Drivers of viral diversity and community compositional change over spatial and temporal scales in coastal British Columbia

by

Julia Anne Gustavsen

B.A., English, University of New Brunswick, 2005
B.Sc. Hons, Biology, University of New Brunswick, 2005

A thesis submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

in

The Faculty of Graduate and Postdoctoral Studies
(Oceanography)

The University of British Columbia
(Vancouver)

April 2016

Abstract

Marine viruses are ubiquitous, abundant, and genetically diverse in natural waters. They play key roles in nutrient and carbon cycles. The composition of marine viral communities changes seasonally and repeats annually, and such patterns can be driven by their hosts in response to environmental changes. Moreover, environmental parameters can also directly affect the viral community through the decay of viruses, and differences in viral infectivity under different conditions. Marine viral communities show changes over time and space, but the mechanisms that drive compositional changes and maintain high diversity are largely unexplored. Determining factors affecting viral community composition and structure is essential to explain how viral diversity is maintained. This dissertation will assess the diversity of marine viral communities, and the role of the environment and putative viral hosts in driving this diversity.

The relationship between environmental parameters and the diversity of viruses and their putative hosts was explored in coastal seawater samples along a transect and over a 13-month time series at a nearshore location. I used PCR amplification to target ecologically-important double-stranded DNA (T4-like myoviruses) and single-stranded RNA (picorna-like) viruses, as well as their putative bacterial (16S rRNA gene) and eukaryotic (18S rRNA gene) hosts were examined. These were interpreted in the context of nutrients, salinity, and temperature.

I observed patchiness in the distribution and diversity of viral communities across space and time (Chapter 2). Chapter 2 greatly increased the known genetic diversity of marine picorna-like viruses with 145 operational taxonomic units (OTUs) occurring within previously seen phylogenetic clades. In Chapter 3 there were temporal shifts in dominance of phylogenetically-related viruses and most viral OTUs were ephemeral. In Chapter 4, I demonstrated that nutrients, salinity, and temperature drive the co-occurrence

of viruses and their putative hosts. Finally, in Chapter 5, I revealed that specific viral and protistan taxa were associated with controlling species composition and the demise of a phytoplankton bloom.

Altogether, this dissertation advances the understanding of the phylogenetic structure of viral communities over time, the drivers of host-virus relationships, and the dynamics of viral and microbial communities during blooms by assessing multiple groups of viruses and microbes.

Preface

One chapter from my thesis has been published elsewhere:

Chapter 2 has been previously published as:

> Gustavsen, Julia Anne, Danielle M Winget, Xi Tian, and Curtis A Suttle. "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, no. 703 (2014). doi:10.3389/fmicb.2014.00703.

I was the lead investigator on this project. Danielle Winget and Xi Tian, and I had many discussions related to the analysis of these data. Xi Tian wrote analytical scripts that were used to analyse the data. I performed all of the lab work, phylogeny and generated all of the figures. Curtis Suttle and I conceived the experiments and we wrote the manuscript for the published paper with the input of the other co-authors.

All other chapters represent original, unpublished, independent work by the author, J. A. Gustavsen, with the supervision of Curtis Suttle.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Adonis Analysis of variance using distance matrices.

ARISA Automated Ribosomal Intergenic Spacer Analysis.

AVS Algal virus specific, *Glossary:* AVS.

BC British Colombia.

BLAST Basic Local Alignment Search Tool.

chl *a* Chlorophyll *a*.

CRT Conditionally rare taxa.

DGGE Denaturing gradient gel electrophoresis.

e-value Expect value.

gp23 Gene product 23, *Glossary:* gp23.

JP Jericho Pier.

LSA Local similarity analysis.

NGS Next Generation Sequencing, *Glossary:* NGS.

NMDS Non-metric multidimensional scaling, *Glossary:* NMDS.

OTU Operational taxonomic unit.

PCR Polymerase chain reaction.

PFGE Pulse field gel electrophoresis.

qiime Quantitative Insights Into Microbial Ecology.

RAxML Randomized Axelerated Maximum Likelihood.

RdRp RNA dependent RNA polymerase, *Glossary:* RdRp.

SOG Strait of Georgia.

T-RFLP Terminal Restriction Fragment Length Polymorphism.

Glossary

α, β diversity The species diversity (or richness) of a local community or habitat (α diversity); the difference in diversity associated with differences in habitat or spatial scale (β diversity).

abiotic Having to do with the chemical, geological, and physical aspects of an entity; i.e., the nonliving components.

amplicon sequencing Sequencing of all the variants in a targeted region of a genome amplified using PCR. In this dissertation the primers used to amplify the targeted region were designed to encompass a large proportion of the communities. Thus the sequence reads are from many different species/strains amplified from the sample. Also known as: amplicon deep sequencing, ultra-deep sequencing, eDNA, pyrotags.

AVS Algal virus specific primer set. Described in Chen and Suttle (1995).

Bacillariornavirus Viral genus of ssRNA viruses infecting diatoms. In the viral order *Picornavirales*.

bank theory Within a community of viruses or cells there are a few abundant viruses, however, most viruses are rare. These rare viruses can rapidly become abundant based on the environment (or hosts or other interactions) that are available. Also known as: seed bank theory.

biogeochemistry The scientific study of the physical, chemical, geological, and biological processes and reactions that govern the cycles of matter and energy in the natural environment.

biotic Having to do with or involving living organisms.

bloom A population outbreak of microscopic algae (phytoplankton) that remains within a defined part of the water column.

bottom-up control The regulation of ecosystem structure and function by factors such as nutrient supply and primary production at the base of the food chain, as opposed to "top-down" control by consumers.

burst size Number of viruses produced per infected cell.

capsid Protein shell of viruses composed of many subunits.

Caudovirales Viral order of double-stranded DNA viruses that includes the tailed bacteriophage.

cloning Replication of DNA inside bacterial cell using a vector into which the target DNA has been placed.

coevolution A process of reciprocal evolutionary change in two interacting species, driven by natural selection.

coexistence The indefinite persistence of two or more species within the same community; this involves species that will continue to persist in the face of perturbations in their abundances. Species that co-occur may or may not be stably coexisting, because one or more of them may be on the way to local extinction at a time scale that is too slow to be immediately apparent.

community An assemblage of species found together in a specific habitat at a certain time.

community structure Determined by species composition and relative abundance.

degree In graph theory, the degree of a node is how many edges are connecting this node to other nodes.

direct effect The immediate impact of one species on another's chance of survival and reproduction, through a physical interaction such as predation or interference.

disturbance An episodic event that results in a sustained disruption of an ecosystem's structure and function. This may be a physical disturbance, a biological disturbance, or an anthropogenic disturbance.

dynamics The changes through time in the size of a population, or in a related measure such as density.

edge In graph theory, refers to the connection between two nodes in a network. The strength of the connection can be displayed. Also called: link.

endemic A species that has a relatively narrow geographic range, such as one that is found only in a particular body of water or in a particular habitat or region.

ephemeral Short time period, also: transient.

evenness Similarity in number or in proportion of the species in a community.

fingerprinting Molecular biology techniques used to quickly survey the genetic diversity of a sample.

gp23 Gene product 23. Codes for the major capsid protein in some families of bacteriophage.

heterotroph An organism that must consume organic compounds as food for growth.

homologues Genes with shared ancestry.

Illumina Company that makes High-throughput sequencers such as the Miseq, Hiseq and NextSeq.

Killing the Winner Frequency-dependent selection on host types as mediated by viruses. "Winners" are hosts that have achieved sufficient numbers that they are capable of supporting virus population growth to densities that can drastically reduce in number these same virus-susceptible hosts.

Labyrnavirus Viral genus of ssRNA viruses infecting thraustrochytrids. In the viral order *Picornavirales*.

library Collection of genetic material prepared for sequencing.

Lotka Volterra Model composed predator–prey dynamics. Model composed of two equations that describe the dynamics of biological systems in which two species interact, one as a predator and the other as prey. The model describes oscillations in the population size of both predator and prey, with the peak of the predator's oscillation lagging slightly behind the peak of the prey's oscillation.

Marnaviridae Viral family of ssRNA viruses infecting Raphidophyte alga. In the viral order *Picornavirales*.

mesocosm Experimental water enclosures. outdoor experiment of the natural environment that is controlled. Examples: 60L bags can be used for aquatic experiments, can be floated in-situ, but it is a closed system.

microbe Another term for a microorganism.

microeukaryotes Microscopic eukaryote.

mineralization The microbially mediated conversion of organically bound nutrients such as nitrogen and phosphorus to soluble inorganic forms that can be taken up by plants.

Miseq A type of Illumina sequencing platform.

monophyletic Describing a group of species that are more closely related to each other than any of them are to other species outside the group. Thus, monophyly.

Myoviridae Viruses infecting bacteria and archaea that are members of the viral order *Caudovirales*.

NGS Next Generation Sequencing. High-throughput sequencing. Generally refers to data from Illumina and Roche 454 sequencers.

niche The specific role and requirements of a particular population or species within a larger community.

NMDS Non-metric multidimensional scaling. An ordination used to visualize the variation from multiple variables in 2-3 dimensions.

node In a network, refers to an object that has been compared to the other objects. In this dissertation a node is an OTU or an environmental parameter. Also called: vertex.

partial redundancy analysis Analysis like redundancy analysis, but can try removing the effect of one variable on the summarised variation.

persistent Enduring for a long period.

Picornavirales Viral order of positive-sense ssRNA viruses.

plankton A collective term for various drifting organisms of the pelagic zone. Phytoplankton are photosynthetic primary producers, and zooplankton are consumers.

Podoviridae Viral family of *Caudovirales* with short tailed bacteriophage.

population A group of individuals of the same species occupying a certain geographic area over a specified period of time.

productivity Rate of generation of biomass in an ecosystem.

prokaryote A single-celled organism lacking membrane-bound a nucleus, and organelles. Not the preferred term because it implies that there are only two types of organisms.

protist Unicellular eukaryote.

protozoa Unicellular eukaryotic non-photosynthesis organism.

pyrosequencing High-throughput sequencing method following "sequencing by synthesis" model. Complementary strands are formed and each time a new base is added light from the reaction is detected. The most common type of this sequencing was Roche's 454 sequencing.

quasispecies Group of viruses related by mutations. Viral "species" is often an average of all of these mutant (relative to ancestral) sequences.

Q value Minimum false discovery rate where the test is deemed significant.

rank abundance curve Plot type used to display relative species abundance. X axis is the rank of the species and Y-axis is the relative abundance of the species.

rare biosphere Where most species are rare and few are abundant. Recent discussion has been precipitated by high-throughput sequencing approaches.

rarefaction curve The statistical expectation of the number of species in a survey or collection as a function of the accumulated number of individuals or samples, based on resampling from an observed sample set.

RdRp RNA dependent RNA polymerase. Enzyme that catalyzes the replication of RNA from RNA. This enzyme is generally well-conserved in viruses, but also found in eukaryotes as part of some RNA Interference pathways. It has not been found in bacteria and archaea.

read normalizing The number of reads from different libraries must be made similar. This is done to account for uneven reads per library as a result modern sequencing techniques. Each library is rarefied (reads picked randomly without replacement) to the number of reads from sequencing library with the lowest number of reads.

reads Sequence reads from DNA sequence, usually refers to raw or quality trimmed sequences.

redundancy analysis Technique to summarise the variation in object (in this dissertation OTUs or groups of OTUs) by another set of explanatory variables.

relative abundance The quantitative pattern of rarity and commonness among species in a sample or a community.

resilience The ability of an ecosystem to recover from or resist disturbances and perturbation, so that the key components and processes of the system remain the same.

Sanger sequencing Original sequencing platform where sequence is determined by chain terminating dideoxynucleotides that are incorporated during a replication with DNA polymerase. Low throughput, but low errors and long read length.

Siphoviridae Viral family of *Caudovirales* with long tailed bacteriophage.

species area curve Relationship between the number of species found and the area. also known as: effort curve.

species richness The number of species in a community, or in a region.

top-down control Regulation of ecosystem structure and function by consumers rather than factors such as nutrient supply and primary production at the base of the food chain.

trophic level The position of a given species in the chain of energy or nutrients.

Acknowledgements

It truly "takes a village" to complete a PhD and while I cannot thank everyone who has helped me, I would like to take a bit of space to try to thank a few of the people who have helped me along the way.

I would like to acknowledge the support and trust of my supervisor, Curtis Suttle, throughout this journey.

My supervisory committee, Steven Hallam, François Jean and Maite Maldonado, have challenged my thoughts, experiments and analysis during my PhD. I am grateful for their high expectations and their help.

I would like to thank all Suttle lab members and visitors past and present for their help, constructive criticism, and laughs. I would especially like to thank Amy Chan, Renat Adelshin, Anwar Al-Qattan, Christina Charlesworth, Caroline Chénard, Anna Cho, Cheryl Chow, Jessie Clasen, Jenn Cook, Chris Deeg, Jan Finke, Matthias Fisher, Jingze Jiang, Jessica Labonté, John Li, Jérôme Payet, Emma Shelford, Alvin Tian, Rick White III, Danielle Winget, Christian Winter, Jennifer Wirth, Marie-Claire Veuilleux, Marli Vlok, and Kevin Zhong. Thanks to Chris Payne for processing nutrient samples, giving guidance on cruise-prep and general ocean-going expertise, and thanks to Rich Pawlowicz for use of his lab's YSI.

Many people have helped me with editing, writing and presentations. Some stand-outs include: Caroline Chénard, Cheryl Chow, Jessie Clasen, Chris Deeg, Jan Finke, Brenda Gustavsen, Thierry Heger, Elisabeth Hehenberger, Bernhard Konrad, Emma Shelford, Xi (Alvin) Tian, Marli Vlok, Xu (Kevin) Zhong.

I would also like to thank Software Carpentry for providing the best timed-workshop of my doctoral degree. Thanks to Greg Wilson, for being a good mentor, and to Jenny Bryan for enriching opportunities.

I have received generous funding while at UBC which made it possible to complete my studies. I have been supported by a National Sciences and Engineering Research Council PGS-M, a PGS-D, a UBC Four Year Fellowship, a UBC Earth, Ocean and Atmospheric Sciences scholarship top-up, and a Beaty Biodiversity Biodiversity Research Integrative Training and Education travel award. The lab has been supported by Canadian Institute for Advanced Research, NSERC discovery grant and CFI.

Thanks to all the friends who had time for coffee, beer, dinners, hikes, bike touring, camping, backpacking trips, skiing, swimming, and yoga.

I would also like to thank my parents and stepmom for their continued support and encouragement as I pursued this degree four time zones and 4000 kilometers away from them. Finally, I thank Thierry Heger for his overall support, ideas, editing, much love and kindness, and zest for adventures that helped me to keep going throughout this degree.

CHAPTER 1

Introduction

1.1 SYNOPSIS

1.1.1 *Overall importance of marine viruses*

Viruses are obligate, intracellular parasites. They range in diameter from 20 nm to 750 nm (Fuhrman, 1999; Arslan *et al.*, 2011), and have genome sizes from 2 kb (*Circoviridae*) (Gorbalenya *et al.*, 2006) to 2.5 Mbp ("Pandoravirus") (Philippe *et al.*, 2013). They contain either DNA or RNA as genetic material which is used to hijack host replication for the production of new viruses. In the ocean, viruses have high abundances (Bergh *et al.*, 1989; Proctor and Fuhrman, 1990; Suttle *et al.*, 1990), a cosmopolitan distribution (Angly *et al.*, 2006; Liang *et al.*, 2014),and play an important role through their effect on biogeochemical cycles (Fuhrman, 1999; Wilhelm and Suttle, 1999). Aquatic viruses harbour some of the highest genetic diversity on earth (Suttle, 2007; Brum and Sullivan, 2015). Molecular surveys of viral marker genes (Short and Suttle, 2002; Payet and Suttle, 2014) and metagenomic surveys of viral genomes (Breitbart *et al.*, 2002; Culley *et al.*, 2006; Winget and Wommack, 2008) have enabled glimpses of this high diversity and dynamics under changing conditions.

1.1.2 *Temporal dynamics of aquatic viruses and microbes*

Viruses fluctuate over time in abundance and community composition (Hewson *et al.*, 2006b; Needham *et al.*, 2013). Viruses show annually and seasonally repeatable patterns in abundance (Parsons *et al.*, 2012). Compositionally, marine bacteriophage communities are more similar across seasons than years (Chow and Fuhrman, 2012; Marston

and Sallee, 2013) and are more similar among connected waters than among isolated bodies of water (Marston and Sallee, 2013). Temporal dynamics have been examined extensively in bacterial communities using the small subunit (SSU) of the ribosomal RNA (rRNA) gene 16S (Horner-Devine and Bohannan, 2006; reviewed in Nemergut *et al.*, 2013) and to a lesser extent in eukaryotic communities using the 18S rRNA gene (Massana *et al.*, 2015). Despite these studies, very little is known about the dynamics of viral community composition and structure across multiple viral families and the phylogenetic-relatedness of viral communities during stochastic or repeatable events in the environment.

To gain more insights into viral community structure and dynamics, I surveyed changes in viral and putative host communities using a year-long study at a coastal site. Changes over time were examined every two weeks using high-throughput sequencing of well-established marker genes of ecologically important viral families, and with domain-specific primers of bacteria and eukaryotes. These datasets will illuminate the important role of viruses over time as modulators of the environment. More specifically, the viral phylogenies will elucidate the identity and ecology of these viruses, and how viral phylogeny could be related to host phylogeny. The community structure demonstrates how communities of viruses and other organisms are regulated.

Marker genes come with certain caveats, and this will be discussed later in section 1.11, but targeting specific communities, is invaluable as the community can be assessed more deeply than with metagenomics (random sequencing of all genetic viral material). Initial feasibility of this approach was done using two timepoints and three spatially proximate samples (Chapter 2). The overall approach of the dissertation is unique in that it combines a holistic view of microbial and viral communities, plus it uses high-throughput sequencing to gain insights about these communities.

1.2 MICROBIAL INFLUENCE ON OVERALL FOOD WEB

1.2.1 *What is the microbial loop?*

Microbes are responsible for recycling large amounts of the material in aquatic ecosystem (Azam *et al.*, 1983; Worden *et al.*, 2015) (Figure 1.1). Microbes cycle the dissolved organic material (DOM) up to higher trophic levels by taking up DOM. When organisms die, material is released and the DOM is recycled when it is taken up by microbes.

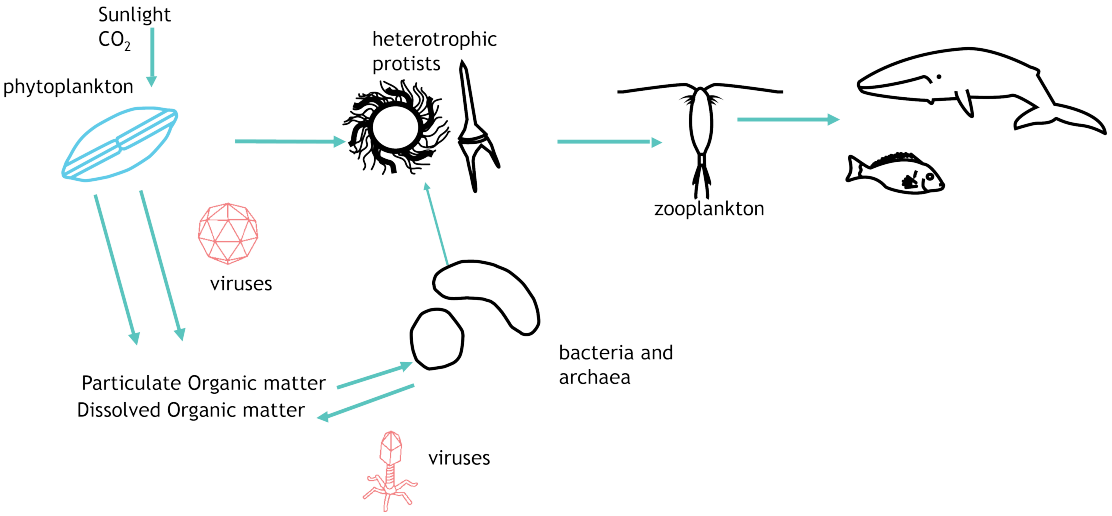

FIGURE 1.1: Marine food web from Azam *et al.* (1983), Wilhelm and Suttle (1999) and Worden *et al.* (2015).

1.2.2 *Viral shunt*

The classical food web, which ignored microorganisms and viruses, was changed with the addition of viruses because the carbon and energy that would be transferred to higher trophic levels is "shunted" to lower trophic levels as viruses lyse organisms which will subsequently release their cellular contents (Wilhelm and Suttle, 1999; Weitz and Wilhelm, 2012). This material can then be taken up by the microbial loop and transferred up the food web.

1.3 WHY STUDY TEMPORAL DYNAMICS?

Determining how organisms change over time has long been an important ecological question. The consideration of time has shaped many ecological theories such as Darwin's ideas about the origin of species and processes involved in Island Biogeography (MacArthur and Wilson, 1967). Community dynamics help explain more about the ecology of organisms and perhaps even what drives their diversity. Furthermore, some dynamics are only visible over long timescales. For example, bacterial communities can be predictable at a monthly scale, but not at a daily scale (Fuhrman *et al.*, 2015). Much can be learned about the ecology of organisms by examining their dynamics over time, and how they change together or with the environment or both (Levin, 1992; Chesson and Huntly, 1997).

1.3.1 *Models describing viral and microbial dynamics*

Although it is intuitive that viruses influence host populations through cell lysis, most marine viruses' hosts are unknown. Viral influence on the host communities can be very substantial, as viruses play a large role in killing hosts. As well, they exert selective forces on the host community composition. Viruses may increase the diversity of bacterial communities (Bouvier and del Giorgio, 2007; Middelboe *et al.*, 2009). Another potential mechanism is reduction in number or mutations in bacterial receptors, thus minimizing virus-receptor contact rates (Lenski, 1988).

Many predator-prey or virus-host relationships have been described using Lokta-Volterra models (Holt and Pickering, 1985). These models describe how the abundance of a predator lags behind the rise in abundance of the "prey" or host. The Lokta-Volterra model inspired many theories governing dynamics between hosts and parasites, including one popular in viral ecology called "Killing the Winner" (Thingstad, 2000; Winter *et al.*, 2010; Thingstad *et al.*, 2014). The "Killing the Winner" hypothesis states that for prokaryotes, in addition to constant protozoan grazing, viruses will kill the most quickly growing prokaryote (via increased encounter rate) therefore promoting host community

succession and diversity (Thingstad, 2000; Winter *et al.*, 2010; Storesund *et al.*, 2015). Thus, different viruses predominate at different times and the host community turnover promotes diversity within viruses. However, strong evidence from laboratory experiments and field studies is sparse since some studies show these dynamics (e.g. Hewson *et al.*, 2003; Schwalbach *et al.*, 2004; Rodriguez-Brito *et al.*, 2010) while others do not (e.g. Hewson *et al.*, 2006b). This theory can also be adapted to examine how the strain-level specificity is operating to maintain the diversity of the system (Thingstad *et al.*, 2015).

Another development from the Lokta-Volterra theories includes the "Red Queen" hypothesis. This hypothesis states that a host is always doing as much as it can to escape the predatory pressure of a parasite and to remain in its current ecological space (Van Valen, 1973; Benton, 2009). With the discovery of the viruses infecting SAR11 (Zhao *et al.*, 2013), a bacterial clade containing some of the most dominant marine bacteria, there have been new theories proposed to explain SAR11's continued persistence in spite of the presence of viruses. One such idea is the "King of the Mountain" theory named in Giovannoni *et al.* (2013)'s reply to Våge *et al.* (2013). This theory examines differences among hosts and the "King of the Mountain" (KoM) is the superior resource competitor in a system. Thus, with high abundances, the population avoids being decimated by viruses. This forms a positive feedback loop. Consequently, these "Kings" contribute more to geochemical cycles than the lesser competitors.

1.4 DIVERSITY OVER TIME

1.4.1 *Measurements of richness*

There are several ways to describe how viruses and microbes are distributed across environments. For example, the count of different species or operational taxonomic units (OTUs -used with microbes to approximate a division of an organism based on sequence identity) in a sample is the richness, which can be used to compare communities from different environments and sampling dates. Moreover, diversity is calculated from both

the richness, and the relative abundance of species (evenness) in a sample using methods such as the Shannon diversity index (Shannon, 1948), and the Simpson index (Simpson, 1949). For most organisms the richness and diversity of communities may be influenced by factors such as latitude, productivity, and habitat (Gaston, 2000). For viruses, hosts likely influence their distribution and genetic diversity (Mizumoto *et al.*, 2006; Yang *et al.*, 2010), yet, with more than one type of virus infecting one host, and some viruses infecting multiple types of hosts (e.g. Sullivan *et al.*, 2003), the hosts might not be singularly determining viral diversity.

1.4.2 *How is richness and diversity measured in viral and microbial communities?*

One way to examine the richness of viral and microbial communities is by using molecular techniques. Before the advent of high-throughput sequencing (HTS) of PCR-amplified marker genes, researchers used molecular fingerprinting methods such as denaturing gel gradient electrophoresis (DGGE)(e.g. Frederickson *et al.*, 2003; Payet and Suttle, 2014), and terminal restriction fragment length polymorphism (T-RFLP) (e.g. Pagarete *et al.*, 2013; Chow *et al.*, 2014) to quantify the diversity of viral genotypes. To examine the diversity of the overall communities based on viral genomes, pulsed-field gel electrophoresis (PFGE) (e.g. Steward *et al.*, 2000) was used. PFGE is very low resolution because there could be many overlapping sizes of viral genomes. These methods provided very useful comparisons of samples, but were limited to the most abundant members of the communities. Furthermore, the sequences of the community members were unknown unless additional steps were performed such as cutting out bands from the gel, cloning, and Sanger sequencing. These techniques have been used similarly in bacterial (Crump *et al.*, 2004) and eukaryotic communities (Diez *et al.*, 2001). Much of this work has moved to using high-throughput sequencing (a.k.a. amplicon sequencing) of the PCR-amplified marker genes.

1.4.3 *Temporal dynamics of viruses and putative hosts at Jericho Pier*

Looking for temporal shifts in viral communities, Short and Suttle (2003) examined a coastal site, Jericho Pier (JP), BC, at one-week intervals for 14 months. Using marker genes for algal viruses (AVS) and potential hosts (18S rDNA), they examined changes in the community composition using DGGE. The viral community was relatively stable throughout time, whereas the potential host community showed greater temporal variation. This demonstrates that environmental changes have a greater effect on microeukaryotes than on viruses, or that the viral primers were specific for a small group of viruses, and the potential host primers amplified a less specific group of eukaryotes and thus showed greater fluctuations. The shifts in viral composition were uncorrelated to the hosts, however, some shifts were associated with tide height, salinity or chlorophyll *a* concentration.

1.5 COMMUNITY STRUCTURE OF MICROBIAL COMMUNITIES

Amplicon sequencing and metagenomic efforts directed towards marine bacteria (using the 16S rRNA gene) have revealed the possibility of more bacterial species than previously recognized (Sogin *et al.*, 2006; Brown *et al.*, 2015). Amplicon sequencing revealed these communities were dominated by a small subset of genotypes, but most genotypes were rare (Pedrós-Alió, 2012). Furthermore, in a controlled freshwater aquaculture lake and in three solar salterns in Southern California, the abundant microbes and DNA viruses persisted over time, however, the rarer members of the community were not detectable in every sample based on the metagenomic surveys (Rodriguez-Brito *et al.*, 2010). The fluctuations in both microbial and viral communities showed a shuffling of dominant types, but generally no extinguishing of specific viral genotypes. This suggests that although viral communities tend to be dominated by few genotypes, they still maintain high richness. These examples of different spatial and temporal fluctuations in the composition of viral communities lead to the question of whether, for viruses,

"everything is everywhere, the environment selects," or if there are endemic viruses. Therefore, examining viral communities over time, and during different environmental conditions will illuminate how these factors influence viral diversity and community composition.

1.5.1 *Uneven structure of viral communities*

The structure of viral communities, i.e. how their genotypes are distributed in an environment, appears to be very uneven (for both DNA and RNA viral communities) (Culley *et al.*, 2006; Djikeng *et al.*, 2009; Rodriguez-Brito *et al.*, 2010). In RNA viral communities at two British Columbian sites, Jericho Pier (JP), and in the Strait of Georgia (SOG), most sequences had no homologues and were rare (Culley *et al.*, 2006). The two communities were dominated by different genotypes and there was no sequence overlap between these two communities, however, it must be noted that only 277 sequences were recovered, limiting the coverage of this study. Similarly uneven were RNA viral communities in a lake near Rockville, MD, sampled in June and November (Djikeng *et al.*, 2009). The lake was dominated by 11 genotypes and many more rare genotypes. Some viruses were abundant at both sampling dates, however, some viruses were abundant one month and then low in abundance or undetectable in the other month. These changes in community composition were not examined in the context of potential hosts or environmental parameters which could have correlated with community changes.

1.5.2 *Seed bank community structure*

The unevenness in viral communities is described by the Bank model (or seed bank model) of viral community structure (Breitbart and Rohwer, 2005). The Bank model is formed from the seed bank idea where the rank abundance curves from a community follow a log-normal type distribution (Chow and Suttle, 2015). Within the community there are a few abundant viruses, however, most viruses are rare. These rare viruses

form a "seed bank" where they can rapidly become abundant based on the environment or hosts that are available.

1.5.3 *Temporal shifts in bacterial and eukaryotic community composition*

Marine bacterial and eukaryotic communities are dynamic over time with shifts related to changes in day length (Gilbert *et al.*, 2011) and composition related to salinity (Lozupone and Knight, 2007). Viruses are dependent on their hosts and thus influenced by the availability of hosts, however, as stated earlier (p. 4) these viruses also exert a selective pressure on their host communities. Shade *et al.* (2014) found that conditionally rare taxa (CRTs), that is those that are usually rare but can quickly become abundant, can be deemed responsible for many of the temporal dynamics seen in bacterial communities. It is has not been investigated how these dynamics relate to viral communities, however, it could be that viruses in the "Bank" lyse these CRTs following "Killing the Winner" type dynamics.

1.5.4 *OTU clustering*

High-throughput sequencing methods generate large amounts of sequence data. To be able to analyse the data appropriately and to correct for sequencing error, sequences are clustered into operational taxonomic units (OTUs), meaning they are grouped together based on their similarity. There are many different types of sequence clustering algorithms (reviewed in Schmidt *et al.* (2015)), however, many have their roots in the hierarchical clustering algorithms used in community analysis such as single linkage and nearest neighbour clustering.

The general process is that from all the sequence reads the unique sequences are chosen (speeds things up computationally), then ordered by length. Next, sequences are compared, and if they are more than a set percentage similar they are grouped into an OTU. Then the next unique read is compared to this OTU and others, if the read is similar (specifically determined based on the algorithm) to an OTU already seen it is

included in this OTU. If it is less similar than the cut-off it will form a new OTU. The process continues until all of the unique reads are clustered into OTUs. For bacteria and eukaryotes the cut-off of 97% sequence similarity is used for these studies to approximate species level clusters (Stackebrandt and Goebel, 1994). The method used in this dissertation is USEARCH (Edgar, 2010)– a centroid based approach for calculating the similarity to the sequence read. The OTUs are then used for further analyses requiring the DNA sequence (e.g. phylogeny, taxonomic classification, etc.).

1.5.5 *Taxonomic classification*

OTUs can be determined to resemble a reference sequence. This is called "classification." The Ribosomal database project (RdP) created a classifier (Wang *et al.*, 2007). The classifier uses a naive (meaning independent) Bayesian approach by evaluating the frequency of sequences cut up into a specific word size (8 was used in the paper because of memory constraints, but also for specificity). Thus each OTU and reference sequence are represented by a collection of the counts of "words." This allows the quick matching of OTUs to the reference database and also gives confidence limits to the assignment at different levels of taxonomic specificity (Wang *et al.*, 2007).

OTUs are classified using a database of reference sequences that have a specific taxonomic assignment. For rRNA genes popular databases are Greengenes (DeSantis *et al.*, 2006), RdP (Cole *et al.*, 2014), Silva (Quast *et al.*, 2013) and for 18S rRNA the Protist Ribosomal Reference database (PR^2) (Guillou *et al.*, 2013). For viral sequences there is no such database since there are comparatively few sequenced viruses, and do not have shared genes among them.

1.5.6 *Phylogenetic context of OTUs*

To compare genetic distances, phylogenetic methods can be used to examine OTUs. This has the advantage of not being reliant on a reference database, however, the outputs are different and can be complementary to the taxonomic classification when both

are available. Taxonomic classification, however, is not generally based on phylogeny. Phylogeny-based approaches use the context of sequences on a phylogenetic tree as a way to give a context to OTUs. Phylogenetic trees based on large amounts of sequence reads require the alignment of reads using multiple sequence alignments. The sequences are either only used to make a tree and then the genetic distances of OTUs are examined, or they can be added to the tree along with reference sequences. As well, they can be placed onto tips of a well-characterized reference tree using a placement algorithm such as evolutionary placement of short reads (EPA)(Berger *et al.*, 2011) or pplacer (Matsen *et al.*, 2010). These placement algorithms are useful with well-defined reference trees and short sequences.

Oftentimes, if all the OTUs are used (and there can be thousands), the trees are either not visualized or are heavily collapsed. A useful metric using phylogenetic distance is Unifrac (Lozupone and Knight, 2005) which compares the phylogenetic similarity of communities, although this distance metric is not without some issues (Long *et al.*, 2014; Lozupone and Knight, 2015).

1.6 RELATEDNESS OVER TIME

1.6.1 *Phylogenetic relatedness over time*

Bacterial communities are more phylogenetically diverse than expected over time (Horner-Devine and Bohannan, 2006). Little is known, however, about the phylogenetic diversity of viral communities neither in general and over time, nor how it is maintained. Phylogenetic diversity is usually related to species richness but little is known about why it can be decoupled, only that it can happen under high levels of regional richness (Tucker and Cadotte, 2013). Phylogenetic relatedness can be correlated to ecological relatedness in other organisms (Harvey and Purvis, 1991; Srivastava *et al.*, 2012).

1.7 HOST-VIRUS INTERACTIONS OVER TIME

1.7.1 *Co-occurrence networks*

Organisms can co-occur over time, and the more often this happens, the more likely there is a type of relationship (e.g. a symbiosis, or the same preferred niche). When species do not occur together they may have an antagonistic relationship, either through allelopathy, predation, or different preferred niches. All of these co-occurrences can be examined as a network. A network is a diagram of relationships, whereby the properties of all the pair-wise associations are examined together and the emergent properties of the overall interactions in the communities can be analysed.

For microbial associations, three steps are required to generate networks. First, a matrix is generated by pairwise comparison of all organisms or those that are most abundant. Many different kinds of species abundance and molecular data can be compared with co-occurrence/correlation analysis such as presence-absence, species counts, relative abundance, high-throughput sequencing, DNA fingerprinting patterns, and microarray data. These pairwise association matrices can be generated using a variety of different methods including Spearman or Pearson correlation coefficient, any distance metric (e.g. Bray-Curtis, Euclidean, etc.), linear regressions (Faust and Raes, 2012), local similarity analysis (LSA) (Ruan *et al.*, 2006; Xia *et al.*, 2011) which can include time-lagged relationships or others. LSA can be used with time series data to look for the strongest correlation among timepoints with a set maximum amount of lag.

Second, the pairwise associations are filtered for strength of association or significance or both. The significance can be tested by examining the p-value (if available, often calculated in pairwise associations by permutation, i.e. rearranging the matrix many times and testing the results), and also by using the Q value which tests for false discoveries when the p-value is deemed significant (Storey *et al.*, 2005).

Third, the significant associations are assembled together into a network which can be analysed for its emergent properties and also can be visualized. These networks, also called graphs (a collection of pairwise relationships between objects), consist of

two parts: nodes and edges. Nodes are the objects or variables used in the pairwise comparison. When nodes are visualized they are drawn as points (can also be called vertices). Edges are the associations between nodes and are often drawn as lines (can also be called links, connections, correlations). Edges can represent positive or negative, time-lagged, strong or weak associations. Properties of the individual nodes that can be determined are the degree of each node, which is number of nodes it is connected to by its edges.

The properties of the overall network can be considered, which can be useful for comparing networks (Cram *et al.*, 2013). The overall degree distribution of the nodes, (e.g. how many have degree n) can be useful for classifying networks. For example, the betweenness, which is the number of shortest paths going through a node, can be calculated for one node, and also as an average over the whole network. The network density (edges per node), the network diameter, and the clustering coefficient (number of groups of three nodes) can be examined. Triangles or triplets, which are three nodes and their edges can be used as a way to examine specific groups of interactions. Finally, modules, which are structures within the graph that are highly connected (often biological networks show a high degree of modularity), can be detected.

Networks can be compared to each other using some of the aforementioned properties, although it is also often useful to compare these networks to randomly generated networks (as a way to classify networks and to test for difference from randomly generated network). Random networks can be generated based on some algorithms following network properties such as the scale-free (where degree distribution follows the power law, where many vertices have a degree greater than average) and small-world. Generally most microbial networks are classified as small-world networks where most nodes are not neighbours of one another, but most nodes can be reached from every other node through a small number of other nodes. Network visualization has been very popular in recent years, however, care must be taken when interpreting biological meaning from the layout of a network. Networks can often resemble "hairballs." They also can be

visualized using force-directed layouts which attempt to lay out the nodes and edges so that the edges are all about the same length and all nodes can be seen.

At the San Pedro Ocean Time series researchers used network analysis of viruses, bacteria and protists to look for associations in their monthly coastal time series (Steele *et al.*, 2011; Chow *et al.*, 2014). They found some expected associations between organisms such as associations between cyanobacteria and cyanophage (Chow *et al.*, 2014) and some associations in the cyanobacterial communities suggestive of functional redundancy. The networks of viruses had different properties as opposed to the bacterial and protistan communities. These studies examined T4-like myoviruses with bacteria and protists, however, these types of studies have not yet been done with other groups or viruses.

1.7.2 *Co-occurrence and co-evolution of viruses and microbes over time*

In a chemostat experiment Marston *et al.* (2012) examined the temporal dynamics of a *Synechococcus* sp. strain and one phage infecting it, and observed rapid diversification of both hosts and viruses over time (167 days). In replicates they saw slightly different amounts of diversification, but all hosts and viruses showed cycles of co-evolution with rapid changes and extinction of some genotypes. There was also the diversification of phenotypic traits such as increased infectivity and resistance. With evidence of such rapid co-evolution of viruses and hosts over time, it is important to examine co-occurrence and change of genotypes over time. Examining these co-occurrences in a natural setting will determine if there are any other factors that drive the diversification of these viruses and of their hosts. Furthermore, there was evidence for viral control of bacteria diversity in Sandaa *et al.* (2009)'s mesocosm experiment. They saw that the number of virus-hosts pairs was similar between mesocosms, but viral composition differed between mesocosms. This observation gives credence to the idea proposed alongside the "Killing the Winner" theory that the number of niches are limited for virus-hosts pairs and a different mechanism controls the composition of the virus-host pairs (Thingstad, 2000).

1.8 INFLUENCE OF ENVIRONMENT ON VIRAL COMMUNITIES

The environment is an important driver of bacterial and eukaryotic communities over time. With seasonality it is easily imaginable that these communities are driven by environmental factors. As well there may also be large biotic factors that drive the community composition. There are many environmental factors that influence the infectivity of viruses such as temperature, salinity, and nutrients (as reviewed in Mojica and Brussaard (2014)). This occurs even though viruses are often physically able to withstand much greater ranges of environmental parameters than their hosts. For example, there are temperatures where certain viruses are better able to infect hosts compared to other viruses (reviewed in Mojica and Brussaard, 2014; Kendrick *et al.*, 2014). Thus, environmental parameters may be driving part of the diversity of the viral community by affecting the infectivity of the viruses.

