleic acids. In fact, there are no data at all on the efficiency with which nucleic acids can be extracted from microorganisms collected on AAO filters.

Given the utility of these membranes for harvesting microorganisms and viruses (Culley and Steward 2007a, Culley et al 2009a, Needham et al 2013), we felt an assessment of their performance was worthwhile. In this study, we tested the flow characteristics and loading capacity of commercially produced AAO filters with 0.02 μm pores size, and systematically tested a number of factors that we predicted would influence the recovery of nucleic acids from microorganisms collected on the filters. This included a comparison of DNA yields when

extraction buffer was pushed through the filter in the same direction as filtration (forward flush) vs. being pulled through in the opposite direction (back flush), and a comparison of recoveries using SDS- vs. guanidinium-based extraction buffers. We also compared the efficiency with which DNA and RNA were recovered from microorganisms when they were collected on filters vs. being suspended in liquid or pelleted, and looked at the effects of lysozyme treatment on recovery of nucleic acids from samples containing cells or only viruses. We then evaluated the linearity of DNA yield for varying volumes of sample filtered, and evaluated the tendency of the filters to trap dissolved DNA. From these experiments we provide an assessment of the performance of AAO filters for nucleic acid extractions and a recommended protocol for conducting such extractions.

4.3 MATERIALS AND METHODS

4.3.1 Filters

We used 25 mm diameter filter membranes sealed by thermal welding in polypropylene housings with integral luer-lock fittings. Depending on the experiment, the filter material was either AAO (0.02 or 0.2 μm pore size; Anotop 25, Whatman) or polyethersulfone (PES, 0.2 μm; Puradisc 25, Whatman). For one experiment comparing flow rates, an AAO membrane disc (0.02 μm pore size, 25 mm diameter) with plastic support ring (Anodisc, Whatman) was mounted in a reusable polypropylene filter housing (Advantec, MFS). The details of the flow rate and extraction experiments are presented in the following sections and are summarized Table 4.1.

4.3.2 Filter flow rate and capacity

The relationship between pressure and flow rate for 0.02 μm Anotop 25 AAO filters was tested by pumping purified water (NanoPure, Barnstead) through the filters at varying temperatures (7, 22, and 35°C) at pressures ranging from 0.26 to 2.1 bar using a peristaltic pump (MasterFlex Digital Drive and L/S 14 PharMed tubing, Cole-Parmer). To evaluate filter-to-filter variability, flow rate at 22°C was tested for three separate filters (Whatman Cat. No. 6809-4002, Lot No. D123932). Pressure was monitored with a flow-through pressure transducer (GE Healthcare Life Sciences) connected immediately upstream of the filter. To account for pressure pulsing from the peristaltic pump, mean pressure was calculated from the highest and lowest pressure readings

observed at a given pump speed. Flow rates were determined by measuring the volume of filtrate collected in a timed interval (1 to 3 minutes) in a 25 ml graduated cylinder (accuracy ± 0.1 ml).

The practical capacity of the filters was tested by pumping whole seawater through AAO filters (0.02 µm Anotop 25) in duplicate from each of two locations differing in their productivity. The first sample was collected at 25 m depth in the oligotrophic open ocean (Station ALOHA, located 100 km north of Oahu, Hawai‘i; 22° 45' 00" N, 158° 00' 00" W) and the second from the upper decimeter of a eutrophic estuarine drainage canal (Ala Wai Canal located in Honolulu, Hawai‘i; 21° 16' 31.3" N, 157° 49' 03.8" W). Pressures and flow rates were recorded during the course of the filtration. To characterize the differences in microbial concentrations in these two habitats, we quantified prokaryotes and viruses by epifluorescence microscopy (Patel et al 2007) and chlorophyll *a* by fluorescence after extraction in acetone (Ala Wai Canal) or by in situ fluorescence using a sensor calibrated to extracted chlorophyll (Station ALOHA) as described in the Hawaii Ocean Time Series Field & Laboratory Protocols (http://hahana.soest.hawaii.edu/hot/protocols/chap12.html).

4.3.3 Preparation of samples for filter extraction tests

a. Coastal Seawater – Seawater samples were collected in polycarbonate carboys from a pier in Kāne‘ohe Bay on the windward side of O‘ahu, Hawai‘i (21° 25’ 46.80” N, 157° 47’ 31.51” W) on multiple occasions between February 2010 and January 2013. For some experiments, whole seawater (10 mL to 1 L) was directly filtered through Anotop 25 (0.02µm) filters. Replicate samples were filtered in parallel using a multi-channel peristaltic pump. In other cases, the seawater (200 mL) was first filtered through 0.22 µm PES membrane filter capsules (Sterivex, Millipore) prior to filtration through Anotop filters. Samples were filtered with a peristaltic pump as described above at approximately 14 ± 2 mL min^{-1}. Residual sample was evacuated from all filters using air pressure from a hand-operated syringe, then the filters were wrapped in Parafilm M (Pechiney), placed in Whirl-Paks (Nasco), and stored at -80°C until they were extracted.

b. Cultivated Bacteriophage – Vibrio phage VvAW1, a podovirus with a double-stranded DNA genome (38 kb) and a capsid diameter of 43-45 nm (Nigro et al 2012), was purified from plate lysates in a CsCl equilibrium buoyant density gradient (Lawrence and Steward 2010), buffer

exchanged into 3 mL SM (0.4 M NaCl, 0.02 M $MgSO_4$, 0.05 M Tris, pH 7.5), then filtered onto Anotop 25 (0.02 μm) filters in triplicate for each method and stored as noted above.

c. Bacterial culture – A strain of *Vibrio vulfnificus* (V93D1V) isolated from coastal waters of O'ahu (Nigro 2012) was grown at 37 °C with shaking, for approximately 3 hours, or until culture reached stationary phase as determined by the OD reading. For experiments with washed cells, a volume of 12 ml of culture was pelleted via centrifugation at 4,000 × *g* for 10 minutes and washed twice with 1 × PBS (Phosphate Buffered Saline; 13.7 mM NaCl, 0.27 mM KCl, 1 mM Phosphate Buffer; 0.02 μm-filtered), and then resuspended in 1 × PBS. For experiments with unwashed cells, 500 μl of liquid culture was diluted with 3 ml of 0.02 μm-filtered SM buffer. Suspended cells were filtered onto AAO (0.02 μm or 0.2 μm Anotop 25) or PES (0.2 μm Puradisc 25) syringe-tip filters with a sterile syringe in triplicate for each method and stored as noted above.

d. Cultivated protist – A uni-algal, non-axenic, culture of a dinoflagellate (*Gymnodinium* strain AL-DI06 isolated from Station ALOHA) was grown to a cell density of ~1.2 x 10^5 cells ml^{-1} in f/2 media at 25 °C under a 12:12 hour light:dark cycle. Subsamples of the culture (1, 2, or 4 ml, depending on the experiment) were filtered onto Anotop 25 (0.02 μm) filters with a sterile syringe in triplicate for each method and stored as noted above.

4.3.4 Back flush versus forward flush

AAO filters through which whole seawater, 0.22 μm-filtered seawater, purified virus, bacterial culture, or dinoflagellate culture had been filtered were extracted using an SDS-based extraction buffer using the reagents in a commercially available kit (MasterPure, Epicentre). The back-flush protocol was conducted as described previously (Steward and Culley 2010), in which case the buffer was introduced through the outlet side of the filter and recovered via the inlet, opposite to the direction in which the sample was loaded. For the forward-flush protocol, the extraction buffer was introduced through the inlet side of the filter and recovered via the outlet.

For each filter, 1 ml of Tissue & Cell Lysis Solution (MasterPure) containing 100 μg ml^{-1} proteinase K was loaded into a 3 ml-capacity sterile syringe (the injection syringe) and gently pushed through the filter via the appropriate opening until liquid just reached the other opening.

The other opening was then sealed with a second 3 ml syringe (the aspiration syringe) and liquid pulled through to further saturate the filter membrane. The sealing of the filter on both ends with syringes allowed for thermal expansion without leakage of the buffer during the high-temperature incubation. The assembly was incubated at 65°C for 15 min while attached to a rotisserie in a hybridization oven rotating at full speed (ca. 16 rpm). The lysis buffer was then recovered by drawing it into the aspiration syringe. The extract was transferred to a 1.7 ml microcentrifuge tube and placed on ice. Salt-induced precipitation of protein-detergent complexes followed by an alcohol precipitation was then carried out following the manufacturer's instructions. All AAO filters from this study were extracted using this protocol, unless otherwise stated.

4.3.5 Trapping of dissolved DNA on AAO filters

To determine whether dissolved DNA is trapped on the AAO membrane filters, we filtered 1 μg of a 5 kb DNA ladder (CHEF DNA size standard, BioRad) resuspended in 500 μl of 1 × TE through Anotop 25 (0.02 μm) filters in triplicate using a hand-operated syringe. We measured the quantity of DNA in the initial flow-through and in 3 × 500 μl washes of 1 × TE. We then extracted the filters using the back flush protocol and measured the quantity of DNA in the extract to estimate how much of the material retained on the membrane would be released. To determine whether there is a difference in the size distribution of the DNA trapped on the filter versus the material that passes through, we analyzed the initial flow-through, the first wash and the extracted material via pulsed-field gel electrophoresis (PFGE). Because the yields of DNA in the extract were low, we first concentrated the extracts by centrifugal ultrafiltration. To account for any losses resulting from the extract concentration step, we diluted triplicate aliquots of the 5 kb ladder in extraction buffer, then concentrated them in the same manner as the extract. We loaded approximately equal total masses of DNA (49-51 ng) of the 5 kb control ladder (in duplicate) and the concentrated control ladder, filtrates, washes, and extracts (all in triplicate) onto a 1% agarose gel with Blue Juice as the loading buffer (final concentration 1 ×, Life Technologies). The ladders were resolved by PFGE in lithium borate buffer (1 × LB, Faster Better Media, LLC) at 14°C, 6 V cm^{-1} and a 120° included angle for 18 hrs with a 1-6 s switch time. The gel was post-stained with 1 × SYBR Gold (Life Technologies) and imaged with Molecular Gel Logic 200 Imaging System (Kodak). We measured the relative intensities of the

bands using Gaussian curve fitting with Molecular Imaging Software v.4.0.3 (Kodak), then scaled the intensities within a sample according to the average recovery determined by fluorometry. We then calculated the percent recovery for each band in a sample by dividing its scaled intensity by that of the corresponding band in the appropriate control ladder.

4.3.6 Effects of filter loading on nucleic acid recovery

Varying volumes of whole seawater (10, 50, 250, 500, 1000 ml) from Kāneʻohe Bay or a dinoflagellate culture (1, 2, or 4 ml) were filtered onto Anotop 25 (0.02 μm) filters in triplicate and extracted by the back flush method as described above. For whole seawater, extraction efficiency was determined relative to two different controls using parallel subsamples of the same seawater: 1) direct extraction (1.8 ml) and 2) extraction of a cell and virus pellet formed via ultracentrifugation (11.7 ml) under the conditions specified in the section *Estimating extraction efficiencies*. Controls for the dinoflagellate culture were either directly extracted from liquid culture or a cell pellet produced by centrifugation at 16,000 × *g* for 10 min. All nucleic acid extractions were performed using the MasterPure kit, following the manufacturer's instructions.

4.3.7 Effects of buffer chemistry and filter material nucleic acid recovery

a. Direct extraction controls – Equal portions of washed cells of *V. vulfnificus* (V93D1V) were transferred into 1.7 ml microcentrifuge tubes and half were extracted using the MasterPure kit, following the manufacturer's instructions for cell samples. The MasterPure kit is based on the protocol by Miller et al (1988) and uses a SDS-containing lysis buffer (Watson et al 1998). The others were extracted with the DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer's instructions for cultured cells under the Purification of Total DNA from Animal Blood or Cells Protocol. The DNeasy kit uses a guanidinium-based lysis buffer (guanidinium chloride, GuHCl).

b. Filter extractions – Aliquots from the same batch of washed cells were filtered onto AAO filters (0.2 μm and 0.02 μm; Anotop 25, Whatman) or PES filters (0.2 μm; Puradisc 25) with a sterile syringe. Half of the filters were extracted in triplicate with the MasterPure kit buffers using both the back- and forward-flush methods described above. The rest of the filters were extracted in triplicate using the back- and forward-flush methods described above, but with

buffers from the DNeasy Kit, following the manufacturer's reaction conditions for cultured cells. For each filter, one-ml of 0.02 µm filtered lysis buffer (500 µl Buffer AL and 500 µl 1 × PBS) containing 40 µl Proteinase K (DNeasy Kit) was loaded into a 3 ml-capacity sterile syringe (the injection syringe) and gently pushed through the filter in the same manner as previously described for either the back- or forward-flush filter extraction method. The other opening of the filter was sealed with a second syringe (the aspiration syringe) and the assembly was incubated at 56°C for 10 min while attached to a rotisserie in a hybridization oven rotating at full speed (ca. 16 rpm). The lysis buffer was recovered with the aspiration syringe, transferred into a 1.7 ml microcentrifuge tube, and 500 µl 100% ethanol was added to the sample. The mixture was vortexed and loaded onto a DNeasy Mini spin column and centrifuged at 10,000 × *g* for 1 min. Buffer washes were then performed according to the manufacturer's protocol and nucleic acids were recovered by eluting with 2 × 200 µl Buffer AE.

4.3.8 Effects of proteinase K digestion time on nucleic acid yields from cells

Replicate sub-samples of water were collected from the Ala Wai canal in Honolulu, Hawai'i on multiple occasions in 2013 and 2014. For some experiments, either whole or filtered (< 0.22 µm) seawater was filtered through Anotop 25 (0.02 µm) filters as described above. Other samples of 2 ml were centrifuged at 20,000 × *g* for 20 min to pellet cells. The supernatant was discarded, samples were frozen at -80°C, thawed, and total nucleic acids were extracted from the pellets and filters in triplicate using the MasterPure kit, but varying the proteinase K digestion time (15 or 30 min, 1, 3, 6, 12, 18, or 24 hr).

4.3.9 Effects of lysozyme on nucleic acid yields from cells

To test the combined effects of lysozyme and proteinase K digestions on nucleic acid recoveries, the proteinase K time series described in the previous section was conducted with and without a 30-minute lysozyme treatment. Replicate sub-samples of water collected from the Ala Wai canal were filtered or centrifuged as described above to collect plankton or pellet cells. Total nucleic acids were extracted from the filters or pellets in triplicate by one of four treatments using combinations of two variables: 1) with and without an initial lysozyme treatment (250 U sample^{-1}, Ready-Lyse Lysozyme Solution, Epicentre) and 2) with a short (15 min) or long (1 or 12 hr) proteinase K digestion. For the filter extraction with lysozyme treatment, the back flush method

described above was followed with a slight modification. After thawing the filters, 500 µl of TES (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 100 mM NaCl) with 1 µl Ready-Lyse Lysozyme (250 U μl^{-1} in TES buffer) was carefully loaded onto the filter with a 3 ml syringe. The aspiration syringe was connected to the outlet to act as a seal for the buffer, and the filter was incubated at room temperature for 30 min. An aliquot of 500 µl 2 × Tissue & Cell lysis solution (MasterPure) with 100 µg proteinase K was then added to the filter using the same injection syringe and the filter-syringe assembly incubated at 65°C for the indicated duration of proteinase K digestion. The rest of the protocol is the same as described above for extracting from filters or pelleted cells.

4.3.10 Comparison of a proprietary and a published salting-out extraction procedure

Replicate subsamples of seawater were filtered onto Anotop 25 (0.02 µm) filters or centrifuged to pellet cells as described above. One set of filters and cell pellets were extracted using the back flush method with the MasterPure kit as described above, with or without lysozyme treatment. Another set was extracted in the same manner (same volumes, temperatures and centrifugation conditions), but using defined buffers derived from the literature, and with or without lysozyme treatment. The first lysis buffer (LB3) contained a relatively high SDS concentration (50 mM Tris-HCl, 5 mM EDTA, 3% SDS; pH 8). This was intended to mimic the concentration of SDS in the MasterPure lysis buffer, (estimated by the volume of the SDS pellet after salting out) and is the same buffer formulation as that used by Yu & Mohn (1999). The second lysis buffer (LB1) contained a lower SDS concentration (50 mM Tris-HCl, 5 mM EDTA, 1% SDS; pH 8). The nucleic acid yield with LB3 was determined with no lysozyme treatment. The yield using LB1 was determined with or without an initial lysozyme treatment (Ready-Lyse Lysozyme Solution, Epicentre). In all cases, the extraction buffers were supplemented with proteinase K (100 µg ml^{-1} final concentration) just prior to use in the extractions. Salt-induced precipitation of protein-detergent complexes was accomplished by adding ammonium acetate (2M final concentration) and centrifugation at 16,000 × *g* for 10 min. The supernatant was transferred to a fresh 1.7 ml microcentrifuge tube and linear acrylamide (20 µg ml^{-1} final concentration; Ambion) was added to the sample just prior to the addition of one volume of isopropanol. Nucleic acids were pelleted by centrifugation at 16,000 × *g* for 30 min. The nucleic acid pellet was then washed with 2 × 500 µl of 70% ethanol, dried completely, and resuspended in 20 µl TE.

4.3.11 Estimating extraction efficiencies

To test for losses of DNA attributable specifically to the filter matrix (e.g., trapping and adsorption), we determined the yield of nucleic acids extracted from microorganisms collected on filters relative to the yields in the absence of a filter (liquid suspensions or pellets). Extraction efficiencies were also determined for all other filter extraction tests in which a control extraction was performed, as indicated in Table 4.1.

a. Whole seawater – Subsamples (11.7 ml) of seawater from Kāne'ohe Bay were transferred to polyallomer tubes and centrifuged at 35,000 × *g* for 105 min in a swinging bucket rotor (SW 41) in an ultracentrifuge (Beckman Optima XL-80K) to pellet cells and viruses. These conditions are sufficient to completely pellet particles with a sedimentation coefficient of ≥ 100S. While those samples were in the centrifuge, additional triplicate subsamples (11.7 ml), were filtered onto Anotop 25 (0.02 μm) filters and extracted using the back flush protocol as described above. For the centrifuged samples, the supernatant was removed and treated as described below. The resulting pellet was immediately extracted using the MasterPure kit, as described by the manufacturer's instructions for cell samples. Nucleic acids remaining in the supernatant (presumed to be dissolved) were assayed as previously described (Brum et al 2004). The supernatant was transferred to a new tube with 2 ml of 0.5 M EDTA and concentrated in a centrifugal ultrafiltration device (30 kDa cut-off; Amicon 15, Millipore) by centrifuging at 2,500 × *g* for 25 minutes. The concentrates were then washed by adding 1 ml of 1 × TE (10 mM Tris-HCl [pH 8], 1 mM EDTA) and centrifuged at 2,500 × *g* for 20 minutes. The sample was then further concentrated by transferring to a smaller centrifugal ultrafiltration unit (30 kDa cut-off; Amicon Ultra-0.5, Millipore) and centrifuged at 10,000 × *g* for 10 minutes. The concentrate was washed with 500 μl of 1 × TE, followed by centrifugation at 10,000 × *g* for 30 min. The final concentrate was recovered by inverting the filter cartridge into a fresh tube and centrifuged at 1,000 × *g* for 3 minutes.

b. Bacterial cultures – Cells in a culture of *V. vulnificus* (V93D1V) were washed and resuspended in 1 × PBS, filtered onto Anotop 25 (0.02 μm) filters in triplicate, then extracted as described above. Equal volumes of the same washed, resuspended cells were transferred into 1.7

ml microcentrifuge tubes for direct nucleic acid extraction using the MasterPure kit, following the manufacturer's instructions for suspended cell samples.

4.3.12 Quantification of nucleic acids

Fluorometric assays were used to estimate the yields of nucleic acids from all of the experiments. DNA was quantified using a Quant-iT DNA Assay Kit (Life Technologies) and a PerkinElmer 2030 Multilabel Reader. To quantify RNA, samples were first treated with TURBO DNase (Life Technologies) to eliminate cross-reaction from DNA (Steward et al 2013a). RNA was then quantified using a Quant-iT RNA Assay Kit (Life Technologies) and a TD-700 fluorometer.

4.4 RESULTS

4.4.1 Filter flow rate and capacity

The filtration rate of pure water through an Anotop 25 (0.02 μm) filter increased linearly with pressure (Fig. 4.1A) and temperature (Fig. 4.1A inset) over the ranges we tested (0.26-2.1 bar and 7-35°C). The maximum flow rate in the tested range of conditions was 25 ml min^{-1} (2 bar, 35°C). Under conditions similar to those that we use for routine collection of field samples (1.5 bar, 22°C), the flow rate was approximately 13 ± 1.4 ml min^{-1} (mean ± s.d., n = 3). From the area of the filter, we calculated a flux at standard ambient temperature and pressure (25 °C, 1 bar) of 1,180 ± 130 L m^{-2} h^{-1} (mean ± sd). The normalized flux for an Anodisc mounted in a reusable housing was about 2.4 times greater than that of the Anotop (data not shown).

Filtration of water from two environments differing in plankton concentrations resulted in increasing pressures over time at the fixed initial flow rate of approximately 14 ml min^{-1}. For the estuarine sample, pressure quickly exceeded the recommended maximum for the tubing (2.7 bar), so pump speed was reduced and filtration continued (total filtration time < 20 min for 100 ml). Pump speed was held constant for the open ocean sample and pressure steadily increased from 1.4 to 2.7 bar over the 75 min filtration time, during which approximately 1 liter of sample was filtered. By normalizing all flow rates for both samples to a constant pressure of 1.5 bar, we found that flow rate declined linearly as a function of the volume filtered and the decline was much steeper for estuarine water (Fig. 4.1B) , which had higher concentrations of plankton, compared to open ocean water (Fig. 4.1B inset). The normalized flow rate dropped by an order

of magnitude after filtering approximately 0.1 L of the estuarine sample, but was reduced by only half after filtering 1 L of open ocean water.

4.4.2 DNA yield by forward versus back flush

The back flush extraction method for Anotop 25 (0.02 μm) filters yielded significantly more DNA than the forward flush method when the sample consisted of cultured cells of a bacterium (*t* test; $P = 0.034$) or a dinoflagellate (*t* test; $P = 0.017$; Fig. 4.2). There was no significant difference in DNA recovered from Anotop 25 (0.02 μm) filters by the forward and back flush methods when the sample was whole seawater (*t* test; $P = 0.473$), 0.22 μm-filtered seawater (*t* test; $P = 0.214$), or a phage isolate (*t* test; $P = 0.920$; Fig. 4.2). Although there was no evidence of any improvement from back flushing when a sample was dominated by viral DNA (0.22 μm-filtered seawater and the phage isolate), the back flush resulted in a higher mean recovery in every trial where cells were present in the sample. The improvement was sometimes small and not statistically significant relative to the variability in yields among triplicates within a given experiment. To improve the statistical power, we normalized the yields of the extractions of all cell-containing samples by expressing them as a percent of the push-through average within each experiment. When the cell extraction experiments (cell cultures and whole seawater) were then considered in aggregate, the higher yields from back flushing were found to be significant (*t*-test: $P = 0.003$) with the difference between forward and back flush increasing in the order phage < bacterium < protist (Fig. 4.3).

4.4.3 Trapping of DNA

The 5 kb DNA ladder had similar masses of DNA in discrete bands ranging in size from 4.9 to ca. 98 kb, which together accounted for 78% of the total mass. Some additional higher molecular weight DNA (> 80 kb; 16% of the total) migrated into the gel but was not well resolved, and some DNA (6% of the total) remained trapped in the well (Fig. 4.4A). Based on bulk estimates of DNA mass by fluorometry, approximately 77±13% (mean ± sd) of the DNA passed through the filter (64% in the initial filtrate and the remainder in two washes) and 7±2% (mean ± sd) was recovered from the filter by the back flush extraction protocol (Table 4.2). The filtrates, washes, and extract together accounted for 84±12% (mean ± propagated error) of the total loaded. Analysis of the DNA size distribution in the ladders, filtrates, washes and extracts showed that

for the smaller bands (5 to 54 kb) a similar large percentage of the DNA was recovered in the filtrate (≥ 80%, Fig. 4.4) and there was no significant relationship between size and recovery (Spearman rank correlation, r = -0.021, P >0.5). An additional small percentage (≤ 11%) of the material in those bands was recovered in the extract. For larger DNA that entered the gel (from 49 to >80 kb) there was a negative correlation between size and the percentage of DNA in the filtrate (Spearman rank correlation, r = -1.00, P < 0.001). The missing material from these bands was not accounted for in the extract (Fig. 4.4B). For the form of DNA that remained trapped in the well under the specified separation conditions, 45% was accounted for in the filtrate plus wash and 46% in the extract.

4.4.4 Effect of sample volume on extraction yield

There was a positive, linear relationship between the volume of whole seawater sample added to an Anotop 25 (0.02 μm) filter and the amount of DNA extracted from it (Fig. 4.5A). The percent recoveries were variable, with an average of 101±24%, and there did not appear to be a consistent trend in extraction efficiency with volume (Fig. 4.5B). The sample-to-sample variability and the high variance in the smaller volume samples may be a result of heterogeneity in the natural sample. However, even with volumes of up to 1 L, recoveries were high relative to the reference estimate of the DNA concentration in the plankton (extraction from pelleted cells and viruses).

To minimize the problems of heterogeneity seen in the natural sample, the experiment was repeated using a culture. We again observed a positive, linear relationship between the volume of culture loaded on the AAO filters and the amount of DNA extracted (Fig. 4.5C). Although there appeared to be a trend of decreasing efficiency with increasing culture volume, this trend was not significant because of the high variance (Fig. 4.5D). The calculated minimum efficiencies for the sample ranged from 72-89% relative to the reference estimate of DNA in the culture obtained by direct liquid extraction (cells plus dissolved DNA).