Viruses influence their host communities by forcing them into a co-evolutionary arms race through selective pressure. Co-evolution, however, is not the only pressure on viral diversity and community composition. Changes in community composition have been associated with, but not limited to environmental parameters such as, salinity, temperature, chlorophyll *a*, the mineral jarosite, carbon, and nutrient fluxes (Short and Suttle, 2003; Kyle *et al.*, 2008; Sandaa *et al.*, 2009). Although some aquatic viruses may be endemic to specific environments (Williamson *et al.*, 2008), some genotypes appear to be cosmopolitan with the most abundant types being widely distributed (Short and Suttle, 2005; Angly *et al.*, 2006). Combining data on compositional changes in specific groups of viral families, potential host communities, and in abiotic parameters will enable the determination of overall and specific drivers of genotypic change in viral communities.

1.9 VIRAL AND MICROBIAL DYNAMICS DURING ECOLOGICAL DISTURBANCES

1.9.1 *What are ecological disturbances?*

Disturbances are ecological perturbations that occur on scales of either days-weeks, pulse disturbances, and those that occur over weeks-months, press disturbances. Disturbances of whole communities and specifically the microbes have been important to study from a community ecology standpoint since the changes in one population can influence other populations and the overall community (Shade *et al.*, 2012a,b; Banks *et al.*, 2013). One community disturbance in which the viral communities are often deeply implicated is during phytoplankton blooms (Bratbak *et al.*, 1996; Wilson *et al.*, 2002a). Phytoplankton blooms are events whereby the phytoplankton biomass increases very quickly and there are high numbers of cells, often from one species of phytoplankton. These blooms can last for days to months depending on the species and the system. These blooms have a large influence on the abiotic and biotic parameters.

1.9.2 *Viruses during blooms*

Viruses have often been associated with the demise of blooms in mesocosms and in natural waters (Bratbak *et al.*, 1996; Wilson *et al.*, 2002a,b). For instance, viruses infecting *Emiliania huxleyi* have been found in high abundance at the termination of blooms. The viruses associated with the bloom can be composed of one or multiple genotypes (Highfield *et al.*, 2014). It is unclear what causes the bloom to terminate with one or several viral genotypes. During a bloom of alga, *Heterosigma akashiwo*, Tarutani *et al.* (2000) saw the evolution of both viral and host strains based on infectivity patterns. Over the progression of the bloom there were different host strains present and the infectivity of the host strains decreased over the course of the bloom.

1.9.3 *Effect of disturbance on other microbial communities*

These disturbance events, specifically phytoplankton blooms, can have large effects on the bacterial communities as there are heterotrophic bacteria that can efficiently and effectively use products from the phytoplankton (such as amino acids, carbohydrates, organic acids, polysaccharides, proteins, nucleic acids, and lipids) and thus quickly increase in abundance during a phytoplankton bloom (Buchan *et al.*, 2014; Teeling *et al.*, 2012). The bacteria that tend to respond to these events are predominantly the Flavobacteria, Gammaproteobacteria and some of the Alphaproteobacteria. These bacteria tend to be good at degrading phytoplankton products (Buchan *et al.*, 2014). Interestingly, the bacterial communities tend to remain even during a bloom, with no one type dominating the community (Delmont *et al.*, 2014). Thus the evenness of the bacterial community can remain high even though the bacterial community composition can change drastically throughout the bloom.

1.10 CHOICE OF GENETIC TARGETS

1.10.1 *Use of marker genes for studying microbial communities*

Unlike cellular organisms, viruses have no universal shared genes that can be used for overall phylogeny like the well-conserved small subunit 16S ribosomal RNA (rRNA) gene for prokaryotes and the 18S rRNA gene for eukaryotes (Woese and Fox, 1977; Woese *et al.*, 1990). Nevertheless, using the polymerase chain reaction (PCR), several well-conserved viral marker genes (such as DNA pol (Chen and Suttle, 1995), major capsid protein (Filée *et al.*, 2005), and the RNA dependent RNA polymerase (Culley *et al.*, 2003)) have been used to illuminate the genetic richness of different groups of aquatic viruses. These genes have highly conserved regions, which are used to determine the presence of viral families in a sample. However, these genes also have variable regions, which determine the richness of a related group of viruses. These conserved genes are

good surrogates for whole genomes since the phylogeny of the conserved viral genes is congruent with the phylogeny of the whole viral genomes (Filée *et al.*, 2005).

1.10.2 *Metagenomics vs. marker genes*

Viral metagenomics is the characterization of the viral community by sequencing the total viral nucleic acid in a sample. Metagenomics of viruses has revealed a huge amount of information about the kinds of viruses in the ocean (Breitbart *et al.*, 2002; Angly *et al.*, 2006; Culley *et al.*, 2006; Hurwitz and Sullivan, 2013; Brum *et al.*, 2015). This approach is very useful for examining unknown viruses and for discovering unknown functions from different viruses since no previous knowledge of the communities is needed. While viral metagenomic data generally show high richness, many of the sequences have no significant homologues in Genbank. In coastal viral communities 65 % of the dsDNA (Breitbart *et al.*, 2002), and 63-81 % of RNA metagenomes had no significant matches to Genbank (Culley *et al.*, 2006). Additionally, since certain classes and families of viruses are only found infecting specific hosts and specific taxonomic host ranges (Koonin *et al.*, 2008) the viral metagenomic approach could eventually be used as a proxy for determining which hosts are present in the environment.

In recent years this metagenomic approach was useful for characterizing the most abundant viruses in communities. For communities with high richness, however, the rarer viruses may be less well characterised. The amount of sequencing has increased rapidly in the past few years coupled with decreased costs. If the present studies were to be conducted today, the metagenomic approach could be appropriate for the questions related to community diversity and structure. Metagenomics, at that time (circa 2010), provided a broad and shallow look at the viral communities compared to marker gene approaches with high-throughput sequencing. Using targeted marker gene sequencing allows a deeper look into the composition, and structure of specific communities over time. However, with continued increases in sequencing output, this argument is being revisited for these types of ecological questions (Sullivan, 2015).

1.10.3 *Description and distribution of the T4-like myoviruses (*Caudovirales*) using the marker gene gp23*

Many different groups of viruses have been found to be ecologically important and numerous in the environment (Adriaenssens and Cowan, 2014). These include viruses infecting eukaryotes such as the *Phycodnaviridae*, and *Picornavirales*, and viruses infecting bacteria including *Caudovirales*, and the *Microviridae* (*Gokushovirinae*). Two gene markers were chosen that target viruses infecting two different host domains. The first gene marker targets a group of viruses infecting bacteria (bacteriophage). There have been observations of bacteriophage in aquatic systems since the 1950's (Spencer, 1955, 1960). Many of the viruses that infect bacteria fall within the order *Caudovirales*. These are double-stranded DNA (dsDNA) viruses whose genomes range in size from 33-244kb, and have a morphology of a head (viral capsid) and tail. Within the *Caudovirales* there are 3 main groups: the *Myoviridae* with contractile tails, the *Siphoviridae* with long tails, and the *Podoviridae* with short tails.

One diverse and well-studied group of bacteriophage in the marine environment is the T4-like myoviruses. Filée *et al*. (2005) designed degenerate primers for gp23, the gene which encodes the major capsid protein in T4-like myoviruses. Filée *et al*. (2005) found a large diversity of environmental viruses that were distantly related to the cultured viruses in this group. Filée *et al*. (2005) found high genetic richness in 3 geographically separated viral samples from British Columbia, the eastern Gulf of Mexico and the western Arctic Ocean. Within this group there are many ecologically important viruses infecting cyanobacteria including those infecting the crucial marine species of cyanobacteria *Synechococcus* and *Procholorococcus*. These viruses can have a broad or specific host range (Sullivan *et al*., 2003). This group also includes some of the pelagiphage which infect a bacterial clade that includes some of the most abundant bacteria in the ocean, SAR11 (Zhao *et al*., 2013). Temporal studies have highlighted seasonal patterns in the T4-like myoviruses; furthermore some T4-like myoviruses are constant members of

the community while others have a more ephemeral nature (Chow and Fuhrman, 2012; Pagarete *et al.*, 2013; Needham *et al.*, 2013).

1.10.4 *Studying aquatic RNA viruses using the marker gene RdRp*

Another important group of viruses in the marine environment are a subset of the viral order the *Picornavirales*. This order of viruses is a group of single-stranded RNA viruses that infect eukaryotes, specifically plants, invertebrates and protists (Le Gall *et al.*, 2008; Tomaru *et al.*, 2015). They are small, icosahedral viruses with a capsid diameter of ~25nm, no overlapping reading frames, a conserved RNA-dependent RNA polymerase, all RNAs are translated into a polyprotein before processing, and have a genome size of ~9kb (Le Gall *et al.*, 2008). In the marine environment, these viruses are important pathogens of protists (Tomaru *et al.*, 2015), and include isolates from the viral family the *Marnaviridae*, and the genera *Bacillariornaviridae* and *Labyrnaviridae*. Sequences in marine viral RNA metagenomes have hit reference genomes from families within this order such as the *Tombusviridae* and the *Dicistroviridae* (Culley *et al.*, 2006, 2014). The isolated viruses are pathogens of ecologically important marine species. The type virus of the *Marnaviridae* is the *Heterosigma akashiwo* RNA virus (Tai *et al.*, 2003) which infects a widespread alga, *Heterosigma akashiwo*, which can form massive blooms in the coastal ocean. Other isolated viruses infect two ecologically important chain-forming diatoms that can be important in blooms in the coastal ocean: 1) the *Rhizosolenia setigera* RNA virus 01 infecting *Rhizoselenia setigera*, and 2) *Chaetoceros tenuissimus* RNA virus 01, and *Chaetoceros socialis* f. radians RNA virus 01 which infect different species of *Chaetocerous* sp.. The final known isolate in these marine picorna-like viruses is a virus isolated from a pennate diatom, *Asterionellopsis glacialis* RNA virus (Tomaru *et al.*, 2012).

Using degenerate primers amplifying the RNA dependent RNA polymerase (RdRp) gene from positive sense single-stranded RNA picorna-like viruses (order *Picornavirales*), Culley *et al.* (2003) described a high richness of sequences from the Strait of Georgia (SOG), British Columbia which all had low identity to homologues in Genbank.

Later, using redesigned picorna-like primers, Culley and Steward (2007) described 5 new putative genera and 24 new putative species of picorna-like RNA viruses from study sites in Hawaii and Monterrey Bay, California. These studies illustrate the widespread occurrence and high diversity of picorna-like viruses in the ocean. Furthermore, these communities of marine picornavirads were highly uneven (Culley *et al.*, 2006; Gustavsen *et al.*, 2014). To examine the dynamics of these highly diverse communities, amplicon sequencing is useful for examining the fine-scale dynamics of these viruses, their relatedness, their putative hosts, and their associations with phytoplankton blooms since they infect known bloom-formers.

1.10.5 *Examining marker genes of bacteria and eukaryotes for richness and composition*

Viruses can only replicate within hosts and thus their hosts are of great interest during any study of viral ecology. Within the past ten years there have been large developments in amplicon sequencing for examining the microbes that could be hosts for the viral communities (Sogin *et al.*, 2006; Lozupone and Knight, 2007; Logares *et al.*, 2014). Bacterial communities are assessed using conserved and variable parts of the 16S rRNA gene. The primer set developed by Baker *et al.* (2003) was used in this dissertation. It was chosen because the primer set was compatible with a 454 study that was on-going. When tested against the Silva database (Quast *et al.*, 2013) the primer set had 87% coverage of the bacteria in the database. Similarly, for eukaryotic communities the 18S ribosomal RNA gene was used to examine the dynamics of the eukaryotic communities. A primer set developed by Diez *et al.* (2001) was used to assess microeukaryote communities. The set was used because the short product length was deemed to be a useful and efficient way to take advantage of the Miseq Illumina platform (which was used for sequence data generation in Chapters 3, 4, and 5).

1.11 CAVEATS (METHODOLOGICAL CONSIDERATIONS)

1.11.1 *Water masses*

This dissertation will discuss the temporal dynamics of marine viral, bacterial and eukaryotic communities. However, it must be understood that the same water mass is not being followed through time. These projects use a Eulerian sampling scheme where the water that comes by a point is sampled over time rather than a Lagrangian sampling scheme – where water masses would be followed and sampled through time (Hewson *et al.*, 2006a; Gilbert *et al.*, 2011; Fuhrman *et al.*, 2015). Eulerian studies have shown important and repeatable dynamics of the marine waters (Gilbert *et al.*, 2011). Thus even though it is not following the same water masses there are predictable patterns found in these waters. Also, there is heterogeneity in seawater at the microscale level (Seymour *et al.*, 2006), so the 60L of sub-surface water collected and used in these projects would likely have contained many small niches (distinct habitats), that were effectively averaged during sample collection.

1.11.2 *OTU clustering*

Another consideration is that individual sequence reads need to be grouped together at a certain similarity to examine these communities. The term Operational Taxonomic Unit (OTU) is used to describe these groupings of sequences. For bacteria and eukaryotes many have used OTU sequence similarity of 97 % for putative species (Stackebrandt and Goebel, 1994), although it is arbitrary and can differ for different species of bacteria (Koeppel and Wu, 2013b). New approaches, such as oligotyping (used in Chapter 5) provide complementary approaches to OTU-based methods by analysing the sequence reads without set similarity cut-offs.

1.11.3 *Taxonomic classification*

Taxonomic classification of a sequence is only as good as the database used. Although short sequences of partial genes can be surprisingly effective at identifying organisms, they can also have lower resolution than full-length sequences of the target genes. Bacterial and eukaryotic sequences were classified using the Silva database (Quast *et al.*, 2013). These reference databases are reliant on isolated organisms and can be limited for sequences lacking cultured representatives. Furthermore, viral databases are too sparse (since there is a paucity of isolated marine viruses) to be able to classify viruses. Unknown viral sequences can be examined for patterns by their occurrence in samples and in the context of potential hosts. This dissertation instead built phylogenies to examine the viruses in a genetic context.

1.11.4 *Problems associated with PCR and sequencing*

All of the marker genes examined in this dissertation use PCR to amplify the targets. Although this is a widely-used technique (and goes by many names such as environmental sequencing, eDNA, amplicon deep sequencing, tag sequencing, pyrotags) there are some caveats associated with it. There is known bias associated with PCR amplification where the amount of template can skew the amplification. Making sure to use appropriate amounts of template can help mitigate this bias (Lee *et al.*, 2012). Another bias is in primer mismatch where different species can be preferentially amplified based on their sequence or structure (Acinas *et al.*, 2005; Sipos *et al.*, 2007) and this is especially important with the use of degenerate primers. Furthermore, sometimes with limited databases primers can be designed which might miss relevant and important species in an environment. Such is the case with the Earth Microbiome primers which missed many important bacterial species (Apprill *et al.*, 2015; Parada *et al.*, 2016). As alluded to in the previous section this can be mitigated by checking primers against databases to see what would be amplified.

In this dissertation two different kinds of sequencing were used, 454 Titanium and Illumina Miseq. Both are both high-throughput methods of sequencing that are frequently used for community analysis of microbial communities because of their high amounts of sequence reads returned. As previously discussed, high-throughput sequencing offers a resolution and sensitivity not approached by (Automated) Ribosomal Intergenic Spacer Analysis (ARISA), T-RFLP and DGGE. However, both 454 Titanium and Illumina Miseq sequencing come with their own caveats. The 454 sequencing approach, used in Chapter 2, has the problem of inaccurate calling of homopolymers resulting in inaccurate strings of one base (Quince *et al*., 2009) which tended to cause inflated levels of diversity (Kunin *et al*., 2009). To reduce this potential bias, denoising (Quince *et al*., 2009; Reeder and Knight, 2010) was used to minimize some of the errors associated with 454 sequencing. Denoising attempts to minimize the loss of sequence information by using the flowgrams from the 454 sequencer to distinguish erroneous sequences from real sequences.

In Chapters 3, 4, and 5 Miseq Illumina sequencing was used. Miseq had the lowest error rate (0.1 substitutions per 100 bases) and highest throughput of any of the "benchtop" sequencers at the time (Loman *et al*., 2012). However there are still errors associated with it (Schirmer *et al*., 2015). Ways to minimize the per-base error rate are by filtering sequences to maintain only highest quality sequences (e.g. above a quality score of 20 which represents 1% error), trimming low quality ends and sequencing targets short enough for paired-end sequence reads to overlap, thus decreasing the probabilities for errors (Bokulich *et al*., 2013).

For microbial ecology studies it must be assured that the samples are comparable by normalizing the sampling/sequencing effort. In ecology, one of the few agreed upon ideas approaching a law is that of the species-area or effort curves (Rosenzweig, 1995), e.g. the more you sample the greater richness you see. The same is often true with sequencing where the more you sequence the greater the observed richness until you reach an asymptote. Thus, because of stochasticity in the sequencing process and the

need to have samples on different sequencing runs, the data need to be normalized to be able enable comparison among samples. A common approach for normalizing reads called "rarefying", is to use the number of reads in the smallest library and randomly pick that number of reads (without replacement) from the other libraries (collection of sequence reads from one sample).

This method of examining microbial and viral communities can be considered semi-quantitative and thus able to compare increases and decreases in samples, not absolute numbers. Mock communities, where known amounts of genomic DNA are used with the same protocols, have been used to help improve and validate the use of these techniques (Schloss *et al.*, 2011). Wherever possible these techniques have been incorporated into the studies (e.g. using cloned mock communities to assess error rates and to help assess appropriate percent identity for clustering of sequence reads).

1.12 RESEARCH OBJECTIVES AND OUTLINE OF THESIS

1.12.1 *Overall goals of thesis*

The overarching goals of this dissertation were to document temporal changes in the genetic composition of natural assemblages of marine viruses, and to relate these changes to variations in the physical and biological environment, including the composition of potential host communities. Viral gene markers examined at temporal and spatial time scales will examine whether genotypes abundant in one sample remain abundant over time or if they remain, but become rarer or undetectable.

Overall approach:

- Collect samples from coastal site Jericho Pier every 2 weeks for one year. Additionally sample Jericho Pier every other day for a tidal cycle in the winter and the summer.

- Quantify richness in viral samples by amplicon sequencing of T4-like myoviruses (gp23), and marine picorna-like viruses (RdRp), and in putative host samples of bacteria (16S rRNA gene) and eukaryotes (18S rRNA gene) amplicons.
- Compare communities for compositional overlap, community richness and diversity.
- Determine the drivers of viral diversity and community composition over time from measured environmental parameters to explain how viral diversity is maintained in different viral groups.

1.12.2 *Description of chapters*

CHAPTER 2: HIGH TEMPORAL AND SPATIAL DIVERSITY IN MARINE RNA VIRUSES IMPLIES THAT THEY HAVE AN IMPORTANT ROLE IN MORTALITY AND STRUCTURING PLANKTON COMMUNITIES — Ecological questions about the distribution of marine viruses over time and space have been examined more extensively in bacteriophages, particularly those infecting cyanobacteria.

Hypotheses: If the dynamics of marine bacteriophages and marine RNA viruses are similar, some RNA viral taxa will persist temporally and spatially, while other taxa will be detected sporadically.

Approach: To test this hypothesis, two samples, taken five months apart at the same location, and three samples taken within hours of each other, but 20 km apart in the same coastal basin were examined using 454 Titanium sequencing of the viral marker gene: RNA dependent RNA polymerase (RdRp). Phylogeny and relative abundance of OTUs in the community were used to characterize the temporal and spatial differences among the samples.

CHAPTER 3: EXAMINATION OF MARINE VIRAL COMMUNITY STRUCTURE REVEALS PHYLOGENETIC SHIFTS OVER TIME — The viral gene markers were used to examine high resolution temporal dynamics, phylogenetic distribution over time, and putative changes

in response to changes in composition to the host communities and to disturbances in the environment.

Hypotheses: 1) Related viruses should share similar ecology and thus should show the same occurrence patterns over time. The structure and composition of viral communities has been described to follow seed bank type distribution where there are many rare viruses and few abundant ones. 2) There would be shuffling of individual viral OTUs in the communities, but communities would remain mostly dominated by a few viral types over time. 3) The phylogenetic structure of viral communities would be influenced by the phylogenetic structure of their putative hosts and would show temporal shifts related to the seasonal progression of their hosts.

Approach: Using a one-year time series, sampled every two weeks at Jericho Pier all of the marker genes for viruses and potential hosts were analysed phylogenetically and the community structure was examined.

Chapter 4: Network analysis of Jericho Pier microbial time series — Viruses present at the site must be infecting a host that is also present and thus it would be expected that viruses co-occur with their specific host. The environment has a role in driving the diversity of these viruses beyond just driving the diversity of hosts since environmental parameters can influence the infectivity of viruses.

Hypotheses: At least co-occurrence (with a potential time lag) of viral OTUs with a putative host and potentially others would be expected. Within networks composed of the two groups of viruses associated to their putative hosts it would be expected to observe different patterns of co-occurrence since the viral host ranges and life histories are different. It is hypothesized that that in addition to the influence of host communities, viral communities will also be driven by environmental parameters.

Approach: Using data from the year-long study at Jericho Pier to form local similarity analysis networks to determine which OTUs co-occur strongly and also to determine if there are time-lags associated with the patterns of strongest co-occurrence of the pair-

wise relationships. Use redundancy analysis and variation partitioning to determine the environmental and biotic drivers of diversity and community composition over time.

Chapter 5: Viral and heterotrophic protistan control of a phytoplankton bloom in coastal waters — **Hypotheses:** Considering the strain level differences in infectivity seen during phytoplankton blooms and in overall temporal dynamics of microbial communities it was predicted that during a bloom of eukaryotic phytoplankton there would be a progression of strain level OTUs of the main bloom former attributable to specific strain level viral pressure.

Approach: Examine how viral, bacterial and eukaryotic richness change during a bloom of one eukaryotic species which occurred during a high-resolution time series segment of the Jericho Pier sampling. Use Shannon entropy decomposition (oligotyping) to sub-divide OTUs into finer resolution to determine how the single OTUs that dominate the communities during the bloom evolve and how they compare to the rest of the time series.

1.13 SIGNIFICANCE

These studies provide a detailed genotypic examination of temporal dynamics in the composition of DNA and RNA viral communities and their putative hosts. Notably, this is the first study that uses all of these communities together while using high throughput sequencing. This dissertation will advance the field in several important areas: additional evidence for community-level theories related to environmental viruses such as the Bank theory and Killing the Winner, insight into dynamics of the phylogenetic structure of viral communities, the dynamics of ephemeral compared to persistent OTUs, the determination of which environmental parameters play an important role in the co-occurrence of viruses and hosts, and examination of the effect of disturbances on microbial communities.

CHAPTER 2

High temporal and spatial diversity in marine RNA viruses implies that they have an important role in mortality and structuring plankton communities[1]

2.1 SUMMARY

Viruses in the order *Picornavirales* infect eukaryotes, and are widely distributed in coastal waters. Amplicon deep-sequencing of the RNA dependent RNA polymerase (RdRp) revealed diverse and highly uneven communities of picorna-like viruses in the coastal waters of British Columbia (B.C.), Canada. Almost 300 000 pyrosequence reads revealed 145 operational taxonomic units (OTUs) based on 95% sequence similarity at the amino-acid level. Each sample had between 24 and 71 OTUs and there was little overlap among samples. Phylogenetic analysis revealed that some clades of OTUs were only found at one site, whereas, other groups included OTUs from all sites. Since most of these OTUs are likely from viruses that infect eukaryotic phytoplankton, and viral isolates infecting phytoplankton are strain-specific; each OTU probably arose from the lysis of a specific phytoplankton taxon. Moreover, the patchiness in OTU distribution implies continuous infection and lysis by RNA viruses of a diverse array of eukaryotic phytoplankton taxa. Hence, these viruses are likely important elements structuring the phytoplankton community, and play a significant role in nutrient cycling and energy transfer.

[1]Chapter 2 has been previously published as: Gustavsen, Julia Anne, Danielle M Winget, Xi Tian, and Curtis A Suttle. 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, no. 703 (2014).

2.2 INTRODUCTION

Viruses are highly abundant and widespread in the oceans (Bergh *et al.*, 1989; Suttle, 2005). Beyond their impacts on host mortality, viruses are significant mediators of biogeochemical processes, horizontal gene transfer, and host community diversity in the oceans (Fuhrman, 1999; Wilhelm and Suttle, 1999; Suttle, 2005). Marine viruses are important pathogens of phytoplankton (Brussaard, 2004a) and have been implicated in the termination of blooms (Schroeder *et al.*, 2003; Nagasaki *et al.*, 1994) and with succession in phytoplankton communities (Mühling *et al.*, 2005). Viruses have been characterized that infect a wide variety of phytoplankton such as haptophytes (Bratbak *et al.*, 1993), prasinophytes (Derelle *et al.*, 2008; Mayer and Taylor, 1979; Brussaard *et al.*, 2004), chlorophytes (Van Etten *et al.*, 1981), diatoms (Shirai *et al.*, 2008), and dinoflagellates (Tomaru *et al.*, 2004). Viruses infecting eukaryotic phytoplankton generally have very narrow host ranges (Short, 2012). Viruses infecting marine phytoplankton have genomes comprised of double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), double-stranded RNA (dsRNA), and single-stranded RNA (ssRNA) (as reviewed in Short, 2012). Their genomes and particle sizes range from very large dsDNA viruses in the *Phycodnaviridae* to very small ssDNA and ssRNA viruses belonging to the genus *Bacilladnavirus* and order *Picornavirales*, respectively. The order *Picornavirales* is comprised of positive-sense, ssRNA viruses that infect eukaryotes (Le Gall *et al.*, 2008), including ecologically important marine protists. These viruses are small (25-35nm), icosahedral, and have a conserved genomic organization that includes a replication area comprised of a type III helicase, a 3C-like proteinase, and a type I RNA dependent RNA polymerase (Sanfaçon *et al.*, 2009). Isolates in the *Picornavirales* that are pathogens of marine protists infect a wide diversity of hosts including the bloom-forming raphidophyte *Heterosigma akashiwo* (Tai *et al.*, 2003) (viral family *Marnaviridae*), the thraustochytrid *Aurantiochytrium* sp. (Takao *et al.*, 2005; Yokoyama and Honda, 2007) (viral genus *Labyrnavirus*) and the cosmopolitan diatoms *Rhizosolenia setigera* (Nagasaki *et al.*, 2004) and *Chaetoceros socialis* (Tomaru *et al.*, 2009) (viral genus *Bacillarnavirus*).

Viruses in the *Picornavirales* appear to be common and widely distributed in coastal waters (Culley *et al.*, 2003; Culley and Steward, 2007).

Metagenomic and targeted gene studies are uncovering the diversity of marine RNA viruses. For example, phylogenetic analysis of RNA-dependent RNA polymerase (RdRp) sequences from seawater samples supports a monophyletic marine group within the *Picornavirales* (Culley *et al.*, 2003; Culley and Steward, 2007; Tomaru *et al.*, 2009; Culley *et al.*, 2014) and several divergent clades within this marine group (Culley *et al.*, 2003; Culley and Steward, 2007; Culley *et al.*, 2014). Additionally, metagenomic analyses reveal that there are numerous sequences from aquatic RNA viruses that cannot be assigned to known taxa (Culley *et al.*, 2006; Djikeng *et al.*, 2009; Steward *et al.*, 2012; Culley *et al.*, 2014). Despite the high diversity of marine RNA viruses (Lang *et al.*, 2009), the spatial and temporal distribution of different phylogenetic groups remains unreported, although there is evidence that the taxonomic structure of marine RNA viral communities is highly uneven. For example, in one sample from a metagenomic study from the coastal waters of British Columbia, 59% of the reads assembled into a single contig, while in a second sample 66% of the reads fell into four contigs, with most falling into two genotypes (Culley *et al.*, 2006). However, with only a few hundred reads in total from the two samples, the coverage of the communities was low. Similarly, RNA viral metagenomic data from a freshwater lake (Djikeng *et al.*, 2009) showed little identical sequence overlap among communities, although there was broad taxonomic similarity over time within a location.

Ecological questions about the distribution of marine viruses over time and space have been examined more extensively in bacteriophages, particularly those infecting cyanobacteria. For example, some data reveal no clear patterns of biogeography in cyanophage isolates locally (Clasen *et al.*, 2013), regionally (Jameson *et al.*, 2011) or more globally (Huang *et al.*, 2010). Other data have shown patterns at a regional scale (Marston *et al.*, 2013), where communities in basins that were connected were most similar and those that were separated by land or current boundaries were the least similar. Other data

for marine bacteriophages have shown temporal variability (Chow and Fuhrman, 2012; Clasen *et al.*, 2013; Marston *et al.*, 2013; Chen *et al.*, 2009; Wang *et al.*, 2011). If the dynamics of marine bacteriophages and marine RNA viruses are similar, some RNA viral taxa will persist temporally and spatially, while other taxa will be detected sporadically. To test this hypothesis, we examined two samples, taken five months apart at the same location, and three samples taken within hours of each other, but 20 km apart in the same coastal basin.

We used high-throughput 454 pyrosequencing to obtain deep coverage of RdRp amplicon sequences and compare the richness of viruses in the *Picornavirales* among samples from the coastal waters of British Columbia, Canada. The results revealed a phylogenetically diverse and spatially variable community of viruses, suggesting that taxon-specific lytic events are important in shaping the phytoplankton community.

2.3 MATERIALS AND METHODS

2.3.1 *Sampling locations*

To assess viral communities from different coastal habitats, we collected samples from three sites in the Strait of Georgia (49° 14.926N 123° 35.682W, 49° 17.890N 123° 43.650W, and 49° 23.890N 123° 59.706W), and from Jericho Pier (49° 16'36.73N, 123° 12'05.41W) in British Columbia, Canada (Figure 2.1). The Strait of Georgia (SOG) is an estuarine-influenced basin that is on average 22 km across, 222 km long and 150 m deep. The upper 50m of SOG is where most of the variability in physical and chemical parameters occurs. Jericho Pier (JP) is adjacent to the shoreline, in a well-mixed location with mixed semidiurnal tides.

2.3.2 *Sample collection*

On 28 July 2010, a rosette equipped with a Seabird SBE 25 CTD (equipped with Seabird SBE 43 dissolved oxygen sensor, WET Labs WETStar fluorometer, WET Labs C-Star

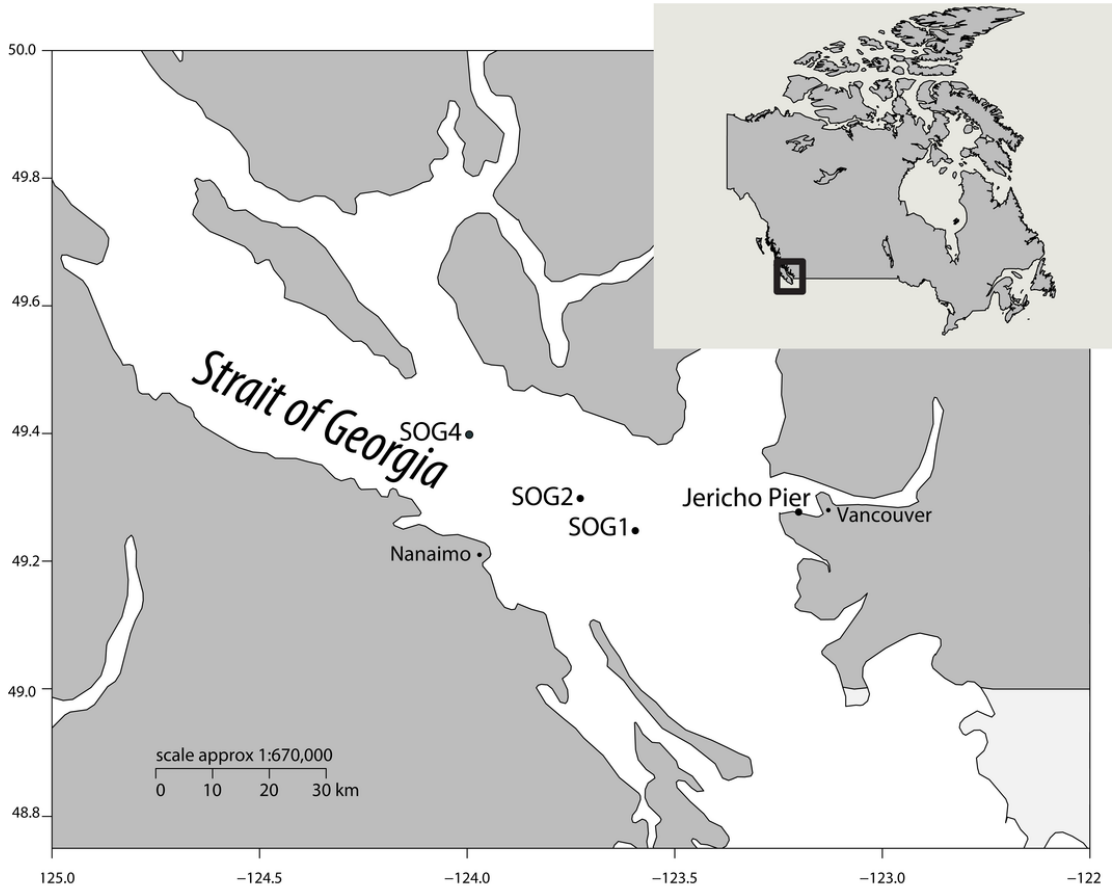

FIGURE 2.1: Location of sampling sites. Map showing the location of sampling sites within the Strait of Georgia (SOG) and Jericho Pier, adjacent to Vancouver, British Columbia, Canada. Jericho Pier was sampled in summer (JP-S) and fall (JP-F).

transmissometer, Biospherical Instruments QSP2200PD PAR sensor) and General Oceanics GO-FLO bottles was used to collect 12 L of water from five depths between 2 and 16 m at each of three stations in SOG. The five depths from each station were combined to obtain three integrated water samples. The depths were selected to attempt to encompass the viral diversity at each station, and consisted of a near-surface sample, a sample from the isothermal zone below the mixed layer, and three samples spanning the chlorophyll maximum.

At JP, 60 L of water was pumped from the 1-m depth on 10 July and 12 October 2010. Salinity and temperature at JP were measured using a YSI probe (Yellow Springs, Ohio, USA). For all samples, the water was filtered through 142-mm diameter, 1.2-μm nominal pore-size glass-fiber (GC50 Advantec MFS, Dublin, CA., USA) and 0.22-μm pore-size polyvinyldine (Millipore Bedford, MA, USA) filters. The viral size fraction in the filtrate was concentrated to ~500 mL (viral concentrate) using tangential flow ultrafiltration using a 30kDa MW prep-scale Spiral Wound TFF-6 cartridge (Millipore) (Suttle *et al.*, 1991). Phosphate, silicate and nitrate+nitrite concentrations were determined in duplicate 15-mL seawater samples filtered through 0.45-μm pore-size HA filters (Millipore) and stored at -20°C until air-segmented continuous-flow analysis on a AutoAnalyzer 3 (Bran & Luebbe, Norderstedt Germany). Chlorophyll *a* (Chl *a*) was determined in triplicate by filtering 100 mL of seawater onto 0.45 μm pore-size HA filters (Millipore), and storing the filters in the dark at -20°C until acetone extraction and then analysed fluorometrically (Parsons *et al.*, 1984). The average and standard error of the replicates was calculated for each sample.

2.3.3 *Nucleic-acid extraction and PCR*

The viral concentrate was filtered twice through 0.22-μm pore-size Durapore PVDF filters (Millipore) in a sterile Sterivex filter unit (Millipore). The filtrate, containing virus-sized particles, was pelleted by ultracentrifugation (Beckman-Coulter, Brea, California, USA) in a SW40 rotor at 108 000 *g* for 5 h at 12°C. The pellet was resuspended overnight

in 100 μL of supernatant at 4°C. To digest free DNA, the pellets were incubated with 1U/μl DNAse with a final concentration 5 mM $MgCl_2$ for 3 h at room temperature. Nucleic acids were extracted using a Qiamp Viral Minelute spin kit (Qiagen, Hilden, Germany) according to the manufacturer's directions. To remove DNA, the extracted viral pellets were digested with DNase 1 (amplification grade) (Invitrogen, Carlsbad, California, USA) and the reaction was terminated by adding 2.5 mM EDTA (final concentration) and incubating for 10 min at 65°C. Complementary DNA (cDNA) was generated using Superscript III reverse transcriptase (Invitrogen) with random hexamers (50ng/μl) as per the manufacturer. PCR was performed using primer set MPL-2 for a targeted set of the marine picornavirus-like RdRp (Culley and Steward, 2007). Each reaction mixture (final volume, 50 μl) consisted of 50 ng of cDNA, 1x (final concentration) PCR buffer (Invitrogen), 2 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 1 μM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with the following conditions: 94°C for 75 s, followed by 40 cycles of denaturation at 94°C for 45 s, annealing at 43°C for 45 s, and extension at 72°C for 60 s and a final extension step of 9 min at 72°C. Negative controls were run with every PCR performed. PCR products were cleaned using the Minelute PCR purification kit (Qiagen).

2.3.4 *Library prep and pyrosequencing*

Libraries for each site were prepared for sequencing using NEBNext DNA Library Prep for 454 kit (New England Biolabs, Ipswich, Massachusetts, USA) following the manufacturer's directions, and using Ampure beads (Beckman-Coulter) for size selection and purification using a bead ratio of 0.8:1 beads:library. PCR amplicons were barcoded and sent for 454 Titanium pyrosequencing (Roche, Basel, Switzerland) at the Génome Québec Innovation Centre at the McGill University (Montreal, QC, Canada).

2.3.5 *Sequence analysis*

Sequencing reads were quality trimmed using length settings between 100 and 600 bp, with a maximum of 3 primer mismatches for the specific primer, and denoised using the denoiser algorithm in QIIME (version 1.7) with default settings for Titanium data (Caporaso *et al*., 2010; Reeder and Knight, 2010). Sequences were checked for chimeras using UCHIME (Edgar, 2010) against a nucleotide database of RdRp sequences built using NCBI-BLAST+ by retrieving nucleotide sequences of the RdRp from *Picornavirales* viral isolates (accessed: 8 August 2012), and also using the denovo chimera check in UCHIME. The overlap of results from these two methods was defined as chimeric sequences, although none were found in this study. Non-chimeric sequences were queried using BLASTx (Altschul *et al*., 1990) with an e-value of $1e^{-3}$ against the database of RdRp viral isolates. All sequences with hits were retained and all sequences with no hits were then queried against the non-redundant (nr) Genbank database (Benson *et al*., 2007) using BLASTx with an e-value of $1e^{-5}$. All the sequences identified as contaminants or as unknown with only 1 read were removed. Remaining sequences were translated to amino acids using FragGeneScan with the 454_10 training option (Rho *et al*., 2010). Sequences were grouped into operational taxonomic units (OTUs) using UCLUST at a range of similarities from 50% to 100% (Figure S 1) using the original seed sequences (centroids) as the output (Edgar, 2010). A similarity of 95% was chosen for this analysis for the following two main reasons: 1) When the NCBI conserved domain alignment for the RdRp region (all *Picornavirales*) was analyzed for percent similarity, the only sequences that displayed greater than 95% similarity in this region were strains of the same virus. 2) The sequences from the control libraries (consisting of 1 clone) clustered into 1 sequence at this percentage (see Supplemental methods). Thus clustering at 95% similarity was a way to use biological and sequence-based information to inform our choice of cut-off to collapse strain level variation and as a conservative approach to avoid variation that may be present because of the sequencing platform. Control sequences obtained by cloning and Sanger sequencing (see Supplemental Methods) were used to

verify the sequence processing methodology. Raw and processed sequence data were deposited in the NCBI BioProject database ID: PRJNA267690.