4.4.5 Guanidinium- versus SDS-based buffers

When DNA was extracted directly from washed bacterial cells suspended in liquid, the DNA yield was approximately 55% lower for the DNeasy kit than for the MasterPure kit. We assumed

that the higher value obtained by the latter kit most closely represented the true amount of nucleic acids in the culture and that number was used as the control value relative to which efficiencies for all of the filter extractions were calculated, regardless of the chemistry used.

The amount of DNA recovered from the AAO filters (0.2 or 0.02 μm Anotop 25) using the MasterPure Kit was significantly higher than when using the DNeasy Kit, whether the forward or back flush method was used (*t*-test; $P = 0.008$ for forward flush on 0.02 μm Anotop 25; $P \leq 0.005$ for all other tests; Figure 4.6). The difference found between the two kits when extracting from PES filters was significant as well (*t*-test; $P \leq 0.005$). When comparing the forward versus back flush methods in the AAO filter extractions for the DNeasy kit, the amount of DNA recovered from the back-flush method was significantly higher (*t*-test; 0.02 μm, $P = 0.005$; 0.2 μm, $P = 0.038$). However, this was not true for the PES filters (0.2 μm) using the DNeasy kit (*t*-test; $P = 0.198$).

4.4.6 Proteinase K digestion time and lysozyme effects on nucleic acid yields from cells

Yields of nucleic acids from pelleted cells after 15 minutes of digestion were not improved by increasing proteinase K digestions up to 6 hours (ANOVA, post-hoc Tukey; $P > 0.05$) without addition of lysozyme treatment. The yields after 12 h of digestion were significantly higher than those at 15 min by 23-25% (ANOVA, post-hoc Tukey; $P < 0.05$). In another experiment with the addition of lysozyme treatment, yields of nucleic acids were not significantly improved from 15 min up to 18 h proteinase K digestion times, but the effect was significant after 24 h (Tukey; $P < 0.05$).

In a second test comparing interacting effects of lysozyme and proteinase K digestion, the effect of increasing proteinase K digestion time from 15 min to 12 h was not significant for RNA or DNA, except when an initial lysozyme digestion was included. An initial lysozyme digestion did not improve the yields of DNA or RNA when the proteinase K digestion was 15 min, but improved yields of DNA and RNA by 40 and 50% when the subsequent proteinase K incubation time was 12 hours (ANOVA, post-hoc Tukey; $P = 0.024$ for DNA, $P = 0.006$ for RNA).

The effects of lysozyme and proteinase K were also tested on filters containing whole seawater or filtered (< 0.22 μm) seawater with proteinase K digestion times of 15 min and 1 h. The

addition of lysozyme was significant only for the whole seawater in the 15 min proteinase K incubation, increasing yields of DNA by 23% (ANOVA, post-hoc Tukey; $P = 0.031$). Because of the variable results among individual experiments, the data from multiple experiments was considered in aggregate by normalizing the data (yields expressed as a percentage of the average nucleic acid yield from the 15 min proteinase K incubation without lysozyme treatment). For the larger pooled data set, the lysozyme treatment significantly improved yields for the whole seawater samples (*t*-test; $P = 0.012$) but not the filtered (0.22 μm) seawater samples (*t*-test; $P = 0.493$). For the same set of normalized data, the duration of proteinase K digestion (15 min, 1 h, 12 h) with or without lysozyme was not significant (ANOVA, $P = 0.402$).

4.4.7 MasterPure vs. modified salting-out extraction

In a direct comparison of the DNA yield from AAO filters using the proprietary MasterPure buffers vs. the yield using non-proprietary buffers derived from the literature, we found the yields with the MasterPure buffers to be consistently higher. The yield from the MasterPure buffers was 2.3 times higher (*t*-test; $P = 0.032$) than the first salting-out method with a higher SDS concentration (3%, LB3). However, when using the second method with a lower SDS concentration (1%, LB1), we found that the yield with the MasterPure buffers was not significantly different, with or without lysozyme (ANOVA, $P = 0.462$). When extracting from pelleted cells, the MasterPure extraction method with lysozyme treatment resulted in significantly higher yields than the LB1 modified salting-out method without lysozyme treatment (ANOVA, post-hoc Tukey; $P = 0.021$). The addition of lysozyme resulted in significantly higher yields (by 27%) for the LB1 modified salting-out method on the pelleted cells (ANOVA, post-hoc Tukey; $P = 0.018$). When the data for the pelleted cell extraction and filter extraction procedures both with and without lysozyme were pooled together, the MasterPure method and second non-proprietary method with 1% SDS were not significantly different (ANOVA, $P = 0.231$).

4.4.8 DNA and RNA extraction efficiency

In an experiment designed to constrain relative extraction efficiency, the masses of DNA and RNA recovered from total seawater plankton collected on Anotop 25 (0.02 μm) filters were 112±5% and 134±13%, respectively (t-test; for DNA P = 0.014; for RNA P = 0.006), of the

yields recovered from plankton pelleted from an equal volume of seawater by ultracentrifugation. To test whether the higher yields from the filters might have been attributable to dissolved DNA that was trapped by the filters, but which failed to pellet, we directly measured the nucleic acids in the supernatant after ultracentrifugation. When we summed the quantity of nucleic acids detected in the supernatant of the controls and the amount recovered from the extracted pellet, the total mass of nucleic acids was indistinguishable from that recovered on the filter (t-test; $P = 0.088$).

In a second experiment using a cultured bacterium, the estimated recoveries of nucleic acids from the Anotop 25 (0.02 μm) filters were 84±8% for DNA and 73±5% for RNA relative to an equal volume of cell culture extracted directly. In this experiment the cells were unwashed so the sample would contain nucleic acids from the cells plus any nucleic acids dissolved in the medium.

In addition to these two tests designed solely to test extraction efficiency, we also estimated extraction efficiencies in all other extraction tests where a control extraction was conducted (Table 4.3). For all tests in which a cell (or cell + virus) pellet was used as a control, the average extraction efficiency was 102±19% for DNA. When the bacteria culture extraction data were combined, the average percent recovery of DNA was 94±9% (Table 4.3; B, H, I). The average percent recoveries of DNA and RNA for all experiments using liquid culture extracts as a control were 93±24% and 82±12%, respectively

4.5 DISCUSSION

In this study, we 1) characterized the flow characteristics and capacity of commercially produced AAO filters, 2) empirically verified the hypothesis that both physical trapping and adsorption can be important sources of loss when extracting nucleic acids from microorganisms harvested on these filters, and 3) demonstrated that a back flushing protocol previously reported (Steward and Culley 2010) can effectively avoid those losses. The results of our experiments to test the effects of extraction chemistry, flow direction, and enzyme treatment on yields of nucleic acids provide insights that may be useful for researchers developing extraction protocols for other types of samples or filters. We first discuss our results and their implications then provide our

recommended protocols for extractions of nucleic acids from AAO filters using proprietary or non-proprietary buffers (Fig. 4.6).

4.5.1 Filter flow characteristics

Our estimate of the flux of 0.02 μm Anotop 25 filter units is only about 30% of the reported inherent flux for the Anopore membranes,in the filter units (Manufacturer's technical literature). We presume that the lower flux is a result of the manner in which the membranes are mounted and supported in the housing. We observed higher flux rates for an Anodisc membrane mounted in a reusable housing (ca. 70% of the reported Anopore flux), which implies that there is something specific about the manufacture of the Anotop units that leads to occlusion of a large fraction of the pores. Nevertheless, we find that the flux is sufficient for our applications and the small disposable units are especially convenient for field sampling and subsequent extraction. Another type of filter (modified hydrophilic PVDF) with the same nominal 20 nm pore size is offered in a very similar format (Optiscale-25 with Viresolve NFP membrane, Millipore). We considered these as an alternative to the Anotop 25, but they have a normalized flux that is four times lower (unpubl. observation) and cost several times more per filter, so we did not evaluate them further.

Our tests of the Anotop 25 filter capacity illustrate how different types of samples will influence the amount of water that can be filtered and how long one can expect filtration to take. Extrapolation of the lines in Fig. 4.1B suggest that the practical capacity of the 0.02 μm Anotop 25 filter was about 100 ml for the highly productive, turbid waters of an urban estuary, but around 2 L for lower productivity open ocean waters. This is consistent with the observation that 1 to 2 L of surface water at Station ALOHA can be filtered through a 0.02 μm Anotop 25, with the amount varying seasonally (C. Schvarcz, personal communication). In one experiment for this study, we were able to filter a liter of water from Kāne'ohe Bay through a 0.02 μm Anotop 25, but on other occasions we have only managed to filter 200 to 550 ml of water from the same location in the bay (Culley and Steward 2007a). This is consistent with the variable conditions in the bay, which range from nutrient poor with low biomass during periods of low rainfall, to nutrient enriched with large plankton blooms following terrestrial runoff from rain events (De Carlo et al 2007).

4.5.2 Physical trapping

The relationship we observed between DNA size and filter capture is consistent with other studies (Rudinger and Blazek 2002), but we were still surprised that there was no discernible loss of DNA with a forward flush extraction of the bacteriophage or a natural viral assemblage. Even though viruses have comparatively small genomes, the persistence length of double-stranded DNA (ca. 35 nm; ref. Brinkers et al 2009) is greater than the pore size and we anticipated significant losses for the viral DNA. The relative high efficiency with which DNA of all tested microorganisms and a DNA ladder passed through the filters is likely a result of DNA elongation driven by the converging flow fields as the molecules approach the pores (Latulippe and Zydney 2011).

Our experiments with the DNA ladder suggest that there was size selection occurring during filtration of DNA in the range typical of viruses, but only for DNA > 50 kb. Since the virus isolate we used for testing was ca. 38 kb (Nigro et al 2012) and the majority of viral genomes in seawater are ≤ 50 kb (Steward et al 2000), this would explain why we found a small, but statistically insignificant, difference in overall yield of DNA from viruses in the forward vs. back flushing tests. We presume that the material in the 5 kb ladder that did not enter the gel was of unusually high molecular weight (Gurrieri et al 1999). We do not know why this material was recovered with higher efficiency than material of intermediate size in the 60 to ca. 100 kb size range, but based on the observed patterns of recovery, one plausible scenario is that smaller DNA molecules readily pass the filter and end up primarily in the filtrate, intermediate sizes become trapped within the pores of the filter and remain associated with the membrane, and a greater proportion of very large molecules are trapped on the upper surface of the filter, such that they can be more efficiently released into the extract by back flushing. Regardless of the mechanism, our data suggest that there is size selection that occurs when DNA is forced through the AAO filter membrane.

In addition to the effect of size, differences in DNA topology (circular vs. linear or degree of supercoiling) can have large influence on filter passage (Higuchi et al 1996, Watson et al 2006). A likely consequence is that viruses having circular or very large genomes could be underrepresented in an extract if the nucleic acids are forced to pass through a small pore size

filter. Forward flushing will also generally reduce the yield of DNA from cells. Therefore, for most routine extractions we recommend back flushing to minimize the chance of representational bias. This is particularly important for quantitative analyses of community composition. One must bear in mind, however, that the trapping and release of at least some fraction of dissolved DNA over a broad size range means that not all of the nucleic acids recovered from a filter will necessarily derive from intact viruses or cells. This is an issue that has been recognized even with large pore size filters (Boström et al 2004, Liang and Keeley 2013, Sørensen et al 2013).

4.5.3 Adsorption

The reversible adsorption of nucleic acids to silica in the presence of the chaotropic salts (e.g., idodide or guanidinium ions) is a well-known phenomenon (Boom et al 1990, Vogelstein and Gillespie 1979) and serves as the basis for some of the most popular DNA and RNA purification kits. However, the binding of nucleic acids to AAO in the presence of guanidinium appears to be irreversible (Gerdes et al 2001). This phenomenon might be exploited for solid-phase amplifications (Dames et al 2006), but for our application it is a possible source of loss.

The changes in yields we observed in response to varying 1) extraction buffer chemistry, 2) direction of buffer flow, 3) filter material or 4) filter pore size were are all consistent with nucleic acids binding to the AAO in the presence of guanidinium. In contrast to the relatively small losses of DNA when forward flushing in the SDS-based buffer (apparently a function of physical trapping), the loss when forward flushing in the guanidinium-based buffer was dramatic, but only for AAO filters. Presumably forward flushing was a particular problem with the guanidinium buffer, because it increases the contact of the DNA with the filter, dramatically increasing adsorption. The loss was less pronounced, but still significant, when forward flushing with the 0.2 μm Anotop 25 filter, which we attribute to decreases in both physical trapping and lower adsorption resulting from the larger pore size. The absence of significant loss when forward flushing through PES membranes, regardless of buffer, suggests that the adsorptive losses are specific as hypothesized.

4.5.4 Buffer chemistry

Although back flushing effectively eliminated the severe adsorption loss when extracting DNA from AAO filters with the guanidinium-based DNeasy kit, we found that the yields were consistently lower than with the alternative SDS-based MasterPure kit by about 50%. In another preliminary experiment we found that the yields between the two kits we used were similar for direct extraction of cultured cells (data not shown), so the difference between the kits appears to vary. The lower yields obtained with the DNeasy kit are not likely a result of exceeding the binding capacity of the silica, since the reported capacity (30 μg) is ten times the yield we achieved. Although we do not have an explanation for them, our observations are consistent with others who have also reported relatively poor yields using the Qiagen spin-column kit compared to a liquid SDS-proteinase K extraction protocol (Boström et al 2004).

We used two commercial extraction kits for our experiments out of convenience and because of their popularity. However, we recognize that a reliance on proprietary kits is not always desirable and their availability and cost can be an issue for many labs. We therefore decided to compare the MasterPure kit with non-proprietary buffers derived from the literature. We tested a buffer with 3% SDS initially, because it appeared to be most similar to concentration of SDS in the MasterPure buffer (roughly estimated from the volume of the SDS pellet after salting out). That produced poor results, so we reduced the SDS concentration to 1%. This appears to have the advantage of reducing the inactivation of proteinase K itself, while still sufficiently denaturing other proteins to achieve near maximal protein degradation rates (Hilz et al 1975). A co-precipitant was added to the non-proprietary protocol to enhance the yield of nucleic acids during the final alcohol precipitation step. A co-precipitant was not added to the MasterPure protocol, because the buffers already include an unspecified reagent to enhance recoveries of low amounts of nucleic acids (Watson et al 1998). The yields using the MasterPure kit buffers were consistently higher than those achieved with the non-proprietary buffers, but the difference between the second formulation with 1% SDS (LB1) and the MasterPure buffer was relatively small and not statistically significant. The failure to reject the null hypothesis (no difference in yields between the two methods), may be a result of the relatively large sample-to-sample variance among the triplicate samples, but for investigators averse to using proprietary buffers or for whom access to the MasterPure kit is difficult, our protocol with a non-proprietary 1% SDS

lysis buffer seems a reasonable alternative. We prefer using the commercial kit, however, because of convenience, consistency in preparation in dedicated clean facilities, and a cost per extraction that is still quite reasonable compared to spin-column kits.

4.5.5 Extraction efficiency

A previous comprehensive study of DNA extraction from microorganisms collected on 0.2 μm PES membranes (Boström et al 2004) identified a number of key variables affecting extraction yields. Two of these were enzymatic digestion steps (lysozyme and proteinase K). Although our tests are not directly comparable to those of Boström et al (2004), because of differences in experimental design, the results of the two studies are consistent and suggest that using an initial lysozyme digestion followed by a long incubation with proteinase K and SDS (up to 24 h) can significantly increase yields of DNA from pelleted cells. We further show that these conditions also increased yields of RNA and found that increasing proteinase K digestion time in the absence of lysozyme was of marginal value. We found some degradation of RNA with these increased proteinase K digestion times (data not shown), and therefore recommend only increasing the proteinase K incubation to about 1 h (or at most ≤ 6 h). As expected, the lysozyme treatment did not result in a significant increase of yields in the absence of cells, so this step could be eliminated if one wishes to extract nucleic acids only from viruses. For quantitative analyses of community composition, on the other hand, we would recommend including a lysozyme digestion, since we found no evidence that it reduces total yields and it may facilitate the lysis of some populations that would be otherwise underrepresented.

Because our primary goal in this study was to test specific hypotheses about the interaction between nucleic acids and the 0.02 μm AAO filters, we employed the simpler protocol with no initial lysozyme digestion and a 15 min proteinase K digestion, the standard conditions recommended by the manufacturer of the MasterPure kit. The yields under these conditions are likely not as high as they could be, but were sufficient for the comparative extraction tests to determine the extent to which the filter matrix influenced yields. However, even these simple comparisons of yields between microorganisms on filters versus those in suspended in liquid or in a pellet are complicated by the presence of dissolved DNA in the samples, a variable fraction of which is physically trapped by the filters.

Despite our inability to definitively distinguish contributions of truly dissolved nucleic acids on the filters or in microorganism pellets, we can constrain the relative extraction efficiencies. In cases where the direct extraction of suspended microorganisms included all of the dissolved DNA in a sample, our filter extractions were underestimates (93±24%), and in cases where the direct extraction was first depleted in dissolved DNA by pelleting, our filter extraction efficiencies were overestimates (102±19%). Given these ranges for the under- and overestimates, the relative extraction efficiency appears to be very high with yields close to 100% of those from direct extraction. This suggests that our recommended protocol effectively avoids interference from the filter matrix.

4.5.6 Conclusions

We have shown that Anotop 25 filters with a 0.02 μm AAO membrane can be used to capture microorganisms from tens to thousands of milliliters of water, depending on the load of microorganisms in the sample, and at flow rates on the order of 10 to 20 ml min^{-1} or more depending on water temperature and desired pressure. When extracting nucleic acids from cells and viruses trapped on the filters, yields can be significantly affected by the extraction buffer, filter material, direction of flow, and nucleic acid size. The most serious losses occurred when a guanidinium-containing extraction buffer was combined with AAO filters and the extract was forced through the filter. Among the conditions tested, the highest relative yields were achieved using the buffers of the MasterPure Total DNA and RNA extraction kit, and back flushing the filter (yields using non-proprietary buffers were similar). Using this preferred protocol, which employs an SDS-based extraction buffer, we showed that nucleic acids could be extracted from microorganisms collected on AAO filters at the same efficiency as from microorganisms in pellets or in liquid suspension. For maximum absolute yields from bacteria, the inclusion of a lysozyme digestion step and increased proteinase K digestion (1-6 h) is recommended. Because yields of nucleic acids were linear over a broad range of mass and volume filtered, collection of microorganisms from liquid samples onto AAO filters is a simple and effective way to increase the mass of nucleic acids harvested without sacrificing efficiency.

4.6 ACKNOWLEDGMENTS

The authors would like to thank O. Nigro for kindly providing cultures of the bacterium *Vibrio vulfnificus* (V93D1V) and its phage (Vibrio phage VvAW1), and C. Schvarcz for providing cultures of the *Gymnodinium* dinoflagellate strain (AL-DI06). This work was funded by NSF grants to GFS and AIC (OCE 08-26650), GFS (ANT 09-44851), and by the Center for Microbial Oceanography: Research and Education (EF 04-24599).

Table 4.1. Summary of samples extracted, variables tested, and controls used. Experiments for which extraction efficiency was calculated are assigned an identification (ID) letter (A-K) for cross-referencing to Table 3. For buffer chemistry, SDS refers to extractions using the MasterPure kit, GuHCl refers to extractions using the DNeasy kit, and LB1 and LB3 refer to non-proprietary lysis buffers. The SDS vs. LB1 extraction was conducted with and without lysozyme. For environmental samples, locations were Ala Wai Canal (AW), Station ALOHA (ALOHA), and Kaneohe Bay (KB). Processing consisted of filtration (F) or centrifugation (C). Controls for extraction efficiency tests consisted of direct extractions of cells or viruses, including the liquid in which they were suspended (Liquid) or after centrifuging to sediment primarily cells (*) or cells and viruses (**) as indicated (Pelleted).

Filter Tests	Variables	ID	Sample Type	Volume Processed (ml)	Controls
Flow rate	Temperature (7-35°C) and Pressure (0.26-2.1 bar)		Seawater, AW	96-97 (F)	-
			Seawater, ALOHA	1000-1100 (F)	-
			Pure water	Variable	-
Extraction efficiency	Flush direction		Seawater (<0.22 µm), KB	200 (F)	-
			Seawater, KB	200 (F)	-
		A	Phage	2 (F)	Liquid
		B	Bacteria culture	3 (F)	Liquid
		C	Protist culture	1 (F)	Pelleted*
	Percent Recovery	D; E	Seawater, KB	11.7 (F)	Liquid; Pelleted**
	Volume loaded	F; G	Seawater, KB	10, 50, 250, 500, 1000 (F)	Liquid; Pelleted**
		H	Protist culture	1, 2, 4 (F)	Liquid
	Buffer chemistry (SDS vs. GuHCl) and filter material (AAO, PES)	I; J	Bacteria culture	3 (F)	Liquid
	Proteinase K, Lysozyme		Seawater, AW	2 (C), 50 (F)	Pelleted*
	Buffer chemistry (SDS vs. LB3)		Seawater, KB	500 (F)	-
	Buffer chemistry (SDS vs. LB1)	K	Seawater, AW	50 (F)	Pelleted**
DNA trapping	DNA size		dDNA, 5kb ladder	0.5 (F)	Unfiltered ladder

Table 4.2. Recovery of DNA in a 5kb ladder recovered after filtration through AAO (0.02 μm Anotop 25) filters. Percentages of DNA in the filtrate and washes represent material passing through the filters. The extract represents material recovered from the filters using a back flush protocol. Values are means and standard deviations of triplicate assays.

Sample	**Percent Recovery**
Filtrate	64 (±12)
Wash 1	10 (±2)
Wash 2	2 (±0.5)
Wash 3	0.6 (±0.3)
Extract	7 (±2)
Total	84 (±12)

Table 4.3. Recovery of nucleic acids from microorganisms collected on AAO (0.02 μm Anotop 25) filters expressed as a percentage of controls. Extraction efficiency controls consisted of direct extractions of microorganisms, including the liquid in which they were suspended (Liquid) or after centrifuging to pellet primarily cells (*) or cells and viruses (**) as indicated (Pelleted). For environmental samples, locations were Ala Wai Canal (AW) and Kaneohe Bay (KB). The extraction methods refer to the standard MasterPure protocol with 15 min Proteinase K incubation and no lysozyme treatment (MP), unless otherwise indicated by 60 (60 min Proteinase K incubation), + Lys (lysozyme treatment) or LB1 (non-proprietary extraction method). Values are mean and standard deviation of triplicate assays.

Control	Sample	ID	Extraction Method	Variable	Percent Recovery	
					DNA	RNA
Liquid	Phage	A	MP	Flush direction	73 (±10)	-
	Bacteria	B	MP	Flush direction	96 (±7)	-
		I	MP	Chemistry 1	84 (±8)	73 (±5)
		J	MP	Chemistry 2	102 (±4)	91 (±4)
	Protist	H	MP	Volume 1 ml	89 (±12)	-
		H	MP	Volume 2 ml	82 (±14)	-
		H	MP	Volume 4 ml	72 (±11)	-
	Seawater, KB	D	MP	Percent recovery	117 (±23)	-
		G	MP	Volume 10 ml	130 (±52)	-
		G	MP	Volume 50 ml	68 (±53)	-
		G	MP	Volume 250 ml	93 (±30)	-
		G	MP	Volume 500 ml	113 (±36)	-
		G	MP	Volume 1 L	95 (±33)	-
All Liquid					***93 (±24)***	***82 (±12)***
Pelleted*	Protist	C	MP	Flush direction	144 (±12)	-
	Seawater, AW	K	MP, 60	Chemistry +Lys	91 (±11)	-
		K	MP, 60 + Lys	Chemistry +Lys	92 (±14)	-
		K	LB1, 60	Chemistry +Lys	110 (±38)	-
		K	LB1, 60 + Lys	Chemistry +Lys	65 (±19)	-
Pelleted**	Seawater, KB	E	MP	Percent recovery	112 (±7)	134 (±15)
		F	MP	Volume 10 ml	132 (±33)	-
		F	MP	Volume 50 ml	69 (±49)	-
		F	MP	Volume 250 ml	95 (±6)	-
		F	MP	Volume 500 ml	115 (±7)	-
		F	MP	Volume 1 L	97 (±15)	-
All Pelleted					***102 (±19)***	***134 (±15)***
Pelleted** + Supernatant	Seawater, KB	E	MP	Percent recovery	106 (±7)	-

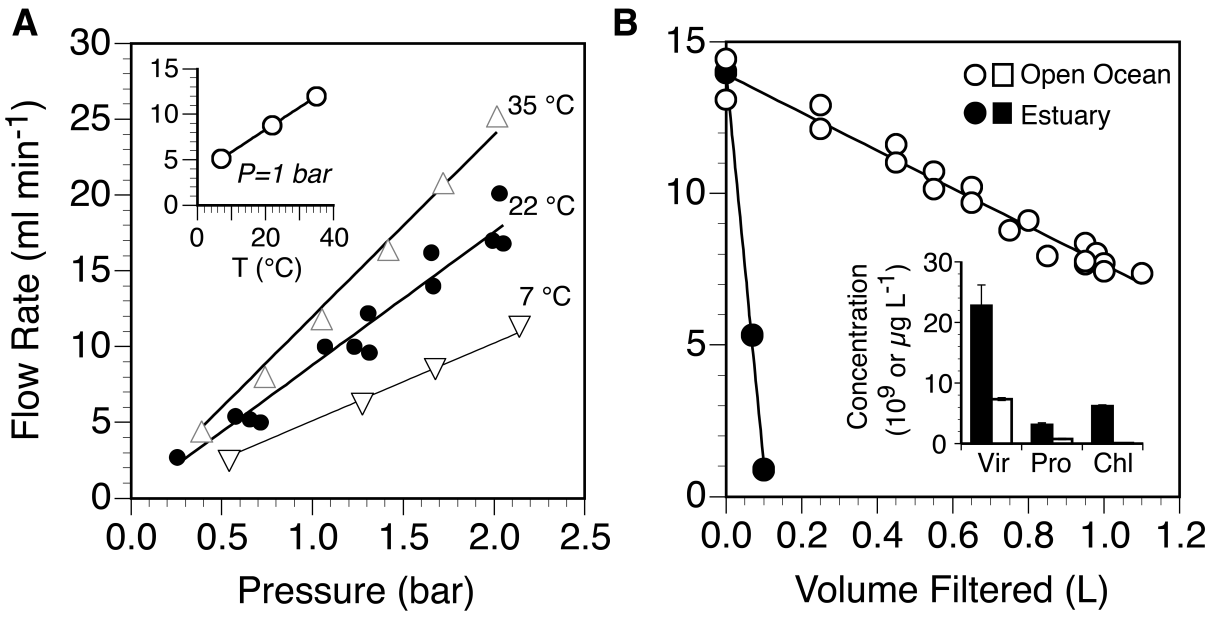


Figure 4.1. Flow rate of pure water (A) or environmental water samples (B) through 0.02 μm Anotop 25 filters. The flow rate of pure water is shown as a function of pressure at various temperatures (Panel A) and as a function of temperature at a fixed pressure of 1 bar (Inset, Panel A). Data for 22 °C is from triplicate filters to show variability among filters. The flow rates of environmental samples from the open ocean and an urban estuary (Panel B) are normalized to a pressure of 1.5 bar and plotted as a function of the volume of sample filtered. Also shown are the concentrations of viruses (Vir) and prokaryotes (Pro) in units of 10^9 L^{-1}, and chlorophyll *a* (Chl) in units of μg L^{-1}, in the two samples (Inset, Panel B).

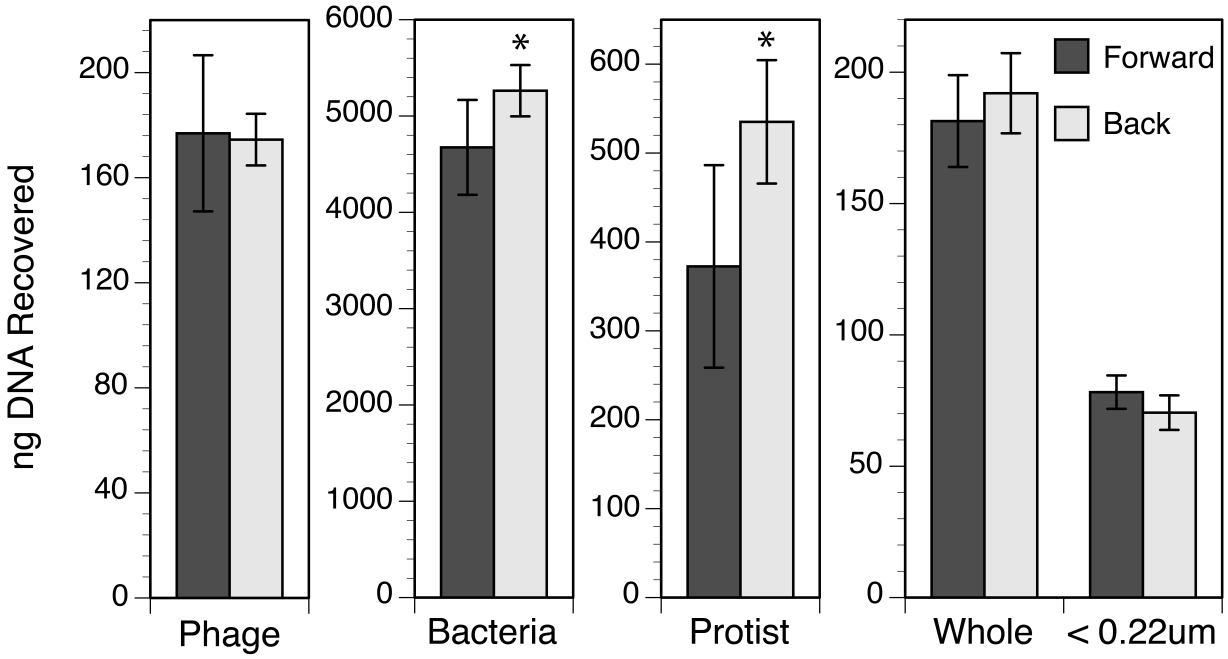


Figure 4.2. Total DNA recovered from cultures and whole or filtered (<0.22 μm) seawater using the forward- and back-flush methods of extraction. Values shown are mean and standard deviation from triplicate assays conducted once for the phage and seawater, and two times (effectively six replicates) for the bacteria and protist. An asterisk denotes the situations where the recoveries for the two methods were significantly different.

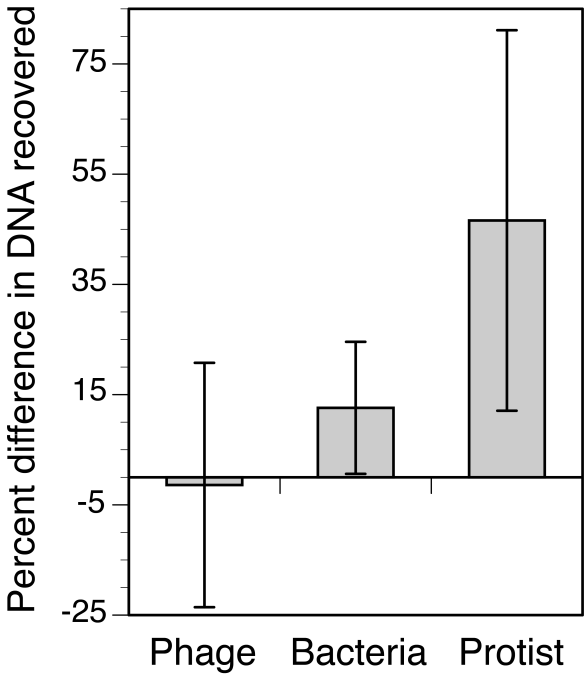


Figure 4.3. Improvement in DNA recovery for phage, bacteria, and protist cultures when using the back-flush relative to the forward-flush method. Shown are the average yields from the back-flush method expressed as a percentage of the forward-flush average for each experiment. Error bars are the standard deviations from triplicate assays conducted once for the phage and two times (effectively six replicates) for the bacteria and protist.

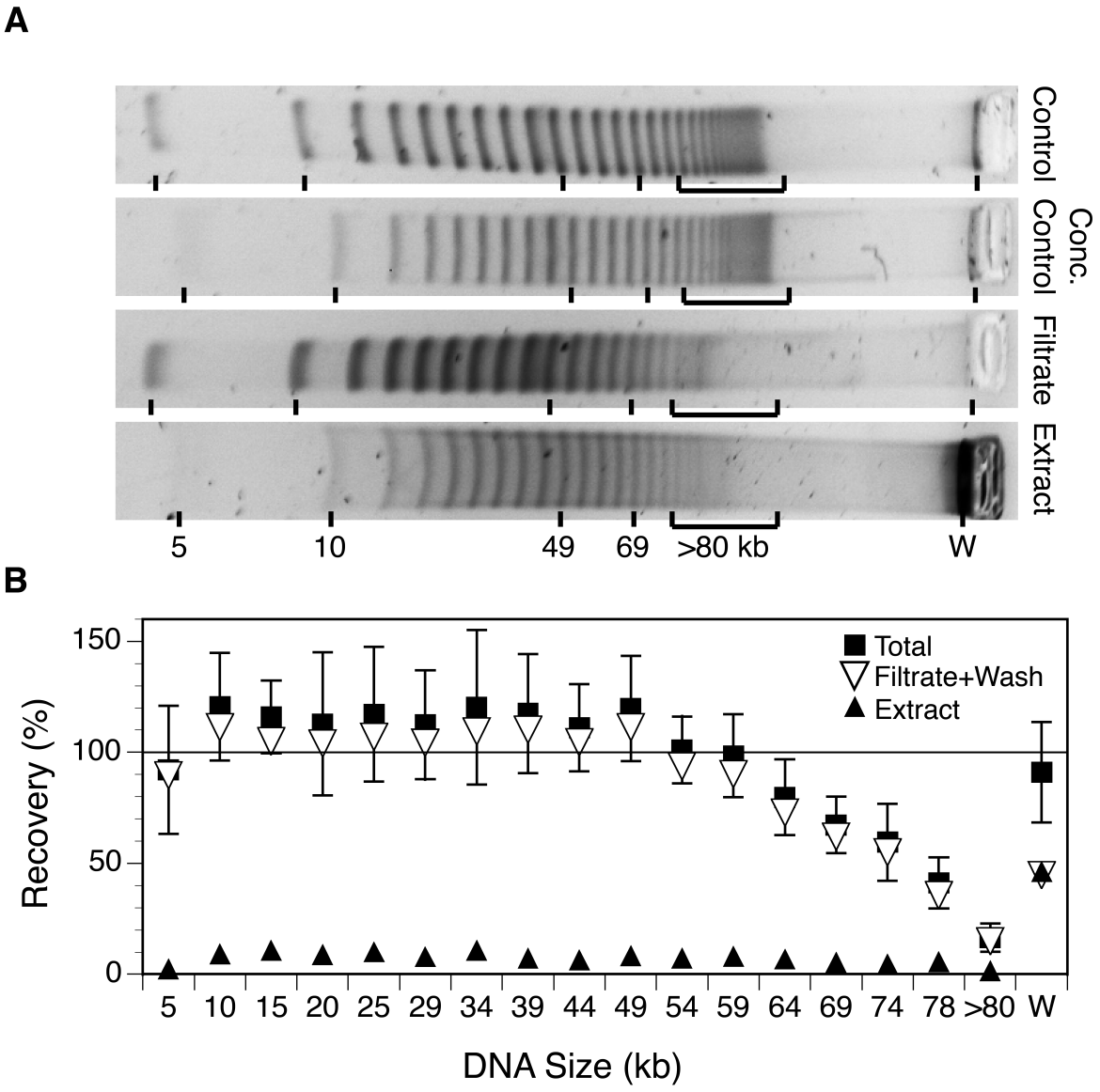


Figure 4.4. Changes in the size distribution of the DNA in a 5 kb ladder as a result of filtration through, and extraction from, a 0.02 µm Anotop filter shown as pulsed-field gel images (A) and as calculated percent recoveries for each band size (B). In panel A, representative gel images are shown of unconcentrated and concentrated control ladders, the material passing through the 0.02 µm filter, and the material extracted from the filter. An approximately equal mass of material was loaded in each lane on the same gel and separated using conditions indicated in the text. Recoveries for each size band in panel B were calculated as a percentage of the intensity in the corresponding control ladder after scaling intensity to the total DNA recovered as determined by fluorometry.

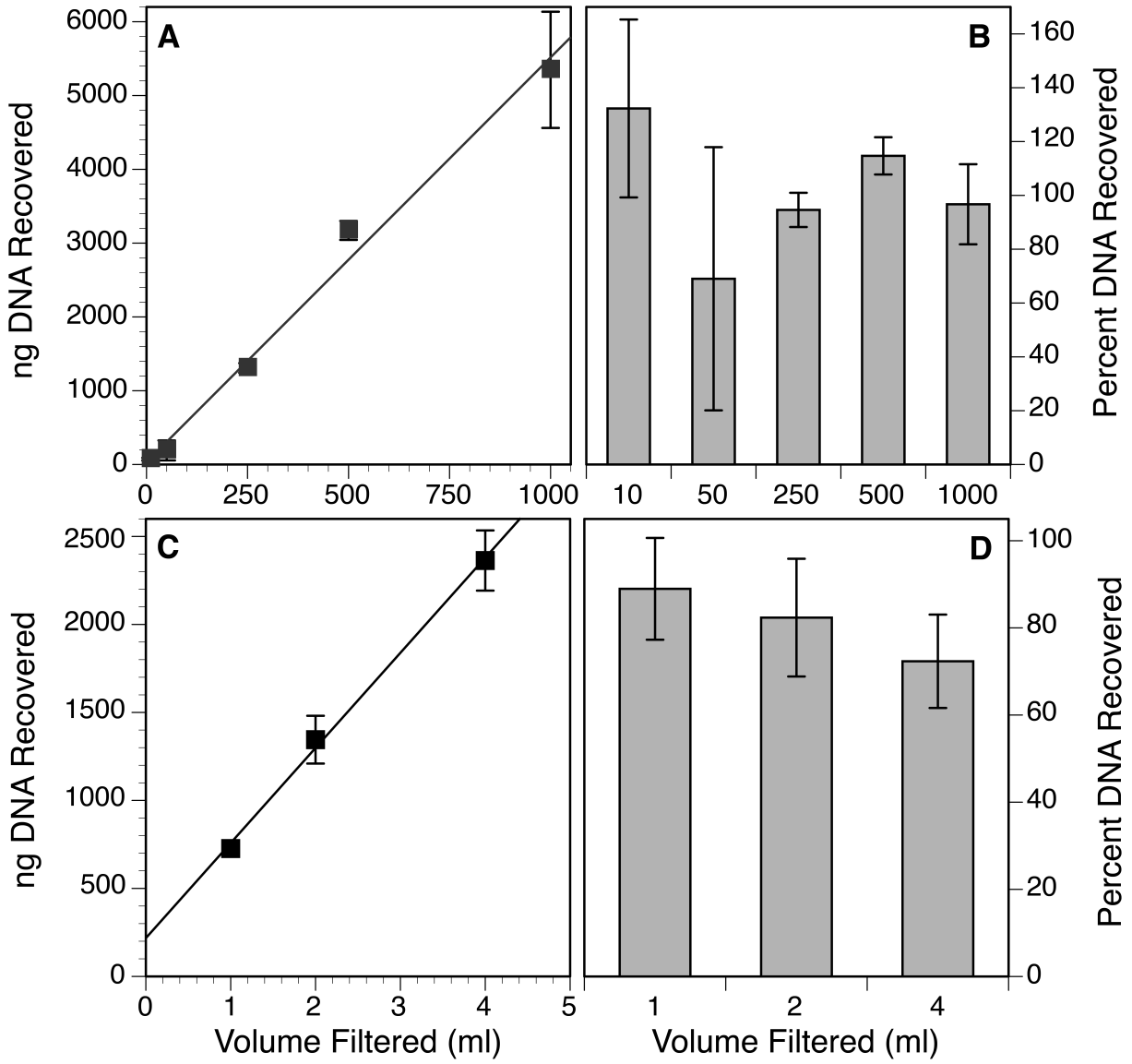


Figure 4.5. Recovery of DNA from AAO filters as a function of volume of sample filtered expressed as either the total (A) or the estimated percentage (B) of DNA recovered from seawater, or the total (C) or percentage (D) of DNA recovered from a dinoflagellate culture. Percentage recoveries for the seawater samples were calculated relative to an estimate of the concentration of planktonic DNA in the seawater obtained by direct extraction of cells and viruses that were pelleted by ultracentrifugation from a subsample of the same water. To calculate percent recoveries from the dinoflagellate culture, the concentration of DNA in the culture was estimated by direct liquid extraction of a culture sub-sample. Values shown are mean and standard deviation of triplicate assays.

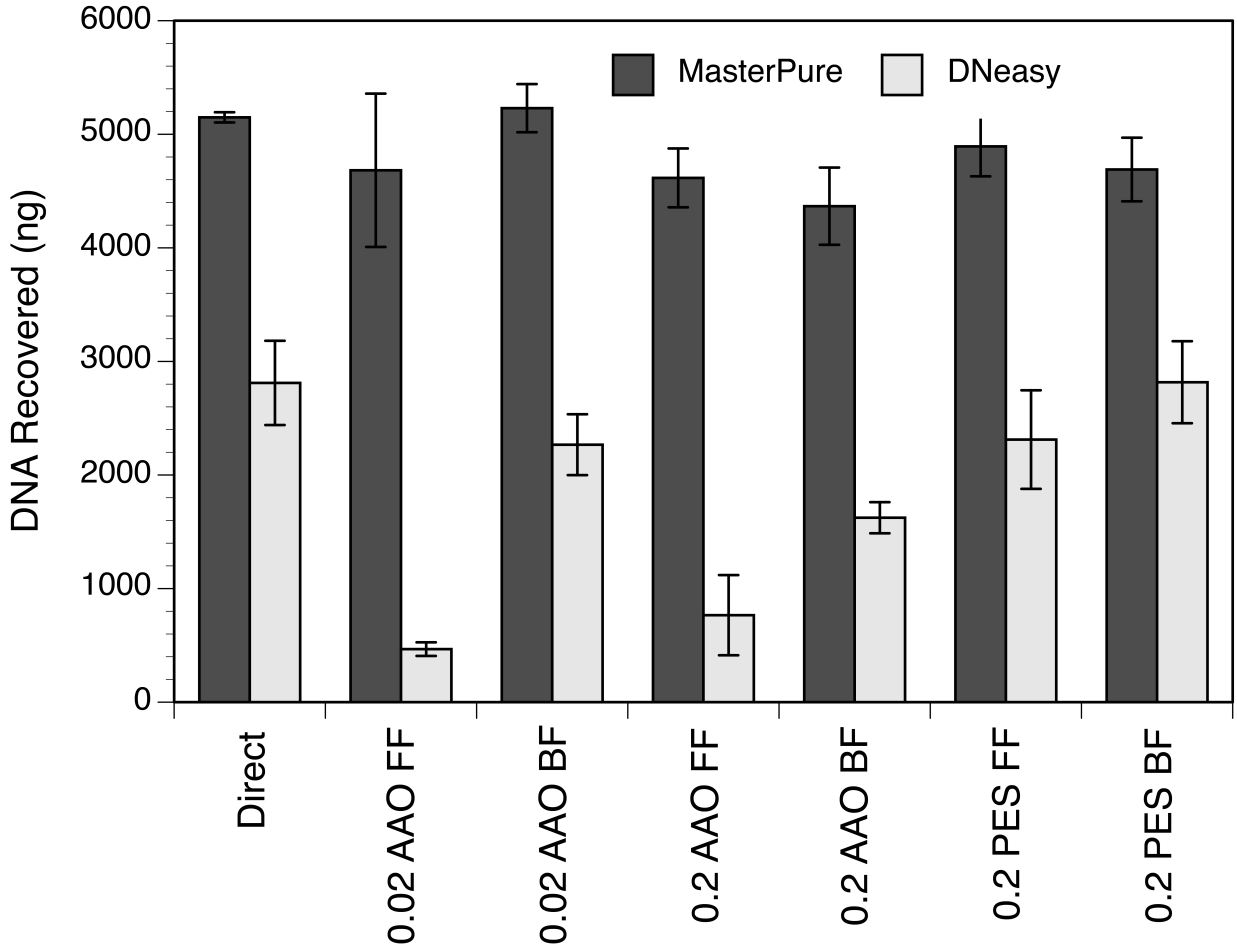
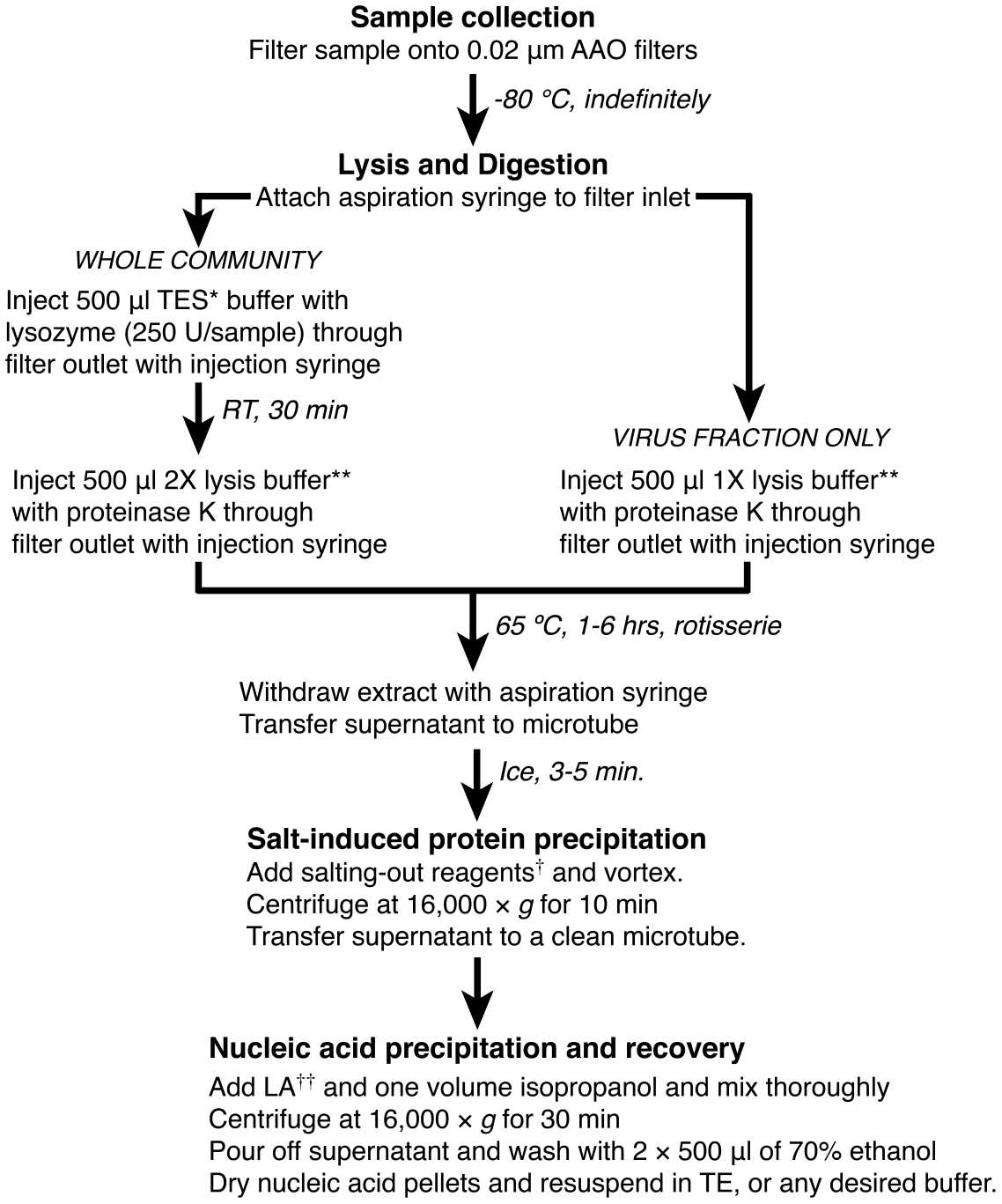


Figure 4.6. Total mass of DNA recovered by direct extraction of a liquid bacterial culture or extraction of an equivalent volume of culture collected on three types of filters (0.02 µm or 0.2 µm AAO, and 0.2 µm PES). These assays were conducted with the forward-flush (Forward) and back-flush (Back) methods using either the MasterPure Complete DNA & RNA Purification kit (Epicentre) or DNeasy Blood & Tissue Kit (Qiagen). Values shown are mean and standard deviation of triplicate assays.



* TES (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 100 mM NaCl)
** Either MasterPure Tissue & Cell Lysis Solution (MasterPure) or non-proprietary lysis buffer (1X: 50 mM Tris-HCl, 5 mM EDTA, 1% SDS; pH 8) with proteinase K at 100 μg ml^{-1}
† MPC Protein Precipitation Reagent or ammonium acetate (2 M final conc.)
†† Linear acrylamide (20 μg ml^{-1} final conc.); only if conducting non-proprietary extraction

Figure 4.7: Recommended extraction protocol for obtaining nucleic acids from AAO filters.

4.7 REFERENCES

Aljanabi SM, Martinez I (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25:** 4692-4693.

Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, Van der Noorda J (1990). Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* **28:** 495-503.

Boström KH, Simu K, Hagstrom Å, Riemann L (2004). Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol Oceanogr Meth* **2:** 365-373.

Brinkers S, Dietrich HRC, de Groote FH, Young IT, Rieger B (2009). The persistence length of double stranded DNA determined using dark field tethered particle motion. *The Journal of Chemical Physics* **130:** 215105.

Brum JR, Steward GF, Karl DM (2004). A novel method for the measurement of dissolved deoxyribonucleic acid in seawater. *Limnology and Oceanography: Methods* **2:** 248-255.

Brum JR, Schenck RO, Sullivan MB (2013). Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME J***:** 1-14.

Culley AI, Steward GF (2007a). New genera of RNA viruses in subtropical seawater inferred from polymerase gene sequences. *Appl Environ Microbiol* **73:** 5937-5944.

Culley AI, Steward GF (2007b). New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Appl Environ Microbiol* **73:** 5937-5944.

Culley AI, Asuncion BA, Steward GF (2009a). Detection of inteins among diverse DNA polymerase genes of uncultivated members of the Phycodnaviridae. *ISME J* **3:** 409-418.

Culley AI, Asuncion BF, Steward GF (2009b). Detection of inteins among diverse DNA polymerase genes of uncultivated members of the Phycodnaviridae. *Isme J* **3:** 409-418.

Dames S, Bromley LK, Herrmann M, Elgort M, Erali M, Smith R *et al* (2006). A Single-Tube Nucleic Acid Extraction, Amplification, and Detection Method Using Aluminum Oxide. *The Journal of Molecular Diagnostics* **8:** 16-21.

De Carlo EH, Hoover DJ, Young CW, Hoover RS, Mackenzie FT (2007). Impact of storm runoff from tropical watersheds on coastal water quality and productivity. *Appl Geochem* **22:** 1777-1797.

Ferrara GB, Murgia B, Parodi AM, Valisano L, Cerrano C, Palmisano G *et al* (2006). The assessment of DNA from marine organisms via a modified salting-out protocol. *Cell Mol Biol Lett* **11:** 155-160.

Fuhrman J, Comeau D, Hagstrom A, Chan A (1988). Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Applied and Environmental Microbiology* **54:** 1426-1429.

Gerdes JC, Marmaro JM, Roehl CA (2001). Nucleic acid archiving. Xtrana: US.

Grimberg J, Nawoschik S, Belluscio L, McKee R, Turck A, Eisenberg A (1989). A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. *Nucleic Acids Res* **17:** 8390.