2.3.6 *Phylogenetic analysis*

All OTUs with less than 5 reads were removed and the remaining OTUs were aligned using profile alignment in Muscle (Edgar, 2010) to seed alignments of viral RdRps from the NCBI Conserved Domain Database (Marchler-Bauer *et al.*, 2010). Sequences from other environmental surveys were clustered in the same manner as the reads in this study (using UCLUST at 95%). The clusters are cluster number followed by the Genbank accession numbers contained in that cluster. 0: 33520549, 33520547, 33520541, 33520533, 33520527, 33520521, 33520519, 33520517, 33520515, 33520513, 33520511, 33520509; 1: 157280772; 2 : 33520525; 3: 157280768; 4: 157280770; 5: 568801536, 568801534, 568801530, 568801528, 568801510; 6: 157280786; 7: 157280774; 8: 157280780; 9: 157280788; 10: 568801494, 568801492, 568801488, 568801482, 568801480, 568801474, 568801470, 568801466, 568801464, 568801462, 568801458; 11: 157280776; 12: 157280778; 13: 568801516; 14: 568801616, 568801614, 568801606, 568801588, 568801586, 568801584, 568801582, 568801580, 568801574, 568801572, 568801564, 568801550, 568801548, 568801540, 568801532, 568801522, 568801520, 157280784; 15: 568801508; 16: 568801542, 568801538, 568801518, 568801486, 568801484, 568801478; 17: 568801612, 568801610, 568801608, 568801604, 568801602, 568801600, 568801598, 568801596, 568801594, 568801592, 568801590, 568801578, 568801570, 568801568, 568801552, 157280782; 18: 568801576, 568801566, 568801546, 568801544; 19: 568801562, 568801556; 20: 568801526, 568801524, 568801514, 568801512, 568801506, 568801504, 568801502, 568801500, 568801498, 568801496, 568801490, 568801472, 568801460; 21: 568801476, 568801468; 22: 568801560, 568801558, 568801554; 23: 157280744; 24: 157280758; 25: 157280748; 26: 157280746; 27: 157280742; 28: 157280766; 29: 157280756; 30: 157280762, 157280754, 157280752; 31: 157280750; 32: 157280760; 33: 157280764; 34: 33520545, 33520543, 33520535, 33520531, 33520529; 35: 33520539, 33520537, 33520523, 33520507. Alignments were masked using trimAl with the automatic heuristic (Capella-Gutierrez *et al.*, 2009) and edited manually. ProtTest 3.2 was used for

amino-acid model selection (Darriba *et al.*, 2011) before building the initial phylogenetic tree using FastTree (Price *et al.*, 2010). Final maximum likelihood trees were done with RAxML using sequences belonging to viruses in the *Sequiviridae* as the outgroup, and the BLOSUM62 amino-acid model with 100 bootstraps (Stamatakis *et al.*, 2008). The tree was visualized in R (R Core Team, 2014) using the ape package (Paradis, 2012) and edited in Figtree (Rambaut, 2014).

2.3.7 *Statistical analysis*

Generation of rarefaction curves by random resampling of OTU abundances was performed using the vegan package (Oksanen *et al.*, 2013) in R (R Core Team, 2014). Relative abundances were normalized by randomly resampling 10 000 times using vegan, normalizing to the library with the lowest number of reads and then taking the median. Rank-abundance curves were generated with ggplot2 (Wickham, 2009) using the OTUs per site normalized by the library with the lowest number of reads by taking the median of 1000 random rarefications of the OTU abundance data. Scripts used in this project are available as part of QIIME and custom user scripts used to process the data are available on github (Gustavsen, 2015).

2.4 RESULTS

2.4.1 *Environmental parameters*

The environmental parameters ranged widely among samples (Figure 2.2). Chlorophyll *a* values were lowest at Jericho Pier Fall (JP-F) at 0.16 μg L-1 (± 0.04) and highest at the Strait of Georgia Station 2 (SOG-2) at 3.2 μg L-1 (± 0.4). Silicate values for all samples were similar (range of 25.5 to 37.8 μM), except for Jericho Pier Summer (JP-S) when silicate was lower at 6.2 μM (± 0.01). Phosphate ranged between 0.90 and 1.3 μM at JP-F, SOG-1, and SOG-4, but was lower at SOG-2 (0.58 μM (± 0.03)) and lowest at JP-S (0.06 μM (± 0.04)). Nitrate + nitrite values were more variable than the other nutrients

and ranged from 1.35 μM (± 0.75) at JP-S to 14.6 μM (± 0.04) at JP-F. The SOG sites were highly stratified with SOG-4 being the most stratified with a calculated mixed-layer depth of 2 m, while SOG-1 and SOG-2 were similar with a mixed-layer depth of 6 m (Table 2.1).

TABLE 2.1: Description of samples and resulting sequencing information. * after quality filtering and matching to the RdRp primer set.

Location of sampling	Date of sample collection	Latitude, Longitude	Reads*	Mixed layer depth (m)
Total reads			300180	
Jericho Pier Summer	10 July 2010	49° 16'36.73N, 123° 12'05.41W	74096	-
Jericho Pier Fall	12 October 2010	49° 16'36.73N, 123° 12'05.41W	84907	-
SOG 1	28 July 2010	49° 14.926N, 123° 35.682W	55197	6
SOG 2	28 July 2010	49° 17.890N, 123° 43.650W	12269	6
SOG 4	28 July 2010	49° 23.890N, 123° 59.706W	73044	2

2.4.2 *Analysis of RdRp sequences*

After quality filtering to remove homopolymers and contaminating reads, 300 180 reads were recovered from the 5 libraries of RdRp amplicons. At all sites the rarefaction curves plateaued indicating that the depth of sampling was adequate to assess the communities (Figure 2.3). From these reads, 265 unique OTUs (at 95% similarity) were identified,

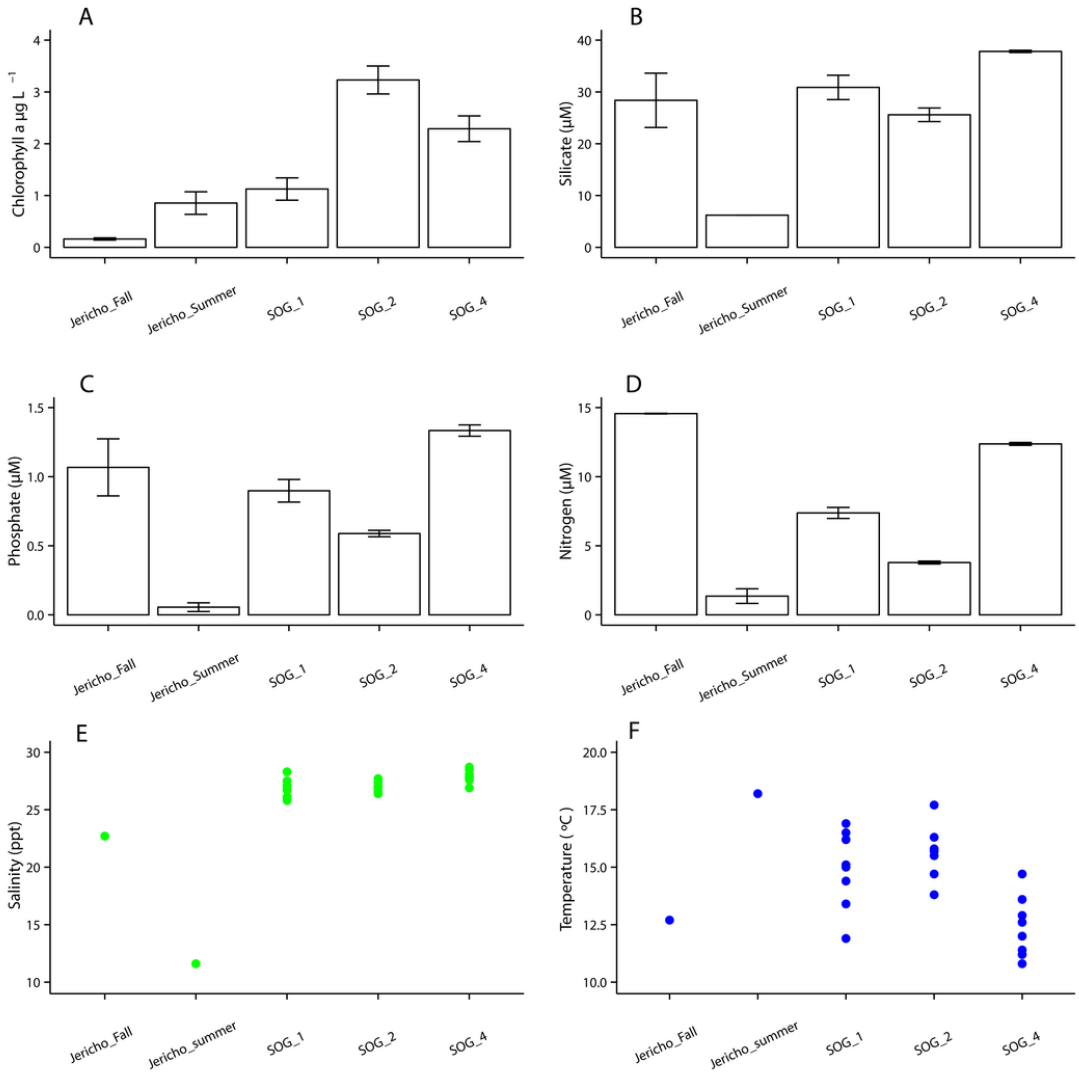

FIGURE 2.2: Environmental parameters. A) Chlorophyll a with standard error of the mean from triplicates. B) Silicate with standard error of the mean from duplicates. C) Phosphate with standard error of the mean from duplicates. D) Nitrate+ nitrite with standard error of the mean from duplicates. E) Temperature with each point as one GO-FLO bottle (SOG samples) or from total seawater sample (Jericho samples). F) Salinity with each point representing a seawater sample as one GO-FLO bottle (SOG samples) or from total seawater sample (Jericho samples).

including 108 singletons. For further analysis OTUs were excluded that did not contain recognizable RdRp motifs (Koonin, 1991; Le Gall *et al*., 2008), generally did not align well with other RdRp sequences and those that were not present in any sample after normalization. Using the above criteria there were 145 OTUs identified in all samples, and between 24 to 71 OTUs per site. The Jericho Pier samples had 116 OTUs of which only 10 (8.6%) were shared between sampling times (Figure 2.4). JP-S had the highest richness with 71 OTUs, of which 59 (83%) were unique, while JP-F had the second highest richness (49 OTUs), of which 45 (92%) were unique. The SOG sites together had 64 OTUs, none of which were shared among all sites. SOG-1 and SOG-2 had the lowest number of OTUs (24). SOG-1 had only three OTUs which were unique. However, 21 (33%) were shared between SOG-1 and SOG-4, and 6 (9%) between SOG-2 and SOG-4. The majority of OTUs (75%) from SOG-2 were unique, whereas most OTUs from the other SOG sites were shared with other sites (87% for SOG-1 and 63% for SOG-4).

Rank abundance curves of the viral OTUs showed that at each site most sequences were assigned to only a few OTUs (Figure 2.5). JP-S had the highest richness but the shallowest slope of these curves, demonstrating more evenness in the abundance of OTUs than at the other sites. SOG-4 and JP-S had similar rank abundance curves that were much shallower that those of SOG-1 and SOG-2 (Figure 2.5).

The OTUs that were observed in more than 5 reads were placed in phylogenetic context using a maximum likelihood RAxML tree (Stamatakis *et al*., 2008) with sequences from previous RdRp gene surveys and isolated viruses (Figure 2.6). OTUs from this study fell within a well-supported clade that includes all the marine isolates belonging to the *Picornavirales*. Within this group there was a well-supported divide between OTUs grouping in the *Marnaviridae* clade and those grouping with sequences from viruses infecting diatoms and a thraustochytrid. The overall tree topology is not well supported, although there are a number of well-supported clades containing OTUs from this study and other environmental sequences. The *Marnaviridae* clade had the greatest number of OTUs (10) associated with it; whereas, very few OTUs (only OTUs 89, 107, 75 and

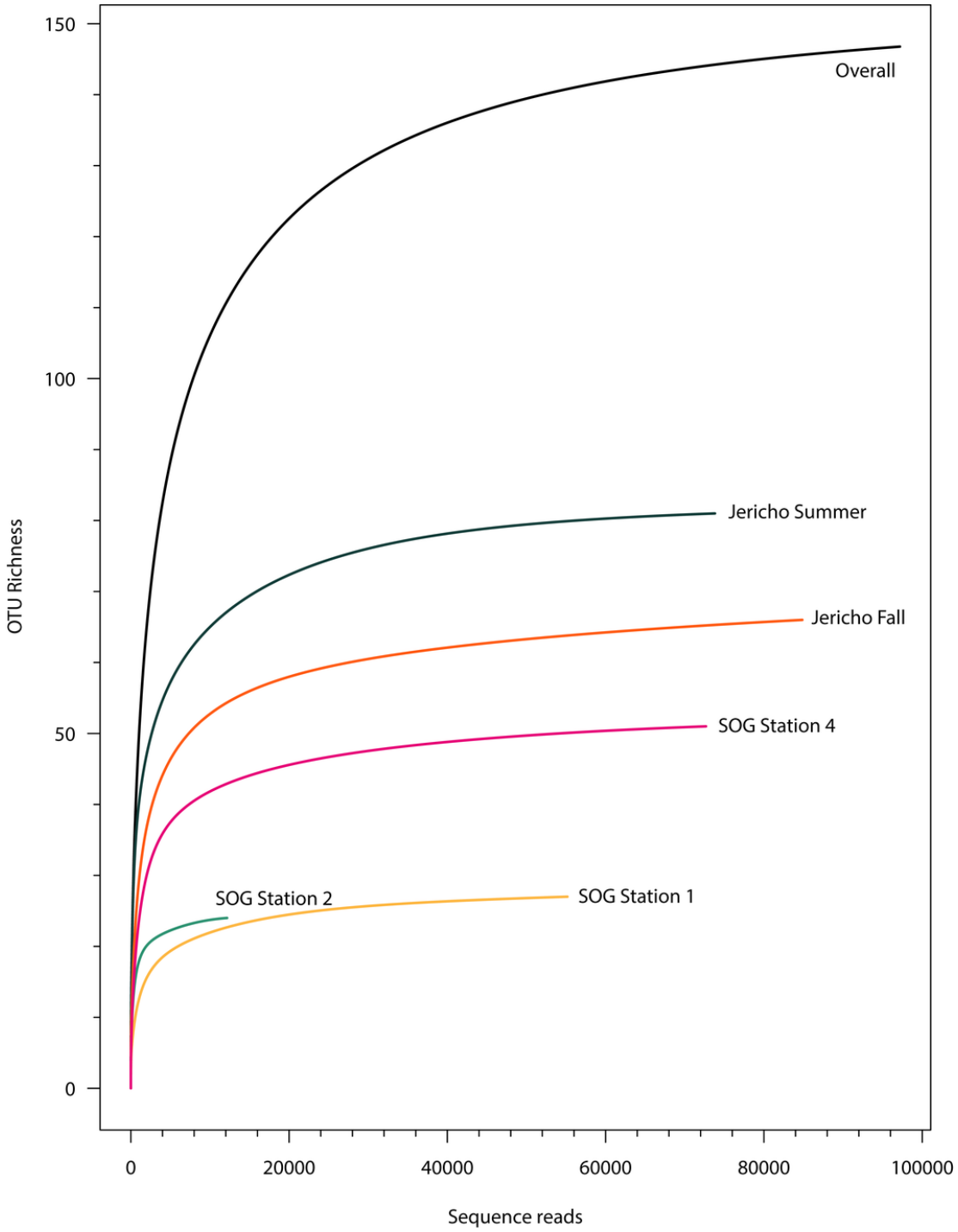

FIGURE 2.3: Rarefaction curves. Rarefaction analysis of RdRp amplicons based on Chao1 richness analysis of operational taxonomic units (OTUs) at 95% similarity. Rarefaction curves were resampled using number of reads recovered per library. Rarefaction curves plateau indicating adequate sequencing for these samples

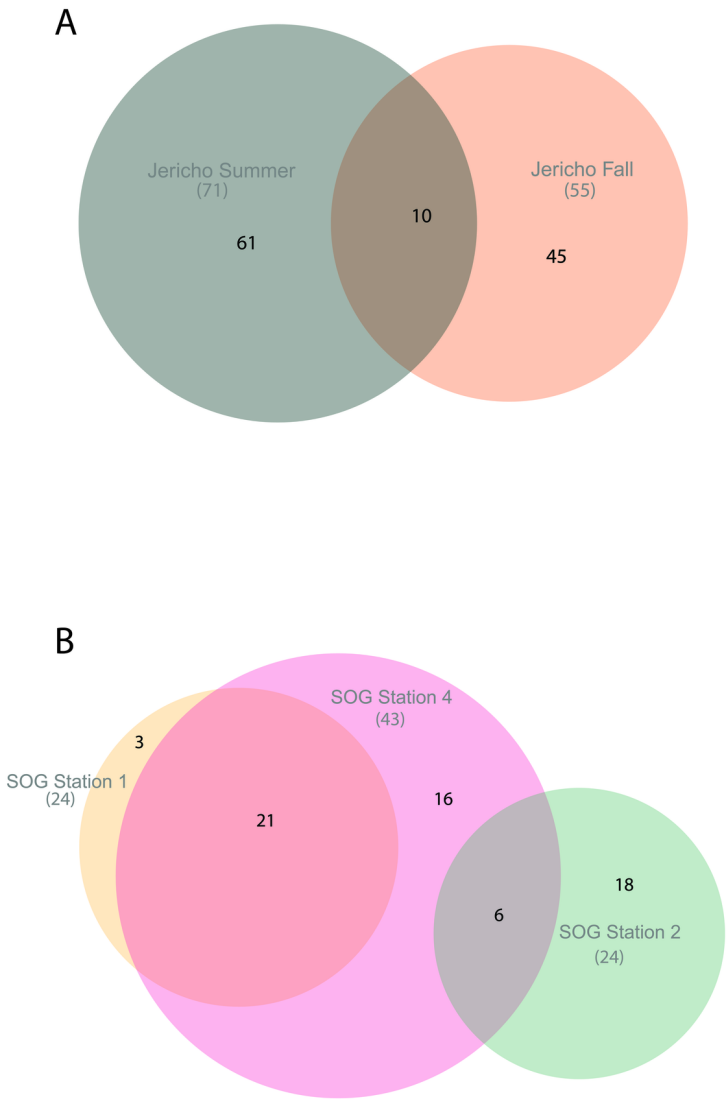

FIGURE 2.4: Euler diagrams of normalized RdRp OTUs. A) Euler diagram of Jericho Pier samples. B) Euler diagram of SOG samples. The OTUs presented were from reads clustered at 95% similarity, comprise only OTUs that could be aligned to the NCBI CDD RdRp alignment, and contained the RdRp motif C. The diagrams were constructed using the venneuler() algorithm (Wilkinson, 2011). The size of the circles is approximately proportional to the number of OTUs recovered per site. The overlap in the diagram describes OTUs that were found at multiple sites and the non-overlapping areas describe OTUs that were unique to that site.

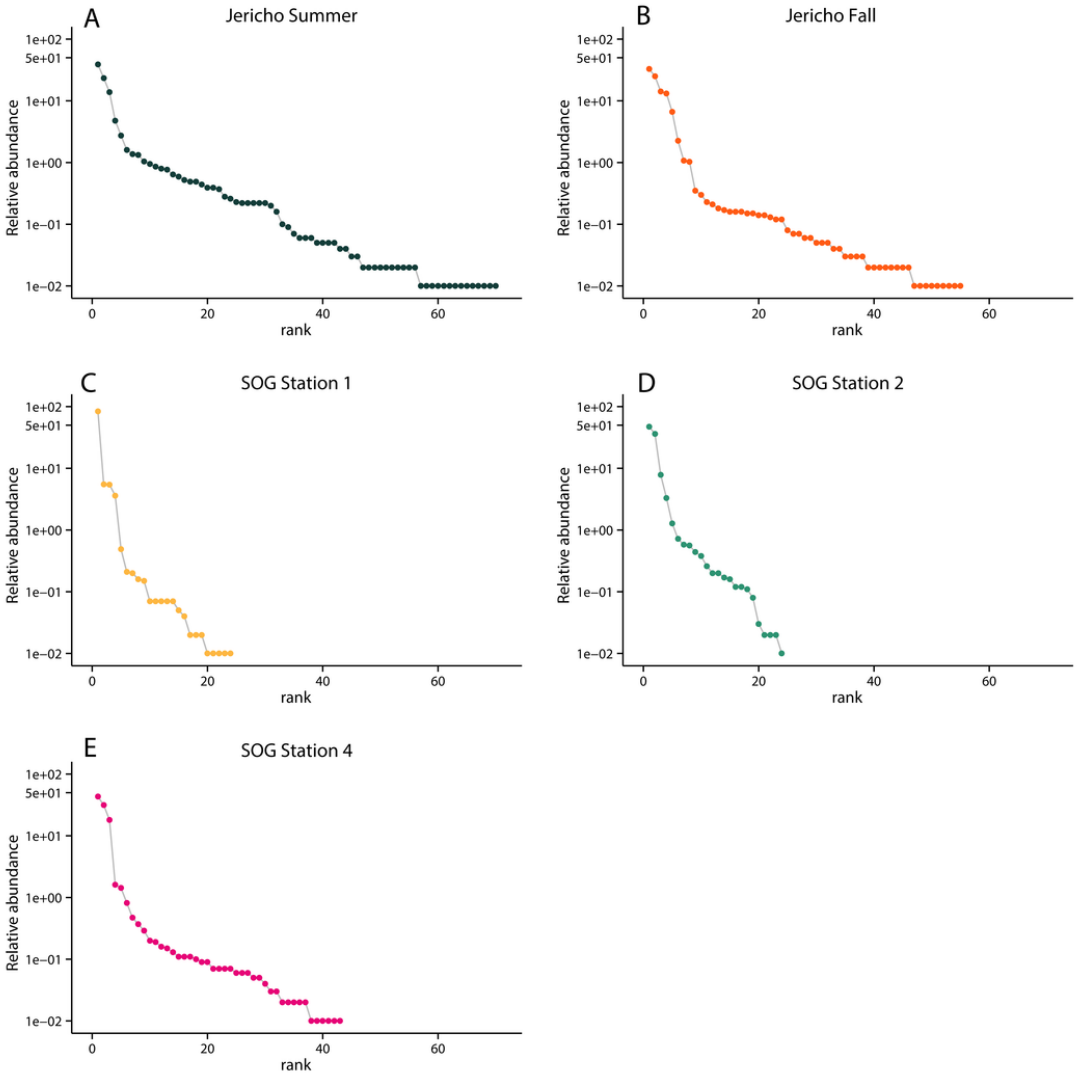

FIGURE 2.5: Rank abundance by site. Relative abundance of OTUs in each sample ordered by rank abundance. OTUs were clustered at 95% amino acid similarity and OTU relative abundances were normalized to the sample with the lowest number of reads

120) from this study were assigned to clades primarily from Hawaii (Culley and Steward, 2007; Culley *et al.*, 2014). No clade contained OTUs from all sites. The Jericho Pier samples were the most phylogenetically diverse (Figure 2.6, Table 2.2), and contained OTUs (e.g. OTUs 6, 7, 35, 31, 84, 47, 14, 4, 34, 39, 44, 23, 8, 20) that fell into clades that did not contain OTUs from any of the SOG samples. Some clades contained OTUs from both JP-S and JP-F samples; however, many OTUs within the clades were unique to one Jericho Pier sample. Phylogenetic diversity differed among samples, except for OTUs from the well-mixed SOG sites, some of which were present in different clades resulting in similar phylogenetic diversity (Table 2.2).

TABLE 2.2: Phylogenetic diversity, species richness. Phylogenetic diversity (PD) is calculated as in (Faith, 1992). OTUs were the same as used in the construction of the phylogenetic tree and must have included five or more reads.

	Phylogenetic diversity	Species richness
Jericho Pier Summer	16.25	46
Jericho Pier Fall	13.62	30
SOG Station 1	3.89	9
SOG Station 2	3.26	4
SOG Station 4	8.35	24

The Strait of Georgia (SOG) sites were sampled within hours of each other, and the water at each site was pooled from multiple depths above, below, and across the chlorophyll maximum. One of the most striking differences among sites was that SOG-1 and SOG -2 had mixed layer depths of 6m; whereas SOG-4 had a mixed-layer depth of 2 m, and much higher richness and phylogenetic diversity. Sites SOG-1 and SOG-2 had the lowest phylogenetic diversity (Table 2.2). All the OTUs found at SOG-1 (33, 12, 16, 9, 10, 27, 29) were within the *Marnaviridae* clade; similarly, all OTUs (5, 82, 2, 1) from SOG-2 were within one distantly related clade. In both cases OTUs from these clades

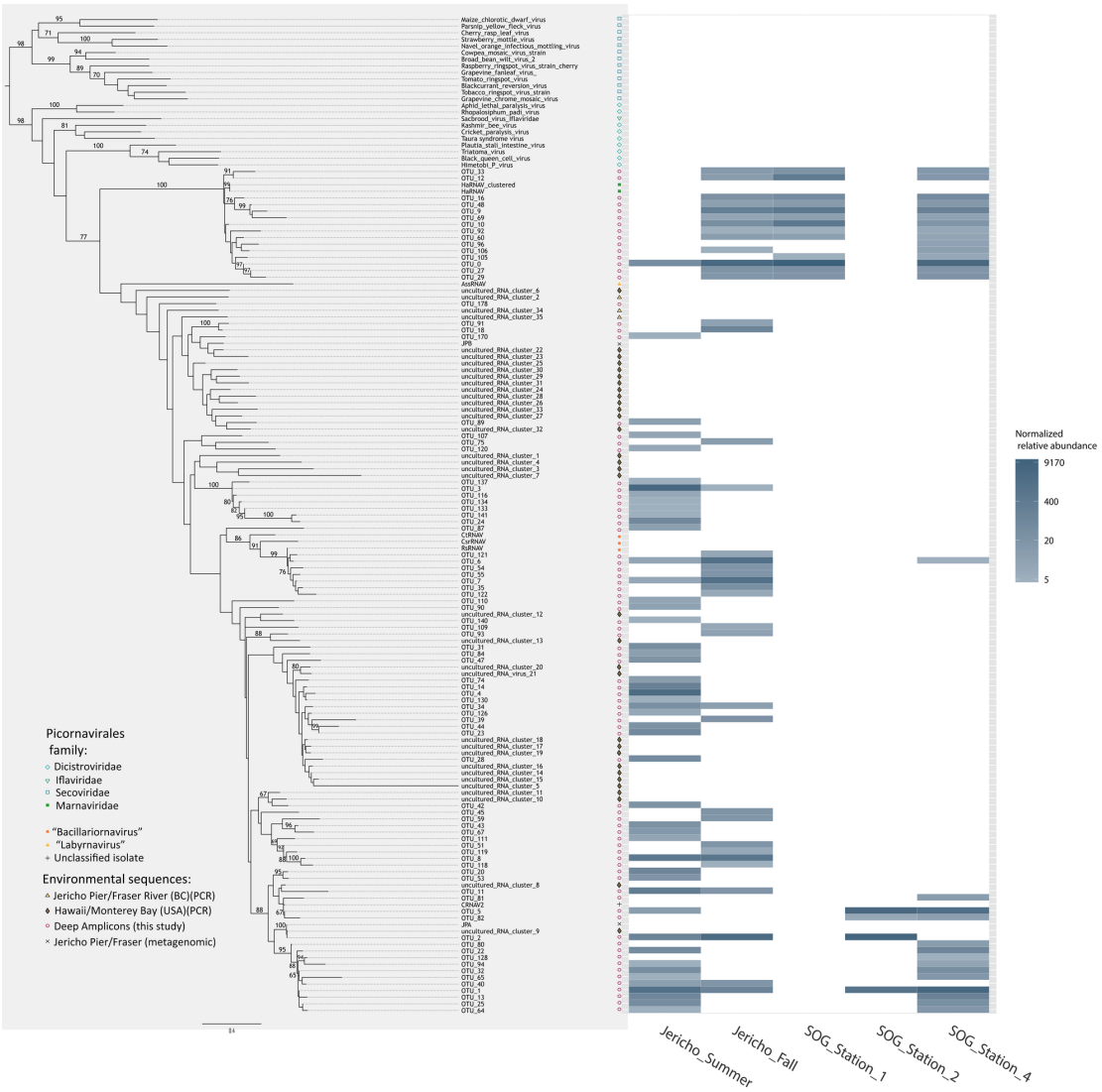

FIGURE 2.6: Phylogenetic tree with heatmap. Maximum likelihood (RAxML) tree of RdRp OTUs that contained more than 5 reads and relevant *Picornavirales* sequences. Bootstrap values above 65% are labelled. Adjacent to the tips of the tree is a heatmap displaying the relative abundance of each OTU at 95% similarity by site. OTU relative abundances were normalized to the sample with the lowest number of reads.

occurred at SOG-4. SOG-2 did not have the high numbers of HaRNAV-related viruses that were found in all other samples.

2.5 DISCUSSION

Pyrosequencing of RdRp gene fragments from coastal samples uncovered much greater genetic diversity than in previous gene surveys (Culley *et al.*, 2003; Culley and Steward, 2007; Culley *et al.*, 2014) and revealed many previously unknown taxonomic groups within the *Picornavirales*. As well, striking differences in the taxonomic richness among samples implies that these viruses infect a wide variety of eukaryotic plankton, but that the mortality imposed on some taxa is highly variable across space and time. Other taxonomic groups within the *Picornavirales* were more widespread, suggesting that infection of some planktonic taxa is more widespread and persistent. These results and their implications are discussed in detail below.

2.5.1 *Expanding the known diversity of* Picornavirales

The high depth of sequencing and limited diversity in each library (Figure 2.3) gives high confidence that the population structure of RdRp amplicons in each sample has been well characterized (Kemp and Aller, 2004). Although some sequences were closely related to those found in previous studies (Culley *et al.*, 2003; Culley and Steward, 2007) (Figure 2.6), many OTUs formed new clades. Many OTUs were related to *Heterosigma akashiwo* RNA virus (HaRNAV) that infects the toxic bloom-forming raphidophyte *Heterosigma akashiwo* (Tai *et al.*, 2003). HaRNAV is the type virus of the family *Marnaviridae* (Lang *et al.*, 2004); it has a genome of about 9.1kb and a high burst size as indicated by the large crystalline arrays of particles in the cytoplasm of infected cells (Tai *et al.*, 2003). HaRNAV was isolated from coastal waters in British Columbia, Canada (Tai *et al.*, 2003) and can remain infectious for many years in sediments (Lawrence and Suttle, 2004). Interestingly, HaRNAV was isolated from the same area as the present study, and

appeared ancestral to many of the recovered sequences based on the phylogeny. For example, OTU 0 was most abundant (18,034 reads after rarefaction) and clustered closely with HaRNAV, although the sequence was only 76.4% similar at the amino-acid level. However, other OTUs in the cluster ranged between 55 and 79% similar to HaRNAV, which is low compared to amino-acid similarities of other RNA viruses within a family that usually have greater than 90% aa similarity (Ng *et al.*, 2012).

2.5.2 *Distinct communities occurred in different seasons at the same location*

While only 8.6% of the OTUs from Jericho Pier were shared between dates, both samples had similar evenness, although the summer sample had greater richness (Figure 2.5, Table 2.2). The small overlap in OTUs between sampling dates is not surprising given the very different conditions between July and October (Figure 2.1), and the dynamic nature of planktonic communities in response to environmental changes. At the same location, dsDNA viruses belonging to the *Phycodnaviridae*, which infect eukaryotic phytoplankton, varied seasonally based on fingerprint analyses of DNA polymerase gene fragments using denaturing gradient gel electrophoresis; however, some OTUs persisted for extended periods (Short and Suttle, 2003). Similarly, the composition of other aquatic viral communities has been shown to be dynamic although some OTUs persist (Djikeng *et al.*, 2009; Rodriguez-Brito *et al.*, 2010), and in some cases have repeatable seasonal patterns (Chow and Fuhrman, 2012; Clasen *et al.*, 2013; Marston *et al.*, 2013). With only single samples from summer and fall, inferences about dynamics cannot be made from our data. One of the few taxonomic groups that occurred in both the summer (JP-S) and fall (JP-F) samples from Jericho Pier was related to HaRNAV, (Figure 2.6). There was greater diversity of OTUs in this clade in JP-F, even though JP-S had higher richness and higher phylogenetic diversity overall. This is unlike bacterial and phytoplankton communities that tend to be more diverse in winter (Zingone *et al.*, 2009; Ladau *et al.*, 2013). However, the RdRp primers target a specific subset of the viral community that does not reflect the overall taxonomic diversity.

Based on genome organization and sequence identity, RNA viruses that infect diatoms have been assigned to the genus *Bacillarnavirus*, that includes *Rhizosolenia setigera* RNA virus (RsRNAV) (Nagasaki *et al*., 2004), *Chaetoceros tenuissimus* Meunier RNA virus (CtenRNAV) (Shirai *et al*., 2008) and *Chaetoceros socialis f. radians* RNA virus (CsfrRNAV) (Tomaru *et al*., 2009). In the JP-F sample, the most relatively abundant cluster grouped with RsRNAV that infects the marine diatom *Rhizosolenia setigera* (Nagasaki *et al*., 2004). This corresponded with the highest levels of nitrate + nitrite, which is often associated with high diatom abundances (Zingone *et al*., 2009); hence, these OTUs are likely associated with viruses infecting diatoms.

2.5.3 *Distinct communities occurred at geographically proximate sites*

Areas of higher habitat diversity, such as stratified water layers, generally have higher biological richness (Klopfer and MacArthur, 1960; Chesson, 2000), and this is consistent with the much higher richness and phylogenetic diversity found at SOG-4, which was the most stratified site and included the most abundant OTUs from SOG-1 and SOG-2. Most OTUs from SOG-1 clustered in the *Marnaviridae* clade, while most SOG-2 OTUs clustered in a phylogenetically distant clade. Given that we have used very conservative clustering, that dsDNA viruses infecting phytoplankton are strain specific and have phylogenies that are congruent with their hosts (Clasen and Suttle, 2009; Bellec *et al*., 2014), and that RNA viruses infecting diatoms and the dinoflagellate, *Heterocapsa circularisquama* (Nagasaki *et al*., 2005) are host-specific, it implies that closely related OTUs infect closely related taxa of phytoplankton. Hence, it suggests that the most abundant viruses at these three locations infect different species.

There are few clear patterns in the spatial distribution of viruses in marine waters where geographically distant sites are connected by currents and mixing. The best examples are for cyanophages. For instance, when looking at local variation in cyanophages isolated at sites in Southern New England, 72% of the viral OTUs were shared between at least 2 sites (Marston *et al*., 2013); however, between Bermuda and Southern New

England only 2 OTUs overlapped and they comprised only 0.6% of the isolates. Yet, clear patterns of cyanophage OTU distribution by depth occurred in areas adjacent to the SOG when assessed using community fingerprinting (Frederickson *et al.*, 2003). The biggest differences with depth occurred in stratified water in which some OTUs were present at all depths, while others were only present at specific depths, even though the samples were collected only meters apart (Frederickson *et al.*, 2003). These viruses infect cyanobacteria, as opposed to the picorna-like viruses, which likely infect protistan plankton. Nonetheless, the factors governing the distribution of cyanobacterial and protistan hosts are likely similar; hence, different OTUs would be expected to occur in environments with different vertical structure (stratification) of the water column.

Rank abundance curves showed that SOG-1 and SOG-2 were the least even communities (Figure 2.5). Overall, at most sites four to five viral OTUs were most abundant (Figure 2.5) similar to other reports for aquatic viral communities in which a few viruses dominate, but most of the diversity comes from rarer viruses (Angly *et al.*, 2006; Suttle, 2007). Our targeted approach showed that the picornavirus-like virus communities at SOG and JP were dominated by only a few genotypes, supporting previous metagenomic results showing that the OTU distributions of RNA viruses in SOG and JP were highly uneven with little overlap between sites (Culley *et al.*, 2006).

2.5.4 *Each OTU likely represents a single lytic event*

Given that the known hosts of marine *Picornavirales* isolates targeted by the primer set (Culley and Steward, 2007) are protists, and that protists are the most abundant eukaryotes in the sea, it is likely that the majority of OTUs recovered in this study are from viruses that infect these unicellular marine eukaryotes. These eukaryotic communities are highly dynamic and change throughout the year based on environmental and biological factors (Larsen *et al.*, 2004). Since viral infection is usually host specific, the diversity in marine viral communities is a reflection of the underlying diversity of the marine eukaryotic hosts. Moreover, viral propagation is dependent on host encounter

rates and is proportional to host-cell abundance (Murray and Jackson, 1992); hence the most abundant taxa will be most likely to encounter and propagate a viral infection, giving the opportunity for rarer species to increase in abundance and promoting diversity (Thingstad, 2000; Winter *et al*., 2010). Since our study was not over time it is difficult to evaluate whether these data support the bank model (Breitbart and Rohwer, 2005), however, some taxa were found at one site, but not a similar nearby site, thus these taxa could be present at background levels at some sites and more abundant in others.

It is probable that the most abundant OTUs in these data are from recent lysis of host taxa. An error rate for replication of RNA viruses of about 1 bp mutation per generation (9000 bp genome x 0.0001 error rate per base pair = 1bp (Holmes, 2009)), and a lower-end burst-size estimate of 1000 particles for marine viruses in the *Picornavirales* that infect protists (Lang *et al*., 2009), would produce about 1000 different genomes from each lysed cell. For the amplified 500bp RdRp gene fragment there is a 0.00056% chance of an error in 1 generation, assuming that mutations are distributed evenly in the genome (Sanjuan *et al*., 2010; Combe and Sanjuán, 2014). Consequently, even with the relatively high error rates of RNA replication, when grouped at 95% similarity at the amino acid level, all of the sequences from a lytic event should fall within a single OTU. The half-life for decay of viral infectivity and particles in the surface mixed layer is typically a few hours (Heldal and Bratbak, 1991; Suttle and Chen, 1992; Noble and Fuhrman, 1997; Bettarel *et al*., 2009); thus the recovered viral OTUs were likely from recent lytic events. Furthermore, considering the specificity of viruses infecting protists (Short, 2012), each OTU probably stems from viruses infecting a single host taxon. Thus, these data imply that infection of marine protists by viruses in the *Picornavirales* is not only pervasive, but likely involves a wide diversity of host taxa; hence, these viruses are likely important structuring elements for phytoplankton communities that influence nutrient cycling and energy flow.

2.5.5 *Amplicon deep sequencing as an approach for estimating viral diversity*

Amplicon deep sequencing is a sensitive and high-resolution approach for examining microbial community dynamics over time and space (Gobet *et al.*, 2012; Caporaso *et al.*, 2011; Gibbons *et al.*, 2013). Careful quality trimming of sequences and removal of singletons is essential for reliable results (Zhou *et al.*, 2011) since errors in sequences will inflate estimates of diversity. With careful data processing and analysis, amplicon deep sequencing is as accurate for assessing community composition and diversity as cloning and Sanger sequencing (Amend *et al.*, 2010), but with much greater depth of coverage of the community.

There are potential biases associated with reverse transcription with random hexamers (which can decrease yield and could inflate diversity) (Zhang and Byrne, 1999), template amplification by PCR (Lee *et al.*, 2012) and with using highly degenerate primers that target a specific part of the community containing many different templates (Culley and Steward, 2007). A danger of the high cycle number can be diversity overestimates which can come from the increasing number of chimeric sequences produced with greater cycle number (Qiu *et al.*, 2001). The sequences were processed with caution considering the high number of PCR cycles employed in this study. Chimera checking denovo was used to look for chimeric sequences originating from two higher abundance reads, and reference-based chimera checking was used a database of RdRps from isolated viruses to correct for this potential error. In addition, a conservative cut-off was used of only OTUs comprising more than 5 reads that aligned to the conserved domain alignment.

Although read abundance of OTUs can be considered semi-quantitative and good for comparisons of richness and diversity among samples (but not for absolute counts of genes)(Amend *et al.*, 2010; Pinto and Raskin, 2012; Ibarbalz *et al.*, 2014). Moreover, by using control sequences obtained by cloning and Sanger sequencing alongside pyrosequenced libraries containing the same sequence (Appendix A) we verified that amplicon deep sequencing and our sequence processing methodology recovered accurate environ-

mental viral sequences and non-inflated estimates of richness like in studies for bacterial amplicons (Sogin *et al.*, 2006; Huse *et al.*, 2008; Kirchman *et al.*, 2010; Caporaso *et al.*, 2011) and clinical viral studies (Watson *et al.*, 2013; Romano *et al.*, 2013).