Gurrieri S, Smith SB, Bustamante C (1999). Trapping of megabase-sized DNA molecules during agarose gel electrophoresis. *Proc Natl Acad Sci USA* **96:** 453-458.

Higuchi A, Kato K, Hara M, Sato T, Ishikawa G, Nakano H *et al* (1996). Rejection of single stranded and double stranded DNA by porous hollow fiber membranes. *Journal of membrane science* **116:** 191-197.

Hilz H, Wiegers U, Adamietz P (1975). Stimulation of proteinase K action by denaturing agents: application to the isolation of nucleic acids and the degradation of proteins. *Eur J Biochem* **56:** 103-108.

Jani AMM, Losic D, Voelcker NH (2013). Nanoporous anodic aluminium oxide: Advances in surface engineering and emerging applications. *Progress in Materials Science* **58:** 636-704.

John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S *et al* (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* **3:** 195-202.

Laitinen J, Samarut J, Holtta E (1994). A nontoxic and versatile protein salting-out method for isolation of DNA. *Biotechniques* **17:** 316, 318, 320-312.

Latulippe DR, Zydney AL (2011). Separation of plasmid DNA isoforms by highly converging flow through small membrane pores. *Journal of Colloid and Interface Science* **357:** 548-553.

Lawrence JE, Steward GF (2010). Purification of viruses by centrifugation. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. Aquatic Sciences Limnology and Oceanography: Waco, TX. pp 166-181.

Liang Z, Keeley A (2013). Filtration Recovery of Extracellular DNA from Environmental Water Samples. *Environmental Science & Technology* **47:** 9324-9331.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16:** 1215.

Needham DM, Chow C-ET, Cram JA, Sachdeva R, Parada A, Fuhrman JA (2013). Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. *ISME J* **7:** 1274-1285.

Nigro OD (2012). Environmental controls on *Vibrio vulnificus* and other pathogenic vibrios in tropical and subtropical coastal waters. Ph.D. thesis, University of Hawaii at Manoa.

Nigro OD, Culley AI, Steward GF (2012). Complete genome sequence of bacteriophage VvAW1, which infects Vibrio vulnificus. *Stand Genomic Sci* **6:** 415-426.

Patel A, Noble RT, Steele JA, Schwalbach MS, Hewson I, Fuhrman JA (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nature Protocols* **2:** 269-276.

Percival SL, Chalmers RM, Embrey M, Hunter PR, Sellwood J, Wyn-Jones P (2004). *Microbiology of waterborne diseases*. Elsevier: London.

Pichard SL, Frischer ME, Paul JH (1993). Ribulose-Bisphosphate Carboxylase Gene Expression in Subtropical Marine Phytoplankton Populations. *Marine Ecology Progress Series* **101:** 55.

Rudinger G, Blazek ER (2002). Fluid mechanics of DNA double-strand filter elution. *Biophysical journal* **82:** 19-28.

Somerville CC, Knight IT, Straube WL, Colwell RR (1989). Simple, rapid method for direct isolation of nucleic acids from aquatic environments. *Appl Environ Microbiol* **55:** 548-554.

Sørensen N, Daugbjerg N, Richardson K (2013). Choice of Pore Size Can Introduce Artefacts when Filtering Picoeukaryotes for Molecular Biodiversity Studies. *Microb Ecol* **65:** 964-968.

Steward GF, Montiel JL, Azam F (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnol Oceanogr* **45:** 1697-1706.

Steward GF, Culley AI (2010). Extraction and purification of nucleic acids from viruses. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*: Waco, TX. pp 154-165.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G (2013a). Are we missing half of the viruses in the ocean? *Isme J* **7:** 672-679.

Steward GF, Culley AI, Wood-Charlson EM (2013b). Marine Viruses. In: Levin SA (ed). *Encyclopedia of Biodiversity*, Second edn. Elsevier: Waltham. pp 127-144.

Suzuki MT, Preston CM, Beja O, de la Torre JR, Steward GF, DeLong EF (2004). Phylogenetic screening of ribosomal RNA gene-containing clones in Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey Bay. *Microb Ecol* **48:** 473-488.

Valentin K, John U, Medlin LK (2005). Nucleic acid isolation from environmental aqueous samples. In: Zimmer EA, Roalson EH (eds). *Molecular evolution: producing the biochemical data*. Elsevier Academic Press: San Diego. pp 15-37.

Vogelstein B, Gillespie D (1979). Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci U S A* **76:** 615-619.

Watson J, Schanke J, Grunenwald H, Meis R, Hoffman L, Lewandowska-Skarbek M *et al* (1998). A new method for DNA and RNA purification. *J Lig Assay* **21:** 394-403.

Watson MP, Winters MA, Sagar SL, Konz JO (2006). Sterilizing Filtration of Plasmid DNA: Effects of Plasmid Concentration, Molecular Weight, and Conformation. *Biotechnology Progress* **22:** 465-470.

Wommack KE, Sime-Ngando T, Winget DM, Jamindar S, Helton RR (2010). Filtration-based methods for the collection of viral concentrates from large water samples. In: Suttle CA, Wilhelm SW, Weinbauer MG (eds). *Manual of Aquatic Viral Ecology*. American Society of Limnology and Oceanography: Waco, TX. pp 110-117.

Yu Z, Mohn WW (1999). Killing two birds with one stone: simultaneous extraction of DNA and RNA from activated sludge biomass. *Canadian Journal of Microbiology* **45:** 269-272.

Chapter 5

Variables influencing the efficiency and interpretation of reverse transcription, real-time quantitative PCR (RT-qPCR): an empirical study using Bacteriophage MS2

Jaclyn A. Mueller

Grieg F. Steward

Center for Microbial Oceanography: Research and Education, Department of Oceanography, University of Hawai‘i at Mānoa, 1950 East-West Road, Honolulu, Hawai‘i 96822

5.1 ABSTRACT

Reverse transcription, quantitative PCR (RT-qPCR) is a sensitive method for quantification of specific RNA targets, but the first step of the assay, reverse transcription, is notoriously variable and sensitive to reaction conditions. In preparation for a study of the dynamics of uncultivated RNA viruses in environmental samples, we investigated some of the reaction conditions that affect RT-qPCR sensitivity and accuracy. Using purified Bacteriophage MS2 genomic RNA as a model virus target, we tested two different RT enzymes (SuperScript II and SuperScript III), two RT-priming strategies (gene-specific primers and random hexamers), and varying background RNA concentrations (0 to 50 ng μl^{-1}) to determine how these variables influence the efficiency of reverse transcription over a range of target concentrations (10^1 to 10^7 copies μl^{-1}). The two tested enzymes had similar efficiencies, but SuperScript II outperformed SuperScript III at high target concentrations. The efficiency of the RT reaction was greatly improved by increasing both background RNA and primer concentrations. At a given target concentration, similar RT efficiencies were achieved with gene-specific primers and random hexamers, but the latter required much higher concentrations. Using random hexamers, we observed a systematic variation in RT reaction efficiency (from 19 to 91%) as a function target concentration. The RT variability was effectively normalized by using an RNA standard curve that was also subject to RT. This should allow accurate determination of relative changes in target concentrations, but the accuracy with which one can determine absolute target concentrations depends critically on the nature of the RNA standard. We recommend use of an RNA standard that is identical to the target in length and sequence whenever possible. If a synthetic RNA standard is to be primed with random hexamers, it should at least extend well beyond the 3' end of the reverse primer site to minimize generation of truncated standard cDNAs, which will lead to overestimation of target concentrations.

5.2 INTRODUCTION

Fluorescence-based reverse transcription, real-time quantitative PCR (RT-qPCR) is a highly sensitive method for quantification of specific RNAs, whether cellular (e.g., rRNA or mRNA) or viral genomes, and has become a common tool in basic research, functional genomics, biotechnology, and medicinal, forensic, and water quality diagnostics (Botes et al 2013, VanGuilder et al 2008). Many studies have been devoted to understanding sources of variability in the underlying enzymatic assays (see reviews by Bustin 2002, Sanders et al 2014, Wong and Medrano 2005), but achieving accurate and reproducible quantification of specific RNAs in a sample can still present challenges. For example, the RT yield can vary by orders of magnitude depending on the reverse transcriptase used, the priming strategy, starting target and background RNA concentrations, and even the type of sequence that is targeted (Levesque-Sergerie et al 2007, Ståhlberg et al 2004a, Ståhlberg et al 2004b, Werbrouck et al 2007).

Syntheses of the findings from the above and other studies has led to some recommended practices for conducting RT-qPCR assays (Nolan et al 2006, Sanders et al 2014) along with exhortations for the community to embrace standardized reporting (Bustin et al 2009, Huggett and Bustin 2011). One salient message from the all of the work on the topic is that the performance of any reverse transcription system (enzymes and associated buffer constituents) is highly context dependent. Despite decades of work to improve this assay, the technology continues to evolve and there remain many gaps in our knowledge.

This study was motivated by our interest in quantifying the genomes of specific, uncultivated RNA viruses in total RNA extracted from environmental samples. Our goal was to determine the effects of changing a suite of assay variables on the sensitivity and accuracy of RT-qPCR when applied to this task. We focused primarily on the efficiency of RT, since this step contributes most to the variability of the assay (Ståhlberg et al 2004a). Our intended application for the assay guided our choices about which variables to test and which specific leads to follow.

Since our eventual intended targets are RNA viruses, we used the single-stranded RNA genome of Bacteriophage MS2 as a model system. We chose to use TaqMan hydrolysis probes for real-time detection, because this technique allows for multiplexing and was reported to have higher specificity and a greater dynamic range of target concentrations (Plaskon et al 2009). We tested

the efficiency with which viral RNA was converted to cDNA using two different RT enzymes, two cDNA priming strategies, and varying target and background RNA concentrations. We also investigated how the choice of material for the standard curves (linearized plasmids, amplicon fragments, or RNA transcripts) influences the interpretation of the final qPCR data. Our results provide some new insights into the performance of RT and RT-qPCR that should be of interest even to those who use them in very different applications.

5.3 METHODS

5.3.1 Target and background RNA sources

Purified genomic RNA from Bacteriophage MS2 (Roche Applied Sciences; catalog number 10165948001) was supplied at a stock concentration of 0.8 μg μl^{-1}. The concentration of the stock RNA was verified by fluorometry (TD-700; Turner Designs) using a Quant-iT RNA Assay Kit (Life Technologies) and by spectrophotometry (DU-800; Beckman Coulter) after dilution in TE (pH 8.0). Since the values from both methods were in reasonable agreement (940±14 and 930±19 ng μl^{-1}, respectively) and the 260/280 absorbance ratio implied clean RNA (≥ 2), the value obtained by spectrophotometry was used to determine the subsequent efficiencies of the RT-qPCR reactions. Purified *Escherichia coli* total RNA (Life Technologies, catalog number AM7940) was supplied at a stock concentration of 1 μg μl^{-1} and used as background RNA in RT and qPCR reactions.

5.3.2 Primers and probes

The primers and probes targeting the lysis protein and RNA replicase β chain of Bacteriophage MS2 (Table 5.1) were selected from previously published primer pairs (O'Connell et al 2006), or created using Primer3 (Koressaar and Remm 2007, Untergasser et al 2012). HPLC-purified primers were obtained from Integrated DNA Technologies (Coralville, IA) and the TaqMan probe (with 5'-VIC and 3'-carboxymethylrhodamine [TAMRA]) was obtained from Life Technologies.

5.3.3 Reverse transcription

MS2 genomic RNA or RNA standard transcripts prepared by in vitro transcription from amplified fragments of the MS2 genome (tMS2-5, tMS2-6) were used as the target RNA. Aliquots of each type of RNA were prepared to avoid the influence of repeated freezing and thawing on the cDNA quality, and dilutions were prepared from a fresh aliquot for each run. All RT reactions were prepared with one master mix for the respective enzyme, and spiked with MS2 RNA template and varying background *E. coli* total RNA, depending on the experiment. Each RT assay was performed in duplicate or triplicate 20-µl reactions. The reaction conditions (cycle temperature and duration) were performed as recommended by the manufacturer for each enzyme (SuperScript II or SuperScript III) and type of primer (random hexamers or gene specific) with an RNase H incubation step.

In detail, each SuperScript II reaction contained 2 pmole gene specific primer or 50 ng random hexamers (unless specified otherwise below), 0.1 pg – 1 µg total RNA, and 10 nmole each dNTP with a final volume of 12 µl in the initial denaturation step incubated at 65 °C for 5 min, followed by a quick chill on ice. The tubes were briefly centrifuged and 1 × First-Strand Buffer (final conc. in 20 µl), 0.2 µmole DTT, and 40 units RNaseOUT were added to the 19 µl reactions, followed by an incubation at 42 °C (gene specific primers) or 25 °C (random hexamers) for 2 min. Then 200 units SuperScript II RT enzyme (final volume = 20 µl) were added to each tube and the reactions were carried out with an initial incubation at 25 °C for 10 min for random hexamers only, followed by 42 °C for 50 min, and 70 °C for 15 min for both primer types. SuperScript III reactions contained 2 pmole gene specific primer, approx. 1 µg total RNA, and 10 nmole each dNTP with a final volume of 10 µl in the initial denaturation step incubated at 65 °C for 5 min, followed by a quick chill on ice. The tubes were briefly centrifuged and 1 × RT Buffer (final conc. in 20 µl), 0.1 µmole $MgCl_2$, 0.2 µmole DTT, 40 units RNaseOUT, and 200 units SuperScript III RT enzyme were added to the reactions (final volume = 20 µl). Tubes were incubated at 50 °C for 50 min, then 85 °C for 5 min. Reverse transcription reactions with either enzyme were followed by an addition of 2 units of *E.coli* RNase H and incubated at 37 °C for 20 min to remove RNA complementary to the cDNA.

a. Background RNA – The tests with differing background RNA concentration were performed

with the SuperScript II RT enzyme and 6.1×10^4 copies of MS2 genomic RNA spiked with 0, 2.5, 5, 25, or 50 ng μl^{-1} (0, 50, 100, 500, or 1,000 ng in 20 µl reactions) *E. coli* total RNA. To test whether background RNA concentrations in the RT reactions would inhibit the subsequent qPCR reactions, we compared the amplification of MS2 plasmid (pMS2-5) standard dilutions (1.4×10^2 to 1.4×10^9 copies per reaction) in the absence or presence of background RNA (5 ng μl^{-1}) to mimic the amount of carry over from the RT reactions with 50 ng μl^{-1} RNA.

b. Enzymes – For the direct comparison of SuperScript II and SuperScript III, 0.1 µM MS2-5R primer was used with 6.1×10^2 or 6.1×10^5 copies of MS2 genomic RNA and 50 ng μl^{-1} background *E.coli* RNA. All other subsequent RT tests were performed using SuperScript II.

c. Primer type and concentration – Priming conditions of the RT reaction were tested using either random hexamers (1.3-168 µM) or an 18-bp gene-specific primer (MS2-5R; 0.1-40 µM) with 5.8×10^5 copies MS2 genomic RNA. Because of the versatility of priming with random hexamers (which allows for the detection of multiple qPCR targets from one RT reaction), we further tested and quantified RT efficiency with this type of primer. For these tests, the higher concentration of random hexamers (168 µM) was used in the RT reactions for both the MS2 genomic RNA and MS2-6 RNA transcript (tMS2-6) standards. Efficiencies were then determined relative to the tMS2-6 standards and pMS2-5 plasmid standards.

5.3.4 Reverse transcription reaction dilutions

We attempted to determine the influence of RT reaction reagents on the qPCR assays using a dilution method, which is often used to reduce RT reaction inhibition. For this experiment, RT reactions were performed with SuperScript II, a background RNA concentration of 50 ng μl^{-1}, and 5.8×10^3 to 5.8×10^7 copies of MS2 genomic RNA. An aliquot of 2 µl was transferred from the undiluted RT reaction, or from a portion of the same RT reaction after diluting 1:1, 1:10 or 1:50 with nuclease-free water, into the qPCR reactions for quantification by the TaqMan assay as described below.

5.3.5 Creating plasmid and amplicon DNA standards

Bacteriophage MS2 genomic RNA was primed with MS2-5R primer (Table 5.1, Fig. 5.1) then reverse transcribed with SuperScript II following the manufacturer's instructions as described

above. A region of Bacteriophage MS2 (aMS2-5; 609 bp) was amplified from the resulting cDNA with MS2-4F and MS2-5R primers (Fig. 5.1). The PCR reaction was performed with the Expand High Fidelity Plus PCR system (Roche) with 2 µl of undiluted cDNA reaction, 1 × reaction buffer, 0.2 mM dNTP, 400 nM each primer, and 2.5 U Expand High Fidelity Plus Enzyme in a total volume of 50 µl. The thermal cycling consisted of an initial denaturation at 94 °C for 2 min followed by 40 cycles of denaturation at 94 °C for 15 s, annealing at 60 °C for 30 s, and extension at 72 °C for 1 min, followed by a final extension step at 72 °C for 7 min. The amplicon was concentrated and purified with a MinElute PCR Purification Kit (Qiagen) and run on a 0.5 % agarose gel in 1 × TAE stained with SYBR Safe. The single band was visualized with a Safe Imager 2.0 Blue-Light Transilluminator (Life Technologies), then excised and purified with a MinElute Gel Extraction Kit (Qiagen).

The amplicon was cloned into the pSMART HC Kan vector (Lucigen) and propagated in E. cloni® competent cells (Lucigen). Purified plasmid was linearized with Swa I restriction enzyme (New England Biolabs), visualized on a 1% agarose gel to confirm complete digestion, then cleaned up with a MinElute Reaction Cleanup Kit (Qiagen) for use as a DNA standard (pMS2-5). A new PCR amplicon was created from the sequenced linearized plasmid for use as another DNA standard (aMS2-5). Both the amplicon and linearized plasmid were quantified with a Qubit 2.0 Fluorometer (Life Technologies). Standards were prepared by serial dilution in 50 ng μl^{-1} background *E. coli* total RNA with nuclease-free water to mimic the sample background RNA concentration, after the ideal concentration was determined as 50 ng μl^{-1} RNA per RT reaction.

5.3.6 Creating RNA standards by in vitro transcription

Reverse transcription was performed on Bacteriophage MS2 genomic RNA with either MS2-5R or MS2-6R reverse primer using SuperScript II, following the manufacturer's specifications as detailed above. An MS2 forward primer containing a T7 promoter (MS2-T7-4F) was then used in PCR amplification with either MS2-5R or MS2-6R using the Expand High Fidelity PCR Plus System (Roche) under the same PCR conditions as specified in the previous section, for 30 cycles. The amplicons were electrophoresed in an agarose gel and purified using the MinElute Gel Purification Kit (Qiagen). Transcripts were produced from ~250 ng PCR product in vitro with the MEGAscript T7 Transcription Kit (Life Technologies). Reactions were carried out for 4

h at 37 °C, followed by DNase treatment with TURBO DNase. The RNA was then purified with a QIAamp MinElute Virus Spin Kit (Qiagen) and eluted in TE. The amount of RNA was quantified by spectrophotometry of dilutions in TE (pH 8.0) using a DU-800 spectrophotometer (Beckman Coulter). Aliquots of the two run-off transcript stocks (tMS2-5 and tMS2-6) were stored at -80 °C to avoid the influence of multiple freeze-thaw cycles on RNA integrity. As needed, stock transcripts were thawed, serially diluted, and used immediately in RT reactions with either random hexamers or MS2-5R under the conditions specified for a given experiment.

5.3.7 Quantitative PCR

All quantitative PCR assays were performed in duplicate or triplicate in a thermal cycler equipped for real-time fluorescence monitoring (Mastercycler ep realplex^2; Eppendorf). For quantification of the MS2 RNA replicase gene, the hydrolysis probe technology (TaqMan probes) was used as the detection system. The PCR mixture contained 10 µl TaqMan Fast Advanced Master Mix (Life Technologies), 300 nM forward (MS2-5F) and reverse (MS2-5R) primers, 250 nM probe (MS2-5P), 2 µl cDNA or dsDNA, and nuclease-free water to reach a final volume of 20 µl per well. The final primer and probe concentrations of 300 and 250 nM, respectively, were determined as the optimum concentration of primers and probes (data not shown), similar to the findings of Quan et al (2010). Thermal cycling conditions were as follows: 50 °C for 2 min (UNG incubation) preceding the first denaturation step at 95 °C for 20 s, followed by 45 cycles of 95 °C for 15 s and 60 °C for 30 s. RT-qPCR data were analyzed using the default noiseband from the Mastercycler ep realplex^2 thermal cycler (Eppendorf) as the threshold. The slopes of the calibration curves were calculated from the plot of the log base 10 of initial target copy number versus corresponding C_q and used to determine the number of gene copies in each RT sample. The PCR efficiency (E) was determined from the slope of the curve obtained with the serially diluted standards, as $E = 10^{(-1/\text{slope})}$. The absolute efficiency was determined from the detected MS2 copies relative to calibration curves and represented as a percentage of the expected copy numbers.

5.4 RESULTS

5.4.1 Effect of background RNA on target detection efficiency

Increasing background RNA concentrations with a fixed concentration of MS2 genomic RNA target resulted in decreasing C_q values (increasing efficiency), but the improvement saturated at approximately 25-50 ng μl^{-1} (Fig. 5.2A). To investigate the effect of background RNA carried over from the RT reaction on qPCR efficiency, we added 100 ng of RNA (5 ng μl^{-1}) to a series of pMS2-5 plasmid standards prior to qPCR (equivalent to the carryover from an RT reaction with 50 ng μl^{-1} of background RNA) and compared the results to the same standards with no RNA added. Each of the standards with added RNA had slightly lower C_q values than those without, but the difference was not significant (Fig. 5.2B; ANCOVA, $F = 1.564$, $p = 0.222$).

5.4.2 Relative efficiency of reverse transcription enzymes

In a comparison of RT reactions performed with the SuperScript II and SuperScript III enzymes at their respective recommended reaction temperatures (42 and 50 °C respectively), the reactions with Superscript II resulted in lower mean C_q values in the subsequent qPCR reaction at both high and low target concentrations (Fig. 5.3), but the difference was statistically significant only at the higher target concentration (6.1×10^5 copies; *t* test: $P < 0.001$). Because of this apparent performance advantage, we used the SuperScript II enzyme in the assays for all subsequent tests.

5.4.3 Effect of diluting reverse transcription reactions on subsequent qPCR reactions

In all experiments, addition of undiluted RT reaction into the qPCR resulted in a higher estimate of the concentration of MS2 genomic RNA than any of the dilutions (Fig. 5.4). These differences were significant for the 1:10 and 1:50 dilutions when data was available for both the diluted and undiluted samples (paired *t* test: $P < 0.001$). The copy numbers were calculated relative to the pMS2-5 plasmid standard curve. In the cases where the initial MS2 copy numbers were low (5.8×10^4 or 5.8×10^3 per reaction), no copies were detected when the RT reactions were diluted prior to addition to the qPCR reaction. For the remaining experiments, we therefore chose to quantify the cDNA copies of MS2 genomic RNA with the undiluted RT reaction to determine reaction efficiencies.

5.4.4 RT-priming conditions

For both gene-specific (MS2-5R) and random hexamer primers, increasing primer concentrations resulted in decreasing C_q values with evidence of saturation at the highest primer concentrations (Fig. 5.5A). When these two higher primer concentrations (40 µM for MS2-5R and 168 µM for random hexamers) were used in RT reactions over a range of MS2 genomic RNA concentrations (5×10^6 to 5×10^2 copies per RT reaction) there was no significant difference in the C_q values obtained for each priming condition (ANCOVA, $F = 4.004$, $P > 0.05$; Fig. 5.5B).

5.4.5 Standard curve analysis

The use of serial dilutions of DNA plasmid (pMS2-5), DNA amplicon (aMS2-5), or RNA transcripts (tMS2-5 or tMS2-6) to construct calibration curves generated reliable standards with high qPCR efficiencies (86-94%). The DNA standard curves reliably detected targets down to 2 to 7 copies μl^{-1} of qPCR reaction, while the RNA standard curves detected targets down to 250-300 copies μl^{-1} of RT-qPCR reaction (25-30 copies μl^{-1} of qPCR reaction). The two types of DNA standard curves (plasmid vs amplicon) were similar to one another, but had significantly different slopes (Fig. 5.6A; 95% confidence intervals) resulting from slightly higher C_q values for the plasmid at higher copy numbers. The slopes of the standard curves produced with tMS2-5 or tMS2-6 with random hexamers were the same, but the longer RNA transcript (tMS2-6) produced significantly lower C_q values for each dilution of RNA (Fig. 5.6B). For this reason, we used the longer tMS2-6 RNA transcript standards for calculating the efficiency of the MS2 genomic RNA in RT reactions below.

5.4.6 Effect of target concentration on RT efficiency

Since the amount of MS2 RNA molecules used to spike each RT reaction is known, and a calibration curve is prepared with a known number of gene copies, we can calculate the real efficiencies of the RT reaction (i.e., the percentage of the MS2 genomic RNA converted to amplifiable cDNA). The RT efficiency was found to vary as a function of target concentration ($r = 0.95$; $P < 0.001$) ranging from 19% at low target concentrations to 91% at the highest target concentration tested (Fig. 5.7A). We then tested whether the observed concentration-dependent efficiency of the RT reaction could be normalized by using dilutions of RNA transcripts as the

standard curve. When efficiency was calculated relative to tMS2-6 RNA standards that were also subjected to the RT step, the apparent target yield was 312% on average, but there was no longer a significant difference in efficiency across the range of standards ($r = -0.67$; $P > 0.05$).

5.5 DISCUSSION

We have investigated some of the key variables that influence the efficiency of the reverse transcription reaction and its quantitative interpretation. Some of our results are consistent with previous observations, but others differ markedly, highlighting the importance of context when interpreting assay performance. Below we discuss the implications of each of the specific variables we investigated.