2.6 CONCLUSION

Amplicon deep sequencing of RdRp gene fragments using 454 pyrosequencing revealed the richness and population structure of marine *Picornavirales* in five coastal samples. The known diversity of viruses in this group was greatly increased with 145 OTUs that differed by at least 5% at the amino-acid level. There were between 24 and 71 OTUs in each sample, with distinct patterns of OTU distribution, richness and diversity among samples. There was little overlap between viral OTUs collected at the same site in summer and fall, and among samples collected 20 km apart on the same day. The high temporal and spatial diversity in RdRp sequences is consistent with viral communities that turnover rapidly, and episodic infection of a wide diversity of protistan hosts. The low overlap in OTUs and phylogenetic diversity among samples implies a dynamic landscape of viral infection and supports the idea that marine picorna-like viruses are important pathogens of marine protists that have an important role in structuring marine planktonic communities, and in nutrient cycling and energy transfer among trophic levels. Ultimately, further study is needed to disentangle the temporal and spatial drivers of these communities.

CHAPTER 3

Marine virus and host community structure exhibits temporal phylogenetic dynamics

3.1 SUMMARY

Marine microbes and their viruses are essential parts of the marine ecosystem that form the base of the foodweb, and drive biogeochemical cycles. Studies have shown that marine viral communities display repeatable changes in abundance and community composition with time; however, whether these changes reflect shifts in dominance within evolutionarily related groups of viruses and their hosts is unexplored. To examine these dynamics, changes in the composition and phylogenetic makeup of two ecologically important groups of viruses, and their potential hosts, were followed at a coastal site near Vancouver, Canada, every two weeks for 13 months. Changes in the taxonomic composition within DNA bacteriophages belonging to the T4-like myoviruses and marine picorna-like RNA viruses infecting eukaryotic phytoplankton, as well as bacteria and eukaryotes, were followed using amplicon sequencing of gene fragments encoding the major capsid protein (gp23), the RNA-dependent RNA polymerase (RdRp) and the 16S and 18S ribosomal RNA genes, respectively. The results showed that for the viral groups the dominant groups of phylogenetically related viruses shifted over time, and that there were many transient taxa and few persistent taxa. Yet, different community structures were observed for different marker genes. Additionally, with strong lagged correlations between viral richness and community similarity of putative hosts, the results imply that viruses influence the composition of the host communities, and that

their community structure is dependent on lifestyle, cementing their role as important structuring elements in marine planktonic communities.

3.2 INTRODUCTION

Understanding diversity, its maintenance and drivers is a continued theme in ecology. This is very evident for microbial systems, for which there has been extensive exploration and discussion on the mechanisms responsible for the observed high diversity. Many studies on microbial diversity and dynamics come from the marine milieu, where it has been argued that community composition is driven by environmental factors (DuRand *et al.*, 2001; Morris *et al.*, 2005; Fuhrman *et al.*, 2006; Gilbert *et al.*, 2009). Against this backdrop are viruses, which are obligate and ubiquitous, and the most abundant and diverse biological entities in the world's oceans (Suttle, 2005).

This diversity arises since viruses have many different lifestyles (Paul, 2008), and morphologies (Brum *et al.*, 2013). Diversity is also generated from an assortment of infection strategies as some viruses infect specific strains or species of hosts whereas others have broad host-ranges (Breitbart, 2012). As well, some groups of viruses show particularly high genetic diversity because of their low fidelity of replication (Lang *et al.*, 2009), while others have high rates of horizontal gene transfer (e.g. Moreau *et al.*, 2010). The role of viruses as obligate pathogens with high host specificity implies that they are important drivers of host composition and diversity (Rodriguez-Valera *et al.*, 2009); yet, our understanding of their roles as drivers of marine microbial diversity remains relatively unexplored.

Marine viruses have repeatable seasonal dynamics as revealed by measures of abundance, infectious units, and taxonomic composition. Seasonal studies in coastal waters have reported that viral abundances are higher in summer than in winter (Bergh *et al.*, 1989; Jiang and Paul, 1994), while multi-year time time series data show that viral production and viral abundances are highest in early spring and summer (Winget *et al.*,

2011; Parsons *et al.*, 2012) although viral lysis is highest in winter (Winget *et al.*, 2011). Moreover, viral dynamics can be associated with putative hosts (Parsons *et al.*, 2012) and specific subsets of the coastal viral community can show seasonal community composition dynamics (Chow and Fuhrman, 2012; Pagarete *et al.*, 2013). As well, viruses infecting cyanobacteria are also temporally dynamic (Waterbury and Valois, 1993; Suttle and Chan, 1994), with communities from the same season resembling each other more than communities sampled in the same year (Marston *et al.*, 2013) and winter communities being more stable than in the summer and spring (Clasen *et al.*, 2013).

Viruses affect community composition in laboratory studies by reducing the abundance of the dominant host, allowing others to grow up (e.g. Middelboe, 2000; Middelboe *et al.*, 2001; Bouvier and del Giorgio, 2007; Marston *et al.*, 2012); thus, viruses promote diversity at the strain level and can be responsible for large shifts in bacterial populations (Hewson *et al.*, 2003; Schwalbach *et al.*, 2004; Rodriguez-Valera *et al.*, 2009). These dynamics have been termed "Killing the Winner" (KtW), a theory in which the most actively growing hosts are killed by viruses only to be replaced by another strain or species (Thingstad, 2000; Thingstad *et al.*, 2015). There is evidence that these KtW dynamics occur in the field as illustrated by a study in a solar saltern where coarsely defined bacterial and viral taxa were relatively constant over time, but showed KtW dynamics at a finer scale (Rodriguez-Brito *et al.*, 2010). However, few environmental studies have shown evidence for these dynamics since few have compared hosts and viruses (Winter *et al.*, 2010).

Examining the temporal dynamics of marine viruses and their hosts has yielded insights about their ecology and evolution, yet little attention has been paid to the phylogenetic relationships within these communities and how they are shaped. An exception is a study by Goldsmith *et al.* (2015), near Bermuda, where the phylogenetic makeup of related groups of viruses over time and depth was found to be highly uneven and variable. There were differences between fall and winter attributable to stratification, with much of the variability due to one phylogenetic group of cyanophages (Goldsmith

et al., 2015). Clasen *et al.* (2013) also found that groups of cyanophages belonging to different phylogenetic clades shifted in their relative dominance over time. Knowing more about the phylogenetic diversity of the viral communities will allow us to better interpret these temporal dynamics.

Phylogenetic relatedness can be correlated to ecological relatedness in plants and animals (Harvey and Purvis, 1991; Srivastava *et al.*, 2012) and microbes have shown phylogenetic patterning (Horner-Devine and Bohannan, 2006; Lennon *et al.*, 2012), yet little is known about these patterns in viral communities. To examine these patterns over time, the following was hypothesized. First, since viruses can be following their hosts (Chow *et al.*, 2014) or being driven by Killing the Winner dynamics (Thingstad, 2000), it was hypothesized that phylogenetic patterns would be detected in the viral communities, as was found in the putative host communities. Second, the structure and composition of viral communities follows a "seed bank" distribution where there are many more rare viral operational taxonomic units (OTUs) than abundant ones (Breitbart and Rohwer, 2005; Goldsmith *et al.*, 2015), therefore, a shuffling in rank of related viral OTUs would be predicted over time, but with a few OTUs dominating the community.

To test these hypotheses, the temporal dynamics of the phylogenetic make-up of two ecologically important groups of marine viruses and their potential hosts were followed in samples taken every two weeks for thirteen months, using amplified marker genes and high-throughput sequencing. With these data, the hypotheses were tested by looking at the community similarities, comparing the phylogenetic diversity over time, and observing the relative abundance of phylogenetically-related groups of OTUs over time in viral and putative host communities. The first group of viruses was the T4-like myoviruses, which are DNA viruses that infect bacteria, including cyanobacteria. A structural gene, gp23, which encodes the capsid was used as an amplification target (Filée *et al.*, 2005). The second was a group of eukaryote-infecting RNA viruses: the marine picorna-like viruses, which were targeted by amplifying the RNA dependent RNA polymerase (RdRp) (Culley and Steward, 2007). These RNA viruses infect ecolog-

ically important phytoplankton such as diatoms belonging to the genera *Rhizoselenia* sp., *Chaetoceros* sp., and the toxic bloom-forming raphidophyte *Heterosigma akashiwo* (Tomaru, 2015). The dynamics of putative hosts were examined by sequencing amplified marker genes for eukaryotes (18S rRNA gene) and bacteria (16S rRNA gene). This contribution opens up new avenues of understanding by showing that temporal changes in the phylogenetic make-up of viruses infecting bacteria and eukaryotic algae are related to seasonal fluctuations in the communities of potential hosts.

3.3 MATERIALS AND METHODS

3.3.1 *Sample collection*

Seawater samples were collected from Jericho Pier (49° 16'36.73N, 123° 12'05.41W) in British Columbia, Canada. Jericho Pier (JP) is adjacent to the shoreline, in a well-mixed location with mixed semi-diurnal tides. In order to get a representative sample of water and enough material for viral extraction 60L of water was pumped from the 1-m depth every two weeks at the daytime high tide between June 2010 and July 2011 (33 samples). An additional set of seven samples was collected every two days from 29 January to 10 February 2011 for more high-resolution analysis of dynamics within these two weeks. Salinity and temperature were measured using a YSI probe (Yellow Springs, Ohio, USA). For all samples, the water was pre-filtered through a 65μm Nitex mesh and filtered sequentially through 142-mm diameter, 1.2-μm nominal pore-size glass-fiber (GC50 Advantec MFS, Dublin, CA., USA) and 0.22 μm pore-size polyvinyldine (Millipore, Bedford, MA, USA) filters. The filtrate, containing the viral size fraction, was concentrated to ~500 mL (viral concentrate) using tangential flow ultrafiltration with a 30kDa MW prep-scale Spiral Wound TFF-6 cartridge (Millipore) (Suttle *et al.*, 1991).

3.3.2 *Nutrients*

Phosphate, silicate and nitrate+nitrite concentrations were determined in duplicate 15-mL seawater samples filtered through 0.45 μm pore-size HA filters (Millipore) and stored at -20°C until air-segmented continuous-flow analysis on a AutoAnalyzer 3 (Bran+Luebbe, Norderstedt, Germany). Chlorophyll *a* (Chl *a*) was determined in triplicate by filtering 100 mL of seawater onto 0.45 μm pore-size HA filters (Millipore), and storing the filters in the dark at -20°C until acetone extraction and then analysed fluorometrically (Parsons *et al.*, 1984).

3.3.3 *Enumeration of bacteria and viruses*

Samples for viral and bacterial abundances were taken at each sampling point by fixing duplicate cryovials containing 980μL of sample with final concentration of 0.5% glutaraldehyde (EM-grade), freezing in liquid nitrogen and storing at -80°C until processing. Flow cytometry samples were processed as in Brussaard (2004b). Briefly, viral samples were diluted 1:10 to 1:10 000 in sterile 0.1 μm filtered 1X TE, stained with SYBR Green I (Invitrogen, Waltham, MA, USA) at a final concentration of 0.5 x 10^{-4} of commercial stock, heated for 10 minutes at 80° C and then cooled in the dark for 5 minutes before processing. Bacterial samples were diluted up to 1:1000 in sterile 0.1 μm filtered 1xTE, stained with SYBR Green I (Invitrogen) at a final concentration of 0.5 x 10^{-4} of commercial stock, and incubated in the dark for 15 minutes before processing. All samples were processed on a FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA) with viral and bacterial samples run for 1 min at a medium or high flow rate, respectively. Event rates were kept between 100 to 1000 events per second and green fluorescence and side scatter detectors were used. Data were processed and gated using Cell-Quest software (Becton-Dickinson).

3.3.4 *Extraction of viral nucleic acids*

The viral concentrate was filtered twice through 0.22 μm pore-size Durapore PVDF filters (Millipore) in a sterile Sterivex filter unit (Millipore). The filtrate, containing viral-sized particles, was pelleted by ultracentrifugation (Beckman-Coulter, Brea, California, USA) in a SW40 rotor at 108 000 *g* for 5 h at 12°C. The pellet was resuspended overnight in 100 μL of supernatant at 4°C. To digest free DNA, the pellets were incubated with 1U μL^{-1} DNAse with a final concentration 5 mM $MgCl_2$ for 3 h at room temperature. Nucleic acids were extracted using a Qiamp Viral Minelute spin kit (Qiagen, Hilden, Germany) according to the manufacturer's directions.

3.3.5 *PCR amplification of T4-like myoviral marker gene*

To target the marine T4-like myoviral capsid protein gene (gp23), PCRs were set up as in Filée *et al.* (2005). Briefly, each reaction mixture (final volume, 50 μL) consisted of 2 μL template DNA, 1x (final concentration) PCR buffer (Invitrogen, Carlsbad, California, USA), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 40 pmol of MZIA1bis and 40pmol of MZIA6, and 1 U Platinum Taq DNA polymerase (Invitrogen) and program conditions as in Table 3.1.

TABLE 3.1: PCR parameters used in this study. (continued below)

Marker gene	Target	Primer names	PCR initial	PCR denaturation	Annealing temperature
gp23	T4-like myovirus	MZIA1bis and MZIA6	94°C for 90 s	94°C for 45 s	50°C
RdRp	Marine picorna-like viruses	MPL-2F and MPL-2R	94°C for 75 s	94°C for 45 s	43°C

TABLE 3.1: continued

Marker gene	Target	Primer names	PCR initial	PCR denaturation	Annealing temperature
18S rRNA gene	Eukaryotes	Euk1209f and Uni1392r	94°C for 75 s	94°C for 1 min	65°C touchdown for 10 cycles followed by 55°C
16S rRNA gene	Bacteria	341F and 907R	94°C for 75 s	94°C for 1 min	64°C, 12cycles followed by 54°C

Extension	Cycles	Final extension	Reference
72°C for 45 s	35	5 min at 72°C	Filée *et al.* (2005)
72°C for 60 s	40	9 min at 72°C	Culley and Steward (2007)
72°C for 60 s	10 + 20	9 min at 72°C	Diez *et al.* (2001)
72°C for 60 s	12+25	10 min at 72°C	Baker *et al.* (2003) and Muyzer *et al.* (1995)

3.3.6 *PCR amplification of picorna-like virus marker gene*

Half of each viral extract was used to synthesize cDNA. To remove DNA, the extracted viral pellets were digested with DNase 1 (amplification grade) (Invitrogen). The reaction was terminated by adding 2.5 mM EDTA (final concentration) and incubating for 10 min at 65°C. Complementary DNA (cDNA) was generated using Superscript III reverse transcriptase (Invitrogen) with random hexamers (50 ng μL^{-1}) as per the manufacturer.

PCR was performed using primer set MPL-2 to target the RdRp of marine picorna-like viruses (Culley and Steward, 2007). Each reaction mixture (final volume, 50 μL) consisted of 50 ng of cDNA, 1x (final concentration) PCR buffer (Invitrogen), 2 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 1 μM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1. Products were run on a 0.5X TBE 1% low melt gel, excised and extracted using Zymoclean Gel DNA Recovery Kit (Zymo) as per the manufacturer and a final elution step of 2x10 μL EB buffer (Qiagen).

3.3.7 *Filtration and extraction of marine bacteria and eukaryotes*

One liter of seawater was taken from the sixty liters and filtered through a 0.22μm pore-size Durapore PVDF 47 mm filter (Millipore) in a sterile Sterivex filter unit (Millipore). The filter was either stored at -20°C until extraction or immediately extracted as follows. Filter extraction was as in Short and Suttle (2003). Briefly, filters were aseptically cut and incubated with lysozyme (Sigma-Aldrich, St. Louis, MO, USA) at a final concentration of 1mg mL^{-1} for 2 h at 37°C. Sodium dodecyl sulfate was added at a final concentration of 0.1% (w/v) and each filter was put through three freeze-thaw cycles. Proteinase K (Qiagen) was then added to a final concentration of 100 μg mL^{-1} and incubated for 1 h at 55°C. DNA was sequentially extracted using equal volumes of phenol:chloroform:IAA (25:24:1), and chloroform:IAA (24:1). DNA was precipitated by adding NaCl to a final concentration of 0.3M and by adding 2X the extract volume of ethanol. Samples were incubated at -20°C for at least 1 h and then centrifuged for 1 h at 20 000 *g* at 4°C. Extracts were washed with 70% ethanol and were resuspended in 50 μL EB buffer (Qiagen).

3.3.8 *PCR amplification of bacterial and eukaryotic ribosomal sequences*

PCR targeting eukaryotes used primers Euk1209f and Uni1392r as in Diez *et al.* (2001). These primers target positions 1423 to 1641 and includes the variable region V8. Each re-

action mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 0.3 μM of each primer, and 2.5 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1.

PCR targeting bacteria used primers 341F (Baker *et al.*, 2003) and 907R (Muyzer *et al.*, 1995). These primers target the v3 to v5 regions. PCRs were run with the following conditions: each reaction mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 0.4 µM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1.

3.3.9 *Sequencing library preparation*

Construction — PCR products not requiring gel excision were purified after PCR using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. Cleaned products were resuspended in 30 μL EB buffer (Qiagen). All products were quantified using the Picogreen dsDNA (Invitrogen) assay using Lambda DNA (Invitrogen) as a standard. Sample concentrations were read using iQ5 (Bio-Rad, Hercules, CA, USA) and CFX96 Touch systems (Bio-Rad). Pooled libraries were constructed using one of each of the amplicons at a concentration so that their molarity would be similar and the total product of the pool to be ~700-900 ng. Pooled amplicons were concentrated using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. NxSeq DNA sample prep kit 2 (Lucigen, Middleton, WI, USA) was used as per manufacturer's directions with either NEXTFlex 48 barcodes (BioO, Austin, USA), NEXTflex 96 HT barcodes (BioO), or TruSeq adapters (IDT, Coralville, Iowa). Libraries were cleaned up using AMPure XP beads (Beckman Coulter) at a ratio of 0.9:1 beads:library.

Quantification and quality control of libraries — Libraries were checked for small fragments (primer dimers and/or adapter dimers) using a 2100 Bionanalyzer (Agilent, Santa Clara, CA, USA) with the High Sensitivity DNA kit (Agilent). The concentration of libraries was quantified using Picogreen dsDNA assay as above. The libraries were quantified and checked for amplifiable adapters using the Library Quantification DNA standards 1-6 (Kappa Biosystems, Wilmington, USA) with the SsoFast EvaGreen qPCR supermix (Bio-Rad) using 10 μL EvaGreen master mix, 3 μL of 0.5 μM F primer, 3 μL of 0.5 μm R primer and 4 μL of 1:1000, 1:5000 and 1:10000 dilutions of the libraries in triplicate on iQ5 (Bio-Rad) and CFX96 Touch qPCR machines. Cycling parameters were as follows: 95°C for 30s, 35 cycles of 95°C for 5s, 60°C for 30s, and the melt curve generation from 65°C to 95°C in 0.5°C steps (10s/step). Quantification from both Picogreen and qPCR assays were used to determine final pooling of all libraries before sequencing. Libraries were sequenced using 2x250bp PE Miseq (Illumina, San Diego, USA) sequencing at Génome Québec Innovation Centre at the McGill University (Montreal, QC, Canada), and 2x300bp PE Miseq (Illumina) sequencing at UBC Pharmaceutical Sciences Sequencing Centre (Vancouver, BC, Canada) and at UCLA's Genoseq (Los Angeles, CA, USA).

3.3.10 *Initial sequence processing*

Libraries were either split by the sequencing centre using CASAVA (Illumina) or split by the user using the Miseq Reporter software (Illumina). Sequence quality was initially examined using FastQC (Andrews, 2015). Contaminating sequencing adapters were removed using Trimmomatic version 0.32 (Bolger *et al.*, 2014) and the quality of the sequencing library further examined using fastx_quality (Gordon, 2014). Libraries were further split into individual amplicons (i.e. 18S, 16S, gp23 and MPL) and then, if the expected overlap of the paired-end reads was 40bp or more, the paired reads were merged using PEAR (Zhang *et al.*, 2014). Sequences were then quality trimmed using Trimmomatic with the default quality settings. Sequences were aligned to known sequences

(Silva 119 database (Quast *et al.*, 2013) for 16S and 18S rRNA genes) using align.seqs in mothur 1.33.3 (Schloss *et al.*, 2009) and those not aligned were removed. Viral sequences were queried using BLAST against databases containing the gene markers of interest and sequences with an e-value below 10^{-3} were kept.

3.3.11 *Chimera checking, OTU picking and read normalization*

The 16S and 18S rRNA gene sequences were checked for chimeras using USEARCH version 8.0.1517 reference (Edgar, 2010) with the Gold reference database. Unique, non-chimeric sequences were clustered at 97% similarity. Taxonomy for the 16S and 18S rRNA gene sequences was assigned using mothur (Wang-type algorithm) and the taxonomy in Silva 119 (Quast *et al.*, 2013). For the viral targets sequences were chimera-checked using USEARCH denovo and reference (Edgar, 2010). Viral sequences were then translated using FragGeneScan 1.20 (Rho *et al.*, 2010). Viral reads were clustered using USEARCH (Edgar, 2010) at 95% similarity for MPL, and 95% similarity for T4-like myoviruses. Operational taxonomic unit (OTU) tables for all targets were constructed using USEARCH (Edgar, 2010). Rarefaction curves were generated using vegan (Oksanen *et al.*, 2015). Sequences were normalized for this project by date and by target using vegan (Oksanen *et al.*, 2015).

3.3.12 *Data analysis and multivariate statistics*

Environmental data — Environmental parameters were mean imputed to fill in data missing because of instrument malfunction or unavailability. Day length data were retrieved using R package geosphere (Hijmans, 2015).

Community similarity and Mantel test — Spearman rank correlations using rcorr as part of Hmisc (Harrell, 2014). Adonis was used from vegan (Oksanen *et al.*, 2015) to test whether community matrices showed seasonal differences. Bray-Curtis distance matrices were constructed from the normalized OTU abundance tables. Mantel tests

were performed by comparing the community distance matrices to each other and to distance matrices of environmental parameters using vegan.

3.3.13 *Phylogeny*

NCBI CDD domain alignments for RdRp and for gp23 were retrieved and used as hidden markov models via HMMER (Johnson *et al.*, 2010) to align translated OTUs with Clustal Omega (Sievers *et al.*, 2014). Environmental sequences for both the gp23 (Filée *et al.*, 2005; EF617478 Sandaa and Kristiansen, 2007; Jia *et al.*, 2007; López-Bueno *et al.*, 2009; ACU57502-AC57509 Mabizela and Litthauer, 2009; Comeau *et al.*, 2010; Chow and Fuhrman, 2012; Bellas and Anesio, 2013; Butina *et al.*, 2013) and the RdRP (Culley *et al.*, 2003; Culley and Steward, 2007) were retrieved from Genbank to give context to the OTUs.

Alignments were checked and manually curated with aliview (Larsson, 2014). Automated trimming of the alignment was done using Trimal (Capella-Gutierrez *et al.*, 2009). Model VT was chosen for the RdRP gene and model JTT for the gp23 gene using Prottest (Darriba *et al.*, 2011). Initial phylogenetic trees were built with Fast Tree (Price *et al.*, 2010) and examined in FigTree (Rambaut, 2014). Final maximum likelihood trees were generated using RAxML (Stamatakis, 2014) with 1000 bootstraps, with VT with the PROTGAMMA model for the RdRp gene tree and JTT with the PROTGAMMA model for the gp23 gene tree on the CIPRES webserver (Miller *et al.*, 2010). Faith's phylogenetic diversity (Faith, 1992) was calculated as implemented in picante (Kembel *et al.*, 2010). The package ggtree (Yu *et al.*, 2016) was used for visualizing and annotating trees and ggplot2 (Wickham, 2009) was used for all other plots made in R (R Core Team, 2014). All scripts used for processing the data are available on github (Gustavsen, 2016).

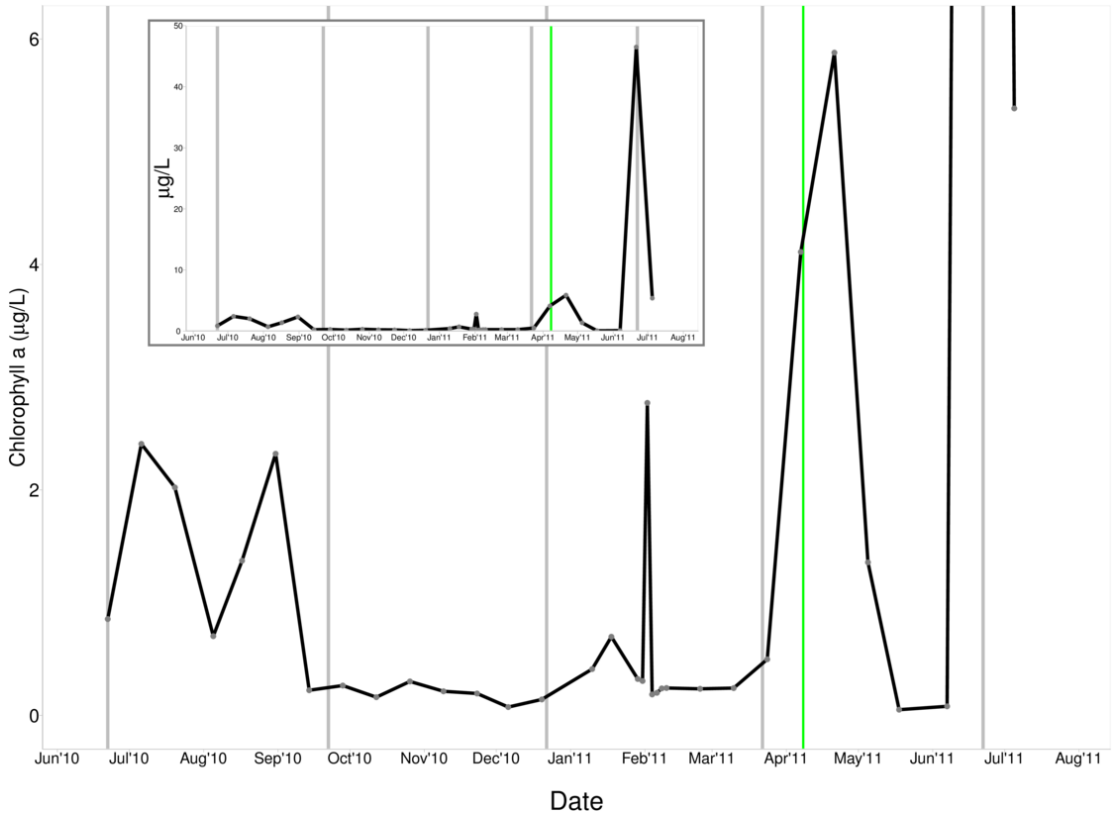

FIGURE 3.1: Chlorophyll *a* concentration over time at Jericho Pier, Vancouver, British Columbia. Inset: Chlorophyll *a* concentration from 0-50 ug/L.

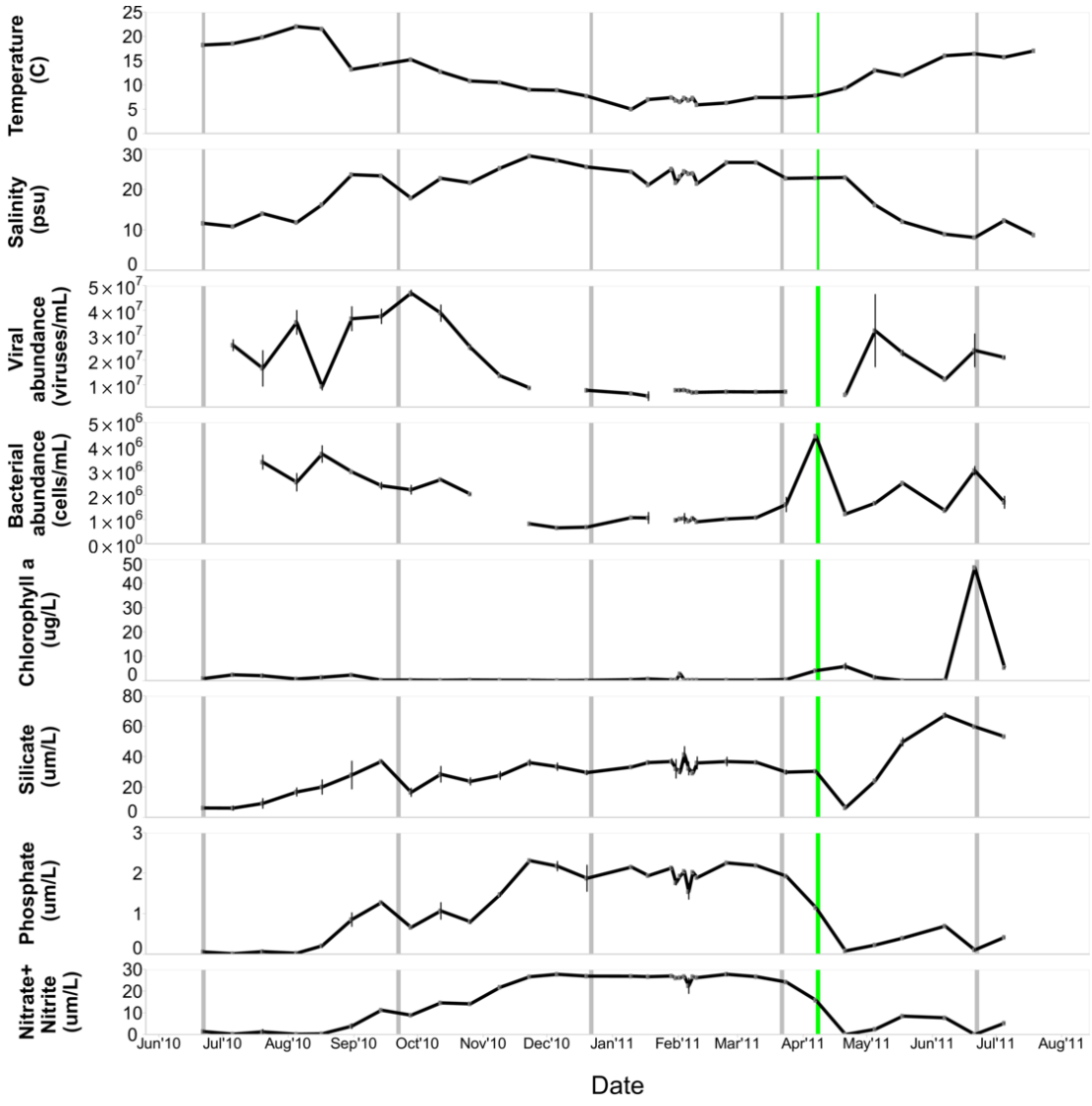

FIGURE 3.2: Environmental parameters during 1-year time series at Jericho Pier. Viral and bacterial abundance were measured by flow cytometry. Grey vertical lines indicate season boundary and green vertical line indicates time of spring bloom. Error bars are standard error based on duplicates.

3.4 RESULTS

3.4.1 *Variability of environmental characteristics*

Chlorophyll *a* (chl *a*) concentrations varied over time with a maximum observed concentration during a eukaryotic phytoplankton bloom (46.5 μg L^{-1} in June 2011)(Figure 3.1). The second highest chl *a* occurred during the annual spring bloom in late April 2011 (5.88 μg L^{-1}) which is mainly composed of diatoms belonging to *Thallassiosira* sp. (Harrison *et al.*, 1983; Allen and Wolfe, 2013). The minimum chl *a* value of 0.05 μg L^{-1} occurred in May and the chlorophyll levels remained below 1 μg L^{-1} from September to March.

Nutrient concentrations were also highly dynamic ranging between 6.1 μM to 67.3 μM for silicate, from below 0.1μM to 2.3 μM for phosphate and from below 0.1μM to 27.7 μM for nitrate+nitrite (Figure 3.2). Overall, nutrient concentrations were high and stable over winter, dipped in late April and then were followed by a large increase in silicate commencing in May.

The viral abundance ranged from 5.41×10^6 to 4.695×10^7 particles mL^{-1} while the bacterial abundance was the expected one order of magnitude lower of 6.59×10^5 to 4.43×10^6 cells mL^{-1} (Figure 3.2).

3.4.2 *Richness and shared microbial and viral OTUs over time*

For each amplicon target representing T4-like myoviruses (gp23), picorna-like viruses (RdRp) bacteria (16S rRNA gene) and eukaryotes (18S rRNA gene), the sequences were translated into amino acids (except 16S and 18S) and normalized to the library with fewest reads. This resulted in 1737 OTUs from T4-like myoviruses at 95% sequence similarity, with a maximum of 495 and a minimum of 149 per timepoint. On average 6% of the T4-like myoviruses OTUs were shared among times. For the marine picorna-like viruses there were 574 OTUs at 95% sequence similarity, with between 60 to 149 OTUs per timepoint. On average 6% of these OTUs were shared. There were 802 bac-

terial OTUs (97% sequence similarity) with an average of 10% shared over time. The highest number of OTUs seen per time point was 270 and the minimum was 82. In the eukaryotic community a total of 1117 OTUs (97% sequence similarity) were found with 6% shared on average, a maximum of 297 and a minimum of 62.

Rarefaction curves for individual samples did not flatten ("saturate") indicating that not all possible OTUs were sequenced in these samples, but when considered together the curves saturated indicating that even if all the diversity was not captured in one sample, the overall community of OTUs was captured (Figure 3.3).

3.4.3 *Community similarity, phylogenetic diversity, and richness over time*

Species richness (SR) and phylogenetic diversity (PD) were stable for the bacteria (Figure 3.4C) and eukaryotes (Figure 3.4D) from the fall to winter months, although there were marked changes in the similarity and richness of the bacterial community during February (Figure 3.4C) even though environmental conditions were relatively stable (Figure 3.2). After this change, bacterial diversity began decreasing through to July. For eukaryotes, the SR and PD decreased after December until February and then climbed again until mid-March and then decreased until mid-May. The marine picorna-like viruses generally had congruent patterns in PD and SR, but during the spring bloom (*Thalassiosira* sp.) the SR and PD diverged and the PD was among the highest values observed. February was the time of highest SR and PD for the T4-like myoviruses, but was followed by the lowest values. In contrast to the bacteria, in which richness decreased after the spring bloom, the richness of the T4-like myoviruses increased. Likewise, the spring and summer T4-like myovirus similarity lagged behind that of the bacteria.

3.4.4 *Dynamics of phylogenetically-related viral OTUs*

To understand their dynamics viral OTUs were placed into a phylogenetic context (Figure 3.5). Well-defined and well-supported phylogenetic groups (A-H) of marine picorna-like viruses (Figure 3.5A and B) showed strong temporal dynamics and differed by season

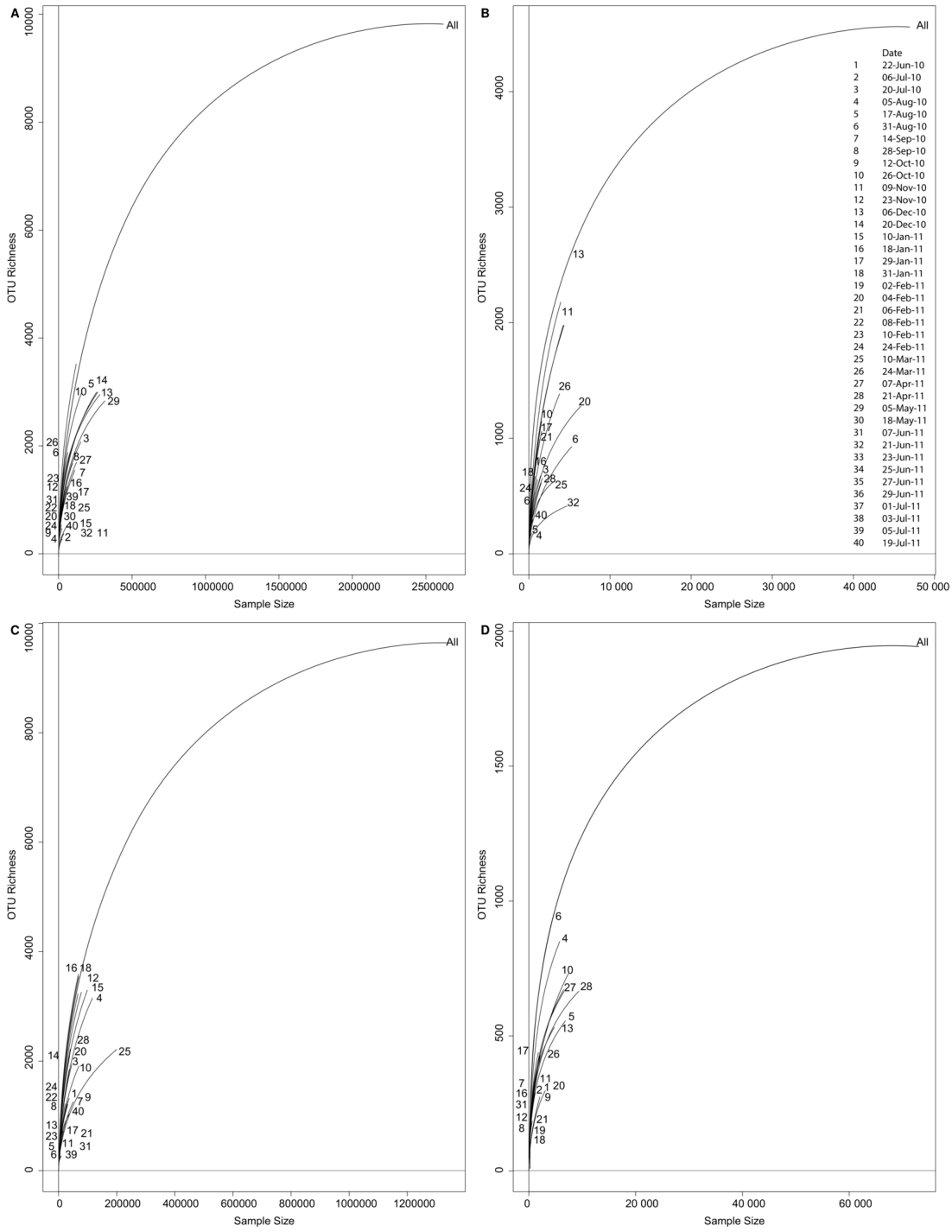

FIGURE 3.3: Rarefaction curves of samples from Jericho Pier time series. A) Rarefaction curve 18S rRNA gene, B) Rarefaction curve 16S rRNA gene, C) Rarefaction curve T4-like myovirus(gp23), D) Rarefaction curve Marine picorna-like(RdRp).