5.5.1 Background RNA concentration

The trend that we observed of improving RT efficiency with increasing amounts of background RNA up to 50 ng μl^{-1} is in contrast to an earlier study in which it was concluded that small amounts of background RNA were beneficial, but more RNA was inhibitory (Pfaffl and Hageleit 2001). However, the RNA concentrations tested in that study were all quite low (maximum of 1 ng μl^{-1}) and it was not clear if the observed effects were statistically significant. In a subsequent study (Ståhlberg et al 2004a), a high, constant concentration of background RNA resulted in more consistent efficiency across a range of target concentrations compared to having no added background RNA. Our results are most similar to those of Levesque-Sergerie et al (2007), who suggested that concentrations of background RNA up to 50 ng μl^{-1} were beneficial to the RT reaction in the presence of both high and low abundance transcripts, but no further improvement was observed from 50 ng μl^{-1} up to 250 ng μl^{-1}. We also tested higher background RNA concentrations of up to 200 ng μl^{-1} background RNA, and likewise found no improvement from the 50 ng μl^{-1} background RNA (data not shown). We also confirmed that the added background RNA would not inhibit the subsequent qPCR reactions. However, the source or quality of this RNA appears to be important, as we saw the opposite effect on RT and qPCR reactions when using RNA extracted and extensively purified from a non-axenic mix of phytoplankton cultures. In these tests, the highest efficiencies were seen without the addition of any background RNA, suggesting that the added RNA may have had an inhibitory effect on the reaction. Therefore, we recommend an addition of 50 ng μl^{-1} background RNA that has been checked for any inhibition

for each cDNA synthesis reaction to obtain the highest yields when using the SuperScript II RT system.

5.5.2 Reverse transcriptase

Of the two enzymatic reactions involved in the RT-qPCR assay, reverse-transcription, the conversion of RNA into DNA, appears to be inherently more variable (Ståhlberg et al 2004a). The RT enzyme chosen for a given study can apparently greatly affect the efficiency of cDNA synthesis. In a study comparing multiple RT systems with genetically modified versions of the enzymes from Moloney murine leukemia virus (MMLV) or avian myeloblastosis virus (AMV), the kit with the commercialized MMLV lacking RNase H activity (SuperScript II) consistently displayed the highest performance (Sieber et al 2010). In general, it appears that low or even absent RNase H activity is more efficient than cloned MMLV or native AMV, since the RNase H could cleave the RNA before the initiation of polymerization in cDNA synthesis (Gerard et al 1997). We originally chose SuperScript II for this study, because it was reported to be significantly more sensitive than six other reverse transcriptases (including SuperScript III) at low target concentrations (Levesque-Sergerie et al 2007). Although we also found SuperScript II to be more sensitive than SuperScript III, the difference was much smaller and only significant at high, rather than low, template concentration.

We did not investigate RT enzymes other than the SuperScript II and III enzymes, both of which contain reduced RNase H activity, but we included the subsequent RNase H digestion after cDNA synthesis, and found this increased qPCR yields compared to no RNase H digestion (data not shown). It appears that the increased thermostability of SuperScript III did not present an advantage in this particular system, but this of course may vary depending on the nature of target RNA secondary structures, where increased temperatures may be necessary to access the target sequence.

5.5.3 Effect of RT reagents on qPCR efficiency

There have been many reports of inhibition of qPCR by RT reaction components, whether the RT enzyme itself (Chandler et al 1998, Chumakov 1994, Suslov and Steindler 2005) or other components of the reaction (Levesque-Sergerie et al 2007). In accordance with these

observations, dilution of the RT reaction prior to qPCR has been reported to improve qPCR efficiencies (Levesque-Sergerie et al 2007). In contrast to previous reports, we found that the RT reaction reagents did not cause inhibition in qPCR reactions under our routine conditions (2 µl of undiluted RT reaction added to 20 µl qPCR reactions) when we used the TaqMan Fast Advanced Master Mix (Life Technologies). Instead, we found the opposite, suggesting something in the RT reaction was improving efficiency of the qPCR. This phenomenon appears to be specific to this particular qPCR formulation, because we did observe inhibition by RT in a test using another qPCR system from the same manufacturer (TaqMan Universal Master Mix II, with UNG; Life Technologies; data not shown). Having the ability to add undiluted RT reaction to a qPCR assay provides the dual advantages of simplifying the assay and maintaining a low detection limit. We therefore adopted the TaqMan Fast Advanced Master Mix for this study.

5.5.4 RT-priming strategy

Our results clearly illustrate that both the type and concentration of the RT primer have a large influence on the efficiency of the RT reaction and the nature of our experiments help to reconcile conflicting conclusions from previous studies. Some researchers have reported, for example, that gene-specific primers produced superior results compared to random hexamers (Lekanne Deprez et al 2002, Zhang and Byrne 1999) and others have reported superior results with random hexamers (Stahlberg et al. 2004). The poor performance of the random hexamers observed by Zhang and Byrne (1999) was a result of the design of their RNA standard rather than the primers per se. Although varying concentrations of primers were tested in that study, quantification was always relative to a standard so the effect of primer concentration on absolute efficiency of the RT reaction could not be determined. The directly opposing conclusions of Lekanne Deprez et al (2002) and Ståhlberg et al (2004a) are best explained by the arbitrarily chosen concentrations for each primer type. The absolute and relative gene-specific and random hexamer primer concentrations used in the former study (4 and 12.5 µM) differed from the latter (1 and 50 µM). Viewed in the context of our results (Fig. 5.5), those concentrations would have naturally favored gene-specific primers in the first case and random hexamers in the latter case.

By testing the effect of primer concentration for both types of primers we have shown that comparable efficiencies can be achieved with gene-specific or random hexamer priming of the

RT reaction, provided the primer concentrations are high enough. We found it surprising that the concentrations recommended by the RT manufacturer (0.1 μM gene-specific primer or 1.3-6.3 μM random hexamers) are too low by at least two orders of magnitude if one wishes to achieve the highest transcription efficiency, and thus the lowest detection limit.

5.5.5 Absolute target quantification

In addition to exploring how to maximize RT efficiency, we investigated how the nature of the standards influences the accuracy of absolute target quantification. Some researchers have argued that DNA standard curves (subjected only to qPCR) are superior to RNA standard curves, (subjected to RT and qPCR), because of their reproducibility, linearity, and sensitivity, even though they do not account for the efficiency of the RT step (Pfaffl and Hageleit 2001). With a DNA standard curve, the assay only quantifies RNA molecules that were successfully converted to cDNA, thereby underestimating the RNA targets by some unknown percentage. If the efficiency of the RT reaction also varies as a function of template concentration (as we observed in Fig. 5.7A), then not only are absolute abundances underestimated, but differences in relative abundance are magnified.

In principle, an RNA standard curve could compensate for both sources of error, but only if standards and targets are reverse transcribed with equal efficiency. There are reasons why this may not be so, such as having a standard that is a poor proxy for the target. In the ideal case, the RNA being used as a standard is identical to the target being quantified (Homs et al 2014, O'Connell et al 2006), but this is not always practical and standards are sometimes produced by in vitro transcription from a cloned or amplified fragment of the intended target (Bowers and Dhar 2011, Costafreda et al 2006, Liu et al 2013, Pang et al 2014). If the resulting transcript has the qPCR reverse priming site at or very close to the 3' end, then priming with random hexamers can result in a high proportion of truncated cDNA molecules missing the primer site that are undetectable by qPCR (Zhang and Byrne 1999). In this latter study, the standard was amplified with much lower apparent efficiency than the target, and the target concentration was severely overestimated by a factor of 19-fold (Zhang and Byrne 1999). As a consequence, the authors concluded that gene-specific priming of the RT reaction is more likely to give the most accurate results.

We observed the same phenomenon reported by Zhang and Byrne (1999). However, we demonstrated that a redesign of the standard to include an additional 134 bases 3' of the qPCR primer site significantly improved the performance of the standard when priming with random hexamers. With the primer site at the end, we would have overestimated target concentrations by a factor of about 13, but this was reduced to a factor of about 3 with our longer redesigned standard. We hypothesize that the reason the efficiency of priming with random hexamers improved was because we provided RT primer binding sites upstream of the qPCR reverse primer site, which reduced the number of truncated, unamplifiable cDNA products. An additional modest increase in the length of this upstream binding region might eliminate the differences in RT efficiency between standard and template. It is possible, however, that RT efficiency will not vary systematically with length, because of influences by secondary structure.

Although the longer tMS2-6 standard had a 3-fold error in absolute RNA target quantification, we did find that the use of this RNA standard compensated for changes in RT efficiency as a function of target concentration, thereby linearizing the response of the assay and allowing for more accurate estimation of relative changes in target abundance. This linearization of response only works if dilutions of the RNA standard are independently reverse transcribed. In some studies an RNA standard is used, but the cDNA from a single reverse transcription reaction is serially diluted for use in the qPCR. This latter approach not only fails to account for variations in RT efficiency as a function of target concentration, but can cause additional biases in qPCR because of variable carryover of the RT reagents into the qPCR (Suslov and Steindler 2005).

Another possible source of bias between standards and samples during RT is differences in the reaction composition. The use of background RNA in the standards, for example, may not perfectly represent the RNA species present in a sample, and that could influence RT efficiency (Pfaffl and Hageleit 2001). However, we added high concentrations of background RNA from the same source to both the samples and the standards to minimize differences in RNA composition. Sample RNA extracts may also contain inhibitors of either RT or qPCR that are absent from the standards, but one can control for presence of inhibitors by adding an external reference RNA target to each sample (Bustin 2010).

5.6 CONCLUSIONS

RT-qPCR is a useful assay with wide-ranging applications, but the accuracy with which it determines relative and absolute concentrations of RNA targets is sensitive to many variables. In this study, we showed that the efficiency of the RT reaction was improved by adding background RNA and by increasing primer concentration in RT reactions well above that usually employed. In contrast to many other studies, we found that a dilution of the RT reaction into qPCR reactions was unnecessary, but this result appears to be qPCR kit-specific. By testing a broader range of conditions, we resolved discrepancies in the literature and demonstrated that random priming can be as efficient as gene-specific priming if primer concentrations are sufficiently high. However, when using random hexamers, one should employ an RNA standard that extends well beyond the 3' end of the reverse qPCR primer site to minimize errors in absolute quantification. Even in cases where the RNA standard is not a perfect proxy for the target RNA, we found that an RNA standard curve helped to normalize for systematic variability in the RT reaction and should therefore provide more accurate representation of relative abundances of RNA targets.

5.7 ACKNOWLEDGMENTS

This work was funded by the Center for Microbial Oceanography: Research and Education (EF 04-24599).

Table 5.1. Primer and probe sequences for reverse transcription quantitative PCR (RT-qPCR) assays for quantitation of MS2 RNA.

Name[a]	Start[b]	Sequence	Reference[c]
MS2-4F	1693	CCT CAG CAA TCG CAG CAA A	A
MS2-T7-4F	1693	TAA TAC GAC TCA CTA TAG GGC CTC AGC AAT CGC AGC AAA	B
MS2-5F	2232	GCT CTG AGA GCG GCT CTA TTG	A
MS2-5P	2255	[VIC]-CCG AGA CCA ATG TGC GCC GTG-[TAMRA]	A
MS2-5R	2301	CGT TAT AGC GGA CCG CGT	A
MS2-6R	2435	TCT GAT GAA AGC ACC GAC CC	B

[a]The numbers correspond to the assays from Table 1 in O'Connell et al., 2006, except for 6, which is only from this study. F = Forward, P = Probe, R = Reverse, T7 = T7 promoter.
[b]Position of first base in each oligonucleotide sequence relative to the whole genome sequence in GenBank accession no. NC_001417.
[c]Reference A: O'Connell et al., 2006; B: this study.

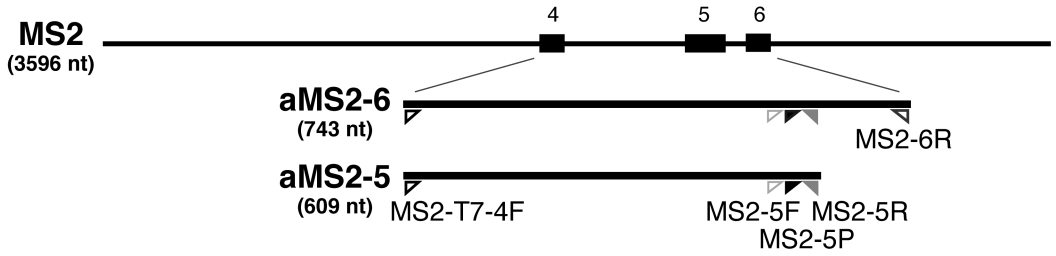


Figure 5.1. Map of MS2 genome and amplicons (aMS2-5 and aMS2-6) that were used to produce RNA transcripts. Thick bars and numbers indicate the approximate locations and sizes of the primers on the MS2 map. Primers indicated by triangles on aMS2-6 and aMS2-5 correspond to those listed in Table 5.1.

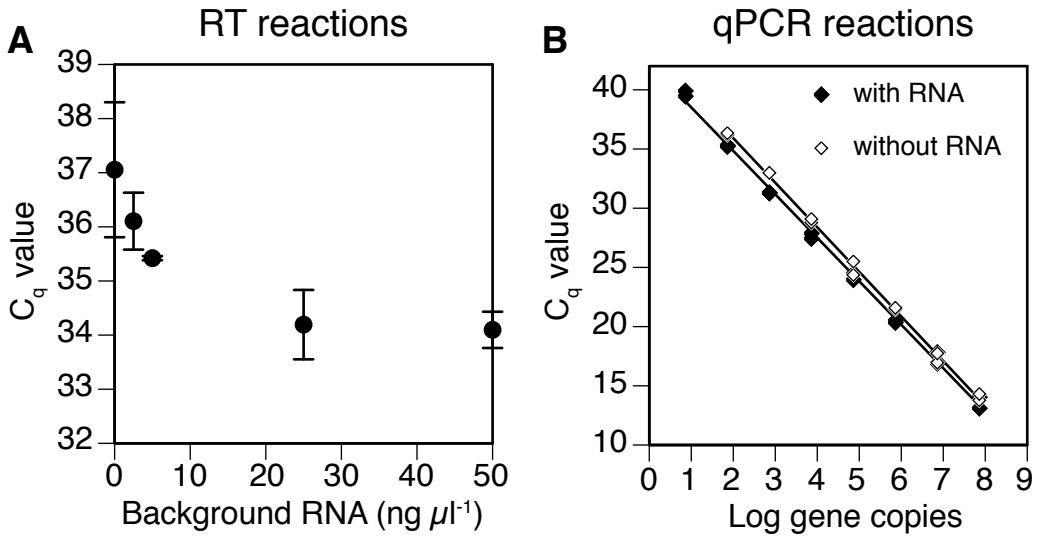


Figure 5.2. Effects of background RNA on RT and qPCR reactions. A. Quantitative measurements of MS2 cDNA from RT reactions with varying background RNA concentrations (*E. coli* total RNA). Data are reported as the mean and standard deviation of C_q values in triplicate qPCR reactions from duplicate RT reactions spiked with 6.1×10^4 copies MS2 genomic RNA. B. Comparison of pMS2-5 plasmid standards in qPCR reactions without RNA and spiked with 5 ng μl^{-1} RNA to mimic the carryover from the RT reactions with 50 ng μl^{-1} RNA.

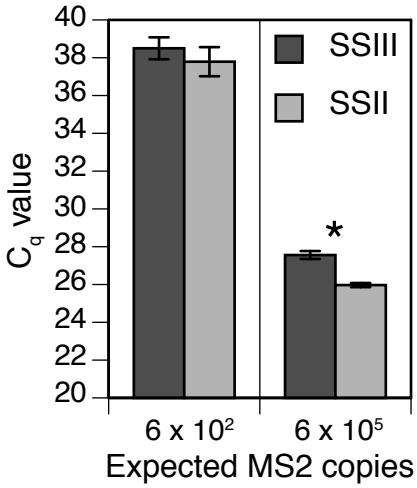


Figure 5.3. Quantitative measurements of MS2 genomic RNA from RT reactions performed with two commercial kits: Superscript II (SSII) and Superscript III (SSIII). Data are reported as the mean and standard deviation of C_q values from duplicate RT reactions spiked with 6×10^2 or 6×10^5 copies of MS2 genomic RNA quantified in triplicate qPCR reactions. RT reactions were conducted with 0.1 μM MS2-5R and a background RNA (total *E.coli* RNA) concentration of 50 ng μl^{-1}. An asterisk (*) indicates a significant difference in detection between the two RT systems ($P < 0.001$).

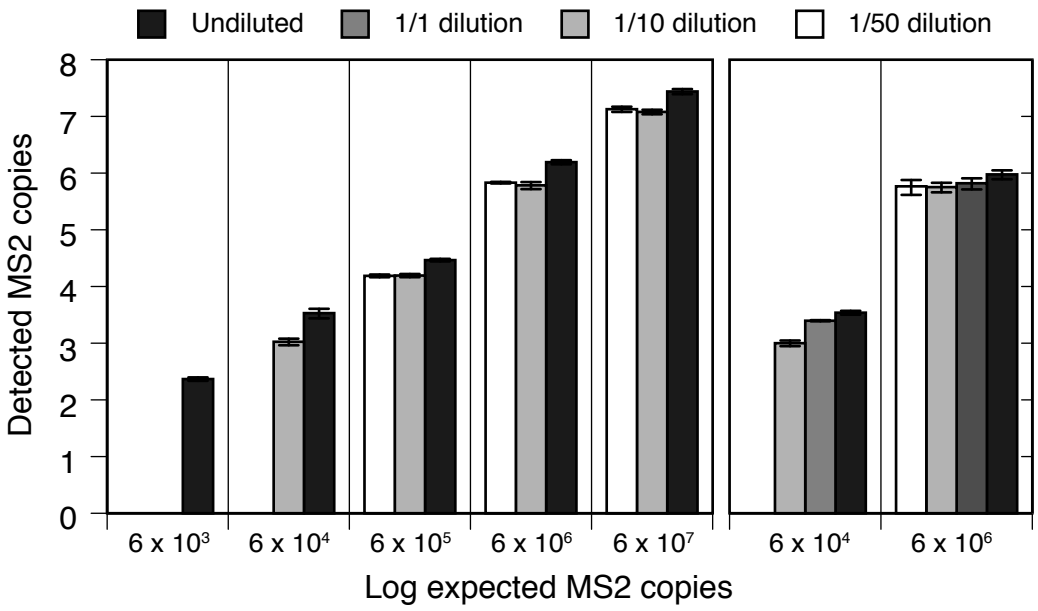


Figure 5.4. Effect of RT dilution on the quantification of amplifiable MS2 cDNA. Quantitative PCR was performed in duplicate on undiluted and diluted (1/10, 1/10 or 1/50) RT samples spiked with MS2 genomic RNA ranging from 5.8×10^3 to 5.8×10^7 copies per reaction. Data from two separate experiments are shown in separate panels, and the mean and standard deviation are reported from duplicate qPCR reactions relative to pMS2-5 plasmid DNA.

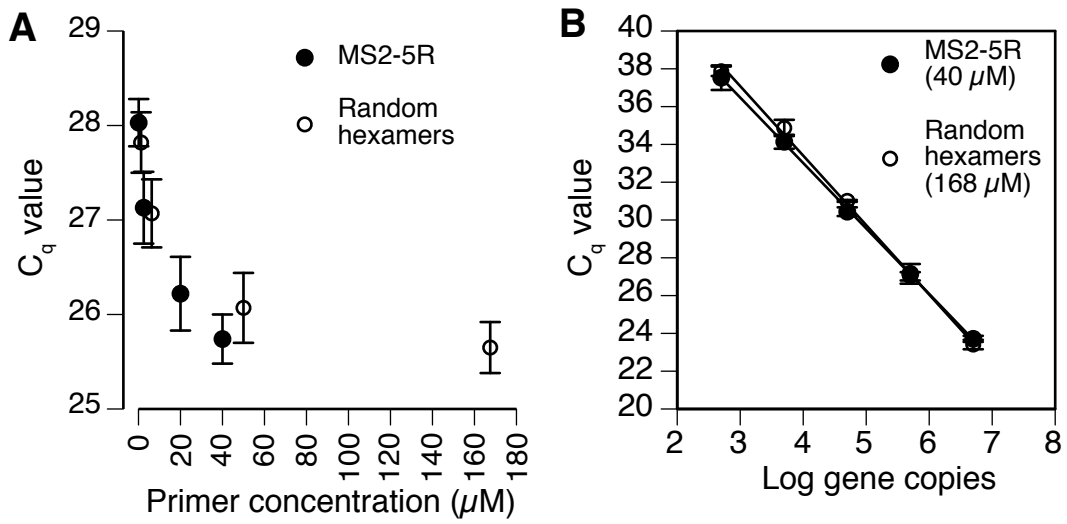


Figure 5.5. MS2 detected from RT reactions performed with either gene specific primers (MS2-5R) or random hexamers. A. Varying concentrations of primer in triplicate reactions spiked with 5.8×10^5 copies MS2 genomic RNA. Data reported are the mean C_q values from triplicate RT reactions performed in the various conditions and quantified in triplicate. B: High primer concentrations (40 μM MS2-5R or 168 μM random hexamers) with 5×10^2 to 5×10^6 copies MS2 genomic RNA. Data reported are the mean C_q values from duplicate RT reactions quantified in duplicate. Error bars represent standard deviations of replicate RT reactions.

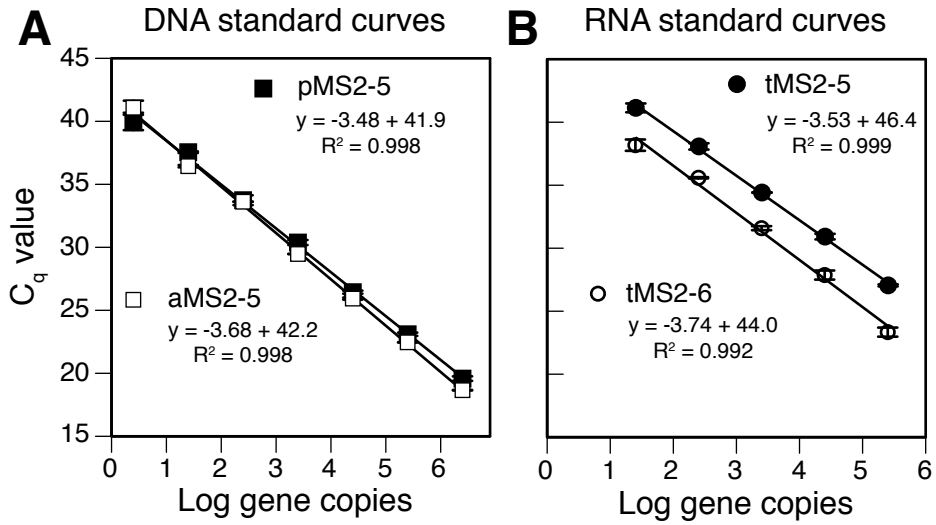


Figure 5.6. MS2 plasmid (pMS2-5) and amplicon (aMS2-5) standard curves (A) and RNA transcript (tMS2-5 and tMS2-6) standard curves (B). RNA standard curves were produced with tMS2-5 and tMS2-6 RNA transcripts using 168 μM random hexamers. Data shown are mean and standard deviation of C_q values from duplicate qPCR (A) or RT (B) reactions.

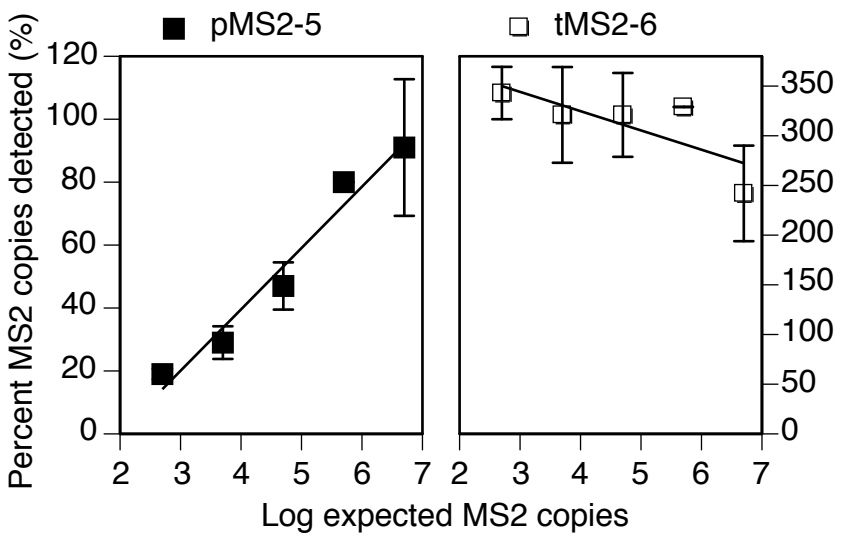


Figure 5.7. Yield of the reverse transcription reactions relative to the pMS2-5 plasmid DNA standards (left) and tMS2-6 RNA transcript (right). All RT reactions were performed in duplicate (except 5×10^5 copies with only one RT reaction) and quantified in triplicate for each amount of target MS2 genomic RNA ranging from 5×10^2 to 5×10^6 copies using 168 μM random hexamers. Data reported are the percent of expected copies of MS2 genomic RNA that were detected as cDNA MS2 copies via qPCR relative to each respective standard curve. Error bars represent standard deviations of duplicate RT reactions.

5.8 REFERENCES

Botes M, de Kwaadsteniet M, Cloete TE (2013). Application of quantitative PCR for the detection of microorganisms in water. *Analytical and bioanalytical chemistry* **405:** 91-108.

Bowers RM, Dhar AK (2011). Effect of template on generating a standard curve for absolute quantification of an RNA virus by real-time reverse transcriptase-polymerase chain reaction. *Molecular and Cellular Probes* **25:** 60-64.

Bustin SA (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of molecular endocrinology* **29:** 23-39.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M *et al* (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55:** 611-622.

Bustin SA (2010). Why the need for qPCR publication guidelines?—The case for MIQE. *Methods* **50:** 217-226.

Chandler DP, Wagnon CA, Bolton H, Jr. (1998). Reverse transcriptase (RT) inhibition of PCR at low concentrations of template and its implications for quantitative RT-PCR. *Applied and Environmental Microbiology* **64:** 669-677.

Chumakov KM (1994). Reverse transcriptase can inhibit PCR and stimulate primer-dimer formation. *PCR methods and applications* **4:** 62-64.

Costafreda MI, Bosch A, Pinto RM (2006). Development, evaluation, and standardization of a real-time TaqMan reverse transcription-PCR assay for quantification of hepatitis A virus in clinical and shellfish samples. *Applied and Environmental Microbiology* **72:** 3846-3855.

Gerard GF, Fox DK, Nathan M, D'Alessio JM (1997). Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA. *Molecular biotechnology* **8:** 61-77.

Homs M, Giersch K, Blasi M, Lutgehetmann M, Buti M, Esteban R *et al* (2014). Relevance of a full-length genomic RNA standard and a thermal-shock step for optimal hepatitis delta virus quantification. *Journal of Clinical Microbiology* **52:** 3334-3338.

Huggett J, Bustin SA (2011). Standardisation and reporting for nucleic acid quantification. *Accreditation and Quality Assurance* **16:** 399-405.

Koressaar T, Remm M (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23:** 1289-1291.

Lekanne Deprez RH, Fijnvandraat AC, Ruijter JM, Moorman AF (2002). Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Analytical Biochemistry* **307:** 63-69.

Levesque-Sergerie J-P, Duquette M, Thibault C, Delbecchi L, Bissonnette N (2007). Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR. *BMC Molecular Biology* **8:** 18.

Liu W, Zhao X, Zhang P, Mar TT, Liu Y, Zhang Z *et al* (2013). A one step real-time RT-PCR assay for the quantitation of Wheat yellow mosaic virus (WYMV). *Virology Journal* **10:** 173.

Nolan T, Hands RE, Ogunkolade W, Bustin SA (2006). SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Analytical Biochemistry* **351:** 308-310.

O'Connell KP, Bucher JR, Anderson PE, Cao CJ, Khan AS, Gostomski MV *et al* (2006). Real-time fluorogenic reverse transcription-PCR assays for detection of bacteriophage MS2. *Applied and Environmental Microbiology* **72:** 478-483.

Pang Z, Li A, Li J, Qu J, He C, Zhang S *et al* (2014). Comprehensive multiplex one-step real-time TaqMan qRT-PCR assays for detection and quantification of hemorrhagic fever viruses. *PLoS One* **9:** e95635.

Pfaffl M, Hageleit M (2001). Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnology Letters* **23:** 275-282.

Plaskon NE, Adelman ZN, Myles KM (2009). Accurate Strand-Specific Quantification of Viral RNA. *PLoS ONE* **4:** e7468.

Quan M, Lourens CW, MacLachlan NJ, Gardner IA, Guthrie AJ (2010). Development and optimisation of a duplex real-time reverse transcription quantitative PCR assay targeting the VP7 and NS2 genes of African horse sickness virus. *Journal of Virological Methods* **167:** 45-52.

Sanders R, Mason DJ, Foy CA, Huggett JF (2014). Considerations for accurate gene expression measurement by reverse transcription quantitative PCR when analysing clinical samples. *Analytical and bioanalytical chemistry* **406:** 6471-6483.

Sieber MW, Recknagel P, Glaser F, Witte OW, Bauer M, Claus RA *et al* (2010). Substantial performance discrepancies among commercially available kits for reverse transcription quantitative polymerase chain reaction: A systematic comparative investigator-driven approach. *Analytical Biochemistry* **401:** 303-311.

Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M (2004a). Properties of the Reverse Transcription Reaction in mRNA Quantification. *Clinical Chemistry* **50:** 509-515.

Ståhlberg A, Kubista M, Pfaffl M (2004b). Comparison of Reverse Transcriptases in Gene Expression Analysis. *Clinical Chemistry* **50:** 1678-1680.

Suslov O, Steindler DA (2005). PCR inhibition by reverse transcriptase leads to an overestimation of amplification efficiency. *Nucleic Acids Research* **33:** e181.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M *et al* (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research* **40:** e115.

VanGuilder HD, Vrana KE, Freeman WM (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44:** 619-626.

Werbrouck H, Botteldoorn N, Uyttendaele M, Herman L, Van Coillie E (2007). Quantification of gene expression of Listeria monocytogenes by real-time reverse transcription PCR: optimization, evaluation and pitfalls. *Journal of microbiological methods* **69:** 306-314.

Wong ML, Medrano JF (2005). Real-time PCR for mRNA quantitation. *Biotechniques* **39:** 75-85.

Zhang J, Byrne CD (1999). Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. *Biochemical Journal* **337:** 231-241.

Chapter 6

Dynamics of RNA viruses in the coastal waters of the Antarctic Peninsula

Jaclyn A. Mueller[1]

Alexander I. Culley[2] and Grieg F. Steward[1]

[1]Center for Microbial Oceanography: Research and Education, Department of Oceanography, University of Hawai‘i at Mānoa, 1950 East-West Road, Honolulu, Hawaii 96822

[2]Département de biochimie de microbiologie et de bio-informatique, Université Laval, 1045, avenue de la Médecine, Quebec, Canada, G1V 0A6

6.1 ABSTRACT

It is likely that viruses play a role in seasonal succession during the summer phytoplankton bloom in coastal Antarctic waters, but there have been few studies of the composition and dynamics of protist-infecting viruses there. In a recent metagenomic study of RNA viruses in the waters of the Western Antarctic Peninsula near Palmer Station, we found evidence that RNA viruses are abundant and appear to be dominated by protist-infecting members of the order *Picornavirales*. In this study, we used reverse-transcription quantitative PCR (RT-qPCR) to study the dynamics of four of the uncultivated viral phylotypes that appeared to be abundant based on metagenomic assembly data. Samples for RT-qPCR were collected by direct filtration approximately bi-weekly over the course of a summer phytoplankton bloom (13 November 2010 to 24 March 2011). Abundance of the RNA viruses ranged from 10^1 to 10^7 viruses ml^{-1} depending on the phylotype and the time of year. The four targets displayed three distinct patterns of abundance. One phylotype of unknown affiliation dominated early in the season and, despite declining by an order of magnitude by mid-season, was also dominant at the end of the study. Two other phylotypes were low in the beginning of the study, but displayed synchronized oscillations in abundance with a generally increasing trend that began in mid-summer. Despite increasing by an order of magnitude, these two were always the least abundant of the four phylotypes. A fourth phylotype that was phylogenetically most similar to known diatom-infecting viruses showed a single prominent peak, rising by more than three orders of magnitude to dominate the community for a month in mid-summer before dropping back to its starting concentration by the end of the study. The high concentration of some of the RNA virus phylotypes relative to epifluorescence counts of viruses provides independent support for our previous conclusion that RNA-containing viruses comprise a major fraction of the total virioplankton in this area. The dramatic differences in the dynamics among the phylotypes suggests that RNA viruses are a persistent force shaping the Antarctic plankton community, and may occasionally cause mass lysis thereby contributing to species succession.

6.2 INTRODUCTION

The Antarctic Peninsula undergoes a highly productive seasonal cycle every year with the advancement and retreat of the sea ice, and this summer period of high productivity supports a diverse array of fauna (Ducklow et al 2007). The bloom in the Western Antarctic Peninsula (WAP) tends to be bimodal, with a succession from diatoms to flagellates through the season (Garibotti et al 2005, Moline et al 2004). This successional pattern is reminiscent of that observed in other natural and experimental marine systems, where bloom termination and phytoplankton community succession were clearly attributable to viral activity (reviewed by Brussaard 2004).

It seems likely that viruses also contribute to the phytoplankton population dynamics in the WAP, but there is little known about the ecology of viruses in Antarctic waters (Pearce and Wilson 2003). Direct counts of total viruses by electron microscopy have shown that viruses are present and generally abundant around Antarctica (Bird et al 1993, Marchant et al 2000, Smith et al 1992). Others have investigated viral genome sizes (Steward et al 2000) and seasonal variation in viral concentrations (Pearce et al 2007) in the seawater around Antarctica, and even concentrations of viruses in sea ice (Gowing et al 2004), but all of these latter studies used methods that detect only larger DNA-containing viruses, and may therefore miss the majority of eukaryotic phytoplankton viruses (Steward et al 2013a).

We recently analyzed metagenomes of planktonic RNA viruses collected on six occasions from November 2010 to March 2011 in waters near Palmer Station, Antarctica (Mueller et al *Submitted*). From the mass ratios of RNA and DNA in the viral fractions, and read coverage of assembled genomes, we concluded that RNA-containing viruses were a large fraction of the virioplankton over much of the austral summer. Most of these viruses appeared to be members of the order *Picornavirales*, and many had sequences similar to those of viruses known to infect other eukaryotic phytoplankton such as diatoms (Mueller et al *Submitted*), but different specific phylotypes dominated at different times over the summer. These data provided the first indirect evidence that RNA viruses are both abundant and dynamic components of the Antarctic summer bloom dynamics.

In this report, we build upon our previous study by tracking changes in concentrations of individual phylotypes of RNA viruses that appeared to be relatively abundant based on assemblies of the metagenomic libraries. We first developed and evaluated a reverse-transcription, quantitative PCR assay (Mueller and Steward *In preparation*), then applied that method to samples collected independently in the same location and over the same time period as the metagenome samples, but at a higher frequency (every 3 to 7 days).

6.3 METHODS

6.3.1 Field sample collection

Seawater samples were collected from coastal waters at the Palmer Long Term Ecological Research (PAL-LTER) Station B (64° 46.45'S, 64° 03.27'W) located south of Anvers Island on the Western Antarctic Peninsula. Samples were collected nearly biweekly (23 total samplings) from November 13, 2010 through March 24, 2011 in 1.7 L Niskin bottles at 5 m using a hand-cranked winch from the side of a zodiac. Seawater was transferred to acid-washed (10%), Milli-Q rinsed 1 L brown amber bottles and transported to the 4°C cold room laboratory within 1 hr. Viruses and other plankton in the water were collected by direct filtration using a peristaltic pump. Whole seawater (approx. 1 L) from 5 m was filtered onto 0.02 μm anodic aluminum oxide (AAO) filters (Anotop 25 mm, Whatman). To detect signal associated with larger eukaryotic cells, whole seawater (approx. 1 L) from 5 m was filtered through a 5 μm polycarbonate filter (Nuclepore PC 25 mm, Whatman) in an Advantec polypropylene filter holder (Cole Parmer). The filtrate was then filtered through a 0.02 μm AAO filter (Anotop 25 mm, Whatman) to detect viruses in the < 5 μm fraction. After processing, the AAO filter inlets and outlets were sealed with Parafilm, and the PC filters carefully removed from the filter holder and transferred into a 2 ml cryovial. Samples were flash-frozen in liquid nitrogen, and stored at -80 °C until extraction. Corresponding temperature and salinity data were recorded with a Seabird CTD, and other associated environmental parameter data accompanying virioplankton samples were determined using standard LTER methods (http://pal.lternet.edu/publications/ documents/protocols/) and obtained from the Palmer LTER data site (http://oceaninformatics.ucsd.edu/datazoo/data/pallter/datasets).

6.3.2 Nucleic acid extraction

Total nucleic acids were extracted from 0.02 μm pore size AAO filters using a modified protocol of the MasterPure Complete DNA & RNA Extraction kit (Epicentre) previously described by Steward and Culley (2010) and rigorously evaluated in Mueller et al (2014). For each filter, 1 ml of Tissue & Cell Lysis Solution (MasterPure) containing 100 μg ml^{-1} proteinase K was loaded into a 3 ml-capacity sterile syringe and gently pushed through the filter via the outlet until liquid just reached the other opening. The other opening was then sealed with a second 3-ml syringe and liquid pulled through to further saturate the filter membrane. The assembly was incubated at 65 °C for 15 min while attached to a rotisserie in a hybridization oven rotating at full speed (ca. 16 rpm). The lysis buffer was then recovered by drawing it into the aspiration syringe. The extract was transferred to a 1.7 ml microcentrifuge tube and placed on ice. Salt-induced precipitation of protein-detergent complexes followed by an alcohol precipitation was then carried out following the manufacturer's instructions. Briefly, one half volume (500 μl) of MPC Protein Precipitation Reagent was added to the 1 ml of lysed sample and vortexed vigorously for 10 s. The precipitate was pelleted by centrifugation at 4 °C for 10 min at 16,000 × *g*, and the supernatant transferred to two clean 1.7 ml microcentrifuge tubes with ~ 750 μl each. Equal volumes of isopropanol (chilled on ice) were added to the recovered supernatant and the tube inverted 30-40 times. The total nucleic acids were pelleted by centrifugation at 4 °C for 30 min at 16,000 × *g*, spinning one tube at a time to concentrate into one pellet. The pellet was then rinsed with 2 × 70% ethanol (chilled on ice), with caution not to dislodge the pellet. Precipitated nucleic acids were resuspended in 100 μl RNA*secure* Resuspension Solution (Life Technologies) by heating the sample to 60 °C for 10 min after addition of preheated solution.

Total nucleic acids were extracted from the 5 μm polycarbonate filters using a similar modified protocol of the MasterPure Complete DNA & RNA Extraction kit. To the flash-frozen filter in a cryovial, 1 ml of Tissue & Cell Lysis Solution (MasterPure) was added, vortexed, and flash frozen in liquid nitrogen again to help break up cells. Samples were heated to 65 °C for 10 min to thaw completely, 100 μg ml^{-1} proteinase K was added, and samples were incubated again at 65 °C for 15 min, vortexing every 5 min. Samples were placed on ice for 5 min and lysis buffer was transferred to a fresh tube before continuing with salt-induced precipitation and alcohol precipitation according to the protocol described above.

6.3.3 DNase treatment and RNA quantification

Sample aliquots (50 µl) were incubated with 2 U TURBO DNase (Life Technologies) and 1 × TURBO DNase buffer at 37 °C for 30 min to eliminate cross-reaction from DNA (Levesque-Sergerie et al 2007, Steward et al 2013a). The reaction was terminated by adding 0.1 volume DNase Inactivation Reagent (Life Technologies) and incubating at room temperature for 5 min, mixing occasionally. Samples were centrifuged at 10,000 × *g* for 1.5 min and transferred into a fresh tube. The RNA was quantified for each sample using the Quant-iT RNA Assay Kit (Life Technologies) and a TD-700 fluorometer (Turner Designs).

6.3.4 Primers and probes

The primers and probes for this study are listed in Table 6.1. Those targeting the lysis protein and RNA replicase β chain of Bacteriophage MS2 for use as an internal standard were selected from previously published primer pairs (Mueller and Steward *In preparation,* O'Connell et al 2006). Bacteriophage MS2 RNA was chosen as an internal control because it is readily available as a purified extract and is not endemic to our area of sampling. Quantitative PCR primers and probes were designed to target the conserved region of the RNA-dependent RNA polymerase (RdRp) gene in marine picorna-like viruses (Culley et al 2006) from virus contigs PAL156, PAL438, PAL473, and PAL_E4 (Mueller et al *Submitted*) using Primer3 (Koressaar and Remm 2007, Untergasser et al 2012). The primers used to produce RNA transcript standards for the assembled viruses were designed outside these RdRp regions targeted by qPCR, using Primer3 (Koressaar and Remm 2007, Untergasser et al 2012) with a length of 825-1050 bp. HPLC-purified primers were obtained from Integrated DNA Technologies (Coralville, IA) and the TaqMan probes (with 5'-VIC or 5'-Carboxyfluorescein [FAM] and 3'-carboxymethylrhodamine [TAMRA]) were obtained from Life Technologies.

6.3.5 Reverse transcription

Extracted, DNase-treated samples were diluted 1:10 with nuclease-free water prior to reverse transcription, as some inhibition was seen in the RT reaction with undiluted samples (data not shown). This is a standard method of minimizing RT inhibition (Kleiboeker 2003). RT reactions were prepared with a master mix using SuperScript II Reverse Transcriptase (Life Technologies)

for the respective replicates of samples, and spiked with background *E. coli* total RNA (Life technologies) and tMS2-6 RNA transcripts produced as described in Mueller and Steward (*In preparation*). Aliquots of the concentrated stocks of each type of RNA were stored frozen (-80 °C) and each was used only once to avoid degradation from repeated freezing and thawing. Dilutions were made from freshly thawed stocks for each run. Each RT assay was performed in duplicate with 3.35 nmol random hexamers (168 µM final concentration in 20 µl), 10 nmol dNTPs, 1 µg *E.coli* total RNA, 1×10^4 copies tMS2-6 RNA, 2 µl extracted RNA sample (1:10 diluted with nuclease-free water) and nuclease-free water to reach a final volume of 12 µl per reaction. The reaction was then heated to 65 °C for 5 min, quickly chilled on ice, and 1 × First-Strand Buffer, 0.2 pmol DTT, and 40 U RNAseOUT added for a final volume of 19 µl. The reaction was incubated at 25 °C for 2 min, then 1 µl (200 U) SuperScript II RT was added, followed by incubations at 25 °C for 10 min, 42 °C for 50 min, and inactivation at 70 °C for 15 min. The reactions were then treated with 1 µl (2 U) RNase H (Life Technologies) and incubated at 37 °C for 20 min to remove RNA complementary to the cDNA.

6.3.6 RNA standard curve synthesis by in vitro transcription

The transcript tMS2-6, which was added at a low constant concentration to all reactions as an internal control for inhibition, and the transcripts for each viral target that were used for standard curves, were produced as outlined in Mueller and Steward (*In preparation*) using SuperScript II (Life Technologies) with 168 µM random hexamers and 50 ng μl^{-1} background RNA. RNA transcripts for PAL targets were produced from PCR amplicons of the target sequences. Amplicons were generated starting from the same cDNA reactions used to create the metagenomic library RNAV_PAL20110107 described in Mueller et al (*Submitted*) using XXX_1F forward primer and XXX_1R reverse primer for each corresponding target where XXX represents the target number (Table 6.1). PCR was conducted with Expand High Fidelity PCR Plus System (Roche) with 2 µl cDNA reaction, 1 × reaction buffer, 0.2 mM dNTP, 400 nM each primer, and 2.5 U Expand High Fidelity Plus Enzyme in 50 µl final reaction volume. The thermal cycling conditions were as follows: initial denaturation at 94 °C for 2 min, followed by 30 cycles consisting of denaturation at 94 °C for 15 s, annealing at 60 °C for 30 s, and extension at 72 °C for 1 min, with a final extension step at 72 °C for 7 min. The amplicon was concentrated and purified with a MinElute PCR Purification Kit (Qiagen) and run on a 0.5 % agarose gel in 1

× TAE stained with SYBR Safe. The single band was visualized with a Safe Imager 2.0 Blue-Light Transilluminator (Life Technologies), excised and purified with a MinElute Gel Extraction Kit (Qiagen), and verified by sequencing. For that purpose, the amplicons were cloned into the pSMART HC Kan vector (Lucigen) according to the manufacturer's instructions. The same plasmid was linearized with Swa I restriction enzyme (New England Biolabs) according to the manufacturer's instructions, visualized on a 1% agarose gel to ensure complete digestion, and cleaned up with a MinElute Reaction Cleanup Kit (Qiagen) for use as template for PCR with T7 primers.

A forward primer containing a T7 promoter (Table 6.1) was then used in PCR amplification with 1R reverse primers using the Expand High Fidelity PCR Plus System (Roche) with 2 µl cDNA reaction, 1 × Reaction Buffer, 0.2 mM dNTP, 400 nM each primer, and 2.5 U Expand High Fidelity Plus Enzyme in a final volume of 50 µl. The thermal cycling conditions were the same as above. The amplicon was concentrated and gel purified as described above, then quantified with a Qubit 2.0 Fluorometer (Life Technologies). T7 in vitro transcription was conducted to produce the RNA standard transcripts with MEGAscript T7 Transcription Kit (Life Technologies). T7 transcription reactions contained ~250 ng PCR product, 1 × Reaction Buffer, 7.5 mM each rNTP, and 2 µl T7 Enzyme Mix in a final 20 µl reaction volume. Transcription reactions were incubated for 4 h at 37 °C, followed by DNase treatment with 1 µl TURBO DNase (Life Technologies) and 15 min incubation at 37 °C. The RNA was then cleaned up with a QIAamp MinElute Virus Spin Kit (Qiagen) and eluted in RNA*secure* Resuspension Solution (Life Technologies). The amount of RNA was estimated by spectrophotometry with dilution in TE (pH 8.0) and measured on a DU-800 spectrophotometer (Beckman Coulter). The transcript RNA was then pooled at the same copy number μl^{-1} for all targets (tPAL156, tPAL438, tPAL473, and tPAL_E4), including MS2 (tMS2-6), and serially diluted. Each dilution was synthesized into cDNA with random hexamers as described above for the samples, but without the spike of and additional tMS2-6.

6.3.7 Quantitative PCR

All quantitative PCR assays were performed in duplicate in a Mastercycler ep realplex^2 thermal cycler (Eppendorf). For quantification of the MS2 RNA replicase gene and PAL target RdRp

genes, the hydrolysis probe technology (TaqMan probes) was used as the detection system. The qPCR was performed in duplex reactions for each target (PAL target plus MS2) containing 1 × TaqMan Fast Advanced Master Mix, 300 nM each forward and reverse primers, 250 nM each probe, 2 µl cDNA, and nuclease-free water to reach a final volume of 20 µl per well. The final primer and probe concentrations of 300 and 250 nM, respectively, were determined as the optimum concentration of primers and probes (data not shown). Thermal cycling conditions were as follows: 50 °C for 2 min (UNG incubation) preceding the first denaturation step at 95 °C for 20 s, followed by 45 cycles of 95 °C for 15 s and 60 °C for 30 s. RT-qPCR data were analyzed using the default noisebands from the Mastercycler ep realplex^2 thermal cycler (Eppendorf) as the threshold for each target (MS2=VIC; PAL targets=FAM). The slopes of the calibration curves were calculated from the plot of the log base 10 of initial target copy number versus corresponding C_q and used to determine the number of gene copies in each RT sample. The qPCR efficiency (*E*) was determined from the slope of the curve obtained with the serially diluted standards, as $E = 10^{(-1/\text{slope})}$.

6.3.8 Cross-reaction tests

In order to confirm that the primers and probes for each target did not cross-react with one another, qPCR assays were run with only one target per reaction and with all the targets pooled together. For these tests, the linearized plasmids containing each target (as described above) were used in simplex (one set of primers and probe) and duplex (PAL target plus MS2 primers and probes) reactions containing 2.5×10^3 µl^{-1} of each target, except for PAL_E4 at 2.5×10^4 µl^{-1}. Reactions were run in duplicate and analyzed as described above.

6.3.9 Relative abundance vs absolute abundance

As a means to compare the absolute abundances determined by RT-qPCR with our previous metagenomic analyses (Mueller et al *Submitted*), we converted the absolute abundances into percent abundances. This rough estimate assumed that the summed abundances of the four virus targets represented the total virus abundance of the community in a given sample. Since samples were not always collected on the same days as the metagenome samples, an average was calculated using the data from the two dates before and after the metagenome sample, and converted to an average percent (±SD). The relative abundance determined from the

metagenomic data (based on reads assembled to each genome; Mueller et al *Submitted*) were converted into a percent as well, assuming that the same four virus targets represented the entire virus community.

6.4 RESULTS

6.4.1 RNA standard curves

The RNA transcripts produced reliable and efficient standard curves (Fig. 6.1). The qPCR standard curve efficiencies ranged from 89-94% for the PAL RdRp target transcripts (tPAL156, tPAL438, tPAL473, and tPAL_E4), and 83-87% for the tMS2-6 transcript. The standard curves reliably detected targets down to 500 copies μl^{-1} of RT-qPCR reaction (50 copies μl^{-1} of qPCR reaction). The C_q values for tMS2-6 were consistently higher than those of the transcripts for the four viral targets for a given copy number (2 to 6 cycles depending on the target). The potential cross-reaction of primers and/or probes between each target was eliminated in qPCR cross-reaction tests with each target individually and pooled together (Fig. 6.2). The C_q values from each reaction with only one target, or with all targets pooled and run in simplex or duplex reactions were indistinguishable for a given PAL target (ANOVA; $P > 0.05$).

6.4.2 MS2-6 internal inhibition control

The C_q value for tMS2-6 inhibition control in the absence of sample was 35.87±0.18 (mean ± 95% C.I., n = 8). The C_q values for the tMS2-6 transcript spiked at the same concentration into each sample as an inhibition control ranged from 34.7 to 38.8, but the range was narrower when considering the set of reactions for each viral target individually (Fig 6.3). The C_q of the inhibition control was significantly negatively correlated with the C_q of the target for PAL156 ($r = -0.77$; $P < 0.001$), PAL438 ($r = -0.62$; $P < 0.01$), and PAL473 ($r = -0.91$; $P < 0.001$) when analyzed individually, and also when the results for these three targets were considered in aggregate ($r = -0.89$, $p < 0.001$). The C_q values of the inhibition control in the PAL_E4 reaction varied the least (± 0.19 cycles of the expected value for no inhibition) and showed no significant correlation with PAL_E4 target concentration ($r = 0.10$; $P = 0.66$).

6.4.3 Environmental parameters

The environmental parameters varied throughout the season (Fig. 6.4). Salinity ranged from 33.2 to 33.9 PSU and temperature ranged from -1.1 to 2.3 °C, with a maximum on January 20, 2011. Phosphate concentrations ranged from 0.6 to 1.7 µM, silicate concentrations ranged from 32.3 to 81.8 µM, and nitrite + nitrate ranged from 5.2 to 28.2 µM. Chlorophyll *a* values ranged from 0.76 to 29 mg/m^3, and primary production ranged from 1.2 to 140 mg/m^3/day.