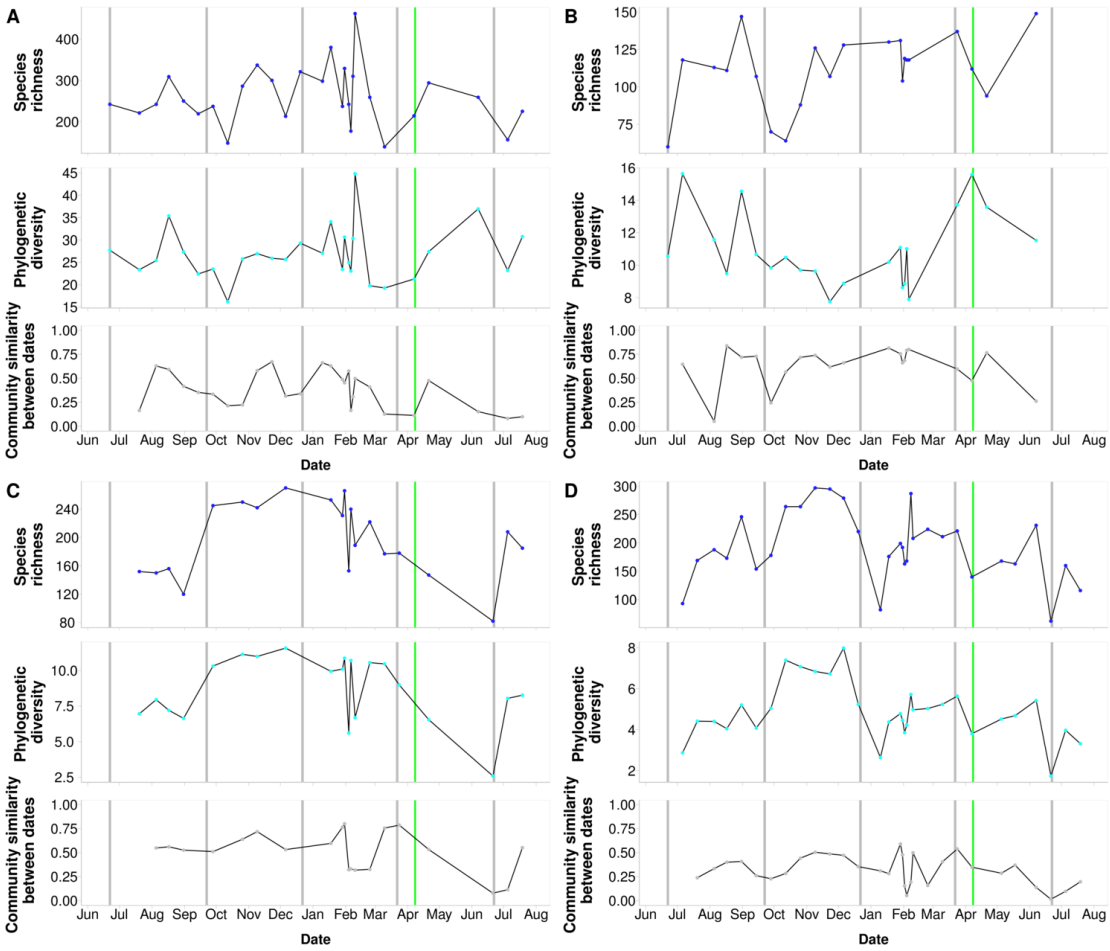

FIGURE 3.4: Species richness, phylogenetic diversity, and community similarity over time. A) T4-like myoviruses (gp23), B) Marine picornalike-viruses (RdRp), C) bacteria (16S rRNA gene), and D) eukaryotes (18S rRNA gene). Faith's phylogenetic diversity (Faith, 1992) was calculated as implemented in picante (Kembel *et al.*, 2010). Community similarity is Bray-Curtis calculated on normalized data between sequential times. Grey vertical lines indicate season boundary and green vertical line indicates time of spring bloom.

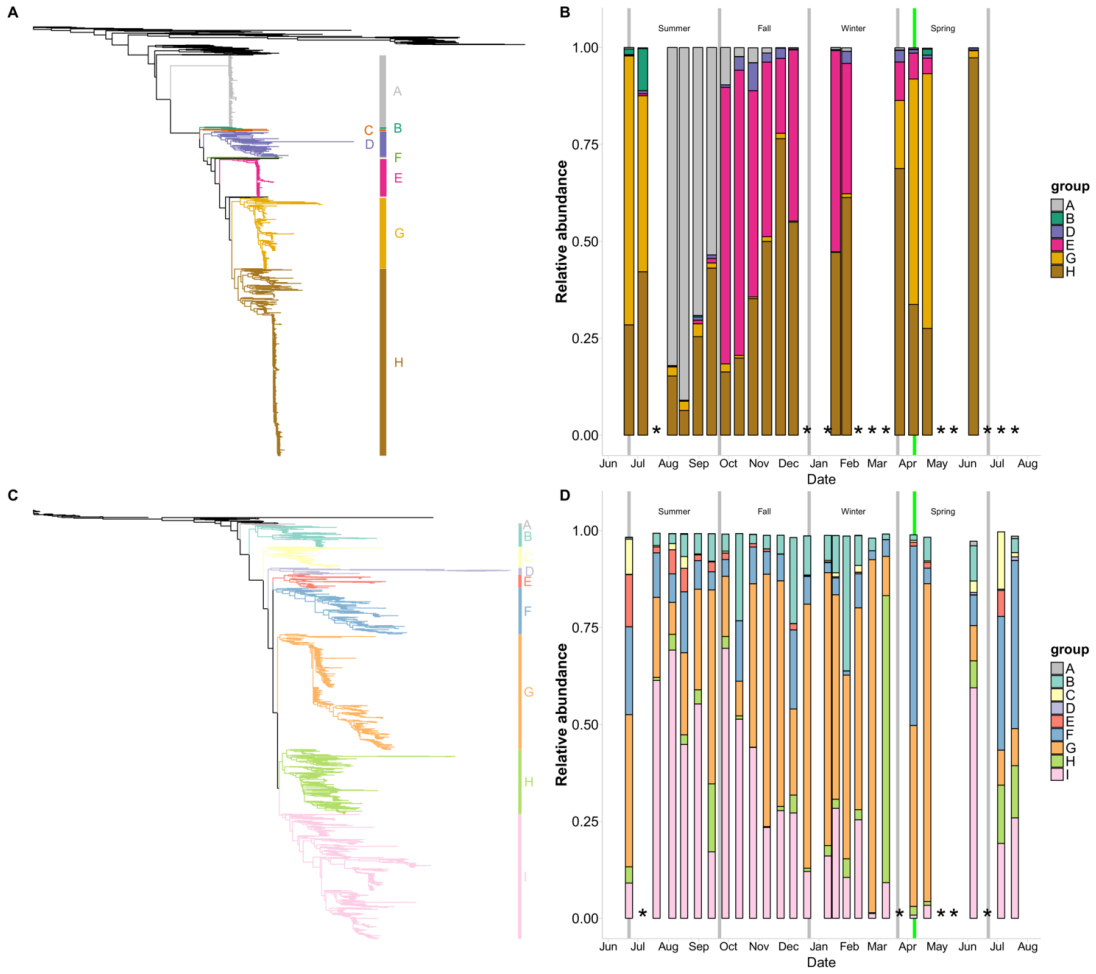

FIGURE 3.5: Maximum likelihood RAxML phylogenetic trees and barplots of closely-related phylogenetic groups of OTUs. A) Tree of marine picorna-like virus RdRp sequences including reference sequences and OTUs generated in this study. Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). B) Barplot of the relative abundances of marine picorna-like virus phylogenetic groups over time. C) Tree of T4-like myovirus major capsid protein sequences including reference sequences and OTUs generated in this study. Outgroup is Enterobacteria phage T4. D) Barplot of the relative abundances of T4-like myovirus phylogenetic groups from over time. Grey vertical lines indicate season boundary and green vertical line indicates time of spring bloom.

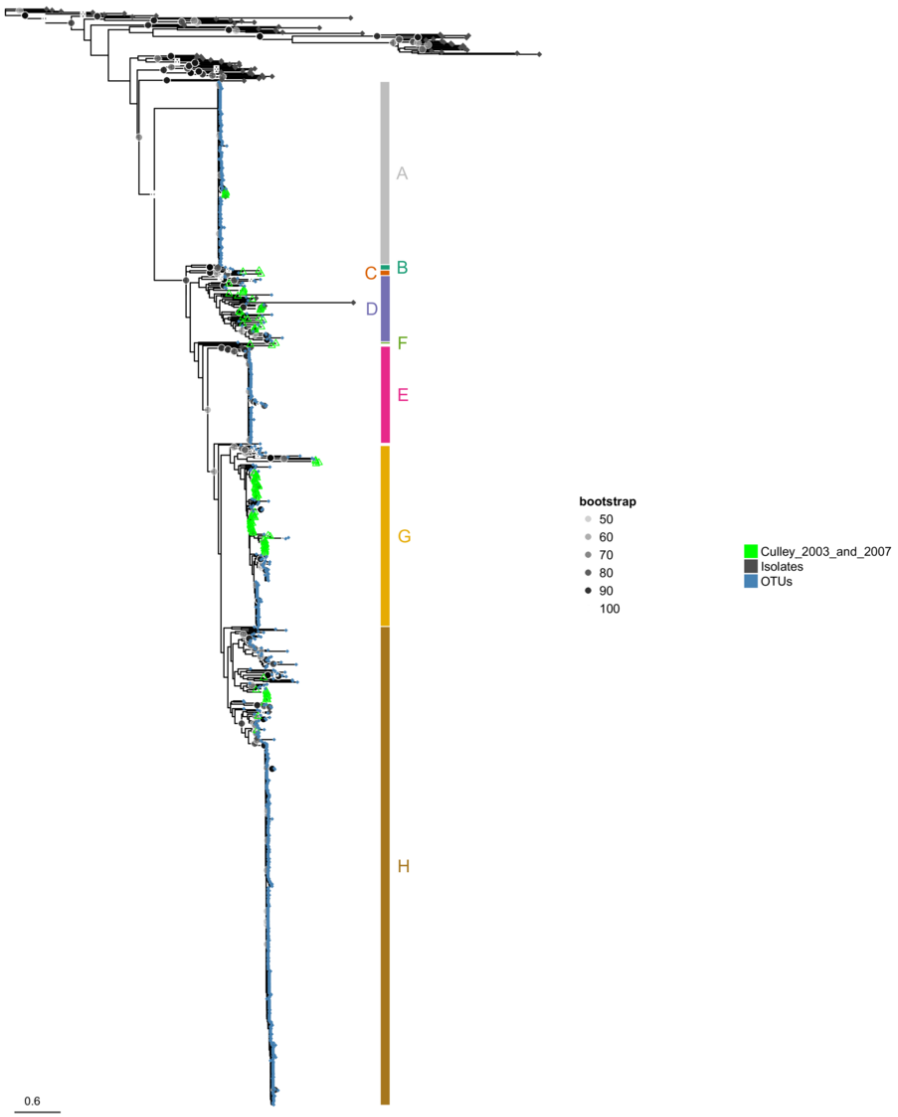

FIGURE 3.6: Maximum likelihood phylogenetic tree (RAxML) of marine picorna-like viruses including reference sequences and OTUs. Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). OTUs at 95% similarity at the amino-acid level. Detailed sub-trees with tip labels are availabe in Appendix B.

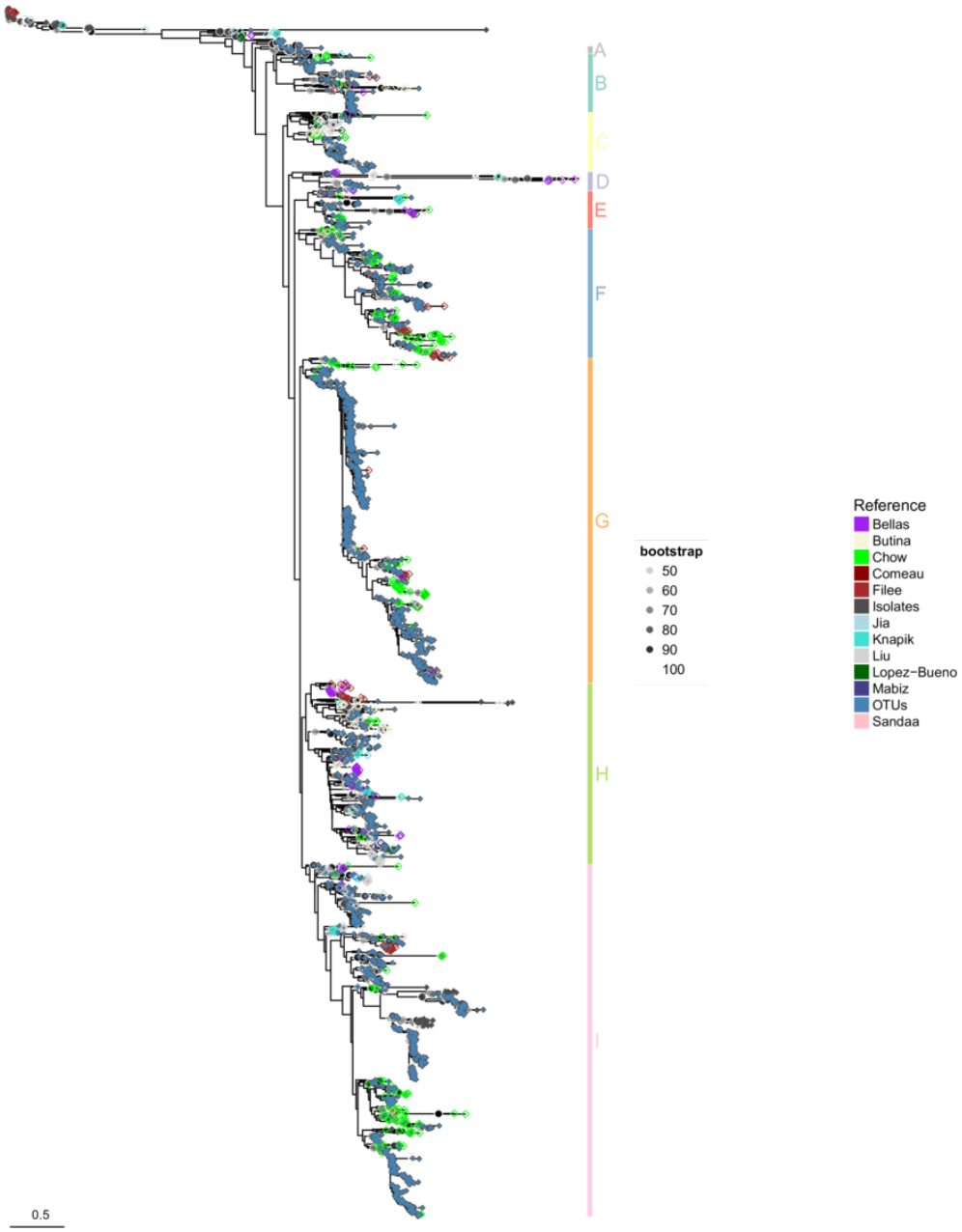

FIGURE 3.7: Maximum likelihood phylogenetic tree (RAxML) of T4-like myoviruses including reference sequences and OTUs. Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level. Detailed sub-trees with tip labels are availabe in Appendix B.

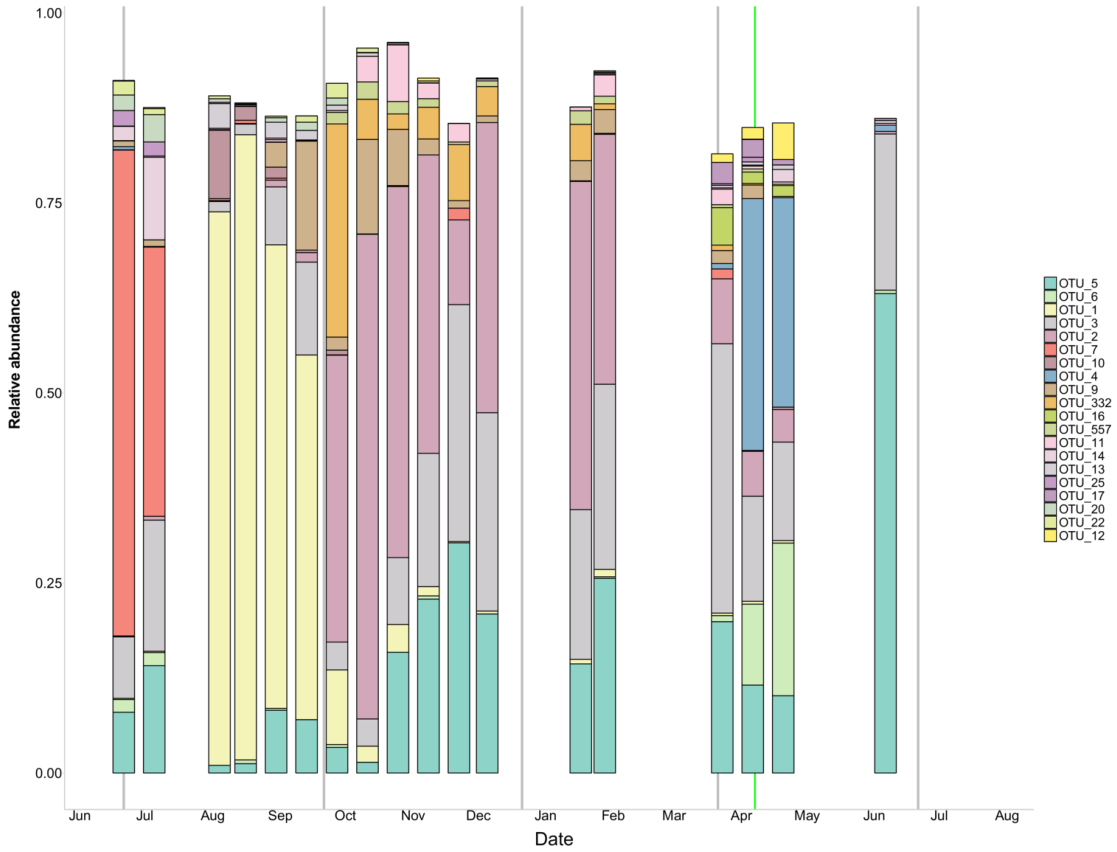

FIGURE 3.8: Barplot of top 20 most relatively abundant marine picorna-like virus OTUs. Relative abundance is the proportion of the community after overall normalization by site. Grey vertical lines indicate boundaries between seasons and the green vertical line indicates the time of the spring bloom.

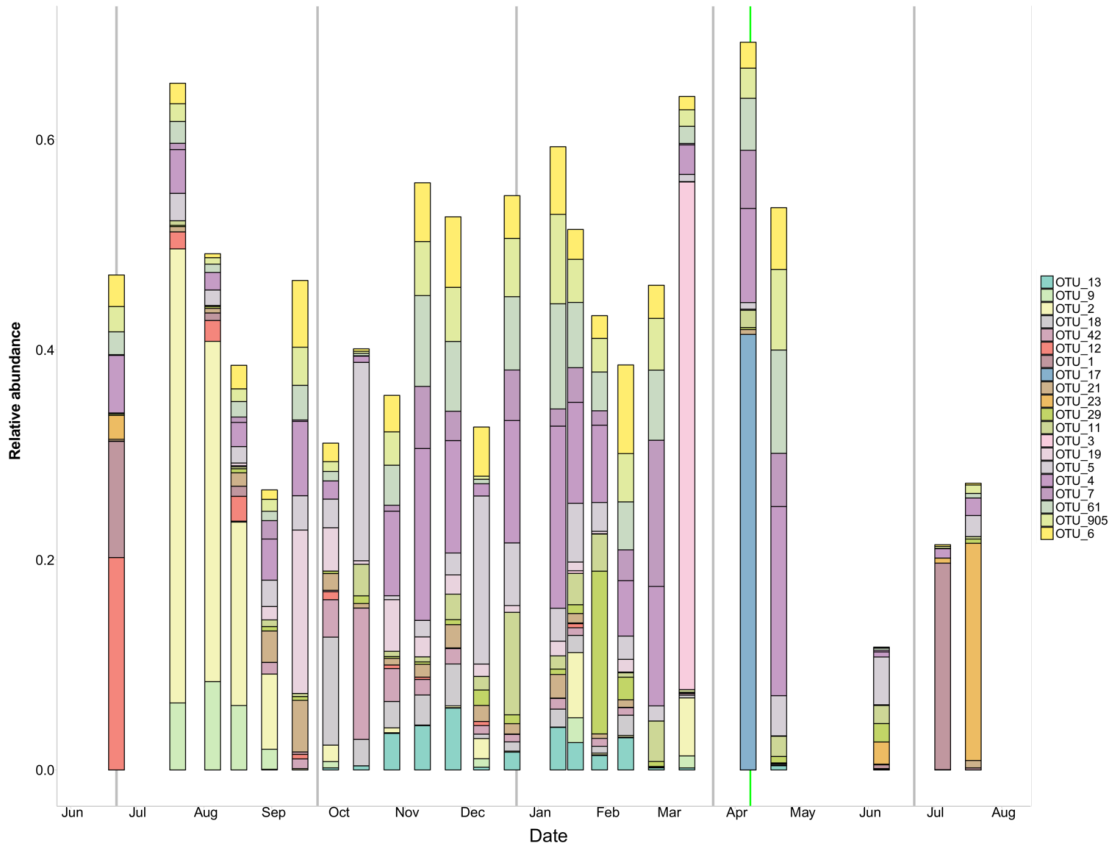

FIGURE 3.9: Barplot of the top 20 most relatively abundant T4-like myovirus OTUs. Relative abundance is the proportion of community after overall normalization by site. Grey vertical lines indicate boundaries between seasons and the green vertical line indicates the time of the spring bloom.

(Adonis R= 0.506, p-value 0.001). Group H, which includes viruses that infect the diatoms *Chaetoceros* sp., and *Rhizoselenia* sp., were constant members of the communities although their relative abundance was highest in late November. Group A was always present and includes many environmental sequences, as well as a virus that infects the raphidophyte, *Heterosigma akashiwo*; the group was most abundant between August and September. The months from October to February were dominated by OTUs in group E, with a smaller contribution by group H. The structure of the MPL OTU phylogenetic groups closely mirrored the structure of the top 20 OTUs found over time (Figure 3.8), demonstrating that this community contained few dominant OTUs.

The T4-like myoviruses OTUs were also placed in a phylogenetic context and categorized into groups of related OTUs (Figure 3.5 C). In the fall, group I dominated the community, followed by group G; both groups include viral isolates infecting cyanobacteria. In January almost half of the relative abundance of the T4-like myoviruses was represented by group B which contains no known isolates. Unlike the marine picorna-like viral community, the T4-like myoviral community had very different patterns among the top 20 OTUs and the phylogenetic groups over time (Figure 3.9). When there was a large increase in nutrients in late September (Figure 3.2), there were shifts in the dominant groups in the T4-like myoviral community. The community returned to its previous state by the next sampling time. The communities showed small differences by season (Adonis R= 0.231, p-value 0.001)

3.4.5 *Taxonomic richness of bacteria over time*

At the phylum level the bacterial communities were relatively stable over time (Figure 3.10). The communities were dominated by Proteobacteria and at times had large proportions of Bacteroidetes. The phylum Bacteroidetes was mostly dominated by members of the Flavobacteria with Cytophaga and Bacteroidia in the fall, and Sphingobacteriia in the winter. The phylum Actinobacteria was present and included stable populations of Acidomicrobia and Actinobacteria, which were mostly absent in March. After June the

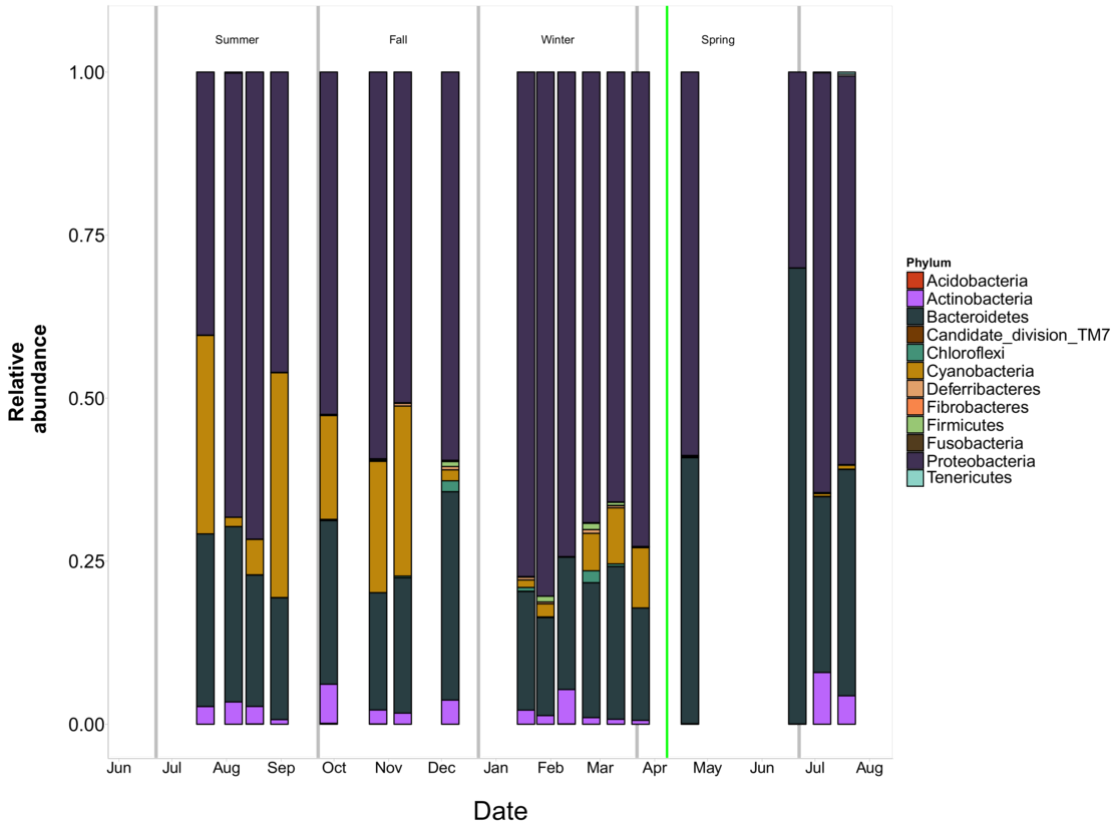

FIGURE 3.10: Bacterial OTUs by phylum over time. Classifications were done using the Wang algorithm as implemented in mothur (Schloss *et al.*, 2009) and using the Silva 119 database (Quast *et al.*, 2013). Grey vertical lines indicate boundaries between seasons and the green vertical line indicates the time of the spring bloom.

Actinobacteria make up a larger proportion of the community. The bacterial community showed small differences by season (Adonis R: 0.361, p-value 0.001).

3.4.6 *Taxonomic richness of eukaryotes over time*

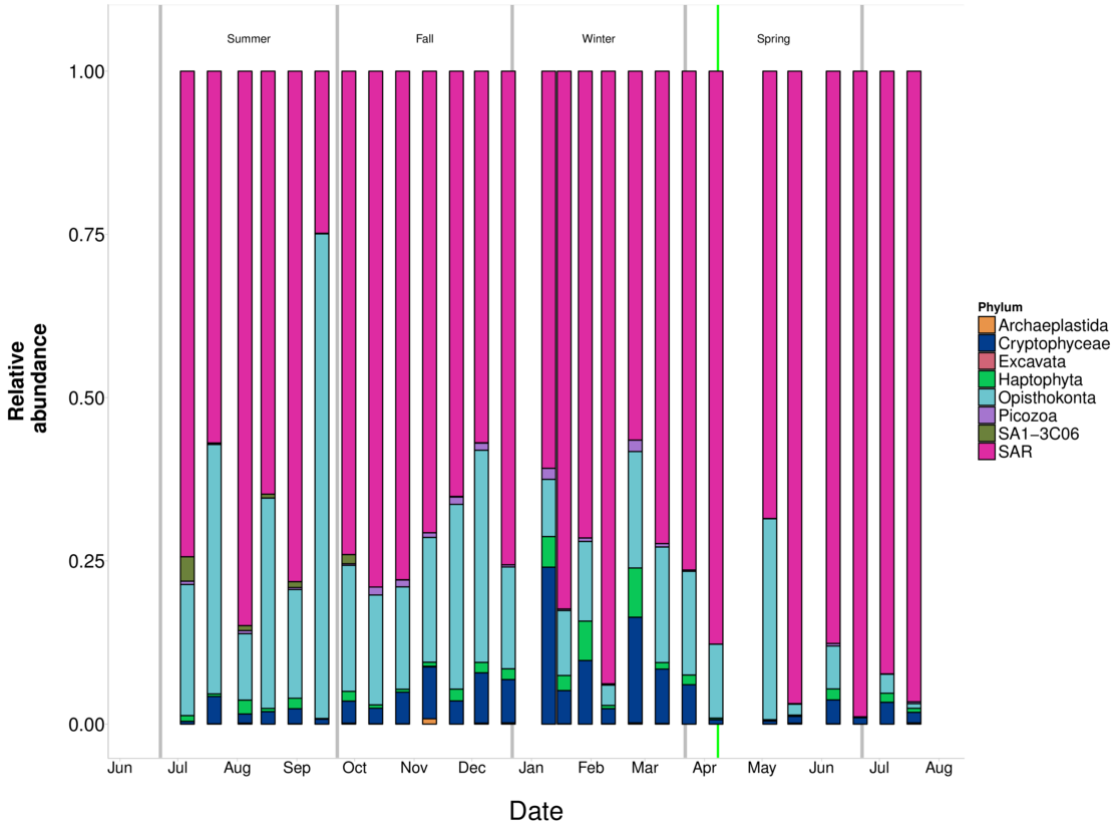

FIGURE 3.11: Eukaryotic OTUs by phylum over time. Classifications were done using the Wang algorithm as implemented in mothur (Schloss *et al.*, 2009) and using the Silva 119 database (Quast *et al.*, 2013). Grey vertical lines indicate boundaries between seasons and the green vertical line indicates the time of the spring bloom.

A large portion of the eukaryotic community was made up of Opistokonts, and also members of the SAR supergroup (Figure 3.11). Haptophytes were also important members of the community from December to January and March to April. Cryptophytes were also relatively abundant in late December to early January and March. At the OTU level the community differed by season (Adonis R=0.23, p-value 0.001).

3.4.7 *Heatmaps of persistent vs. ephemeral OTUs*

FIGURE 3.12: Plot of relative abundance of T4-like myoviral OTUs (95% amino acid similarity) that were present in over 90% of samples (persistent) or in less than 20% of samples (ephemeral). OTUs are ordered by phylogenetic tree. Stars (*) indicate missing samples. A heatmap of all OTU relative abundances are available in Appendix A.

There were persistent OTUs in the T4-like myoviral community, but not in all of the phylogenetic groups (Figure 3.12). Some persistent OTUs remained even when phylogenetically related OTUs were undetectable. In the marine picorna-like viruses, constant OTUs were often phylogenetically similar to ephemeral OTUs (Figure 3.13). There were also clear changes in OTU composition from fall to winter, as was seen for the marine picorna-like viruses (Figure 3.5 B), however some OTUs persisted. Nonetheless, rather than a gradual expansion of the OTUs, changes tended to occur quickly among groups of related viruses.

FIGURE 3.13: Plot of relative abundance of marine picorna-like OTUs (95% amino acid similarity) that were present in over 90% of samples (persistent) or in less than 20% of samples (ephemeral). OTUs are ordered by phylogenetic tree. Stars (*) indicate missing samples. A heatmap of all OTU relative abundances are available in Appendix A.

3.4.8 *Lagged correlations with hosts over time*

RAPHIDOPHYTES AND MARINE PICORNA-LIKE VIRAL GROUP A — The relative abundance of group A (Figure 3.5A), which includes HaRNAV, a virus which infects the raphidophyte *Heterosigma akashiwo*, increased in relative abundance after the increase in relative abundance of eukaryotic sequences classified as raphidophytes (Figure 3.14, correlation: 0.43 and *p* value 0.06). There were further peaks in raphidophytes, but they did not coincide with an increase in the relative abundance of marine picorna-like viral group A.

Of all the OTUs in group A, OTU 1 was the most relatively abundant and differences in other sequences to it are highlighted (Figure 3.15), there were changes in an amino acid from D to E in the palm region of the RdRp (te Velthuis, 2014).

CYANOBACTERIA AND T4-LIKE MYOVIRAL GROUP I — Comparing the T4-like myoviral group I, which contains cyanophage isolates, to the cyanobacterial OTUs (Figure 3.16), showed that the relative abundance of viruses increased in the fall after peaks in the cyanobacterial OTUs (correlation: 0.5, *p* value: 0.03). The lags in relative abundances of putative cyanophages relative to cyanobacteria continued, and after the spring bloom there was a lag before there was an increase in the relative abundance of a different putative cyanophages in group I, showing succession in the viral community.

TABLE 3.3: Spearman correlations among environmental parameters, community richness, and community similarity. Includes correlations that were lagged over time: each time point was compared to the previous timepoint (~2 weeks earlier). * is $p < 0.05$, SR is Species richness. (continued below)

	SR bacteria (16S)	SR eukaryotes (18S)
Bray-curtis bacteria (16S)	-	-0.38

TABLE 3.3: continued

	SR bacteria (16S)	SR eukaryotes (18S)
Bray-curtis bacteria (16S) lagged	-	-0.34
Bray-curtis eukaryotes(18S)	-0.37	-
Bray-curtis eukaryotes (18S) lagged	-0.51	-
Bray-curtis marine picorna-like (RdRp)	0.08	-0.09
Bray-curtis marine picorna-like (RdRp) lagged	-0.45	-0.15
Bray-curtis T4-like myoviruses (gp23)	0.13	0.08
Bray-curtis T4-like myoviruses (gp23) lagged	-0.08	-0.06

	SR marine picorna-like (RdRp)	SR T4-like myoviruses (gp23)
Bray-curtis bacteria (16S)	-	-0.18
Bray-curtis bacteria (16S) lagged	-	-0.07
Bray-curtis eukaryotes(18S)	-0.07	-0.25
Bray-curtis eukaryotes (18S) lagged	0.33	-0.36

	SR marine picorna-like (RdRp)	SR T4-like myoviruses (gp23)
Bray-curtis marine picorna-like (RdRp)	-	-
Bray-curtis marine picorna-like (RdRp) lagged	-	-
Bray-curtis T4-like myoviruses (gp23)	-	-0.28
Bray-curtis T4-like myoviruses (gp23) lagged	-	0.13

	Viral abundance	Bacterial abundance
Bray-curtis bacteria (16S)	-	-0.24
Bray-curtis bacteria (16S) lagged	-0.02	-0.11
Bray-curtis eukaryotes(18S)	-0.19	-0.17
Bray-curtis eukaryotes (18S) lagged	-0.31	-0.11
Bray-curtis marine picorna-like (RdRp)	-0.19	-0.15
Bray-curtis marine picorna-like (RdRp) lagged	-0.22	0.08
Bray-curtis T4-like myoviruses (gp23)	0.16	-0.06

	Viral abundance	Bacterial abundance
Bray-curtis T4-like myoviruses (gp23) lagged	0.12	-0.14

	Chlorophyll *a*	PO4	SiO2	NO3+ NO2
Bray-curtis bacteria (16S)	-0.37 *	0.18	0.18	0.18
Bray-curtis bacteria (16S) lagged	0	0.02	0.06	0.05
Bray-curtis eukaryotes(18S)	-0.19	0.2	0.35 *	0.2
Bray-curtis eukaryotes (18S) lagged	0.17	0.24	0.18	0.16
Bray-curtis marine picorna-like (RdRp)	-0.13	0.2	0.31	0.07
Bray-curtis marine picorna-like (RdRp) lagged	0.01	0.24	0.17	0.26
Bray-curtis T4-like myoviruses (gp23)	0.12	-	-0.07	0.02
Bray-curtis T4-like myoviruses (gp23) lagged	0.07	-0.01	-0.18	-

	Temperature	Salinity	Dissolved oxygen (%)
Bray-curtis bacteria (16S)	0.07	-0.02	-0.2

	Temperature	Salinity	Dissolved oxygen (%)
Bray-curtis bacteria (16S) lagged	-0.05	0.04	0.02
Bray-curtis eukaryotes(18S)	-0.16	0.17	-0.1
Bray-curtis eukaryotes (18S) lagged	-0.26	0.12	-0.06
Bray-curtis marine picorna-like (RdRp)	-0.2	0.06	-0.06
Bray-curtis marine picorna-like (RdRp) lagged	-0.32	0.1	-0.03
Bray-curtis T4-like myoviruses (gp23)	-0.15	0.09	-0.05
Bray-curtis T4-like myoviruses (gp23) lagged	0.01	0.1	-

	pH
Bray-curtis bacteria (16S)	-0.11
Bray-curtis bacteria (16S) lagged	-0.05
Bray-curtis eukaryotes(18S)	-0.1
Bray-curtis eukaryotes (18S) lagged	-0.07
Bray-curtis marine picorna-like (RdRp)	-0.11
Bray-curtis marine picorna-like (RdRp) lagged	-0.35

	pH
Bray-curtis T4-like myoviruses (gp23)	-0.02
Bray-curtis T4-like myoviruses (gp23) lagged	0.01

CORRELATIONS BETWEEN COMMUNITY SIMILARITY AND RICHNESS — The community similarity of the T4-like myoviruses had a strong lagged negative correlation to the richness of the bacterial community (Table 3.3 and Table 3.9). The correlations to the MPL community were much stronger to other communities when lagged than when directly compared. Viral abundance was negatively correlated to bacterial community similarity.

TABLE 3.9: Mantel tests among community similarity matrices and distance matrices of environmental data. * is $p < 0.05$, ** $p < 0.01$

	Bacteria (16S rRNA gene)	Eukaryotes (18S rRNA gene)	Marine picorna-like (RdRP)	T4-like myoviruses (gp23)
Viral abundance	-0.1	-0.14	0.17	-0.15
Bacterial abundance	0.02	-0.11	0.19	0.08
Chlorophyll *a*	-0.1	-0.19	-0.1	0.28
PO4	-0.03	-0.04	0.07	0.06
SiO2	0	-0.03	0.04	0.09
NO3+ NO2	-0.05	-0.06	0.02	0.03
Temperature	0.02	-0.03	0.15	-0.13
Salinity	-0.01	-0.08	0.17	0.03
Dissolved oxygen (%)	-0.1	-0.14	0	0.02

TABLE 3.9: continued

	Bacteria (16S rRNA gene)	Eukaryotes (18S rRNA gene)	Marine picorna-like (RdRP)	T4-like myoviruses (gp23)
pH	-0.18	0.04	-0.06	-0.05
Bacteria (16S rRNA gene)	-	0.81 **	0.52 **	0.18 *
Bacteria (16S rRNA gene) lagged	-	0.4 **	0.34 *	1 **
Eukaryotes (18S rRNA gene)	0.81 **	-	0.34 **	0.3 **
Eukaryotes (18S rRNA gene) lagged	0.91 **	-	1 **	0.94 **
T4-like myoviruses (gp23)	0.18 *	0.3 **	0.47 **	-
T4-like myoviruses (gp23) lagged	0.01	0.37 **	0.5 **	-
Marine picorna-like (RdRP)	0.52 **	0.34 **	-	0.47 **
Marine picorna-like (RdRp) lagged	1 **	0.37 **	-	0.97 **

MANTEL TESTS AMONG COMMUNITY SIMILARITY — Mantel tests examined concurrent community changes in distance matrices over time (Table 3.9). The bacterial and eukaryotic community compositions fluctuated strongly together. The marine picornalike community fluctuated with viral abundance and with salinity, more strongly with the

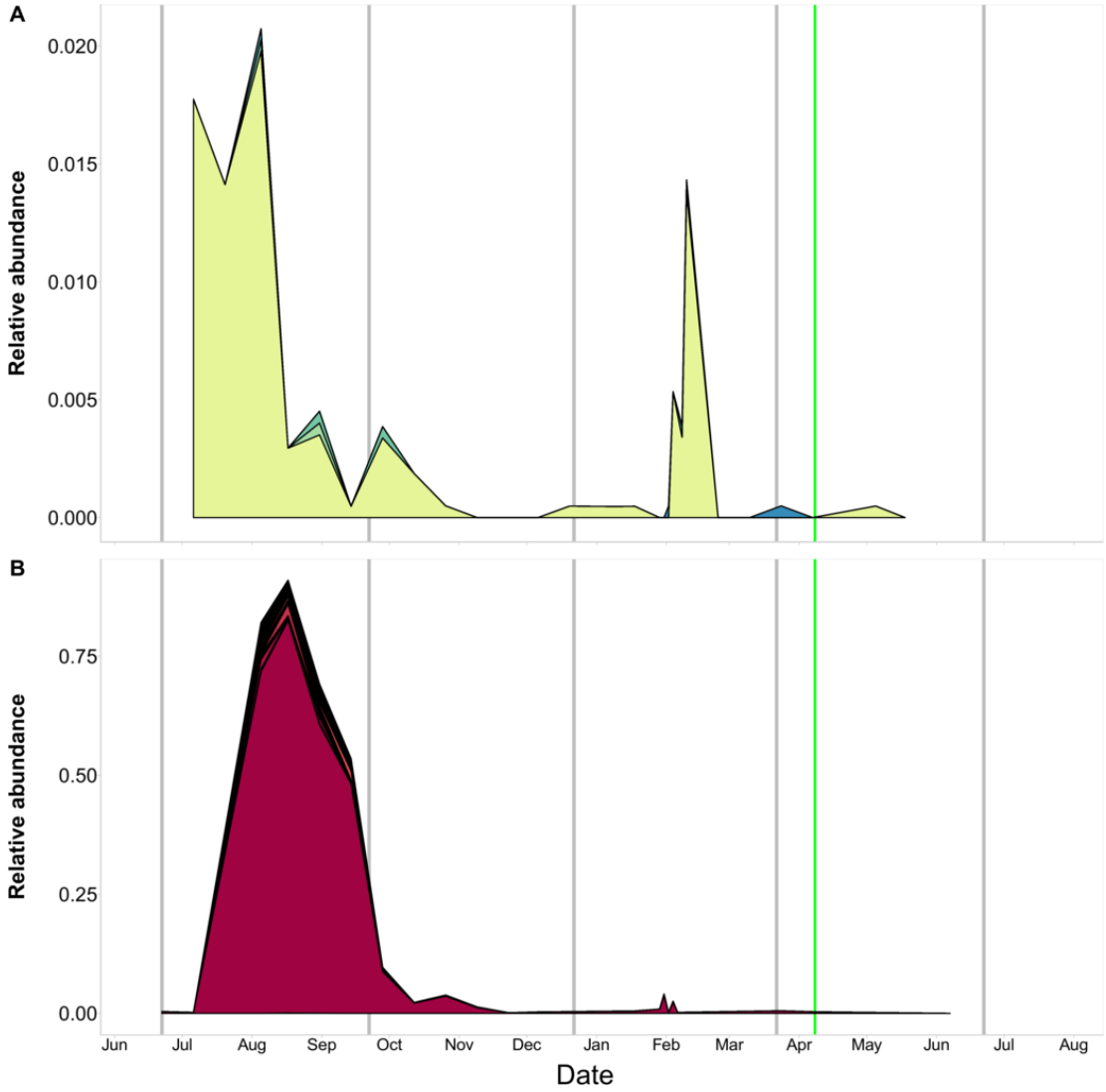

FIGURE 3.14: Marine picorna-like viral group A compared to eukaryotic OTUs classified as raphidophytes over time. A) Relative abundance of eukaryotic OTUs (97%) classified as Raphidophyte (Stramenopiles) over time. B) Relative abundance of marine picorna-like virus group A OTUs (95% amino acid) over time. Each coloured contour represents a separate OTU. Grey vertical lines indicate boundaries between seasons and the green vertical line indicates the time of the spring bloom.