6.4.4 Dynamics of uncultivated RNA viruses

Each RNA virus target displayed dynamic and unique patterns of abundances throughout the season as determined by RT-qPCR (Fig. 6.5). Copy numbers of PAL438 were always the lowest of the four targeted viruses, but the concentration of the target generally increased over the course of the study (late spring to early fall) from 6.0×10^1 to 9.9×10^3 viruses ml^{-1}. The increase was marked by several local maxima that occurred on January 3, February 10, and March 14, 2011. The target PAL473 was consistently higher than PAL438 by a factor of three to ten-fold but otherwise displayed a virtually identical pattern of concentration over time. The abundances of PAL438 and PAL473 were significantly correlated throughout the season ($r = 0.99$; $P < 0.001$; Fig. 6.6). PAL_E4 had the highest abundance at the beginning of the season (10^6 ml^{-1}), and was highest at the end of the season (10^5 ml^{-1}), despite declining in abundance by an order of magnitude. Its abundance was surpassed only during the mid-season by the peak of PAL156. PAL156 was the third most abundant target early in the season (10^4 ml^{-1}), but increased by three orders of magnitude (5×10^7 ml^{-1}) to temporarily become the most abundant of the four targets in January around the mid-summer low of primary production and maximum in chlorophyll. The concentration of PAL156 dropped back to pre-bloom concentrations by the end of the study.

6.4.5 Percent abundance (absolute vs relative)

The percent abundances determined from the absolute and relative abundance data were well correlated for PAL156 and PAL_E4 ($r = 0.89$, $P = 0.018$; $r = 0.84$, $P = 0.04$, respectively; Fig. 6.7). The data were not correlated for the lower abundance targets (PAL438: $r = -0.20$, $P = 0.71$; PAL473: $r = -0.05$, $P = 0.92$). Absolute abundances from RT-qPCR data were converted to

percent abundance, assuming the total abundance of viruses was equal to the sum of the four virus targets. Relative abundances from metagenomic analyses (Mueller et al *Submitted*), were also converted to percent abundance, with the same assumption that the total abundance of viruses was equal to the sum of the four virus targets.

6.4.6 Size fractionation of viruses

The majority of total detectable PAL156 targets (74 to 99%) was found in the < 5 μm fraction throughout the study period, but relative changes in abundance over time were similar in the ≥ 5 μm fraction (Fig. 6.8, upper panel). Although the log concentrations in the small and large size fractions were correlated ($r = 0.96$, $P < 0.001$), the fraction of PAL156 targets in the ≥ 5 μm fraction was significantly higher in early January at the time of maximum total target abundance than at any other time (ANOVA with *post hoc* Tukey test; $P < 0.00001$; Fig. 6.8, lower panel).

6.5 DISCUSSION

Following the discovery that there are diverse picornavirads in temperate coastal seawater (Culley et al 2003, Culley et al 2006), it was shown that these types of viruses were also present in tropical coastal seawater (Culley and Steward 2007), but their numerical contribution to the virioplankton pool was unknown. Subsequent indirect estimates of the relative abundance of RNA viruses, first in tropical waters (Steward et al 2013a) and then polar waters (Mueller et al *Submitted*), suggested that they could often equal or exceed the abundance of DNA viruses in some habitats. It appears that most of these marine RNA viruses belong to a very broad clade within the order *Picornavirales* (Culley et al 2006, Culley et al 2014, Mueller et al *Submitted*) and are likely infecting marine protists (Culley and Steward 2007, Lang et al 2009). Metagenomic and targeted gene analyses of marine RNA viruses suggest that the composition of the RNA viral assemblage is spatially variable and temporally dynamic (Culley et al 2003, Gustavsen et al 2014, Mueller et al *Submitted,* Steward et al 2013a), but these studies provide little detail about how the absolute abundance of individual populations change over time. In this study, we employed RT-qPCR to track for the first time the temporal dynamics of individual members of marine RNA viruses at high resolution over the course of an Antarctic summer phytoplankton bloom. We first discuss the technical aspects of the study, then its contributions to our understanding of viral ecology.

6.5.1 Evaluating the accuracy of RNA virus abundance estimates

RT-qPCR standard curves produced with RNA transcripts from the RdRp region of each of the PAL targets resulted in reliable and reproducible standards that allowed us to estimate the absolute abundances of these viruses in our extracts. A previous evaluation of the virus harvesting and extraction method we used in this study suggests that it is very efficient (Mueller et al 2014), but in the absence of an extraction control, our default assumption of 100% extraction efficiency could result in underestimates of the actual environmental abundances. The accuracy of abundance estimates also depends upon how well the synthetic standard mimics the full-length target and whether the reaction efficiencies are the same in the presence and absence of sample. We discuss these two possible sources of error below.

We found previously that our synthetic transcript tMS2-6 is detected by RT-qPCR with somewhat lower efficiency (ca. 3-fold) than full-length MS2 genomic RNA and provided evidence that the detection efficiency for a transcript primed with random hexamers is strongly influenced by the availability of primer binding sites upstream of the qPCR reverse primer (Mueller and Steward *In preparation*). The upstream regions of the synthetic transcripts we constructed for the PAL standard curves are 1.4 to 3 times longer than that of tMS2-6, and this could explain, in part, why their C_q values are consistently lower for the same number of transcripts. These substantially lower C_q values imply that the PAL standards are unlikely to result in overestimates of the corresponding target abundances like we observed for tMS2-6 (Mueller and Steward *In preparation*). If anything, we may be somewhat underestimating the viral target abundances.

A common source of error in RNA target abundance estimates by RT-qPCR is the presence of RT or PCR inhibitors in samples that are not present in the standard curve, which will result in the underestimation of target concentration (Nolan et al 2006). To determine whether such inhibitors might be influencing our results, we added a known amount of tMS2-6 to every RNA extract. In the absence of inhibition, the C_q values for the inhibition control should be the same as the C_q value obtained for that control when prepared in pure water. This was true in the duplex reactions with one of the targets (PAL_E4), but a systematic variability was observed in the C_q values for tMS2-6 in the duplex reactions with the other targets. Because the same RT reaction

was used in the qPCR reactions to detect all targets, our results demonstrate that the observed variability could not have resulted from inhibition of the RT reaction, but is likely a conditional result of the qPCR assay. Furthermore, the systematic variability of tMS2-6 C_q as a function of target concentration for three of the targets displayed a significant correlation, which strongly suggests that the variability in C_q also did not result from general inhibition of the qPCR (Fig. 6.3). Instead, the inhibition likely resulted from competition between these three viral targets and the control transcript during PCR amplification. We do not know why the C_q for tMS2-6 dropped below the expected minimum at the lowest target concentrations (duplex reactions with PAL438), but this suggests that real sample extracts are not necessarily inhibitory, but can improve efficiency. Even if we were to take the inhibition control data at face value, the implied error suggests that we would have underestimated the highest values for PAL156 by about two cycles, or a factor of four. However, it seems very unlikely that there was a reciprocal inhibition by the control transcript on our viral targets, because the control transcript was added at very low concentrations (C_q values > 34).

From the above arguments, it seems unlikely that we have overestimated the concentrations of the target viruses, but we are likely to have somewhat underestimated abundances, because of losses during extraction and purification. Even so, the possible sources of over- or underestimation appear to be small relative to the orders of magnitude changes seen in copy numbers for the individual phylotypes. Dramatically different patterns displayed by the viral targets also argue strongly that the dynamics we observed are not artifacts stemming from varying efficiencies of viral collection, nucleic acid extraction, or detection efficiency by RT-qPCR. Additional strong support for our RT-qPCR-based estimates of in situ concentrations of the four target viruses comes from our previous metagenomic analyses (Mueller et al *Submitted*). The samples for metagenomics were collected and processed independently of the samples for RT-qPCR analysis using very different methods, but the results are in very good agreement for the two most abundant targets (PAL156 and PAL_E4; Fig. 6.7).

6.5.2 Quantitative contribution of RNA viruses to the virioplankton

The numerical contribution of RNA viruses to the total virioplankton in seawater has been a long-standing question. Indirect estimates based on mass ratios of RNA and DNA from purified

marine viruses, first in tropical waters (Steward et al 2013a) then in polar waters (Mueller et al *Submitted*), suggested that RNA viruses may rival or exceed the abundance of DNA viruses in these locations. Our RT-qPCR data provide additional support for this conclusion using an independent approach. Counts of viruses by epifluorescence microscopy using SYBR Gold ranged from 0.2 to 1.1 x 10^7 ml^{-1} from November 2010 through February 2011 (Brum et al 2011). These are nominally "total virus" counts, but are unlikely to include RNA viruses or small single-stranded DNA viruses (Brussaard et al 2000, Holmfeldt et al 2012, Tomaru and Nagasaki 2007). The sum of the RNA phylotypes that we quantified by RT-qPCR provides us with a minimum estimate of the total RNA virus abundance over this same time period that ranged from 0.2 to 5 x 10^7 ml^{-1}. Comparing the estimates on 4 occasions in this time period suggest that the RNA viruses ranged from 14 to 90% of the total viruses with a mean of around 40%. Both approaches have possible sources of error, so the exact contribution of RNA viruses remains uncertain, but with either approach, RNA viruses appear to make a significant and occasionally dominant contribution to the virioplankton.

6.5.3 RNA virus population dynamics

Although the interactions between viruses and their hosts can be complex, ranging from parasitism to mutualism (Roossinck 2011, Sime-Ngando 2014, Steward et al 2013b), viruses of eukaryotic phytoplankton are often lytic (Short 2012) and display strain-specific, density-dependent propagation that can be modeled according to classical predator-prey dynamics (May 1973). Therefore, although we know neither the identities nor the abundances of the hosts for each of the viral phylotypes we tracked, our data provide some hints about the underlying virus-host interactions. We assume these abundance patterns represent temporal variability of virus-host abundances throughout the bloom. Because there was no obvious correspondence between the abundances of these viruses and the water mass data (as determined by salinity and temperature), we do not think these data represent intrusions of different water masses into the sampling site. However, patchiness within the same water mass could explain some of the temporal variability displayed, but we could not control for this.

Oscillations in the abundance of the two lower abundance phylotypes (PAL438 and PAL473) are reminiscent of stable predator-prey cycles. It is possible that the viruses are driving these cycles

by controlling host abundances. We cannot rule out, however, that the hosts are being controlled by some other mechanism and the viruses are simply tracking host abundance. The phylogenetic distance between these two viruses, and differences in their nearest neighbors (Mueller et al *Submitted*) seems to suggest that they infect different host organisms, so it is curious that they displayed virtually identical patterns of abundance. We eliminated the potential cause due to cross-reaction of the targets, showing nearly identical results from reactions with each individual target and pooled targets (Fig. 6.2). One explanation for these similar patterns is that these viruses infect two host species that have strongly coupled dynamics, whether a result of symbiosis, syntrophy, or the occupation of ecological niches that are linked in some other way. It might be that they are infecting the same host, but it seems surprising that distinct viruses exploiting the same resource would track each other so well over an extended period without one displacing the other. One way this might occur is if PAL438 were a satellite virus of PAL473, reliant upon co-infection of the host cell for its own replication. This could account for the difference in their absolute abundance, but similar relative changes over time. While satellite viruses are well described in some systems (for example, in plants), very few have been identified in the marine environment.

The dynamics of PAL_E4 were particularly confounding. The persistently high relative and absolute abundance of this phylotype implies a continuous and significant source for the virus starting very early in the spring. This seems to argue against an obligate photoautotroph being the host, but perhaps a protistan species that can thrive in the lower light early in the season (mixotroph or heterotrophic bacteriovore) was present at sufficient concentrations to support the observed viruses. PAL_E4 branches so deeply within the order *Picornavirales* that we have little basis for hypothesizing what kind of organism it might infect. It is possible that its source is infected planktonic animals or that it had exogenous sources (introduction from sea ice microbial communities, runoff from land), but it is hard to imagine any of these sources maintaining such high concentrations in the water, unless this virus has an unusually low decay rate.

PAL156, which clusters most closely with viruses infecting a *Thalassiosira* sp. and *Chaetoceros* sp., displayed the most dramatic changes in abundance. A single large peak around mid-summer suggests that the host of this virus was a dominant diatom species that was driven to very low levels by an epidemic viral infection. This occurred prior to the maximum peaks in chlorophyll,

suggesting that the virus may have been involved in an early bloom succession, with the subsequent rise in chlorophyll being driven by other phytoplankton species. It is also possible that a virus-resistant strain of the same species contributed significantly to the bloom that followed.

The very close agreement between our estimates of PAL156 abundance in the whole water and those in the summed large and small size fractions, which were filtered, extracted, and analyzed separately, demonstrates the high degree of reproducibility of our assays. We had assumed that most of the Antarctic eukaryotic phytoplankton cells, especially diatoms (the putative host type for PAL156), would be removed by 5 μm filtration. If this is true, then the size distribution of signal suggests that most of the PAL156 viruses were present as free virions (as opposed to intracellular) throughout the bloom. We had anticipated that we might see an offset in the peak signals in the large and small size fractions, with the large size fraction peaking first as the spread of the infection accelerated, then a "free virus" signal peaking later when the rate of cell lysis began to exceed the rate of new infections. We found instead that the peak signals in the large and small fractions were coincident, but our sampling resolution may simply have been too low to distinguish these stages of the epidemic.

6.6 CONCLUSIONS

Our quantitative data provide additional direct support for the idea that RNA viruses comprise a large fraction of the total virioplankton in Antarctic coastal waters. The unique dynamics of several abundant, uncultivated RNA viruses during the summer phytoplankton bloom in Antarctic coastal waters provide evidence that these viruses make varied contributions to Antarctic plankton ecology. The dramatic rise and fall of one putative diatom-infecting virus in particular suggests that it contributed to succession in the phytoplankton community. Viral lysis of phytoplankton is expected to have very different consequences for fate of primary production compared to consumption by zooplankton. Our results suggest, therefore, that additional work to constrain the magnitude of viral lysis of primary producers will be essential to improving models of the Antarctic food web.

6.7 ACKNOWLEDGEMENTS

We thank Christopher Schvarcz for his invaluable assistance in sample collection and for providing RdRp sequences from his phytoplankton virus isolates. We thank and the Palmer LTER group for assistance with sampling and logistics. In particular, Alice Alpert, Jennifer Brum, Kaycee Coleman, Carolina Funkey, Mike Garzio, and Edgar Woznica provided instrumental assistance with collecting samples. This work was supported by NSF grants to GFS and AIC (ANT 09-44851) and by NSF funding to the Center for Microbial Oceanography: Research and Education (EF 04-24599).

Table 6.1. Primer and probe sequences for reverse transcription, real-time quantitative PCR (RT-qPCR) assays for quantitation of targets.

Target	Primer[a]	Start[b]	Sequence	Size[c]
MS2	MS2-T7-4F	1693	TAATACGACTCACTATAGGGCCTCAGCAATCGCAGCAAA	
	MS2-4F	1693	CCTCAGCAATCGCAGCAAA	742
	MS2-6R	2435	TCTGATGAAAGCACCGACCC	
	MS2-5F	2232	GCTCTGAGAGCGGCTCTATTG	69
	MS2-5P	2255	[VIC]-CCGAGACCAATGTGCGCCGTG-[TAMRA]	
	MS2-5R	2301	CGTTATAGCGGACCGCGT	
PAL156	T7_156_1F	3612	TAATACGACTCACTATAGGGCACGGTCCGAAAATCAATCA	
	156_1F	3612	GCACGGTCCGAAAATCAATCA	1047
	156_1R	4659	CGCATTGCTCCTCCGATATTC	
	156_qF	4325	TCCGCGATTTGGTCTCACTT	149
	156_qP	4353	[FAM]-CGGTGATGACGCAAAGGGATCTGTCCG-[TAMRA]	
	156_qR	4474	GGGATCAGATGTTTTGTCTGGC	
PAL438	T7_438_1F	3517	TAATACGACTCACTATAGGGTGCCTAACGCTAAAGCTGATGA	
	438_1F	3517	TGCCTAACGCTAAAGCTGATGA	850
	438_1R	4367	TCACGGAACGAACCTTCTGC	
	438_qF	4024	TTTCTCAAAATGTTCGTCTTGGC	132
	438_qP	4096	[FAM]-ACACCCACACTTCTTGCCAGGAACAGT-[TAMRA]	
	438_qR	4156	CAGCCATTGTGTAGCCAATATCG	
PAL473	T7_473_1F	1877	TAATACGACTCACTATAGGGTACACCAGCCAAACCTTTCGA	
	473_1F	1877	TACACCAGCCAAACCTTTCGA	1001
	473_1R	2878	CTGGCATGGTGTAAATCTGACC	
	473_qF	2538	TGCGATTACAGCGAGGAAGA	137
	473_qP	2561	[FAM]-TCGGGTGATGGAAGCAATGACTGGTGA-[TAMRA]	
	473_qR	2675	ACGACAGTCAAGGAGTTTCCA	
PAL_E4	T7_E4_1F	3742	TAATACGACTCACTATAGGGTGGTCGTTATGTCACCTGCC	
	E4_1F	3742	TGGTCGTTATGTCACCTGCC	826
	E4_1R	4568	GGTAAACCCGAGCCTAGCAG	
	E4_qF	4042	GAATCCTGCTGAGTCCTGGG	127
	E4_qP	4064	[FAM]-GACCTTCACAAGCGGTTTAGCAGGCT-[TAMRA]	
	E4_qR	4169	GACAGCAGCATCGAGAAGGT	

[a]F = Forward, P = Probe, R = Reverse, T7 = T7 promoter, q = qPCR primer.

[b]Position of first base in each oligonucleotide sequence relative to the whole genome sequence in GenBank accession no. NC_001417 and in assembled contigs.

[c]Size of PCR amplicons for standard transcripts (tMS2-6, tPAL156, tPAL438, tPAL473, and tPAL_E4) and size of qPCR targets.

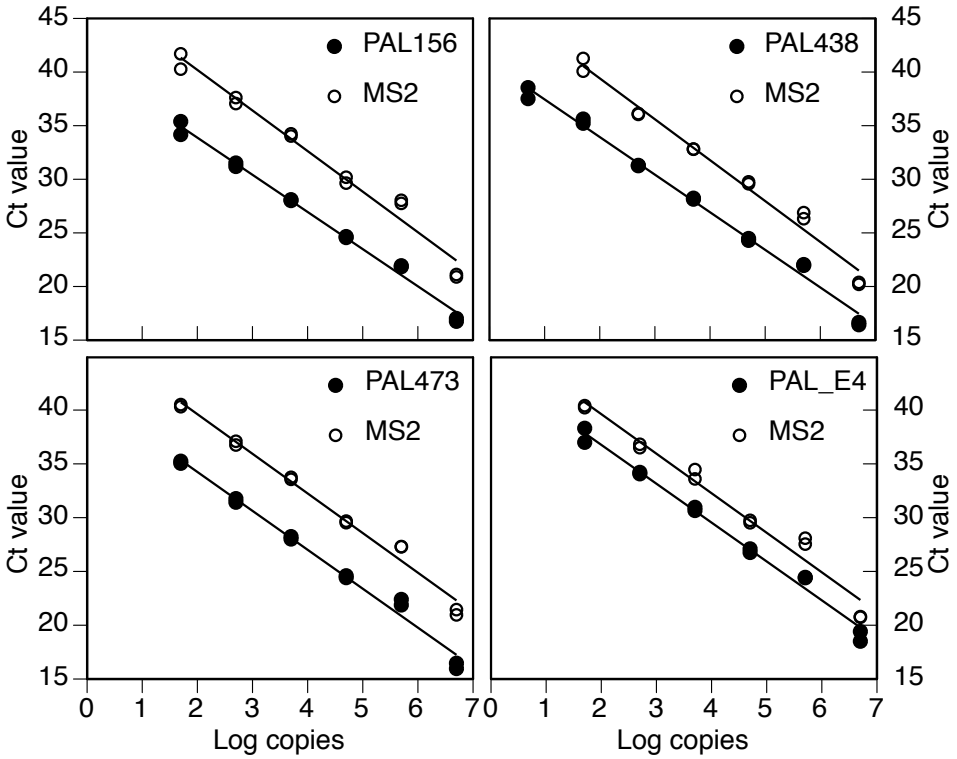


Figure 6.1. Standard curves of RNA transcripts for tPAL156, tPAL438, tPAL473, tPAL_E4, and tMS2-6. Data shown are from duplex qPCR assays performed in duplicate for each RT reaction containing a serial dilution of all five targets (5×10^1 to 5×10^6 copies μl^{-1} qPCR assay).

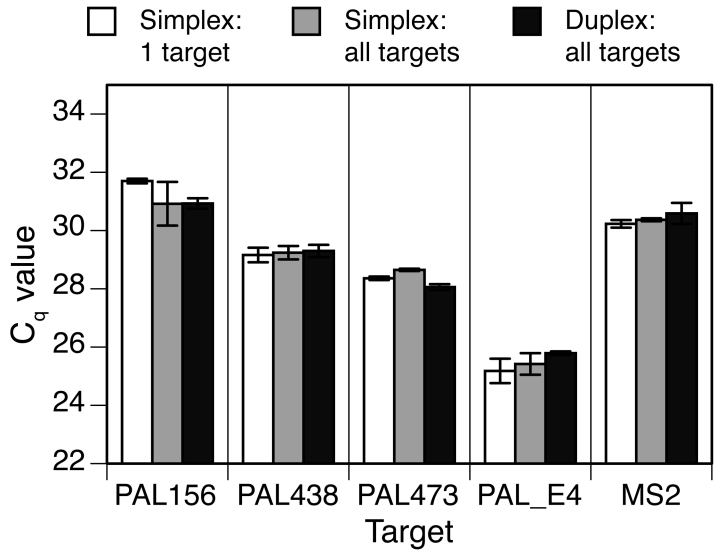


Figure 6.2. Cross reactions for each set of primers with only one target per reaction and with all targets (including MS2) pooled together. Simplex reactions contain primers and probes for only one target, and duplex reactions contain primers and probes for one PAL target, plus MS2. Data shown are the average and standard deviation of C_q values from duplicate qPCR reactions containing 2.5×10^3 copies μl^{-1} of each target (as DNA plasmids), except for PAL_E4 at 2.5×10^4 copies μl^{-1}.

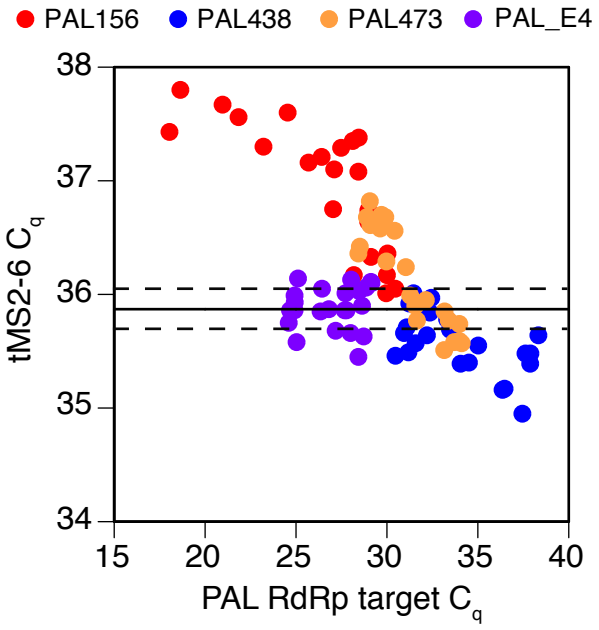


Figure 6.3. C_q values of tMS2-6 transcript internal control relative to C_q values of PAL156, PAL438, PAL473, and PAL_E4 RdRp targets. Each sample RT reaction was spiked with 1×10^4 copies tMS2-6 transcript RNA. Data shown are from duplicate RT reactions quantified in duplicate. The solid line indicates the mean tMS2-6 C_q value (± 95% confidence in dashed lines) from RT reactions spiked with 1×10^4 copies tMS2-6 quantified in simplex qPCR reactions.