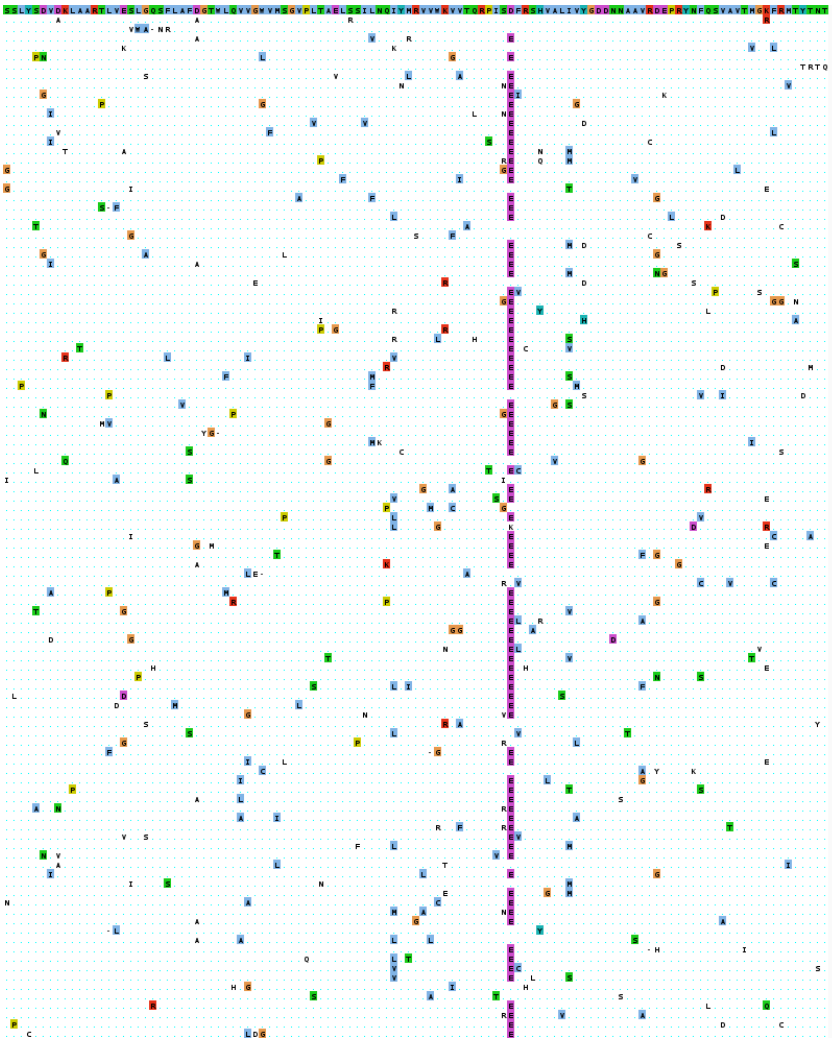

FIGURE 3.15: Marine picorna-like group A sequences aligned. Differences to the most abundant OTU are highlighted.

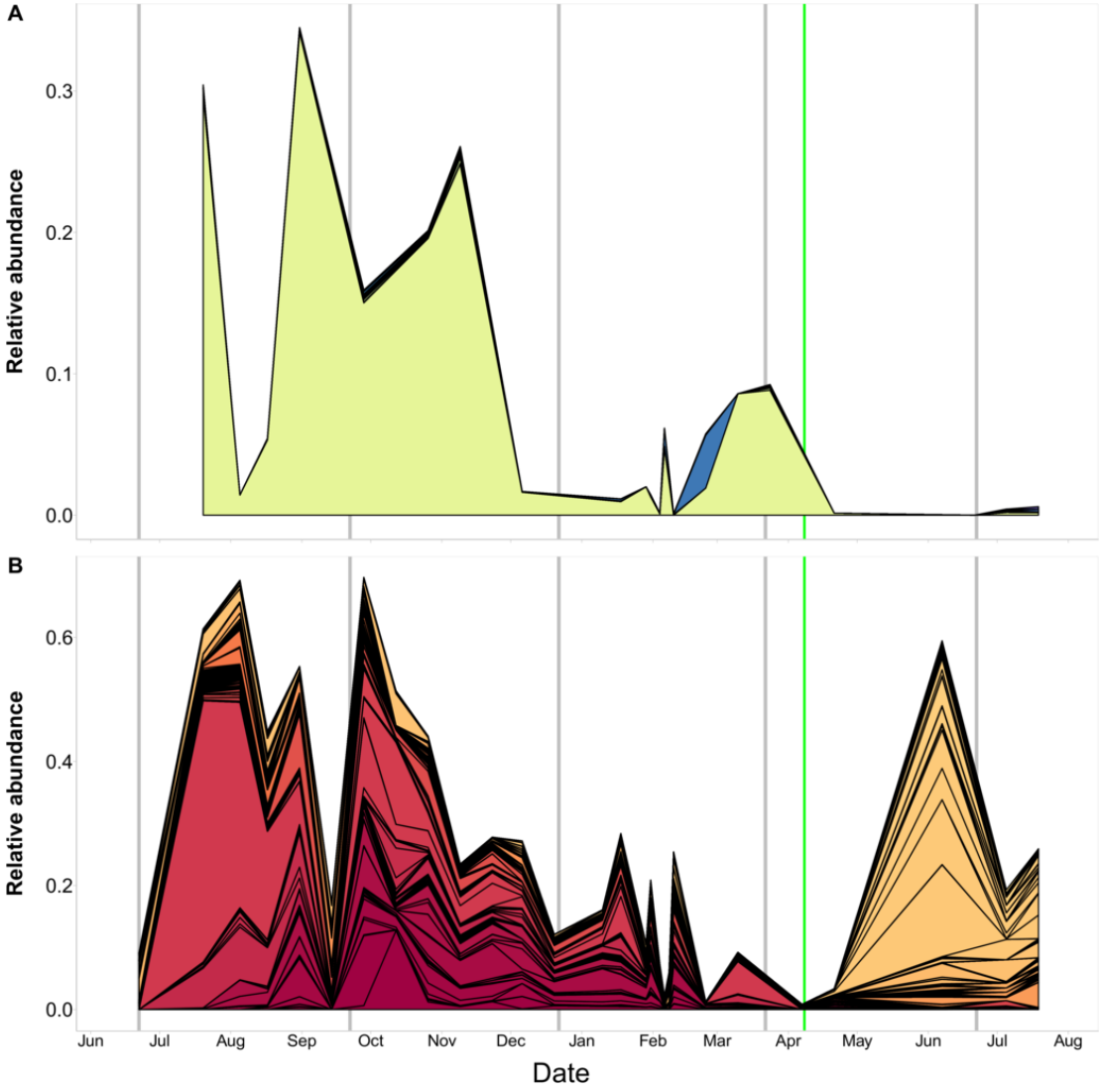

FIGURE 3.16: T4-like myoviral group I compared to bacterial OTUs classified as cyanobacteria over time. A) Relative abundance of bacterial OTUs (97%) classified as cyanobacteria over time. B) Relative abundance of T4-like myovirus group I OTUs (95% amino acid) over time. Each coloured contour represents a separate OTU. Grey vertical lines indicate boundaries between seasons and the green vertical line indicates the time of the spring bloom.

bacterial community, and slightly less strongly with the eukaryotic community. T4-like myoviruses showed changes with the eukaryotes and with the marine picornalike viruses, but not with environmental parameters.

3.5 DISCUSSION

3.5.1 *Major findings*

Using high-throughput sequencing of samples collected at a coastal site every two weeks for one year the dynamics of two groups of ecologically important groups of viruses were described in the context of their putative hosts and the environment. Community dynamics revealed differences by season. There was a large diversity of viruses and putative hosts in this study, and groups of phylogenetically-related viruses showed temporal dynamics in dominance. While some members of these related groups persisted throughout time, others were more ephemeral. These findings were put in context of potential quasispecies behaviour, of the dynamics of putative hosts, and of the seed bank and Killing the Winner theories.

3.5.2 *Shifts in dominance of viral communities by related viruses over time*

A major theme in microbial ecology is describing temporal shifts in community composition. Observing viral communities using the viral gene marker *pho h*, highly uneven communities were found at Bermuda Atlantic Time-series Study (BATS) using deep amplicon sequencing (Goldsmith *et al.*, 2015). Populations were divided into phylogenetically distinct groups and differences between the fall and winter samples were attributable to a phylogenetic group containing cyanobacterial infecting viruses and to water stratification (Goldsmith *et al.*, 2015). In this study at Jericho Pier, even though there was a large diversity of OTUs in the viral communities, the overall temporal dynamics were driven by shifts in phylogenetically-related OTUs (Figure 3.5). The MPL viruses had similar dynamics between the related viruses (Figure 3.5B) and the top 20

MPL OTUs over time (Figure 3.8) indicating an uneven community as previously seen (Culley *et al.*, 2006; Gustavsen *et al.*, 2014). Conversely, in the T4-like myoviruses the phylogenetically related group patterns did not resemble the patterns of the top 20 OTUs indicating that the community was much more diverse and even. Phylogenetic dynamics are important and fluctuate even though viral richness fluctuated minimally (Figure 3.4).

The dynamics of the phylogenetic groups are in agreement with Rodriguez-Brito *et al.* (2010) where the viral and microbial communities were stable at the genus and above taxonomic levels, but were dynamic at the strain or species level. Furthermore, taxonomically related species can have more similar niches and ecology (Harvey and Purvis, 1991; Srivastava *et al.*, 2012), thus it is possible these related viruses could be infecting similar hosts or responding to environmental cues in similar ways or both.

3.5.3 *Groups of related viruses contain ephemeral OTUs and constant OTUs (quasispecies)*

In Figure 12 and Figure 13 there are OTUs that are persistent and thus continually successful, and others that are ephemeral. In the T4-like myoviruses there are phylogenetic groups that do not contain persistent OTUs, but still have ephemeral OTUs that periodically dominate the viral community (Figure 3.12). Rodriguez-Brito *et al.* (2010) found that the largest OTUs were persistent throughout time and suggested that transient viruses were washed in from different areas. This seems unlikely for this study since the ephemeral viruses at Jericho Pier were closely related to the persistent viruses. Furthermore, both persistent and ephemeral viruses of phytoplankton were found in a freshwater lake (Short and Short, 2009; Rozon and Short, 2013), and at a coastal site (Short and Suttle, 2003) where some ephemeral viruses were correlated with shifts in environmental parameters. Additionally, in marine T4-like myoviral communities some OTUs were persistent and many more were ephemeral in 3-year (Chow and Fuhrman, 2012) and 2-year (Pagarete *et al.*, 2013) time series. Therefore, both eukaryotic and bacterial marine viruses show this structure of ephemeral and persistent viruses.

The population structure of RNA viruses is proposed to be a mixture of genotypes called quasispecies that are produced through errors and encompass the community of genotypes theoretically produced from one infection (Holmes, 2010; Domingo *et al.*, 2012). Thus a sequenced viral genome is often an "average" of all of these individual genotypes. In an Antarctic lake the ecological setting likely influences the presence of quasispecies in viral RNA metagenomes (López-Bueno *et al.*, 2015) since the number of quasispecies recovered from in the lake water compared to the microbial mats over time was very different. There were more single nucleotide variations (SNVs) in the lake water metagenomes since either there is more turnover and more ecological niches/diversity in these samples, or it is just the result of convergence of water from more locations.

Quasispecies behaviour has been best characterized in RNA viruses which are known to have a higher mutation rate and have higher burst sizes compared to the dsDNA viruses (Milo *et al.*, 2010). However, it is theoretically possible for quasispecies to exist in bacteriophage populations (Weitz *et al.*, 2005). In high confidence viral DNA metagenomes there is heterogeneity in assembled reads beyond just expected sequencing errors (Dutilh *et al.*, 2014), and there is site-specific variation in genomes from DNA viral populations studied in humans (Renzette *et al.*, 2015). As was found in López-Bueno *et al.* (2015), Renzette *et al.* (2015) saw that certain regions of the genomes had more SNVs than others indicating that selection does not appear constant across the genomes. When the OTUs in the marine picorna-like group A are compared in an alignment (Figure 3.15) many of the mutations in the ephemeral OTUs are randomly spread in the gene fragments, but most of them change from D to E in the "palm" section of the catalytic site C (te Velthuis, 2014). This would point to a population that is marginally successful while the most abundant OTU (retaining the D amino acid) remains persistent. This raises questions about whether this phenomenon is prevalent in marine settings and should be incorporated into current ecological theories (e.g. seed bank and KtW).

3.5.4 *Implications for theories related to community structure and dynamics*

In the Killing the Winner theory viruses infect the most active organism (Thingstad, 2000). Hosts compete for limiting resources which determines composition by site and viruses determine the identity and abundance of these hosts (Storesund *et al.*, 2015). With the lagged dynamics in the raphidophyte and viruses related to raphidophyte infecting viruses, the "winner" has been killed (Figure 3.14). There is a later small increase in relative abundance of the Raphidophytes with no associated or lagged increase in RdRp viral group A. One explanation is that the number of susceptible hosts were not abundant enough for a detectable increase in the viral group A, or the host could be targeted by a different subset of viruses (e.g. HAV, a DNA virus, also infects HAKA (Nagasaki and Yamaguchi, 1997)), or the hosts could be being controlled by another protist (ciliate e.g. (Harvey and Menden-Deuer, 2012)). Similar patterns were detected in the cyanophage from group I and the cyanobacteria (Figure 3.16). These types of patterns have been seen in (Chow *et al.*, 2014) where there were many viral OTUs detected with strong time-lagged correlations to bacterial OTUs. Also, in a mesocosm experiment with *Emiliania huxleyi* there was a peak in host abundance and then four days later a peak in *Emiliania huxleyi* Virus (EhV) abundance (Schroeder *et al.*, 2003). Rapid shifts in fine-scale viral dynamics and stability at coarse scale viral dynamics suggest that the Killing the Winner theory is operating at the strain level (Rodriguez-Brito *et al.*, 2010; Emerson *et al.*, 2013) preserving bacterial strain level diversity (Rodriguez-Valera *et al.*, 2009; Thingstad *et al.*, 2015).

In the marine environment, viruses, at a coarse scale of family/genus, could show Killing the Winner dynamics, and then at a finer scale could form a "seed bank" where there is shuffling of phylogenetically-related viruses on a rank abundance curve. This bank or seed bank model (Breitbart and Rohwer, 2005) which explains how high local viral diversity (shuffling of viruses) can be consistent with low overall global diversity (most abundant viruses) by a constant local production of viruses has been supported by many studies (Short *et al.*, 2010; Chow and Fuhrman, 2012; Zhong and Jacquet, 2014;

Brum *et al*., 2015). Our results are concordant with Goldsmith *et al*. (2015) who found that the viral community was mostly dominated (over 50%) by a few successful OTUs and the rest of the OTUs were rare and contained in the "bank". A further layer has been added to this idea by showing that the relatedness of the viruses in this seed bank is crucial to understanding their dynamics. Our data reveals that the viruses in the seed bank can be ephemeral and related to persistent viruses and that groups of related viruses can become abundant through ecological processes (through habitat filtering e.g. Koeppel and Wu (2013a)).

3.5.5 *Caveats*

Although the challenges with viral gene markers has been previously discussed (Gustavsen *et al*., 2014) and PCR in general (Lee *et al*., 2012), this is a useful approach for examining the population structures. Furthermore, especially with the rare and putative quasispecies OTUs, it cannot be ruled out that they are erroneous due to PCR (Pinto and Raskin, 2012) or sampling anomalies (as discussed in Shade *et al*., 2014), but these OTUs were seen multiple times in different samples therefore seem less likely to be spurious. To increase confidence in the results, some libraries with lower numbers of reads were excluded so that more sequence could be used overall when normalizing samples (Pinto and Raskin, 2012). The sequences were checked for chimeras since chimeras can form as a result of high cycle number as used one of the targets (Qiu *et al*., 2001). Read abundance of OTUs can be considered semi-quantitative and good for comparisons of richness and diversity among samples (but not for absolute counts of genes) (Pinto and Raskin, 2012).

3.5.6 *Conclusions*

Related viruses show temporal patterns of dominance over time. There were strong lagged correlations between hosts and groups of related viruses and viral communities show evidence of following host communities. More fine-scale structure of the communities was dependent on the different life strategies of the viral communities examined.

The marine picorna-like viruses exhibited more quasispecies-type behaviour and the T4-like myoviruses viruses, with their theoretically lower burst sizes, appeared to have a different mix of persistent and ephemeral viral dynamics. Overall, viral community dynamics are largely influenced by phylogenetically related groups of viruses over time and these dynamics add another layer to the well-established theories of Killing the Winner and the seed bank model for viral communities.

CHAPTER 4

Network analysis of Jericho Pier microbial time-series

4.1 SUMMARY

Marine microbes play a fundamental role in the dynamics of the marine ecosystem. The co-occurrence between these microbes can show links between organisms or shared niches, and thus describes the structure, dynamics and stability of these communities. Microbial co-occurrence networks at global ocean and coastal ocean scales have found expected, time-lagged and new links. Yet, no studies have included multiple groups of viruses in the association networks. To investigate the ecological patterns and drivers of diversity in microbial communities, high-throughput sequencing was performed for eukaryotic (18S ribosomal RNA gene), bacterial (16S ribosomal RNA gene) and viral (gp23 for T4-like myoviruses and RNA dependent RNA polymerase (RdRp) for marine picorna-like viruses) amplicons from a 1-year time series at a coastal site in British Columbia, Canada. Using local similarity analysis (LSA), co-occurrence networks and the phylogenetic relatedness in these communities were examined. The network topology revealed that within the viral communities there were more links than within the bacterial and eukaryotic communities, and within the viral networks phylogenetically-related viruses separated into tightly-connected subnetworks (modules). Examining the co-occurrence of operational taxonomic units (OTUs) over time, the T4-like myoviral community had the greatest number of links that were strongest when they were compared to the previous time-point. Over time, the eukaryotic and bacterial OTUs were more strongly correlated to environmental factors than the viral OTUs, thus these communities were strongly driven by the environment. Communities sampled in the

fall were more strongly correlated than in other seasons and shared the greatest number of links with the winter timepoints, indicating a time of stability for these communities. Thus, the ecological interpretations of microbial association networks, with the inclusion of viruses, could further our understanding of the drivers of microbial diversity and assembly.

4.2 INTRODUCTION

The high diversity, high abundance, and dynamics of marine microbial communities have begun to be explored, largely through advances in sequencing. Though often overlooked, it is important to integrate viruses into these community analyses since viruses can have a large influence on bacterial and eukaryotic communities. For example, viruses are estimated to be responsible for the daily lysis of 10-50% of heterotrophic bacteria and 5-10% of the cyanobacteria (Wilhelm and Suttle, 1999; Weinbauer, 2004) in plankton. Some studies have found that viruses control the host population composition or abundance or both (Bouvier and del Giorgio, 2007; Storesund *et al.*, 2015). Using microbial association networks, Chow *et al.* (2014) found that viruses may follow their host's abundance rather than control it. Moreover, in a global ocean survey of surface waters, viruses were found to be host-range limited because of large geographic distances (Lima-Mendez *et al.*, 2015).

An important step in understanding why and how marine bacterial, protistan and viral communities show seasonality and repeatability over time in composition and abundance (Gilbert *et al.*, 2011; Chow and Fuhrman, 2012; Chow *et al.*, 2013; Fujiki *et al.*, 2014; Simon *et al.*, 2015) is to explore the co-occurrences of individual taxa in communities over time, and how these relationships are affected by environmental and biotic factors. Characterizing these relationships can also help towards incorporating the role of viruses in maintaining the diversity of host populations and for understanding the dynamics of these systems.

4.2.1 *Network analysis to examine relationships*

Co-occurrence of organisms over time can be used to infer relationships or shared niches; these co-occurrences can be visualized using network diagrams. Network analysis has been used to examine relationships in microbial communities that are otherwise difficult to visualize, to examine known relationships, to propose new putative relationships, and to examine the overall structure of communities (Faust *et al.*, 2015b; Lima-Mendez *et al.*, 2015). Network analysis is used to examine association matrices produced by pairwise distance matrices (e.g. Bray-Curtis), Spearman or Pearson correlations, local similarity analysis (Ruan *et al.*, 2006; Xia *et al.*, 2011), generalized boosted linear models (Faust and Raes, 2012) or other association measures. Associations can be positive or negative, and can be the result of symbiosis (complementary functions), similar niches, competition, different resource use, predation, viral lysis, or grazing (ideas from Schluter, 1984; Chow *et al.*, 2014). Network analysis can be used to examine co-occurrence of organisms without knowing if the relationships are direct (e.g. predation or viral lysis) or indirect (e.g. density-mediated interactions) (Miki and Jacquet, 2010). As well, detection of microbial keystone species, those crucial for the overall ecosystem or community, shows promise by the examination of topological network characteristics (number of connections, and network properties when organism is removed) (e.g. Berry and Widder, 2014; Williams *et al.*, 2014).

Networks have also been used to examine the niches of organisms and be used to reveal the niches occupied by microbes (Steele *et al.*, 2011). Network analysis has also been used to assess temporal associations in marine systems (Gilbert *et al.*, 2011; Chow *et al.*, 2013, 2014; Cram *et al.*, 2015), and spatial associations in oceans (Lima-Mendez *et al.*, 2015), soils (Barberán *et al.*, 2012), and permafrost thaw ponds (Comte *et al.*, 2015). Time series studies at the San Pedro Ocean Time series (SPOT) in California have found repeatable patterns in the bacterial and viral communities, and strong connections in microbial networks illustrating known and new putative interactions (Chow *et al.*, 2014). Detecting such groups of highly-connected organisms can allow inferences to be made

with respect to temporal or seasonal variability (Cram *et al.*, 2015). As well, co-occurrence networks can help identify potential host-virus pairs and illuminate the potential ecology of these viruses. For example, viral and bacterial networks constructed from a global ocean survey showed that 43% of the phage populations were only strongly correlated to one bacterial OTU and the rest (57%) were strongly correlated to a few bacterial OTUs (Lima-Mendez *et al.*, 2015).

Microbes, including viruses, are key members of marine ecosystems and play large roles in geochemical cycling, photosynthesis, and nutrient remineralization (Worden *et al.*, 2015). Despite seasonal forcing, communities exhibit resilience implying that internal factors, such as composition or diversity may account for this resilience (Fuhrman *et al.*, 2015; Faust *et al.*, 2015a). Previous network analysis of time-series data for microbial communities provided important insights, but generally did not examine phylogenetic relationships within these communities, nor which co-occurrences were most important for defining the communities temporally. Phylogeny is important because co-occurring microbes can be phylogenetically closely related (Chaffron *et al.*, 2010). Using networks to analyze the dynamics of marine microbial communities can reveal how these systems respond to change over time, as well as how they maintain resilience and stability.

4.2.2 *Approach*

Using marker genes for two group of viruses (the T4-like myoviruses and the marine picorna-like viruses, as described in the Introduction and Chapter 3), bacteria, and eukaryotes, these communities were examined in a one year time series with samples taken every two weeks at Jericho Pier, in Vancouver, British Columbia. Local similarity analysis (LSA) was performed on the operational taxonomic units (OTU) and environmental parameters to examine the co-occurrence of all possible pairs within the same time-point (no lags) and with lags of up to one month (two timepoints). Additionally, to test the influence of the environment on these communities, redundancy analyses (RDA) and variation partitioning were performed with the community data and environ-

mental parameters. The aim of collecting these data are to determine the relationships of the viral, bacterial, and eukaryotic communities over time and the influence of the environment on driving the changes in these communities.

4.2.3 *Hypotheses*

Patterns of co-occurrence Viral replication depends on infection; hence, viruses must co-occur with their hosts either at the same time or with time-lags. Therefore, strong correlations in co-occurrence should occur between a virus and its host, and potentially between viruses and other organisms. However, different patterns in co-occurrence would be expected given the range of lifestyles of hosts and their viruses. For example, T4-like myoviruses can have narrow or broad host ranges (Sullivan *et al.*, 2003) and burst sizes of about 25 to 200; whereas, marine picornalike viruses are host-specific and have burst sizes > 1000 (Lang *et al.*, 2009). Viruses with high host specificity would produce networks with fewer connections to putative hosts (but does not limit intra-virus connections) than viruses with a wider host range, such as in some of the T4-like myoviruses.

Temporal niches. The environment influences viral diversity beyond regulating host diversity; for example, temperature and salinity can affect the ability of viruses to infect hosts (Kendrick *et al.*, 2014). RDA will be used with community-level data to infer what is driving the diversity of different communities. To observe another aspect of temporal niches, at the OTU-level, subnetworks, composed of a strongly correlated OTU pair and a strongly correlated environmental parameter (environmental triplets), will be used to examine environmental drivers of OTU pairs.

Resilience and stability Most microbial networks have "small-world" properties, meaning that the number of nodes separating two organisms is low, even with a high number of organisms, which suggests that these communities should be resistant to change. Hence, it is hypothesized that the microbial networks will also have few connections between organisms.

4.3 MATERIALS AND METHODS

4.3.1 *Sample collection*

Samples were collected from Jericho Pier (49° 16'36.73N, 123° 12'05.41W) in British Columbia, Canada. Jericho Pier (JP) is adjacent to the shoreline, in a well-mixed location with mixed semi-diurnal tides. Sixty litres of water were pumped from 1m depth every two weeks at the daytime high tide between June 2010 and July 2011, inclusive. Salinity, temperature and dissolved oxygen were measured using a YSI probe (Yellow Springs, Ohio, USA). For all samples, the water was filtered sequentially through 142-mm diameter, 1.2 μm nominal pore-size glass-fibre (GC50 Advantec MFS, Dublin, CA., USA) and 0.22 μm pore-size polyvinyldine filters (Millipore, Bedford, MA, USA). The filtrate, containing the viral size fraction, was concentrated to ~500 mL (viral concentrate) using tangential flow ultrafiltration with a 30kDa MW prep-scale Spiral Wound TFF-6 cartridge (Millipore) (Suttle *et al.*, 1991).

Phosphate, silicate, and nitrate+nitrite concentrations were determined in duplicate 15 mL seawater samples filtered through 0.45 μm pore-size HA filters (Millipore) and stored at -20°C until air-segmented continuous-flow analysis on a AutoAnalyzer 3 (Bran+Luebbe, Norderstedt, Germany). Chlorophyll *a* (Chl *a*) was determined in triplicate by filtering 100 mL of seawater through 0.45 μm pore-size HA filters (Millipore). The filters were stored in the dark at -20°C until acetone extraction and then analysed fluorometrically (Parsons *et al.*, 1984).

4.3.2 *Enumeration of bacteria and viruses*

Samples for viral and bacterial abundances were taken at each sampling point by fixing duplicate cryovials containing 980 μL of sample with final concentration of 0.5% glutaraldehyde (EM-grade, EMS, Hatfield, PA, USA), freezing in liquid nitrogen and storing at -80°C until processing. Briefly, viral samples were diluted 1:10 to 1:10 000 in sterile 0.1 μm filtered 1X TE, stained with SYBR Green I (Invitrogen, Waltham, MA,

USA) at a final concentration of 0.5 x 10^{-4} of commercial stock, heated for 10 min at 80° C and then cooled in the dark for 5 min before processing. Bacterial samples were diluted up to 1:1000 in sterile 0.1 μm filtered 1X TE, stained with SYBR Green I (Invitrogen) at a final concentration of 0.5 x 10^{-4} of commercial stock, and incubated in the dark for 15 min before processing. All samples were processed on a FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA) with viral and bacterial samples run for 1 min at a medium or high flow rate, respectively. Event rates were kept between 100 to 1000 events per second and green fluorescence and side scatter detectors were used. Data were processed and gated using Cell-Quest software (Becton-Dickinson).

4.3.3 *Viral concentration and extraction*

The viral concentrate was filtered twice through 0.22 μm pore-size Durapore PVDF filters (Millipore) in a sterile Sterivex filter unit (Millipore). The filtrate, containing virus-sized particles, was pelleted by ultracentrifugation (Beckman-Coulter, Brea, California, USA) in a SW40 rotor at 108 000 *g* for 5 h at 12°C. The pellet was resuspended overnight in 100 μL of supernatant at 4°C. To digest free DNA, the pellets were incubated with 1U/μL DNAse with a final concentration 5 mM $MgCl_2$ for 3 h at room temperature. Nucleic acids were extracted using a Qiamp Viral Minelute spin kit (Qiagen, Hilden, Germany) according to the manufacturer's directions.

4.3.4 *PCR amplification of T4-like myoviral marker gene*

To target the marine T4-type bacteriophage capsid protein gene (gp23) PCRs were set up as in Filée *et al.* (2005). Briefly, each reaction mixture (final volume, 50 μL) consisted of 2 μL template DNA (approx. 40 ng/μL measured with a Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA)), 1x (final concentration) PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 40 pmol of MZIA1bis, 40pmol of MZIA6, and 1 U Platinum Taq DNA polymerase (Invitrogen). Program conditions as in Table 3.1.

4.3.5 *PCR amplification of marine picorna-like viral marker gene*

Half of the viral extract was used for the generation of cDNA. To remove DNA, the extracted viral pellets were digested with DNase 1 (amplification grade) (Invitrogen). The reaction was terminated by adding 2.5 mM EDTA (final concentration) and incubating for 10 min at 65°C. Complementary DNA (cDNA) was generated using Superscript III Reverse Transcriptase (Invitrogen) with random hexamers (50 ng/μL) as per the manufacturer.

PCR was performed with primer set MPL-2 for a targeted set of the marine picorna-like virus RdRp (Culley and Steward, 2007). Each reaction mixture (final volume, 50 μL) consisted of 50 ng of cDNA, 1x (final concentration) PCR buffer (Invitrogen), 2 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 1 μM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1. Products were run on a 0.5X TBE 1% low melt gel, excised and extracted using Zymoclean Gel DNA Recovery Kit (Zymo) as per manufacturer's directions with a final elution step of 2x10 μL EB buffer (Qiagen).

4.3.6 *Filtration and extraction of marine bacteria and eukaryotes*

One liter of seawater was filtered through a 0.22 μm pore-size Durapore PVDF 47 mm diameter filter (Millipore) in a sterile Sterivex filter unit (Millipore). The filter was either immediately extracted or was stored at -20°C until extraction. Filter extraction proceeded as in Short and Suttle (2003). Briefly, filters were aseptically cut and incubated with lysozyme (Sigma-Aldrich, St. Louis, MO, United States) at a final concentration of 1 mg mL^{-1} for 2 h at 37°C. Sodium dodecyl sulfate was added at a final concentration of 0.1 % (w/v) and each filter was put through three freeze-thaw cycles. Proteinase K (Qiagen) was then added to a final concentration of 100 μg/mL and incubated for 1 h at 55°C. DNA was sequentially extracted using equal volumes of phenol:chloroform:IAA (25:24:1), and chloroform:IAA (24:1). DNA was precipitated by adding NaCl to a final concentration of

0.3M and by adding 2X the extract volume of ethanol. Samples were incubated at -20°C for at least 1 h and then centrifuged for 1 h at 20 000 *g* at 4°C. Extracts were washed with 70 % ethanol and then after drying were resuspended in 50 μL EB buffer.

4.3.7 *PCR amplification of marine bacteria and eukaryotes*

PCR targeting the eukaryotic fraction of the extract was performed using Euk1209f and Uni1392r primers as in Diez *et al.* (2001). Briefly, each reaction mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 0.3 μM of each primer, and 2.5 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1.

PCR targeting the bacterial fractions using primers 341F (Baker *et al.*, 2003) and 907R (Muyzer *et al.*, 1995) was performed with the following conditions: each reaction mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 0.4 μM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1.

4.3.8 *Sequencing library preparation*

Construction — PCR products not requiring gel excision were purified using AM-Pure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. Cleaned products were resuspended in 30 μL EB buffer (Qiagen). All products were quantified using Picogreen dsDNA (Invitrogen) assay using Lambda DNA (Invitrogen) as a standard. Sample concentrations were read using iQ5 (Bio-Rad, Hercules, CA, USA) and CFX96 Touch systems (Bio-Rad). Pooled libraries were constructed using one of each of the amplicons so that their molarity would be similar and totalling ~700-900 ng. Pooled

amplicons were concentrated using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. NxSeq DNA sample prep kit 2 (Lucigen, Middleton, WI, USA) was used as per manufacturer's directions with either NEXTflex 48 barcodes (BioO, Austin, USA), NEXTflex 96 HT barcodes (BioO), or TruSeq adapters (IDT, Coralville, Iowa, USA). Small Libraries were cleaned up using AMPure XP beads (Beckman Coulter) at a ratio of 0.9:1 beads:library.

Quantification and quality control of libraries — Libraries were checked for small fragments (primer dimers and/or adapter dimers) using a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) with the High Sensitivity DNA kit (Agilent). The concentration of libraries was quantified using Picogreen dsDNA assay as above. The libraries were checked for quantification and for amplifiable adapters using the Library Quantification DNA standards 1-6 (Kappa Biosystems, Wilmington, USA) with the SsoFast EvaGreen qPCR supermix (Bio-Rad) using 10 μL EvaGreen master mix, 3 μL of 0.5 μM F primer, 3 μL of 0.5um R primer and 4 μL of 1:1000, 1:5000 and 1:10000 dilutions of the libraries in triplicate on iQ5 (Bio-Rad) and CFX96 Touch qPCR machines. Cycling parameters were as follows: 95°C for 30s, 35 cycles of 95°C for 5s, 60°C for 30s, and the melt curve generation from 65°C to 95°C in 0.5°C steps (10s/step). Quantification values obtained from both Picogreen assays and qPCR assays were used to determine final pooling of all libraries before sequencing. Libraries were sequenced using 2 x 250bp PE Miseq (Illumina, San Diego, USA) sequencing at Génome Québec Innovation Centre at the McGill University (Montreal, QC, Canada), and 2 x 300bp PE Miseq (Illumina, San Diego, CA, USA) sequencing at UBC Pharmaceutical Sciences Sequencing Centre (Vancouver, BC, Canada) and at UCLA's Genoseq (Los Angeles, CA, USA).

4.3.9 *Initial sequence processing*

Libraries were either split by adapter by the sequencing centre using CASAVA (Illumina) or by the user using the Miseq Reporter software (Illumina). Sequence quality was

initially examined using FastQC (Andrews, 2015). Contaminating sequencing adapters were removed using Trimmomatic version 0.32 (Bolger *et al.*, 2014) and sequencing library quality was further examined using fastx_quality (Gordon, 2014). Libraries were further split into individual amplicons (i.e. 18S, 16S, gp23 and RdRp) and then, if the expected overlap of the paired-end reads was 40bp or more, the paired reads were merged using PEAR version 0.9.6 (Zhang *et al.*, 2014). Sequences were then quality trimmed using Trimmomatic with the default quality settings. Reads were annotated by library and then all libraries from individual primers were added together. Sequences were aligned to known sequences (Silva 119 database (Quast *et al.*, 2013) for 16S and 18S and for viruses to alignments built from viral isolates and Sanger sequenced environmental surveys) using align.seqs in mothur 1.33.3 (Schloss *et al.*, 2009). Reads not aligning were removed. Sequences were queried using BLAST against databases containing the gene markers of interest downloaded from Genbank and sequences with an e-value below 10^{-3} were kept.

4.3.10 *Chimera checking, OTU picking and read normalization*

The 16S and 18S rRNA gene sequences were checked for chimeras using USEARCH version 8.0.1517 reference (Edgar, 2010) with the Gold reference database. Unique, non-chimeric sequences were clustered at 97% similarity. Taxonomy for the 16S and 18S rRNA gene sequences was assigned using mothur (Wang-type algorithm) and the taxonomy in Silva 119 (Quast *et al.*, 2013). For the viral targets sequences were chimera-checked using USEARCH denovo and reference (Edgar, 2010). Viral sequences were then translated using FragGeneScan 1.20 (Rho *et al.*, 2010). Viral reads were clustered using USEARCH (Edgar, 2010) at 95 % similarity for MPL, and 95 % similarity for T4-like myoviruses. Operational taxonomic unit (OTU) tables for all targets were constructed using USEARCH (Edgar, 2010). Rarefaction curves were generated using vegan (Oksanen *et al.*, 2015). Sequences were normalized for this project by date and by target using vegan (Oksanen *et al.*, 2015).

4.3.11 *Data analysis and multivariate statistics*

ENVIRONMENTAL DATA — Environmental parameters were mean imputed to fill in data missing because of instrument malfunction or unavailability. Day length data were retrieved using R package geosphere (Hijmans, 2015), and irradiance values from the UBC Measurement Network (Christen, 2013).

HYPOTHESIS TESTING: RDA AND VARIATION PARTITIONING — Redundancy analysis (RDA) was performed on Hellinger transformed OTU tables to test whether the environmental parameters influenced the variability and structure of the communities over time (Legendre and Legendre, 1998) in vegan (Oksanen *et al.*, 2015). Highly co-related variables were removed prior to starting the model. Significance of the overall RDA was determined using function anova.cca with 999 permutations and alpha of 0.05. For each axis the same check for significance was performed using function anova.cca. Variation partitioning (Borcard *et al.*, 1992) was performed to find the percentage of community variation determined from abiotic vs. biotic factors. Only significant fractions were kept in the model.

4.3.12 *Network analysis*

LOCAL SIMILARITY ANALYSIS AND NETWORK VISUALIZATION — OTUs present in at least 30% of sites were used for local similarity analysis (LSA). LSA was run for 29 time points with a maximum lag of two timepoints (equivalent to about a one month lag). *Q* values, used to evaluate the false discovery rate, were determined (Storey *et al.*, 2005) and LSA results with a *Q* value less than 0.05 and *p* value less than 0.05 were kept as significant links. LSAs were formatted into networks and network statistics calculated using igraph (Csardi and Nepusz, 2006) in R and then visualized in Cytoscape 3.3.0 (Smoot *et al.*, 2011). Cyrest (Ono *et al.*, 2015) was used to send the networks to Cytoscape from R. Force-directed layouts were used from AllegroLayout 2.2.2 (Allegro, Santa Clara, CA,

USA) in Cytoscape. Sub-networks were constructed by filtering the overall network by amplicon.

To assess whether the networks had specific properties not only stemming from the number of nodes and edges, LSA networks were compared to simulated random networks. Simulated networks were constructed using a random network generator (erdos.renyi.game) and scale-free network generator (barabasi.game) in igraph (Csardi and Nepusz, 2006).

MODULE PICKING — Modules, which represent groups of highly connected OTUs (sub-graphs), were determined using the cluster walktrap algorithm in igraph (Csardi and Nepusz, 2006). This algorithm looks for highly connected nodes in a network using random walks through the network. Walks with short path lengths indicate that nodes were in the same community.

ENVIRONMENTAL TRIPLETS — To examine connections between OTUs driven by the environment, connections were examined where there were two correlated OTUs that were also each correlated to an environmental parameter giving "environmental triplets"

4.4 RESULTS

4.4.1 *Co-occurrence of OTUs over 1 year-time series*

As seen in Chapter 3, there was high overall diversity and richness in the OTUs in the viral, bacterial and eukaryotic networks. Using local similarity analysis (LSA) (Ruan *et al.*, 2006; Xia *et al.*, 2011) association networks were created from the significant pair-wise associations. Sub-networks containing either only viral, bacterial, eukaryotic or a mixture of nodes were created by filtering the overall network for specific types of OTUs and visualizing the connections.

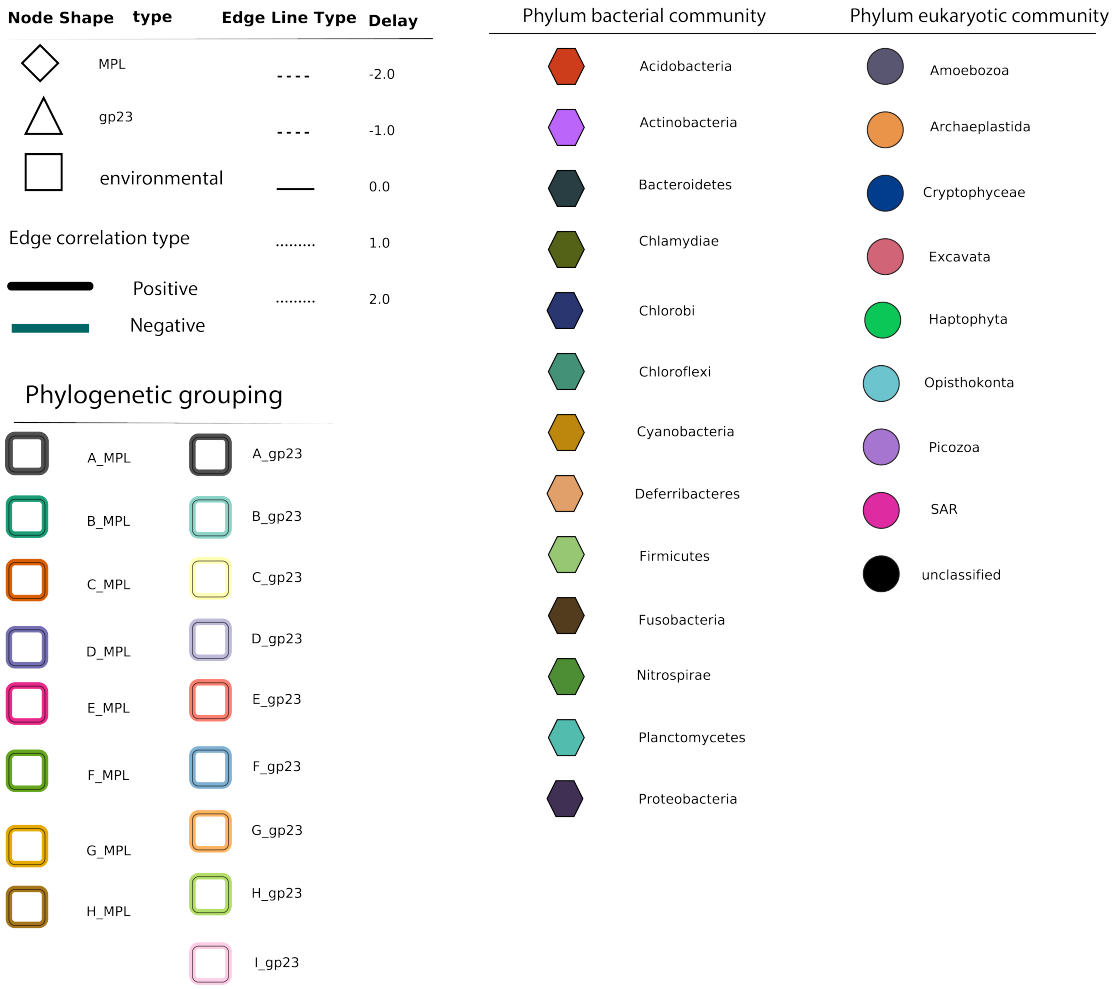

FIGURE 4.1: Legend for networks. Taxonomic phylum classifications for bacteria and eukaryotes based on Silva 119 (Quast *et al.*, 2013). Node shapes for viral OTUs and environmental parameters. Edge line type displays the delay between time points where strongest correlation is found. Edge correlation type displays whether OTUs are negatively or positively correlated. Colour of node outline by phylogenetic grouping for viral OTUs. Colours for these groups are from Chapter 3.

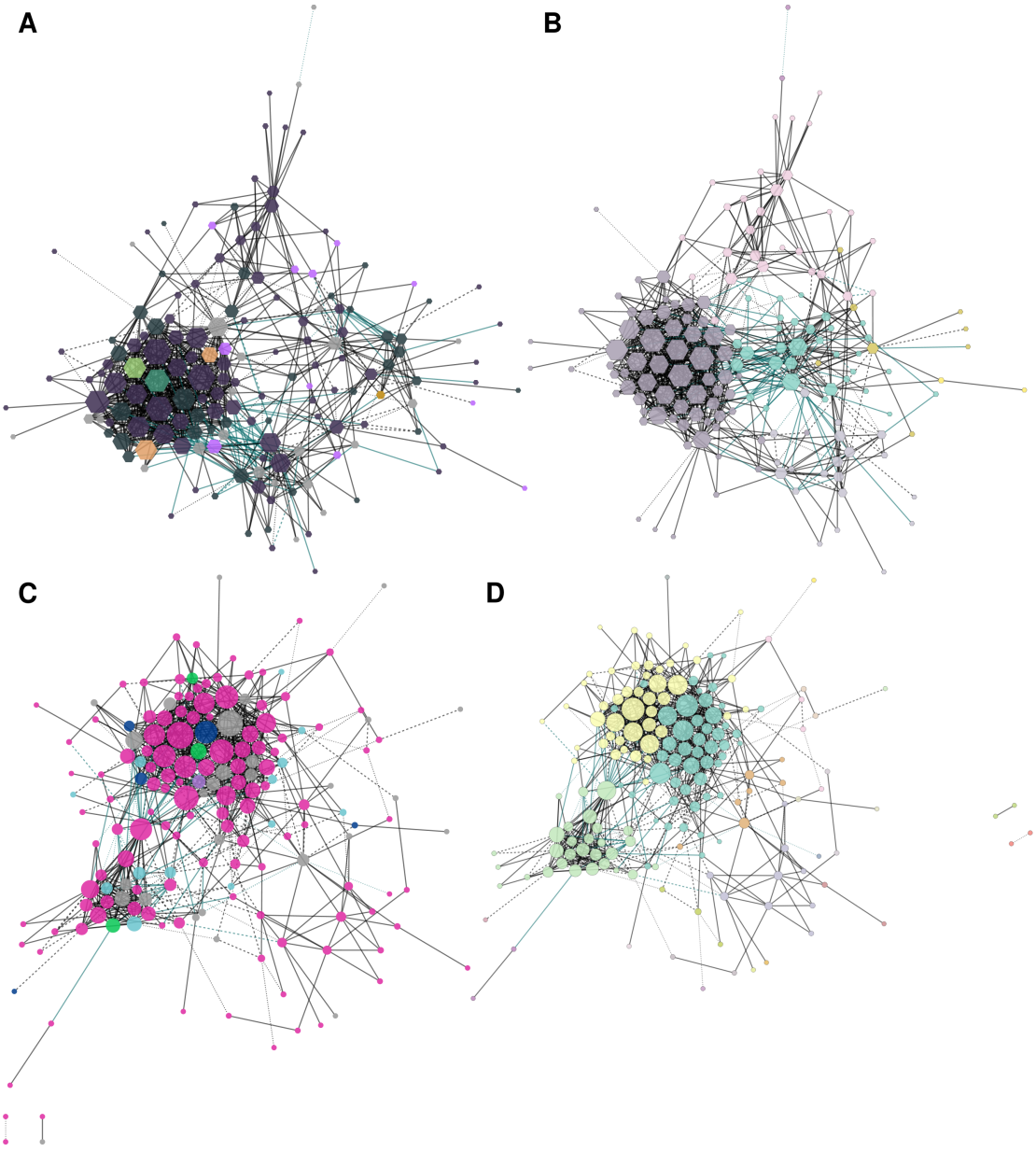

FIGURE 4.2: Bacterial and eukaryotic networks. A) Bacterial OTU network coloured by phylum. B) Bacterial OTU network with modules. C) Eukaryotic OTU network coloured by phylum. D) Eukaryotic OTU network with modules. For legend see Figure 4.1. Modules detected using the cluster walktrap algorithm. Members of the same module are the same colour. Node size is scaled based on degree, i.e. the number of strong co-occurences with that node.

COMMUNITY DETECTION IN NETWORKS — In the bacterial community network modules, highly connected groups of OTUs in individual sub-networks, were visualized and this highlighted two large, negatively correlated, modules (Figure 4.2B). Most of the links between the two modules represented negative correlations among the Proteobacteria and the Flavobacteria. The two modules had similar composition, but one featured more highly connected Actinobacteria (Figure 4.2A). The most well connected OTU in the overall network was classified as *Defluviicoccus* spp. (S16_127, Rhodospiralles, a class of Alphaproteobacteria).

The eukaryotic networks displayed similar patterns as the bacterial communities with two large groups of negatively correlated modules (Figure 4.2D). These two groups had similar composition and were largely connected by one 18S OTU (18S-3) classified as a dinoflagellate (class Gymnodiniphycidae) (Figure 4.2C).

In the T4-like myoviral networks, the groups detected using the module algorithm were very similar to the phylogenetic groups detected in Chapter 3 when overlaid onto the network (Figure 4.3A and B). The most interconnected group of T4-like myoviruses showed many negative correlations to one of the modules that contained T4-like viruses from many different phylogenetic groups.

As with the T4-like myoviral communities, the marine picorna-like viral community contained modules that were very similar to the phylogenetic groupings from Chapter 3 (Figure 4.3C and D). There were strong negative correlations between the largest module and one of the other modules.

In all the networks the links within modules were more often positive than negative and connections between modules were more often negative.

CO-OCCURRENCE BETWEEN TWO TYPES OF AMPLICONS — The association network for both the bacterial and eukaryotic communities contained two main modules that were negatively correlated (Figure 4.4A and B). Overall in this network the node classified as Rhodobacteraceae (16S OTU-323) had the most connections to other OTUs. The association network for the bacterial community and the T4-like myoviruses contained

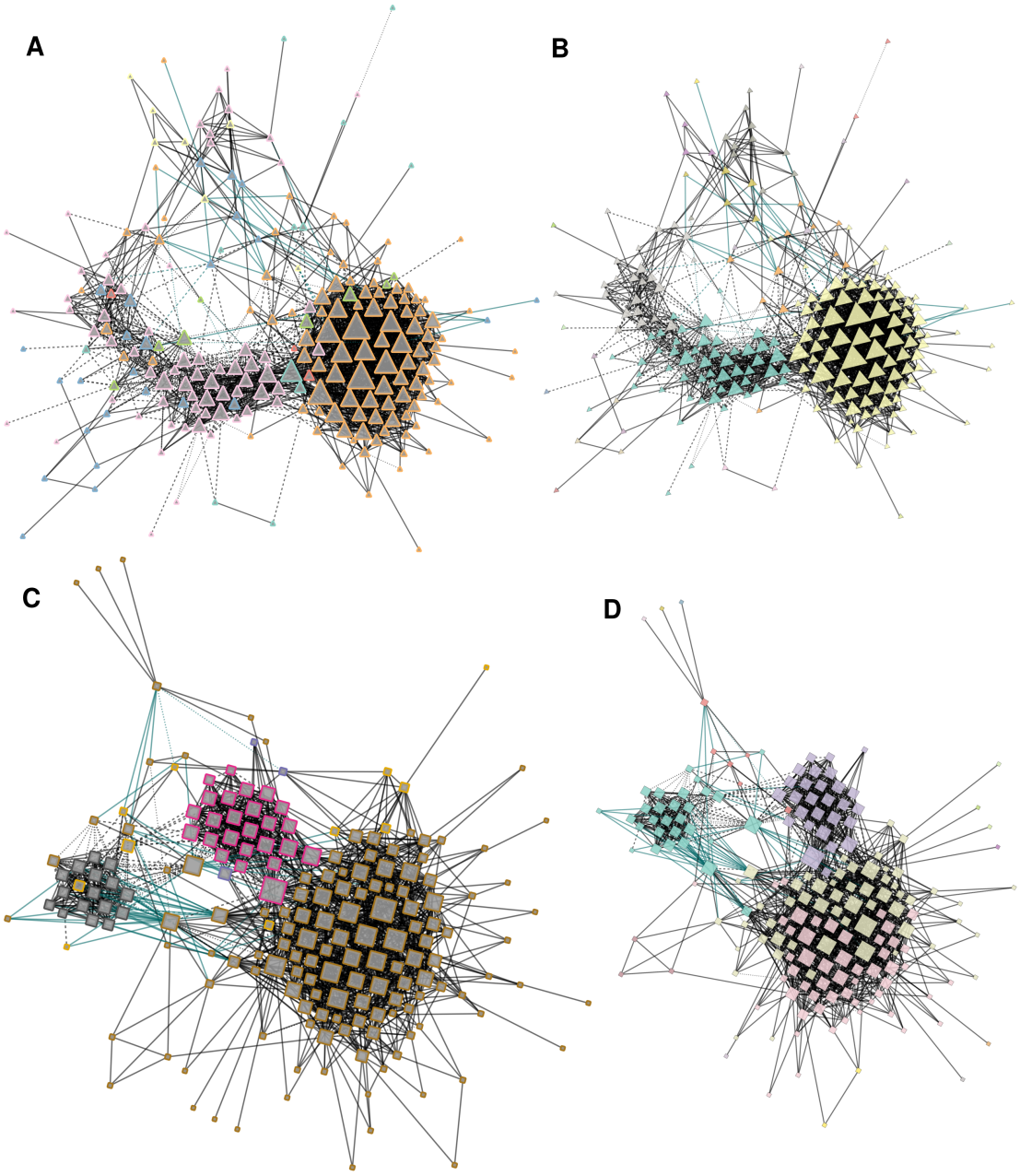

FIGURE 4.3: Viral community networks. A) T4-like myoviral OTU network coloured by phylogenetic grouping. B) T4-like myoviral OTU network with modules. C) Marine picorna-like OTU network coloured by phylogenetic grouping. D) Marine picorna-like OTU network with modules. For legend see Figure 4.1. Modules detected using the cluster walktrap algorithm. Members of the same module are the same colour. Node size is scaled based on degree, i.e. the number of strong co-occurences with that node.

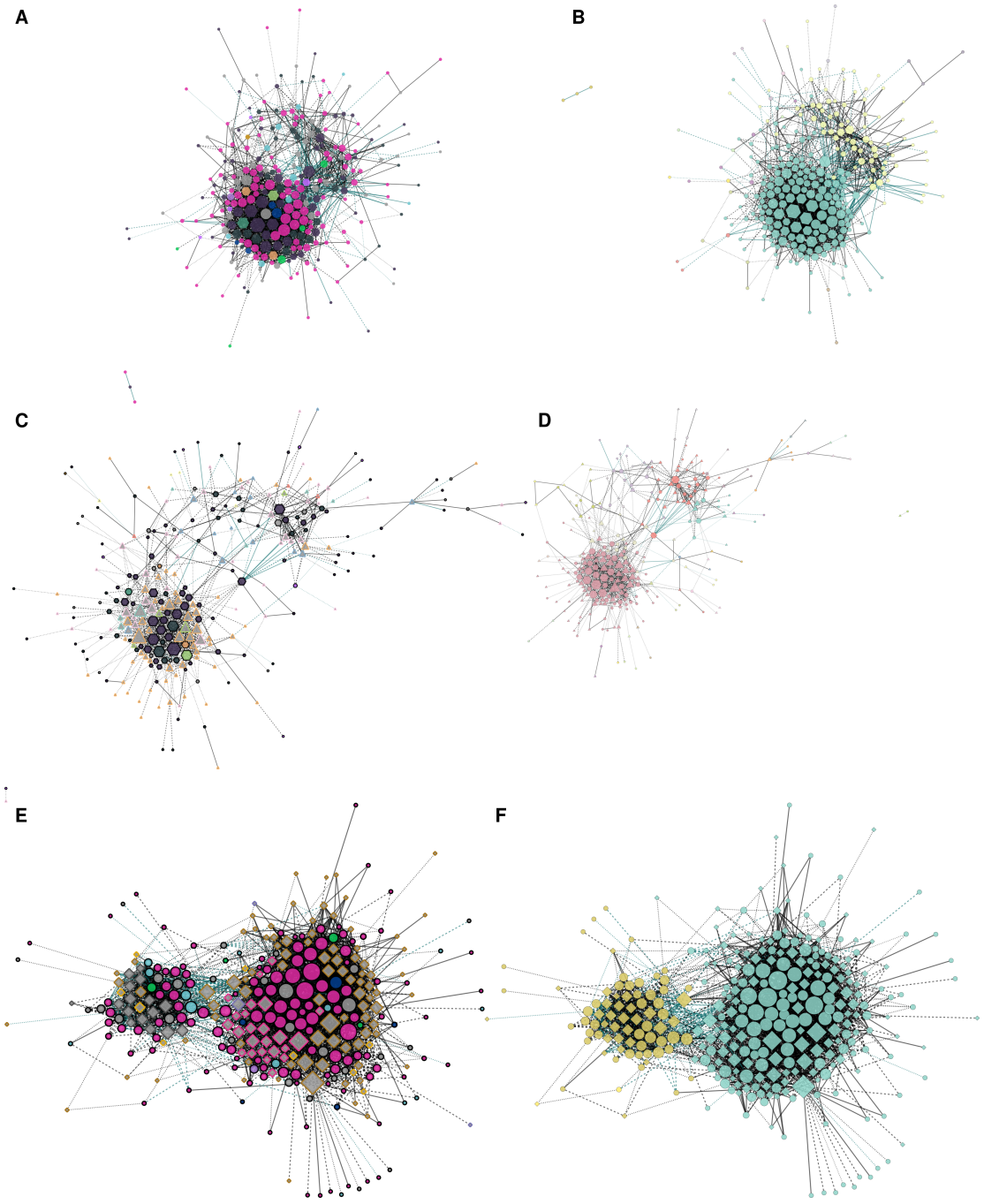

FIGURE 4.4: Networks between different communities. Bacterial and eukaryotic OTU network A) coloured by phylum and B) modules coloured. Bacterial and T4-like myoviral OTU network C) coloured by phylum and D) coloured by modules. Eukaryotic and marine picorna-like OTU network E) coloured by phylum and F) coloured by modules. Legend Figure 4.1. Modules detected using the cluster walktrap algorithm. Members of the same module are the same colour. Node size is scaled based on degree.

two main modules (Figure 4.4C and D). In one tightly clustered module there were mostly the phyla Alphaproteobacteria (Rhodobactierales, Rhodospirales, SAR 11), Betaproteobacteria (Burkoholderiales, Flavobacteria, some other Bacteroides, Chloroflexi (SAR 202 clade), Deferribacteria), Deltaproteobacteria (Desulfobacteriales, SAR 324) Firmicutes (Clostridia), and Gammaproteobacteria (Oceanospiralles, Alteromonadales, Salinispaherales). The two large modules were connected by OTUs classified as Rhodobacteriales which had negative correlations to one module and time lagged positive correlations to the other (Figure 4.4C). In the other large module the bacteria were classified as phyla Flavobacteriia, Alphaproteobacteria (SAR 11 clade, Rhodobacterales, OCS116_clade), and Betaproteobacteria (Methylophilales). When the phylogenetic groups for the T4-like myoviruses were overlaid onto this network the phylogenetic patterns were no longer grouped into distinct modules (Figure 4.4D).

In the association network containing the eukaryotic and marine picorna-like viral communities there were two large modules present (Figure 4.4E and F). In one group there were many marine picorna-like viral OTUs, and many eukaryotic nodes which were classified as Alveolates (Dinoflagellata, Ciliophora, Protalveolata), Cryptomonadales (Hemiselmis, Teleaulax, FV18-2G7), Holozoa (Metazoa). Cryptophyceae (Kathablepharidae) Picozoa (Picomonadea), Haptophyta (Prymnesiophyceae), Rhizaria (Cercozoa), Stramenopiles (Ochrophyta, Peronosporomycetes, MAST2, 4, 6, and 12). The second most highly connected module also had many marine picorna-like viral OTUs, and for eukaryotic nodes it contained nodes from the phyla Alveolates (Dinoflagellata, Ciliophora, Protalveolata), Holozoa (Metazoa), Picomonadea, Haptophyta (Prymnesiophyceae), Rhizaria (Cercozoa), Stramenopiles (Ochrophyta, Peronosporomycetes, MAST 12). The marine picorna-like viral OTUs 3 and 5 connected the modules.

4.4.2 *Temporal associations between communities*

Number of edges between and within by season — The fall had the highest number of shared edges even though the winter season had a higher number of nodes (Figure

4.5A). The summer had the lowest number of significant edges by date and the lowest number of nodes. When comparing the number of edges between seasons, the fall and winter had the greatest number of shared edges and the greatest number of nodes present in both seasons. Summer to winter had the fewest number of edges shared even though the summer to winter had more nodes than in the spring to summer.

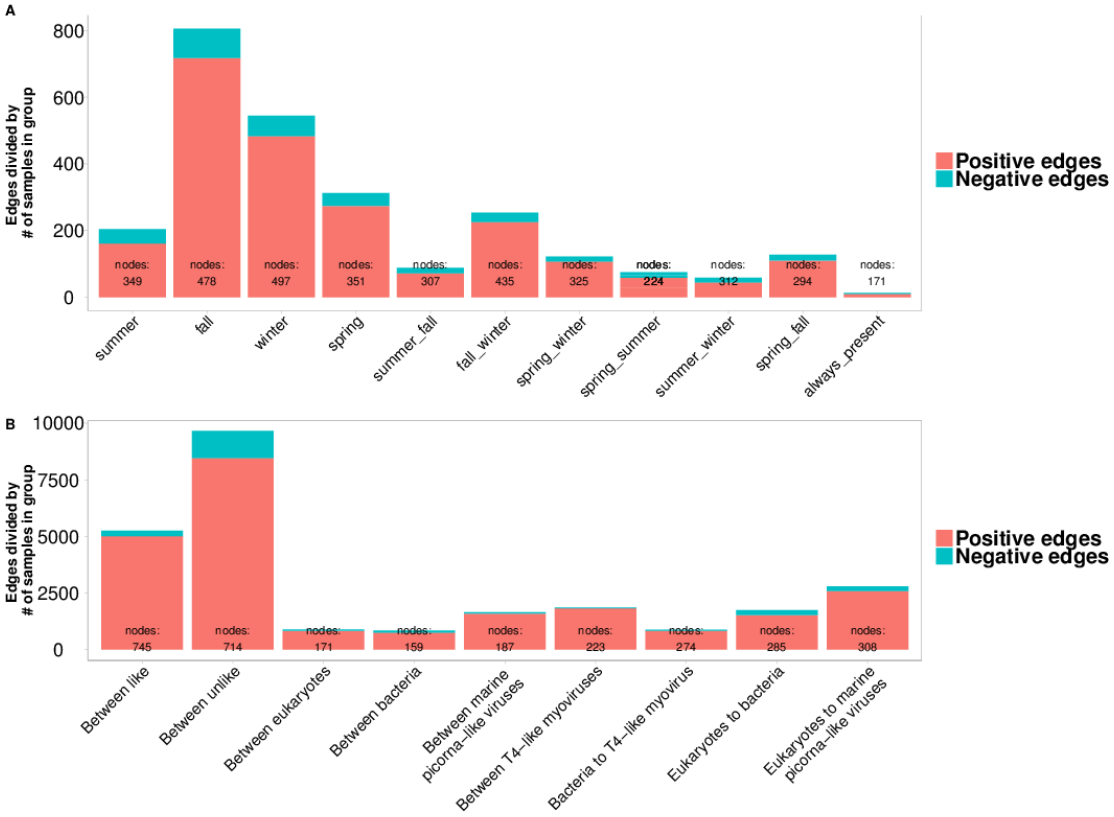

FIGURE 4.5: Interactions over time, between and within communities. A) Number of positive and negative edges using local similarity analysis (LSA). B) Number of edges within and between seasons from local similarity networks. Edges by season divided by numbers of samples by season.

NUMBER OF EDGES BETWEEN AND WITHIN COMMUNITIES — There were more connections between OTUs from different communities than within communities (Figure 4.5B). Most edges were positively correlated and this was consistent across subnetworks. The subnetwork with the greatest number of edges within a single type of community was the

T4-like community and between communities was the eukaryote to marine picorna-like viral network.

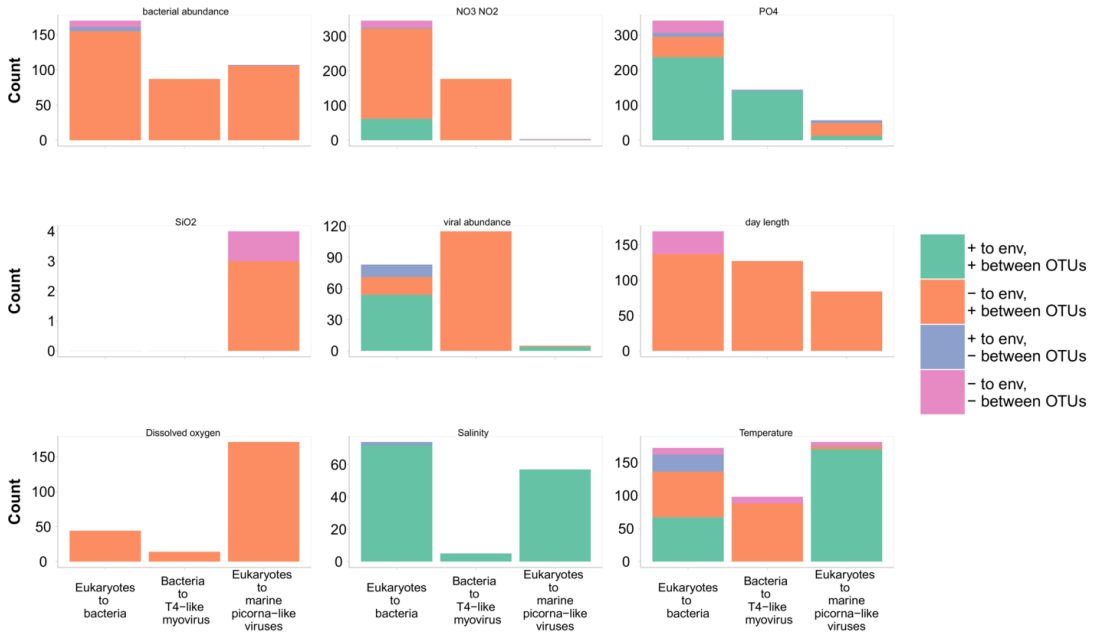

FIGURE 4.6: Counts of triplets by environmental factor. Chlorophyll *a*, pH, and month number were not displayed because of few to no connections. There were no triplets found where one OTU was positively correlated to the environmental parameter and the other OTU was negatively correlated.

EDGES COUNTED BY TRIPLETS BY ENVIRONMENTAL FACTORS — Temperature and phosphate had the highest number of edges in all the communities (Figure 4.6). The T4-like myoviral to bacterial edges had no strong connections to bacterial abundance, viral abundance or dissolved oxygen and very few to nitrate+nitrite. The T4-like to bacterial edges had the highest number of edges associated to silicate concentration and all of these edges were negatively correlated. The marine picorna-like viral community to eukaryotic community edges had few triplets linked to nitrate+nitrite and viral abundance and they had the highest number of edges strongly linked to dissolved oxygen, silicate and temperature. The eukaryotic to bacterial edges had no strong connections to silicate

and few strong links to dissolved oxygen, but in all other categories these communities generally had the most edges.

4.4.3 *Network statistics*

TABLE 4.1: Network statistics. Table counting associations with Q value < 0.05, p value < 0.01, abs(LS) > 0.5. (continued below)

	Overall	Eukaryotic network	Bacterial network	Marine picorna-like virus network	T4-like myovirus network
nodes	789	171	159	187	223
edges	14918	892	846	1655	1861
density	0.024	0.031	0.034	0.048	0.038
modularity	0.37	0.42	0.41	0.47	0.42
average degree	38	10	11	18	17
max degree	172	41	41	55	70
positive edges (%)	90	93	89	96	98
negative edges (%)	10	7	11	4	2
no delay (%)	49	76	92	90	76
delay pos (%)	23	10	4	6	9
delay neg (%)	29	14	3	4	15

	Eukaryote to bacteria network	Eukaryote to marine picorna-like virus network	Bacteria to T4-like myovirus network
nodes	285	308	274
edges	1746	2804	886
density	0.022	0.03	0.012
modularity	0.22	0.3	0.39
average degree	12	18	6
max degree	59	77	34
positive edges (%)	87	92	95
negative edges (%)	13	8	5
no delay (%)	55	38	13
delay pos (%)	17	29	44
delay neg (%)	29	33	42

Of all the pairwise associations 3.77% were significant and passed the false discovery test (*Q* value) (14918 out of 395605 potential associations). The strongest correlations were found within the single community networks which were mostly composed of edges with no delay (between ~75-95% with no delay). This contrasted what was found in the overall network and in the networks composed of two different communities which ranged from 14% to 39% of edges with no delay (Table 4.1).

When considering only the eukaryotic network there was higher network density (0.03), and modularity (0.42) than for the overall graph, but similar percentages of positive (93%) and negative (7%) edges (Table 4.1). The marine picorna-like viral communities had the highest density (0.05) and the gp23 communities had the second highest density (0.04). The viral community networks had a higher average degree (18 and

17) than the eukaryotic and bacterial communities (10 and 11) and a higher percent of positive edges (96% and 98%).

4.4.4 *Network statistics over time*

NETWORKS AMONG TWO COMMUNITIES — In the overall network, the node count with significant edges fluctuated between 50 to almost 300 nodes (Figure 4.8A and see also Appendix B). Conversely, all the subnetworks between communities had more stable nodes and edges over time. There were fluctuations in density over time with the fall and winter being the least dense. However, the bacterial to T4-like myovirus network had a spike in density in late March, and all communities had a spike in density in late June (which is related to the decrease in nodes, edges and diameter at this time). The bacterial to T4-like myovirus network had the highest modularity over time and the eukaryotic to bacterial had the lowest. For most of the communities the median degree over time was stable, however, for the eukaryotic to bacterial network there was a spike in August, but there was no spike in nodes or edges.

NETWORKS AMONG SINGLE COMMUNITIES — For the single community networks the nodes and edges were stable, except for the T4-like myoviruses which had more fluctuations in both edges and nodes over time (Figure 4.8B). All communities generally showed constant diameters up until late June, when there was a large decrease in diameters. The T4-like myovirus community generally had the highest diameter and the marine picorna-like community had the lowest. The marine picorna-like virus network had the highest density over time. Examining the median degree, the eukaryotic network had the highest value in the summer and then it dropped off in the fall. The T4-like myoviruses were the most highly connected for most of the time series.

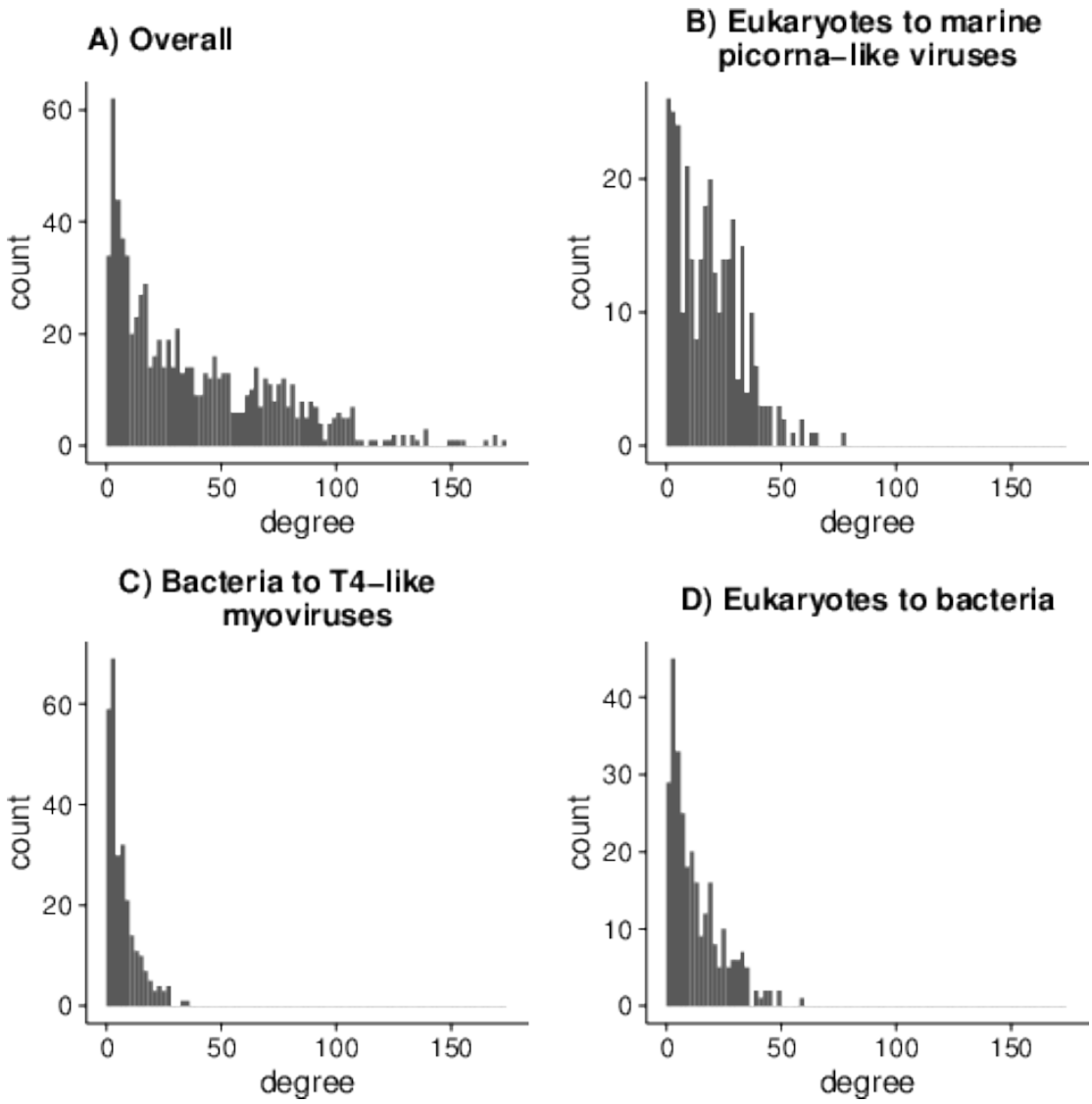

FIGURE 4.7: Degree histograms. A) degree overall, B) degree between eukaryotes and marine picorna-like viruses, C) degree between bacteria and T4-like myoviruses, D) degree between eukaryotes and bacteria,

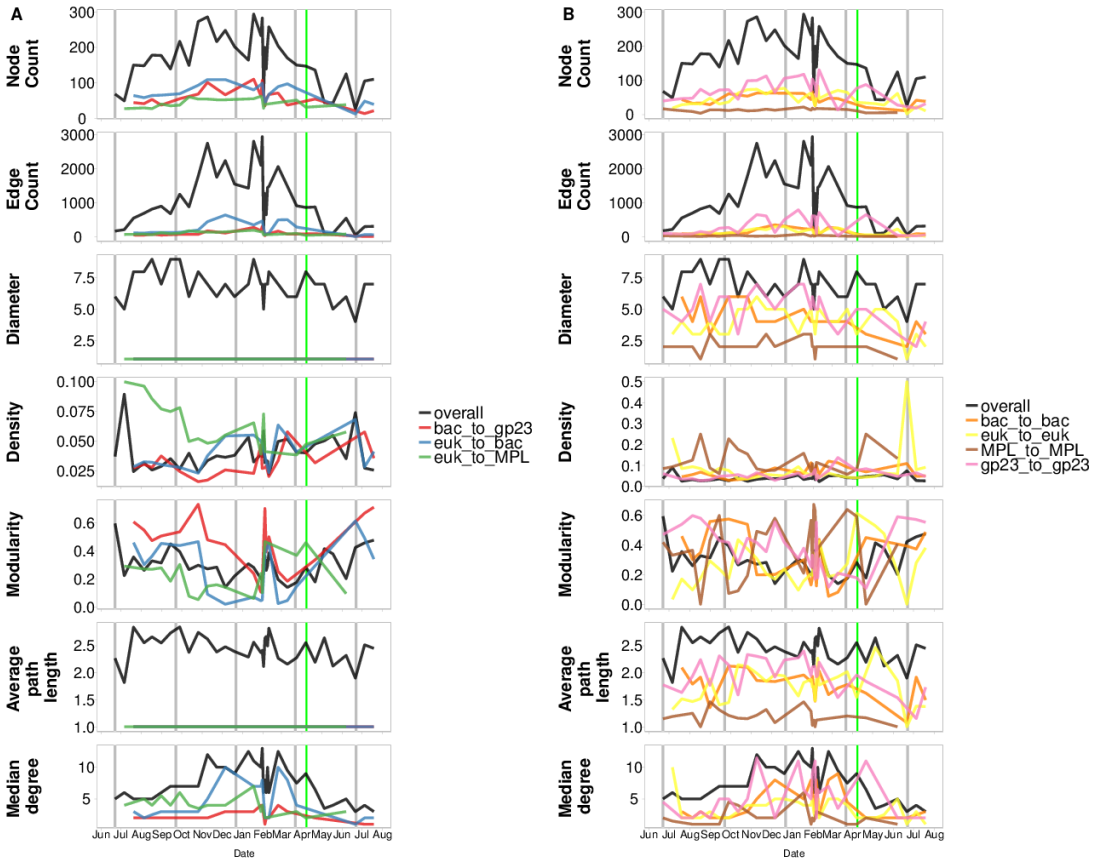

FIGURE 4.8: Network statistics over time of the subnetworks composed of two different communities andof subnetworks within single communities. A) Two community networks B) Single community networks. A and B include: Node counts, Edge counts, Diameter (note for all the combined networks this was 1), Density, Modularity, Average path length (note for all the combined networks this was 1), and Median degree. The green vertical line corresponds to the spring bloom and grey lines correspond to divisions between seasons.