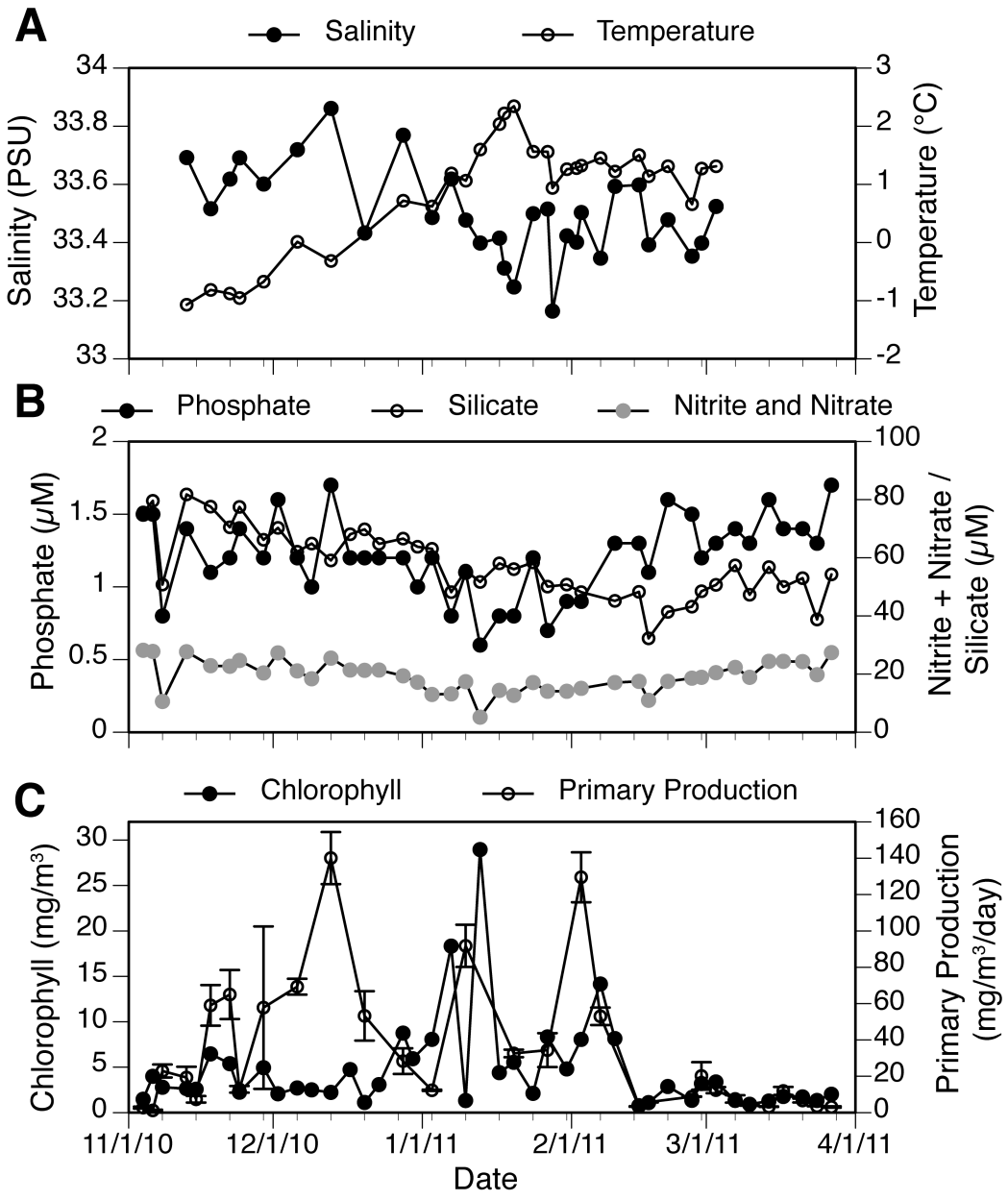


Figure 6.4. Environmental parameters. **A.** Salinity and temperature data collected from Seabird CTD. **B.** Phosphate, silicate, and nitrite + nitrate data. **C.** Chlorophyll and primary production. Data from B and C were determined by the PAL LTER program.

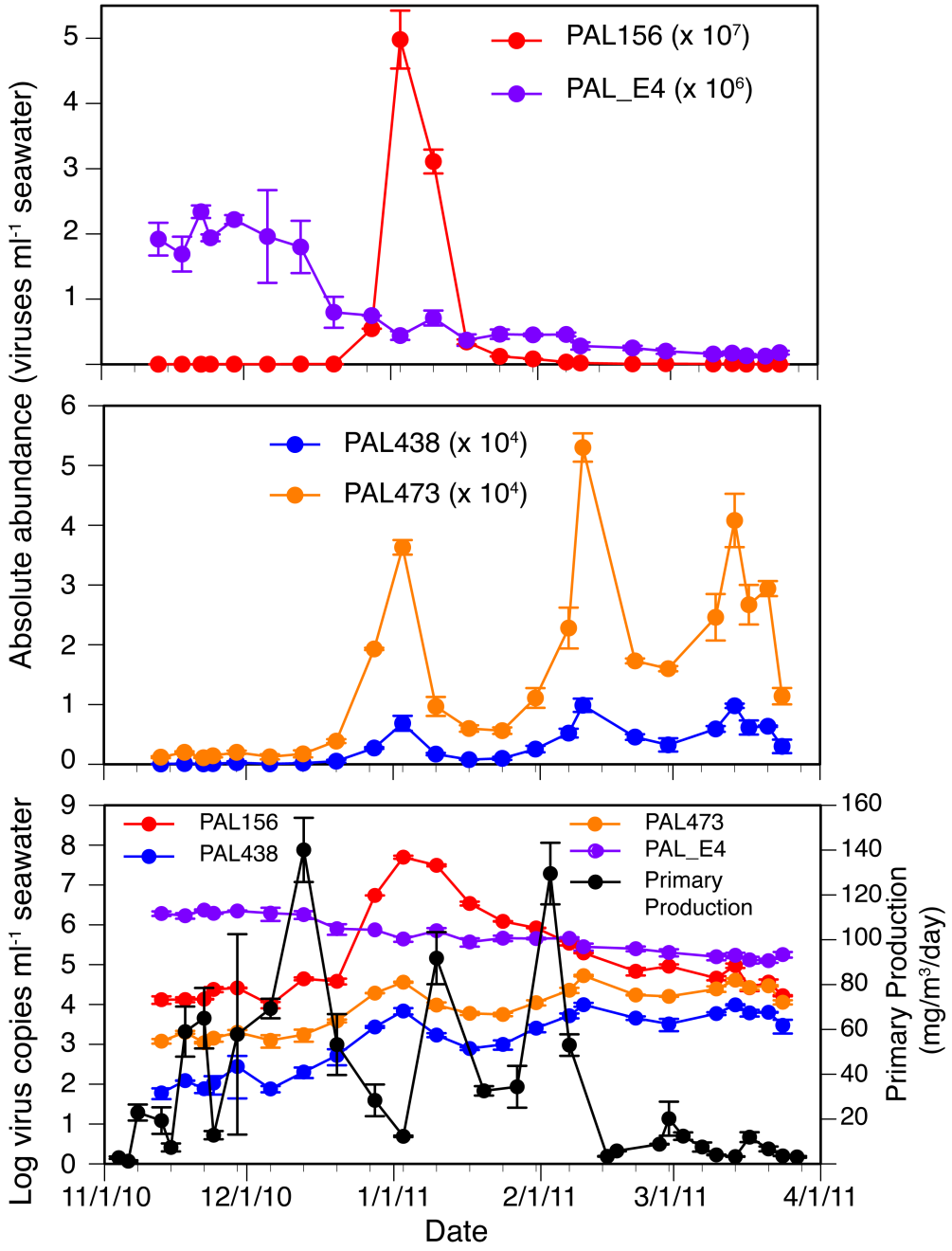


Figure 6.5. Seasonal shifts in RNA virus abundance and primary production. Upper 2 panels: Absolute target abundances are × 10^n ml^{-1} seawater as indicated in each legend, and represent the means and standard deviations from duplicate RT reactions quantified in duplicate qPCR reactions. Bottom panel: Log-transformed concentrations of all four viral targets are plotted with untransformed primary production data to illustrate relative concentrations among the viruses and their relationship to primary production. Primary production data were determined by the PAL LTER program.

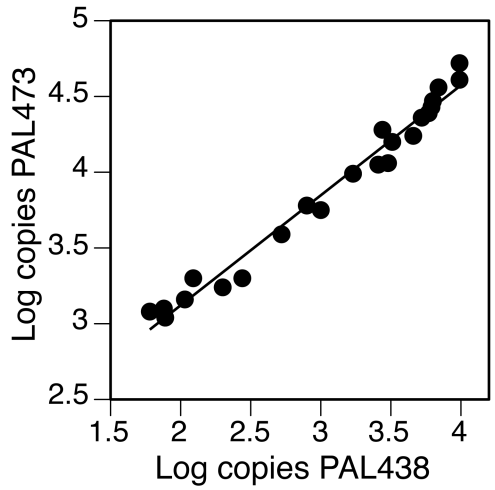


Figure 6.6. Relationship between PAL438 and PAL473. Log copies of PAL438 relative to PAL473 from data collected at Palmer Station throughout the season.

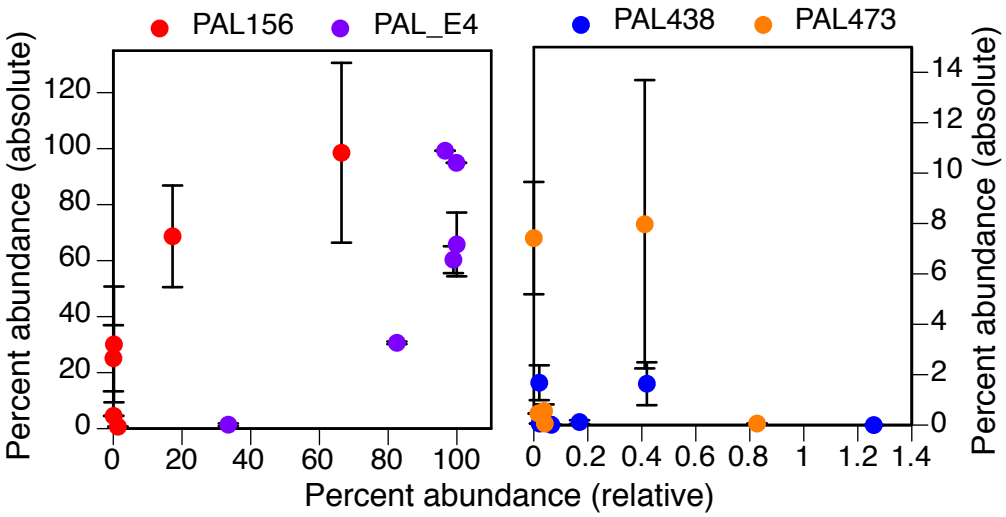


Figure 6.7. Percent abundance of each virus as determined from relative and absolute abundances. Absolute abundances were determined by RT-qPCR, and relative abundances were determined by metagenomic analyses (Mueller et al *Submitted*). Data for PAL156 and PAL_E4 were significantly correlated (left panel), while data for PAL438 and PAL473 showed no correlation.

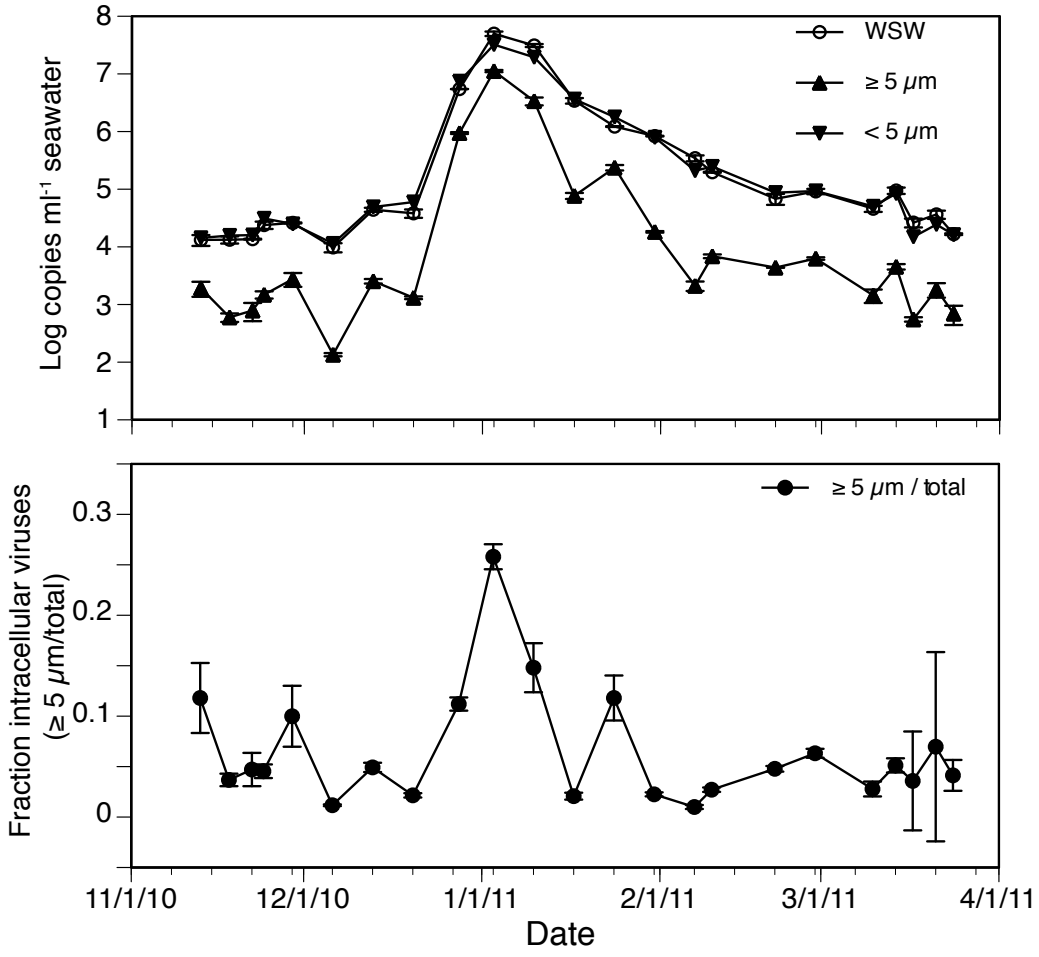


Figure 6.8. PAL156 virus abundance in whole seawater and size fractions. Upper panel. Log transformed target abundance was determined from extracts of whole seawater filtered directly onto 0.02 µm Anotops (WSW) or onto 5 µm polycarbonate filters (≥ 5 µm), as well as the 5 µm filtrate filtered onto 0.02 µm Anotops (< 5 µm). Data shown are mean and standard deviations of duplicate RT reactions quantified in duplicate qPCR reactions. Lower panel. The fraction of intracellular viruses was determined by dividing the intracellular abundances (≥ 5 µm) by the total viruses (WSW). Data shown are the mean and propagated percent error of the ratios.

6.8 REFERENCES

Bird DF, Maranger R, Karl D (1993). Palmer LTER: Aquatic virus abundances near the Antarctic Peninsula. *Antarctic Journal***:** 234-235.

Brum J, Ducklow H, Sullivan MB: Marine viral survival skills: how oceanic microbial viruses succeed in the Southern Ocean. *NSF Site Visit*. 2011.

Brussaard CPD, Marie D, Bratbak G (2000). Flow cytometric detection of viruses. *Journal of Virological Methods* **85:** 175-182.

Brussaard CPD (2004). Viral control of phytoplankton populations--a review. *The Journal of eukaryotic microbiology* **51:** 125-138.

Culley AI, Lang AS, Suttle CA (2003). High diversity of unknown picorna-like viruses in the sea. *Nature* **424:** 1054-1057.

Culley AI, Lang AS, Suttle CA (2006). Metagenomic analysis of coastal RNA virus communities. *Science* **312:** 1795-1798.

Culley AI, Steward GF (2007). New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Applied and Environmental Microbiology* **73:** 5937-5944.

Culley AI, Mueller JA, Belcaid M, Wood-Charlson EM, Poisson G, Steward GF (2014). The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *MBio* **5**.

Ducklow HW, Baker K, Martinson DG, Quetin LB, Ross RM, Smith RC *et al* (2007). Marine pelagic ecosystems: the West Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362:** 67-94.

Garibotti I, Vernet M, Ferrario M (2005). Annually recurrent phytoplanktonic assemblages during summer in the seasonal ice zone west of the Antarctic Peninsula (Southern Ocean). *Deep Sea Research Part I: Oceanographic Research Papers* **52:** 1823-1841.

Gowing MM, Garrison DL, Angela H. Gibson JMK, Jeffries MO, Fritsen CH (2004). Bacterial and viral abundance in Ross Sea summer pack ice communities. *Marine Ecology Progress Series* **279:** 3-12.

Gustavsen JA, Winget DM, Tian X, Suttle CA (2014). High temporal and spatial diversity in marine RNA viruses implies that they have an important role in mortality and structuring plankton communities. *Frontiers in microbiology* **5:** 703.

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

Kleiboeker SB (2003). Applications of Competitor RNA in Diagnostic Reverse Transcription-PCR. *Journal of Clinical Microbiology* **41:** 2055-2061.

Koressaar T, Remm M (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23:** 1289-1291.

Lang AS, Rise ML, Culley AI, Steward GF (2009). RNA viruses in the sea. *FEMS Microbiology Reviews* **33:** 295-323.

Levesque-Sergerie J-P, Duquette M, Thibault C, Delbecchi L, Bissonnette N (2007). Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR. *BMC Molecular Biology* **8:** 18.

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

May RM (1973). Time-Delay Versus Stability in Population Models with Two and Three Trophic Levels. *Ecology* **54:** 315-325.

Moline MA, Claustre H, Frazer TK, Schofield O, Vernet M (2004). Alteration of the food web along the Antarctic Peninsula in response to a regional warming trend. *Global Change Biology* **10:** 1973-1980.

Mueller JA, Culley AI, Steward GF (2014). Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters. *Applied and Environmental Microbiology* **80:** 3930-3942.

Mueller JA, Steward GF (*In preparation*). Variables influencing the efficiency and interpretation of reverse transcription, real-time quantitative PCR (RT-qPCR): an empirical study using Bacteriophage MS2.

Mueller JA, Culley AI, Schvarcz CR, Steward GF (*Submitted*). RNA viruses as major contributors to Antarctic virioplankton.

Nolan T, Hands RE, Ogunkolade W, Bustin SA (2006). SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Analytical Biochemistry* **351:** 308-310.

O'Connell KP, Bucher JR, Anderson PE, Cao CJ, Khan AS, Gostomski MV *et al* (2006). Real-time fluorogenic reverse transcription-PCR assays for detection of bacteriophage MS2. *Applied and Environmental Microbiology* **72:** 478-483.

Pearce DA, Wilson WH (2003). Viruses in Antarctic ecosystems. *Antarctic Science* **15:** 319-331.

Pearce I, Davidson AT, Bell EM, Wright S (2007). Seasonal changes in the concentration and metabolic activity of bacteria and viruses at an Antarctic coastal site. *Aquatic Microbial Ecology* **47:** 11-23.

Roossinck MJ (2011). The good viruses: viral mutualistic symbioses. *Nature Reviews Microbiology* **9:** 99-108.

Short SM (2012). The ecology of viruses that infect eukaryotic algae. *Environmental Microbiology* **14:** 2253-2271.

Sime-Ngando T (2014). Environmental bacteriophages: viruses of microbes in aquatic ecosystems. *Frontiers in microbiology* **5:** 355.

Smith DC, Steward GF, Azam F (1992). Virus and bacteria abundances in the Drake Passage during January and August 1991. *Antarctic Journal of the US* **27:** 125-127.

Steward GF, Montiel JL, Azam F (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnology and Oceanography* **45:** 1697-1706.

Steward GF, Culley AI (2010). Extraction and purification of nucleic acids from viruses. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of aquatic virology*. American Society of Limnology and Oceanography: Waco, TX. pp 154-165.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G (2013a). Are we missing half of the viruses in the ocean? *Isme Journal* **7:** 672-679.

Steward GF, Culley AI, Wood-Charlson EM (2013b). Marine Viruses. In: Levin SA (ed). *Encyclopedia of Biodiversity*, Second edn. Elsevier: Waltham. pp 127-144.

Tomaru Y, Nagasaki K (2007). Flow Cytometric Detection and Enumeration of DNA and RNA Viruses Infecting Marine Eukaryotic Microalgae. *Journal of Oceanography* **63:** 215-221.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M *et al* (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research* **40:** e115.

Chapter 7

7.1 Conclusions and future directions

The modern era of marine viral ecology is now in its third decade, but our understanding of the importance of RNA viruses is still in its infancy. My aim with this dissertation was to provide the first description of RNA viruses in the Antarctic, investigate their relative abundances, and explore their temporal and spatial dynamics. To investigate the genetic diversity of RNA viruses, I analyzed metagenomes prepared from operationally defined viral RNA. By analyzing multiple samples collected during the course of a phytoplankton bloom at Palmer Station (Chapter 2) and along a transect in the Larsen A Ice Shelf region (Chapter 3) I was able to demonstrate that RNA viruses make up a large fraction of the virioplankton and that the composition of RNA virus assemblages varies over time. I next wanted to quantitatively track the abundance of some of the most abundant phylotypes of viruses during the course of the bloom at Palmer Station. To accomplish this I first developed and evaluated of two molecular methods: nucleic acid extraction from microbes (Chapter 4) and the quantification of RNA viruses through reverse transcription, real-time quantitative PCR (RT-qPCR; Chapter 5). I then applied these optimized methods to track RNA virus abundances and provide the first estimates of RNA virus phylotype abundances in the ocean (Chapter 6). The analysis of these RNA virus communities offers insight to the potential role they play in annual blooms of eukaryotic phytoplankton that sustain an entire food web of the Southern Ocean. My findings not only illustrated the prevalence and ecological significance of RNA viruses in polar waters, but also laid technical groundwork for future studies of the importance of RNA virus communities throughout the entire oceanic ecosystem.

Analyses of marine viruses in coastal tropical and now polar seawater suggest that eukaryote-infecting RNA viruses could be a major component of the virioplankton, and contribute to over half of the free-living virus community (Steward et al 2013, Chapter 3). These numbers were obtained from two very different locations using different viral collection techniques: tangential flow filtration vs. iron chloride flocculation. The fact that the data agree suggests that this could be generally true in the ocean, but the data are still limited. Regions dominated by prokaryotic

plankton, for example the oligotrophic open ocean and deep sea, may not support such high relative abundances of these predominantly eukaryotic-infecting viruses. More studies are required in different regions of the ocean to address whether relatively high abundances of RNA viruses are unique to coastal habitats or are a more general phenomenon in the ocean.

An investigation of the temporal variability in specific RNA virus abundances using assembly data from metagenomes (Chapter 2) and targeting specific RNA viruses by RT-qPCR (Chapter 6) suggested that individual RNA viral phylotypes can represent a significant portion of the total virus community (46 to 90%) during bloom conditions. Since these protistan-infecting RNA viruses appear to be predominantly lytic in nature (reviewed by Lang et al 2009), the high abundance of RNA viruses implies a significant contribution of viruses to protistan mortality with a concomitant release of nutrients and carbon back into the dissolved and colloidal pools. However, relative abundances of RNA viruses during the winter months remain unknown. An intriguing open question for future research is how these viruses endure throughout the long dark winters when host communities are almost nonexistent. Tracking the abundances of specific RNA viruses also revealed distinct abundance patterns and suggested unique host ecologies. Additional work to identify the hosts of these viruses would enable the tracking of virus-host dynamics. This will provide a better understanding of the vial contributions to phytoplankton ecology. Additionally, virus production and decay experiments would allow us to understand how the turnover times of RNA viruses compare with DNA viruses. Since RNA viruses are often neglected in our current methods of virus enumeration through flow cytometry and epifluorescence microscopy, our rates of viral turnover times and production likely need to be revisited to incorporate these viruses into biogeochemical cycling models.

The metagenomic and untargeted approach of analyzing the RNA virus communities provides us with the ability to discover novel genomes, and explore these diverse uncultivated virus communities. Studies investigating RNA viruses from marine environments in coastal temperate (Culley et al 2006), tropical (Culley et al 2014, Steward et al 2013), and now polar (Chapter 2 and 3) waters suggest that +ssRNA viruses within the order *Picornavirales* dominate the RNA virus communities, with dsRNA viruses present in very low abundances. Based on the RdRp phylogeny of the Antarctic RNA virus communities, they appear to be dominated by viruses most similar to those that infect diatoms, yet diverse viruses completely unrelated to any other

known viruses within the order *Picornavirales* were also found (Chapter 2 and 3). The reason for this strong bias towards picornavirads in these marine metagenomes remains unknown, but sampling methods and the limited numbers of RNA viral metagenomic surveys have been suggested as potential causes for a biased representation of the viral diversity in the environment (Culley et al 2014, Kristensen et al 2010). As a result of the low concentrations of RNA typically found in the viral fractions of environmental samples, a sequence-independent single primer amplification (SISPA) step is required prior to sequencing. Not only are there potential biases related to the amplification step (Rosseel et al 2013), but the reverse transcription reaction in which RNA is converted to cDNA is highly susceptible to variability (Ståhlberg et al 2004). The apparent dominance of picornavirads persisted in these three environments even with differences in viral harvesting techniques (tangential flow filtration vs. iron chloride flocculation) and sequencing methods (Sanger vs. 454 vs. Illumina). However, a thorough assessment of the RNA virus metagenomic library preparation is needed to ensure that the results do not result from a bias in the amplification methods. Even if dsRNA viruses are present at relatively low concentrations, they may still be making significant ecological contributions. Methods have been developed to specifically isolate double-stranded RNA (dsRNA) viruses from environmental samples (Decker and Parker 2014) and it would be interesting to further investigate their diversity.

The results from my investigations of methods for nucleic acid extraction and quantification represent significant contributions to microbial ecology. The improved efficiency of extracting nucleic acids from total microbial plankton collected on anodic aluminum oxide filters allows for a more thorough investigation of the entire microbial community, including viruses. Many studies require quantification of specific RNAs, whether to determine the abundance of RNA viruses or investigate gene expression by cells. With these methods, I was able to track specific virus phylotypes throughout a season, but the same methods could be adapted to also track the hosts of these viruses and to investigate changes in host gene expression in response to infection. Together, these methods can be applied in future studies to better understand steady-state and bloom dynamics and to investigate functional genomics within marine and other microbial communities.

7.2 REFERENCES

Culley AI, Lang AS, Suttle CA (2006). Metagenomic analysis of coastal RNA virus communities. *Science* **312:** 1795-1798.

Culley AI, Mueller JA, Belcaid M, Wood-Charlson EM, Poisson G, Steward GF (2014). The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *MBio* **5**.

Decker CJ, Parker R (2014). Analysis of double-stranded RNA from microbial communities identifies double-stranded RNA virus-like elements. *Cell reports* **7:** 898-906.

Kristensen DM, Mushegian AR, Dolja VV, Koonin EV (2010). New dimensions of the virus world discovered through metagenomics. *Trends in Microbiology* **18:** 11-19.

Lang AS, Rise ML, Culley AI, Steward GF (2009). RNA viruses in the sea. *FEMS Microbiology Reviews* **33:** 295-323.

Rosseel T, Van Borm S, Vandenbussche F, Hoffmann B, van den Berg T, Beer M *et al* (2013). The origin of biased sequence depth in sequence-independent nucleic acid amplification and optimization for efficient massive parallel sequencing. *PLoS One* **8:** e76144.

Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M (2004). Properties of the Reverse Transcription Reaction in mRNA Quantification. *Clinical Chemistry* **50:** 509-515.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G (2013). Are we missing half of the viruses in the ocean? *Isme Journal* **7:** 672-679.