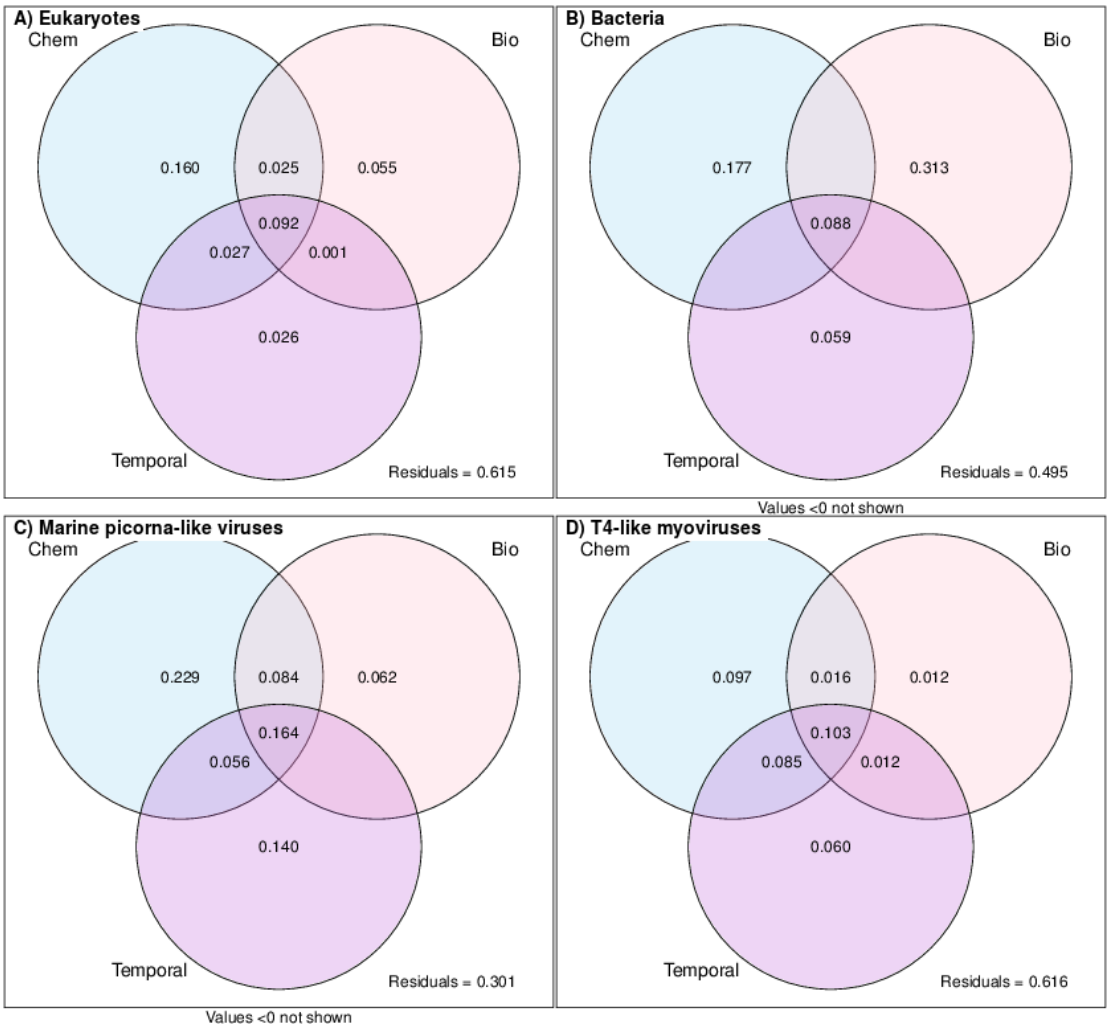

FIGURE 4.9: Variation partitioning based on partial RDA. A) Eukaryotes grouped by taxonomic order, B) Bacteria grouped by taxonomic order, C) Marine picorna-like viruses grouped by phylogenetic clades from Chapter 3, D) T4-like myoviruses grouped by phylogenetic clades from Chapter 3. Parameters were divided into chemical, biotic and temporal parameters before being forward selected and then used in variation partitioning.

4.4.5 *Variation partitioning*

To determine whether biological, chemical or temporal factors (or a combination of these) played a role in structuring the communities, partial redundancy variation partitioning was performed (Figure 4.9). Both the eukaryotic and bacterial communities were not significantly explained by models at the OTU-level, but had significant models at the order and family levels. In the eukaryotic community chemical parameters represented the largest amount of variation in the community (16%), followed by biotic (6%) and temporal factors (3%) (61% unexplained). For the bacterial communities, biological factors explained the most variation (29%), followed by chemical (13%), and then temporal (4%, unexplained 55%).

The viral communities were not significantly explained at the OTU-level by the environmental parameters, but were significant when the viral community relative abundance was summed using the phylogenetic groups from Chapter 3. For the marine picorna-like viral communities, biological parameters explained the most of the variation (23%), followed by temporal (13%) and then chemical (10%). Twenty percent of the variation was overlapping between biological and chemical factors (residual 30%). For the T4-like myoviral communities less of the variation was explained, but the chemical parameters explained the most (8%), followed by temporal (5%), then biological factors (1%), with 10% of overlap (residual 64%).

4.5 DISCUSSION

Amplicon sequencing of eukaryotic, bacterial and viral communities revealed significant relationships in the temporal patterns of co-occurrence. Further analysis of these relationships using network analysis, and the overall communities using redundancy analysis and variation partitioning showed distinct communities (Figure 4.2, Figure 3 and Figure 4.4), that were highly dynamic over time (Figure 4.8).

The modules, which are highly connected communities of OTUs, detected similar groups in the networks as were determined phylogenetically (Figure 4.4). This further strengthens the argument proposed in Chapter 3 that there is phylogenetic structuring to these communities. Also, the viral networks were more densely connected than those for bacteria and eukaryotes (Table 4.1 and Figure 4.8) suggesting a different structure in these communities as was also seen in Chow *et al.* (2014). The overall network had the greatest number of connections per timepoint in the fall. Also when examining strongly correlated triplets of OTUs correlated with environmental factors, the bacterial-eukaryotic networks had more associations than the host-viral networks. One of the viral groups, the marine picorna-like viruses, had a large proportion of its variance attributable to chemical factors, more than any of the other communities. This was surprising, but similar results were seen in T4-like myoviral communities in coastal waters (Chow *et al.*, 2014). Overall, temporal partitioning of the environment was an important driver of these communities.

4.5.1 *Environmental influence on individual OTUs compared to communities*

There was evidence for environmental control of the overall variance of the composition and structure of planktonic communities. The influence of abiotic and biotic factors on the communities was examined since these are important factors for structuring bacterial, eukaryotic and viral communities (e.g. Cullen, 1991; Kirchman *et al.*, 1991; Morris *et al.*, 2005; Fuhrman *et al.*, 2006; Fuhrman, 2009; Gilbert *et al.*, 2009; Chow *et al.*, 2014). Environmental influence at the community level differed from the pairwise LSA associations for the individual OTUs. Phosphate, nitrate+nitrite, temperature and salinity had similarly large influences on the communities at the OTU- and community-levels. Nutrients, day length, temperature, salinity and chl *a* were important in all the viral, bacterial and eukaryotic communities, but not necessarily for the dynamics of the individual OTUs. This is similar to Gilbert *et al.* (2011) who found that change in day length could explain most of the variability in bacterial community diversity, therefore,

seasonal change was deemed more important than trophic interactions. Chow *et al.* (2013) found that the bacterial community variance was related to salinity and chl *a*; protistan community variance was related to day length and bacterial abundance; and T4-like myoviral communities were related to day length, change in day length, salinity and temperature. These examples and our results are in opposition to Lima-Mendez *et al.* (2015) who, based on their global spatial microbial networks, argued that biotic factors are more important than abiotic for structuring communities since a minority of the associations they found could be explained by environmental factors.

In the marine picorna-like viruses to eukaryotes environmental triplets, the dissolved oxygen had the greatest number of edges and most were negatively correlated. This could be because of a specific type of OTU associated with a phytoplankton bloom since the decrease in dissolved oxygen could show the demise of a phytoplankton bloom.

Relationship of niche to environmental drivers of diversity — The niche effects of OTUs correlated to the environment were examined as environmental triplets (Lima-Mendez *et al.*, 2015). The OTUs in these triplets revealed that the bacterial and eukaryotic communities were more frequently strongly linked to each other and to the environment than were the host-virus-environment links. This points to a stronger relationship of putative hosts to the abiotic factors, and that even at the OTU-level some bacteria and eukaryotes could be driven by similar parameters (Figure 4.6). The bacterial to T4-like myoviral communities had their strongest links with viral abundance. This is not surprising as the majority of viruses counted by flow cytometry would be bacteriophage (Brussaard, 2004b). Phosphate and nitrate+nitrite had the greatest number of links in the triplet networks. Differences between community- and OTU-level dynamics were likely from the dynamics of ephemeral OTUs. Changes in viral abundance and the viral community could have a partitioned effect on the bacterial community e.g. these changes might not affect the entire community in the same way (Chow *et al.*, 2013).

Within single networks, the links and modules were stronger than between different types of amplicons. Gilbert *et al.* (2011) found stronger connections within bacterial

networks than between bacteria and eukaryotes. However, they used different resolution data to examine the bacteria and the eukaryotes which could explain the different patterns. Similarly, Lima-Mendez *et al*. (2015) found that biotic factors were more able to predict the composition of communities than abiotic. This hypothesis was examined by looking at connections between communities and found that the highest number of connections were between eukaryotic and bacterial communities (Figure 4.5). The viral networks were more highly connected to themselves than the bacterial and eukaryotic communities (Table 4.1 and Figure 4.5). This would indicate that it is not always the same viruses that are infecting the same hosts, or at least they are not completely dominant since this analysis requires OTUs to be present more than 20% of the time. Lima-Mendez *et al*. (2015) looked at the connections of the viruses and found that 43% only interact with one host OTU and the rest only interact with a few. In the marine picorna-like communities, 8% of OTUs were connected to one putative host OTU and 24% were connected to five or less (Figure 4.7). In the T4-like myoviral to bacterial networks 22% of OTUs were connected to one putative host OTU and 58% were connected to five or less. This is in contrast to the overall network where 4% were connected only to one OTU and 18% to five or less. Considering only the connections between the eukaryotic and bacterial communities, 10% were connected to only one OTU and 38% to five or less. Chow *et al*. (2013) found that protist-bacteria networks were composed of many small hubs whereas the virus-bacteria was one big network. Also using LSA, Chow *et al*. (2014) saw many T4-like myoviral OTUs that were correlated to bacterial OTUs and some T4-like myoviral OTUs correlated strongly to multiple hosts suggesting a broad host range. Together these observations are consistent with the broader host ranges observed for the T4-like viruses compared to the more host-specific picorna-like viruses.

The overall amount of variation explained in the communities was low. Variance partitioning found that the marine picorna-like viral communities were more explained by chemical parameters than the eukaryotic communities; while this was surprising it is concordant with what was found in T4-like myoviral communities at SPOT (Chow

et al., 2014) where the community variation was more explained by the environment than the bacterial or eukaryotic communities. The bacterial communities in our study were more influenced by biotic parameters such as chlorophyll *a* and bacterial abundance than the eukaryotes. The marine picorna-like viral community had the most distinct seasonal progression (see Chapter 3 Adonis tests), so it not surprising that it has the highest variance attributed to time.

TWO LARGE NEGATIVELY CORRELATED MODULES DETECTED FOR PUTATIVE HOST NETWORKS — There were two large negatively correlated modules in the bacterial, eukaryotic and bacterial-eukaryotic networks (Figure 4.2 and Figure 4.4). Cram *et al.* (2015) suggest that modules can show differences in community states and these modules can represent seasonal patterns. In the eukaryotic, bacterial and bacterial to eukaryotic networks there were no taxonomic patterns found in the modules (Figure 4.2), which is similar to results in Williams *et al.* (2014). Conversely, modules detect the main phylogenetic groups in the viral-only networks (Figure 4.3). The viral OTUs sequenced are a smaller subset of the total viral community than the eukaryotic and bacterial OTUs are of their respective communities, so there is a greater chance that they would be temporally linked or are following specific distributions of hosts.

4.5.2 *Network statistics over time: temporal niches*

The overall network was determined to have a "small world topology" because of the degree distribution (Figure 4.7) and short path lengths found throughout the network (Table 4.1). This type of topology is quite common in microbial networks (e.g. Steele *et al.*, 2011; Chow *et al.*, 2014; Cram *et al.*, 2015; Lima-Mendez *et al.*, 2015; Comte *et al.*, 2015).

As with other microbial time series, e.g. Chow *et al.* (2013), the highest richness of OTUs and of community similarity was seen from Nov-Feb (parts of fall and winter seasons). The fall was the most highly connected network (Figure 4.5). In the northern

hemisphere bacterial diversity is highest in winter at higher latitudes (Ladau *et al.*, 2013). Ladau *et al.* (2013) found that the high richness in these communities was predicted by closeness to the thermocline, phosphate concentration, and day length. Short photoperiods are associated with high richness (Gilbert *et al.*, 2011; Ghiglione and Murray, 2012). Tan *et al.* (2013) describes how temporal niches are important for bacteria and can promote diversity in these communities. Thus dynamic communities allow this diversity to persist or accumulate in the fall/winter. One hypothesis is that with similar environmental parameters over the fall/winter there was more time for the organisms to diversify or more growth of organisms adapted to the environmental conditions. In networks, resistance and stability comes from having small world properties (Peura *et al.*, 2015). These small world networks, like most of the microbial networks found, are stable because of these properties and this could explain why there are not dramatic shifts over time even with highly dynamic environmental parameters. If one organism is lost, the network and its connections are not dramatically changed.

4.5.3 *Caveats*

Although the challenges with viral gene markers and PCR in general were previously discussed in the Introduction (p. 22), there are specific caveats for this network analysis of the co-occurrence of OTUs and the variance in communities over time. Caution must be taken in interpreting the associations discovered in co-occurrence studies since association does not necessarily mean interaction (Schluter, 1984). Additionally, only those taxa that occurred more than 30% of time were examined; so although individual OTUs were examined in the network analysis, these were abundant and common members of the community. This excluded ephemeral associations, which could have included important host-virus associations. However, it was preferred to filter the dataset conservatively rather than to add many false positives.

4.5.4 *Conclusions*

A one-year time series at Jericho Pier, in Vancouver B.C., provided a dynamic setting within which to examine the potential relationships of microbes and viruses. Based on this study of the viral, bacterial and eukaryotic communities, their co-occurrence, and co-variance over time, partitioning of the environment is an important driver of microbial community diversity. Overall there were more co-occurrences between different communities than between like communities, reinforcing the idea that associations among different communities are important. Highly connected groups within viral communities matched the phylogenetic groups found in Chapter 3, further strengthening the hypothesis of phylogenetic structuring of these communities. Based on the analysis of the environmental triplets, the environment likely a large role in filtering the host-virus pairs that occur in an environment and that occur seasonally.

CHAPTER 5

Viral and heterotrophic protistan control of a phytoplankton bloom

5.1 SUMMARY

Phytoplankton blooms are important ecological events in the coastal and open ocean that drive large drawdowns of CO_2, stimulate the remineralization of nutrients and promote succession in microeukaryotic and bacterial communities. Algal blooms can be composed of multiple species and summer algal blooms in the Salish Sea can often be dominated by members of the genus *Heterosigma* (raphidophyte). Studies have examined how *Heterosigma* blooms progress over time and the potential allelopathic interactions with other organisms, but not the viral, bacterial and microeukaryotic communities present during these blooms. During a summer algal bloom that was dominated by the genus *Heterosigma*, high-throughput amplicon sequencing was used to examine the dynamics of the viral, bacterial and eukaryotic communities every other day for 18 days. For context these samples were compared to a one-year time series from the same site. At the peak biomass of the summer algal bloom, the bloom was dominated at the peak of chlorophyll by a population of *Heterosigma* that was undetectable two days later, but returned 10 days later to form a smaller bloom. A succession of phytoplankton occurred during the summer algal bloom in which dinoflagellates formed smaller sub-blooms during the 18 days that did not co-occur with *Heterosigma* suggesting an antagonistic relationship with *Heterosigma* or a preference for different environmental conditions or selective predation of these phytoplankton. The bacterial and eukaryotic communities mirrored each other at the class level suggesting that bacterial communities are closely

associated with specific taxa of phytoplankton. The succession of phytoplankton blooms stimulated diversity in all communities and this increased diversity could be a response to disturbance in the communities. Probing deeper into the diversity revealed strain-level dynamics where prominent OTUs in the bloom were decomposed into sub-types using Shannon entropy decomposition (oligotyping). The succession of these oligotypes was linked to viral selective pressure early in the bloom and to protistan predation later in the bloom, thus illuminating the strain-specific succession of phytoplankton species during blooms.

5.2 INTRODUCTION

Phytoplankton are a functional group that performs up to half of the photosynthesis on Earth (Field *et al.*, 1998). Many phytoplanktonic taxa, in addition to being persistent members of the marine ecosystem, can also rapidly increase in cell number and overall biomass to forming "blooms." Blooms tend to occur when there in an influx of nutrients via mixing or terrestrial inputs (Behrenfeld and Boss, 2014) and can cause large draw-downs of CO_2 and can have large effects on biogeochemical cycles (Alkire *et al.*, 2012). Blooms can boost the rate of assimilation of nutrients like nitrogen and phosphorus in the ecosystem and eventually increased bacterial remineralization (Buchan *et al.*, 2014).

Phytoplankton blooms can comprise single or mixed species of diatoms, dinoflagellates, raphidophytes, and cyanobacteria (Harrison *et al.*, 1983; Domingues *et al.*, 2005), and typically occur in the spring, summer, and fall under specific conditions. Some of these blooms can be toxic to fish and other organisms and are a concern for aquaculture and shellfish collection (Nakamura *et al.*, 1998; Oda *et al.*, 1998). One important bloom-forming example is the raphidophyte, *Heterosigma akashiwo*, which is a eukaryotic alga that is toxic to fish (Horner *et al.*, 1997; Lewitus *et al.*, 2012; Powers *et al.*, 2012). The toxicity to fish is mediated by gill-clogging mucus produced either by the alga or the fish itself (Nakamura *et al.*, 1998; Oda *et al.*, 1998), or by alga-produced neurotoxins

(Khan *et al.*, 1997; Ono *et al.*, 2000). *Heterosigma akashiwo* tends to bloom under specific conditions when the water temperature reaches 15°C and often a decrease in salinity below 15 psu (Taylor and Haigh, 1993). *Heterosigma* can form cysts that can stay dormant and eventually seed a bloom (Powers *et al.*, 2012). Blooms of *Heterosigma* have been documented in many studies along the B.C. coast (e.g. Haigh *et al.*, 1992; Taylor *et al.*, 1994), but the communities of microbes and viruses associated with the progression of these blooms is unexplored.

Phytoplankton, especially bloom formers, can heavily influence their environments and other associated organisms. For example, there are often correlations between bacterial production and concentration of chlorophyll *a* (Cole, 1982; Croft *et al.*, 2005; Sher *et al.*, 2011). Shifts in the phytoplankton community (i.e. the phytoplankton in a geographic area) can be followed by shifts in the bacterial community (Pinhassi *et al.*, 2004), including changes in specific lineages such as the Flavobacteria, Bacteroides and Alphaproteobacteria (Buchan *et al.*, 2014). Specific bacterial populations (i.e. the bacterial species occurring in the same area) associated with phytoplankton blooms are frequently seen in 16S rRNA amplicon studies (Buchan *et al.*, 2014) and also in metagenomic samples, providing information on specialized transporters, and metabolic pathways (Teeling *et al.*, 2012). Bacteria can support a bloom by recycling nutrients, but they also compete with phytoplankton for nutrients (Buchan *et al.*, 2014). When there is a large influx of carbon from phytoplankton, it has been estimated that half of the carbon is processed by heterotrophic bacteria while the rest enters the food chain or sinks out of the photic zone (Buchan *et al.*, 2014). During blooms, the structure of the bacterial communities may remain even (Delmont *et al.*, 2014), but often there are shifts in the communities to lineages with increased carbon cycling (Landa *et al.*, 2016) and efficient degradation of by-products of phytoplankton (e.g. dimethylsulfoniopropionate (DMSP), transparent exopolymer particles (TEP), etc.) and decaying cells (Buchan *et al.*, 2014).

Shifts in the communities and populations of eukaryotes can occur during phytoplankton blooms. For example, in blooms of the cosmopolitan dinoflagellate *Alexan-*

drium minutum, genetic diversity within the population can develop quickly (Dia *et al.*, 2014). Conversely, during a bloom of the colony-forming haptophyte, *Phaeocystis*, the number of rare eukaryotic taxa and zooplankton biomass increased (Monchy *et al.*, 2012). Understanding the role of bottom-up controls (e.g. nutrients) vs. top-down controls (grazing, viral lysis) is important for understanding the conditions under which blooms occur and progress. Blooms can be initiated by inputs of nutrients such as nitrogen, phosphorus, iron and silicate (Buchan *et al.*, 2014) and by light availability. Ratios of these nutrients can also direct the establishment of specific bloom species (Buchan *et al.*, 2014). Blooms can be terminated by bottom-up forces such as nutrient limitation (Dale *et al.*, 1999; Blain *et al.*, 2004; Mahadevan *et al.*, 2012; Chiswell *et al.*, 2015), or by top-down forces such as grazing by protists or zooplankton (Rosetta and McManus, 2003) or viral lysis (Bratbak *et al.*, 1993, 1996; Lawrence and Suttle, 2004; Brussaard, 2004a).

High abundances of viruses have often been associated with phytoplankton blooms (Bratbak *et al.*, 1990; Matteson *et al.*, 2012). For example, in the bloom forming coccolithophore, *Emiliania huxleyi*, viruses (*Emiliania huxleyi* viruses (EhV)) are often associated with blooms (Wilson *et al.*, 2002a,b). Conversely, photosynthesis system reduction has been observed when a bloom is controlled by viruses (Kimmance *et al.*, 2014). Blooms can terminate with one viral genotype dominating; whereas, in other blooms there can be multiple genotypes (Schroeder *et al.*, 2003; Martinez Martinez *et al.*, 2007; Sorensen *et al.*, 2009; Highfield *et al.*, 2014). However, it is unknown what influences these different scenarios.

Heterosigma akashiwo strains can be infected by DNA viruses (Nagasaki and Yamaguchi, 1997) and RNA viruses (Tai *et al.*, 2003). The susceptibility of *Heterosigma akashiwo* to viruses is often dependent on the stage of the bloom when the strain was isolated. Tarutani *et al.* (2000) found that during *Heterosigma akashiwo* blooms there were multiple host strains. Viral control of these strains was apparent since strains isolated later in the bloom were more resistant to viruses than those isolated early in the bloom.

Blooms disturb ecosystems by drastically changing the biological or physical environment. Disturbances can disrupt an ecosystem, but can also promote diversity, increased richness and release of resources (Connell, 1978; Holt, 2008). Microbial communities can respond to disturbance events by not changing (stability), by returning to the way they were before the event (resilience), by changing composition but remaining functionally the same, or by becoming markedly different after the disturbance (Allison and Martiny, 2008). As seen in the network analysis in Chapter 4 (p. 123) and references within, microbial communities appear to shift in composition, but diversity is maintained over time, thus, these communities appear resilient to change during a disturbance.

Two main hypotheses were investigated in a summer algal bloom. First, since there is evidence for viral control of phytoplankton blooms in mesocosm studies (Wilson *et al.*, 2002b; Martinez Martinez *et al.*, 2007; Larsen *et al.*, 2008), it suggests that high-throughput sequencing could be used to detect viral control of a naturally occurring summer algal bloom. These blooms can be dominated in this region by *Heterosigma* and viruses that infect *Heterosigma* have been isolated and detected in the coastal waters of British Columbia. Additionally, based on the strain-level specificity of viruses infecting phytoplankton, it was hypothesized that there will be strain-level progression of the bloom-forming phytoplankton as they experience selective pressures from specific viruses. Second, since phytoplankton blooms can be considered to be disturbances, and microbial communities are theorized to have high resilience, it was hypothesized that the bacterial, eukaryotic, and viral communities will behave as disturbed communities during the bloom, show higher richness following the bloom, and then return to their initial state.

To examine these hypotheses a summer algal bloom was sampled every other day during its initiation, peak and demise at Jericho Pier in Vancouver, B.C. and high throughput sequencing of marker genes was used to follow temporal changes in the taxonomic composition of T4-like myoviruses, marine picorna-like viruses, bacteria and eukaryotes, and the richness, evenness, and composition of these communities as described in

Chapter 3 (p. 139), bacteria and eukaryotes. The richness, evenness, and composition of all communities were examined over time in addition to a variety of environmental parameters. The strain-level variation of the bloom-forming taxa was also examined using Shannon Entropy decomposition (oligotyping Eren *et al.*, 2013). These changes were examined in the context of a variety of environmental parameters.

5.3 MATERIAL AND METHODS

5.3.1 *Sample collection*

The samples were collected and processed largely as outlined in Chapter 3 with the following modifications: Samples were collected from Jericho Pier (49° 16'36.73N, 123° 12'05.41W) in British Columbia, Canada. Jericho Pier (JP) is adjacent to the shoreline, in a well-mixed location with mixed semi-diurnal tides. In order to get representative water samples and enough material for viral extraction sixty liters of water was pumped from the 1-m depth every two days at the daytime high tide from 21 June 2011 to 5 July 2011 (8 samples). The samples from the full-year time series June 2010 and July 2011 were used for comparison. Salinity and temperature were measured using a YSI probe (Yellow Springs, Ohio, USA). For all samples, the water was pre-filtered through a 65 μm Nitex mesh and filtered sequentially through 142-mm diameter, 1.2-μm nominal pore-size glass-fiber (GC50 Advantec MFS, Dublin, CA., USA) and 0.22 μm pore-size polyvinyldine (Millipore, Bedford, MA, USA) filters. The filtrate, containing the viral size fraction, was concentrated to ~500 mL (viral concentrate) using tangential flow ultrafiltration with a 30kDa MW prep-scale Spiral Wound TFF-6 cartridge (Millipore) (Suttle *et al.*, 1991).

5.3.2 *Nutrients*

Phosphate, silicate and nitrate+nitrite concentrations were determined in duplicate 15-mL seawater samples filtered through 0.45 μm pore-size HA filters (Millipore) and stored

at -20°C until air-segmented continuous-flow analysis on a AutoAnalyzer 3 (Bran+Luebbe, Norderstedt, Germany). Chlorophyll *a* (Chl *a*) was determined in triplicate by filtering 100 mL of seawater onto 0.45 μm pore-size HA filters (Millipore), and storing the filters in the dark at -20°C until acetone extraction and then analysed fluorometrically (Parsons *et al.*, 1984).

5.3.3 *Enumeration of bacteria and viruses*

Samples for viral and bacterial abundances were taken at each sampling point by fixing duplicate cryovials containing 980μL of sample with final concentration of 0.5% glutaraldehyde (EM-grade), freezing in liquid nitrogen and storing at -80°C until processing. Flow cytometry samples were processed as in Brussaard (2004b). Briefly, viral samples were diluted 1:10 to 1:10 000 in sterile 0.1 μm filtered 1X TE, stained with SYBR Green I (Invitrogen, Waltham, MA, USA) at a final concentration of 0.5 x 10^{-4} of commercial stock, heated for 10 minutes at 80° C and then cooled in the dark for 5 minutes before processing. Bacterial samples were diluted up to 1:1000 in sterile 0.1 μm filtered 1xTE, stained with SYBR Green I at a final concentration of 0.5 x 10^{-4} of commercial stock, and incubated in the dark for 15 minutes before processing. All samples were processed on a FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA) with viral and bacterial samples run for 1 min at a medium or high flow rate, respectively. Event rates were kept between 100 to 1000 events per second and green fluorescence and side scatter detectors were used. Data were processed and gated using Cell-Quest software (Becton-Dickinson).

5.3.4 *Extraction of viral nucleic acids*

The viral concentrate was filtered twice through 0.22μm pore-size Durapore PVDF filters (Millipore) in a sterile Sterivex filter unit (Millipore). The filtrate, containing viral-sized particles, was pelleted by ultracentrifugation (Beckman-Coulter, Brea, California, USA) in a SW40 rotor at 108 000 *g* for 5 h at 12°C. The pellet was resuspended overnight

in 100 μL of supernatant at 4°C. To digest free DNA, the pellets were incubated with 1U μL^{-1} DNAse with a final concentration 5 mM $MgCl_2$ for 3 h at room temperature. Nucleic acids were extracted using a Qiamp Viral Minelute spin kit (Qiagen, Hilden, Germany) according to the manufacturer's directions.

5.3.5 *PCR amplification of T4-like myoviral marker gene*

To target the marine T4-like myoviral capsid protein gene (gp23), PCRs were set up as in Filée *et al.* (2005). Briefly, each reaction mixture (final volume, 50 μL) consisted of 2 μL template DNA, 1x (final concentration) PCR buffer (Invitrogen, Carlsbad, California, USA), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 40 pmol of MZIA1bis and 40pmol of MZIA6, and 1 U Platinum Taq DNA polymerase (Invitrogen) and program conditions as in Table 3.1.

5.3.6 *PCR amplification of picorna-like virus marker gene*

Half of each viral extract was used to synthesize cDNA. To remove DNA, the extracted viral pellets were digested with DNase 1 (amplification grade) (Invitrogen). The reaction was terminated by adding 2.5 mM EDTA (final concentration) and incubating for 10 min at 65°C. Complementary DNA (cDNA) was generated using Superscript III reverse transcriptase (Invitrogen) with random hexamers (50 ng μL^{-1}) as per the manufacturer.

PCR was performed using primer set MPL-2 to target the RdRp of marine picorna-like viruses (Culley and Steward, 2007). Each reaction mixture (final volume, 50 μL) consisted of 50 ng of cDNA, 1x (final concentration) PCR buffer (Invitrogen), 2 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 1 µM of each primer, and 1 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1. Products were run on a 0.5X TBE 1% low melt gel, excised and extracted using Zymoclean Gel DNA Recovery Kit (Zymo) as per the manufacturer and a final elution step of 2x10 μL EB buffer (Qiagen).

5.3.7 *Filtration and extraction of marine bacteria and eukaryotes*

One liter of seawater was taken from the sixty liters and filtered through a 0.22μm pore-size Durapore PVDF 47 mm filter (Millipore) in a sterile Sterivex filter unit (Millipore). The filter was either stored at -20°C until extraction or immediately extracted as follows. Filter extraction was as in Short and Suttle (2003). Briefly, filters were aseptically cut and incubated with lysozyme (Sigma-Aldrich, St. Louis, MO, USA) at a final concentration of 1mg mL^{-1} for 2 h at 37°C. Sodium dodecyl sulfate was added at a final concentration of 0.1 % (w/v) and each filter was put through three freeze-thaw cycles. Proteinase K (Qiagen) was then added to a final concentration of 100 μg mL^{-1} and incubated for 1 h at 55°C. DNA was sequentially extracted using equal volumes of phenol:chloroform:IAA (25:24:1), and chloroform:IAA (24:1). DNA was precipitated by adding NaCl to a final concentration of 0.3M and by adding 2X the extract volume of ethanol. Samples were incubated at -20°C for at least 1 h and then centrifuged for 1 h at 20 000 *g* at 4°C. Extracts were washed with 70 % ethanol and were resuspended in 50 μL EB buffer (Qiagen).

5.3.8 *PCR amplification of bacterial and eukaryotic ribosomal sequences*

PCR targeting eukaryotes used primers Euk1209f and Uni1392r as in Diez *et al.* (2001). These primers target positions 1423 to 1641 and includes the variable region V8. Each reaction mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 0.3 μM of each primer, and 2.5 U Platinum Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1.

PCR targeting bacteria used primers 341F (Baker *et al.*, 2003) and 907R (Muyzer *et al.*, 1995). These primers target the v3 to v5 regions. PCRs were run with the following conditions: each reaction mixture (final volume, 50 μL) consisted of 2 μL template, 1x (final concentration) PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (Bioline, London, UK), 0.4 μM of each primer, and 1 U Platinum

Taq DNA polymerase. The reaction was run in a PCR Express thermocycler (Hybaid, Ashford, UK) with program conditions as in Table 3.1.

5.3.9 *Sequencing library preparation*

Construction — PCR products not requiring gel excision were purified after PCR using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. Cleaned products were resuspended in 30 μL EB buffer (Qiagen). All products were quantified using the Picogreen dsDNA (Invitrogen) assay using Lambda DNA (Invitrogen) as a standard. Sample concentrations were read using iQ5 (Bio-Rad, Hercules, CA, USA) and CFX96 Touch systems (Bio-Rad). Pooled libraries were constructed using one of each of the amplicons at a concentration so that their molarity would be similar and the total concentration of the pool was ~700-900 ng. Pooled amplicons were concentrated using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. NxSeq DNA sample prep kit 2 (Lucigen, Middleton, WI, USA) was used as per manufacturer's directions with either NEXTFlex 48 barcodes (BioO, Austin, USA) , NEXTflex 96 HT barcodes (BioO), or TruSeq adapters (IDT, Coralville, Iowa). Libraries were cleaned up using AMPure XP beads (Beckman Coulter) at a ratio of 0.9:1 beads:library.

Quantification and quality control of libraries — Libraries were checked for small fragments (primer dimers and/or adapter dimers) using a 2100 Bionanalyzer (Agilent, Santa Clara, CA, USA) with the High Sensitivity DNA kit (Agilent). The concentration of libraries was quantified using Picogreen dsDNA assay as above. The libraries were quantified and checked for amplifiable adapters using the Library Quantification DNA standards 1-6 (Kappa Biosystems, Wilmington, USA) with the SsoFast EvaGreen qPCR supermix (Bio-Rad) using 10 μL EvaGreen master mix, 3 μL of 0.5uM F primer, 3 μL of 0.5 μm R primer and 4 μL of 1:1000, 1:5000 and 1:10000 dilutions of the libraries in triplicate on iQ5 (Bio-Rad) and CFX96 Touch qPCR machines. Cycling parameters were as follows: 95°C for 30s, 35 cycles of 95°C for 5s, 60°C for 30s, and the melt curve gener-

ation from 65°C to 95°C in 0.5°C steps (10s/step). Quantification from both Picogreen and qPCR assays were used to determine final pooling of all libraries before sequencing. Libraries were sequenced using 2x250bp PE Miseq (Illumina, San Diego, USA) sequencing at Génome Québec Innovation Centre at the McGill University (Montreal, QC, Canada), and 2x300bp PE Miseq (Illumina) sequencing at UBC Pharmaceutical Sciences Sequencing Centre (Vancouver, BC, Canada) and at UCLA's Genoseq (Los Angeles, CA, USA).

5.3.10 *Initial sequence processing*

Libraries were either split by the sequencing centre using CASAVA (Illumina) or split by the user using the Miseq Reporter software (Illumina). Sequence quality was initially examined using FastQC (Andrews, 2015). Contaminating sequencing adapters were removed using Trimmomatic version 0.32 (Bolger *et al.*, 2014) and the quality of the sequencing library further examined using fastx_quality (Gordon, 2014). Libraries were further split into individual amplicons (i.e. 18S, 16S, gp23 and MPL) and then, if the expected overlap of the paired-end reads was 40bp or more, the paired reads were merged using PEAR (Zhang *et al.*, 2014). Sequences were then quality trimmed using Trimmomatic with the default quality settings. Sequences were aligned to known sequences (Silva 119 database (Quast *et al.*, 2013) for 16S and 18S rRNA genes) using align.seqs in mothur 1.33.3 (Schloss *et al.*, 2009) and those not aligned were removed. Viral sequences were queried using BLAST against databases containing the gene markers of interest and sequences with an e-value below 10^{-3} were kept.

5.3.11 *Chimera checking, OTU picking and read normalization*

The 16S and 18S rRNA gene sequences were checked for chimeras using USEARCH version 8.0.1517 (Edgar, 2010) with the Gold reference database. Unique, non-chimeric sequences were clustered at 97% similarity. Taxonomy for the 16S and 18S rRNA gene sequences was assigned using mothur (Wang-type algorithm) and the taxonomy in Silva

119 (Quast *et al.*, 2013). Resolution of targets and database prevented the assignment of taxonomy below the level of genus for most OTUs (notably *Heterosigma* could not be identified to species level). For the viral targets sequences were chimera-checked using USEARCH denovo and reference (Edgar, 2010). Viral sequences were then translated using FragGeneScan 1.20 (Rho *et al.*, 2010). Viral reads were clustered using USEARCH (Edgar, 2010) at 95% similarity for MPL, and 95% similarity for T4-like myoviruses. Operational taxonomic unit (OTU) tables for all targets were constructed using USEARCH (Edgar, 2010). Sequences were normalized for this project by date and by target using vegan (Oksanen *et al.*, 2015).

5.3.12 *Oligotyping*

Oligotypes were chosen from three eukaryotic OTUs (97% similarity) that dominated the microeukaryotic community during the bloom. Oligotyping partitions the OTU into subdivisions based on variability at positions of high nucleotide variability, thus it can reveal finer-scale dynamics within OTUs. One OTU was taxonomically classified as a raphidophyte and the other two as dinoflagellates. To normalize sequencing effort, 10 000 random eukaryotic reads were selected and classified. From these, the reads from the top OTU classified as a raphidophyte, and the two most abundant dinoflagellate OTUs were used for Shannon Entropy decomposition aka oligotyping (Eren *et al.*, 2013). The same procedure was performed for the dinoflagellate OTUs. A similar procedure was performed with the marine picorna-like viral OTUs, where 2500 random reads were chosen.

5.3.13 *Community similarity and Mantel tests*

Bray-Curtis distance matrices were constructed from the normalized OTU abundance tables. Mantel tests were performed by comparing the community distance matrices to each other and to distance matrices of environmental parameters using vegan.

5.4 RESULTS

5.4.1 *Environmental parameters*

During the two week bloom period there were dramatic shifts in some environmental parameters (Figure 5.1). The most striking were measurements of chlorophyll *a* which were already high at 46.5 μg/L and then increased to 168.7 μg/L (+- error 26.56 μg/L) on 23 June 2011. The chlorophyll *a* measurements were cross-referenced with secchi disk measurements and with chlorophyll estimates from the whole Strait of Georgia where large blooms had also been reported for these dates (Irvine and Crawford, 2012).

5.4.2 *High viral abundance followed high bacterial abundance*

The viral abundance reached its lowest measured abundance (1.65×10^7 viruses per mL) at the chlorophyll peak on 23 June 2011 but increased to a maximum of 7.145×10^7 viruses per mL on 29 June 2011. This peak in viral abundance lagged behind a peak in bacterial abundance two days earlier of 5.135×10^6 cells per mL (Figure 5.1).

Temperature and salinity were stable during this period and ranged from 14.8°C to 18.6°C and 7.6 psu to 13 psu. Nutrients (silicate, phosphate and nitrate+nitrite) were also mostly stable except for a spike in phosphate on 27 June 2011 to 1.02 μM (up from 0.04 μM). There was a similar spike seen in nitrate+nitrite concentration beginning on 29 June 2011.

5.4.3 *Richness and evenness of communities during the bloom*

The richness (Figure 5.3) and Pielou's evenness (Figure 5.4) of viral, bacterial, and microeukaryotic communities fluctuated during the bloom. The lowest richness was observed in the eukaryotic communities during the peak of the *Heterosigma* blooms (21 June and 1 July 2011) and the highest was observed on 3 July, nine days after the peak of the first *Heterosigma* bloom. The T4-like myoviruses had the highest richness on 27 June (2 days before the peak in viral abundance) and the lowest on 3 July 2011.

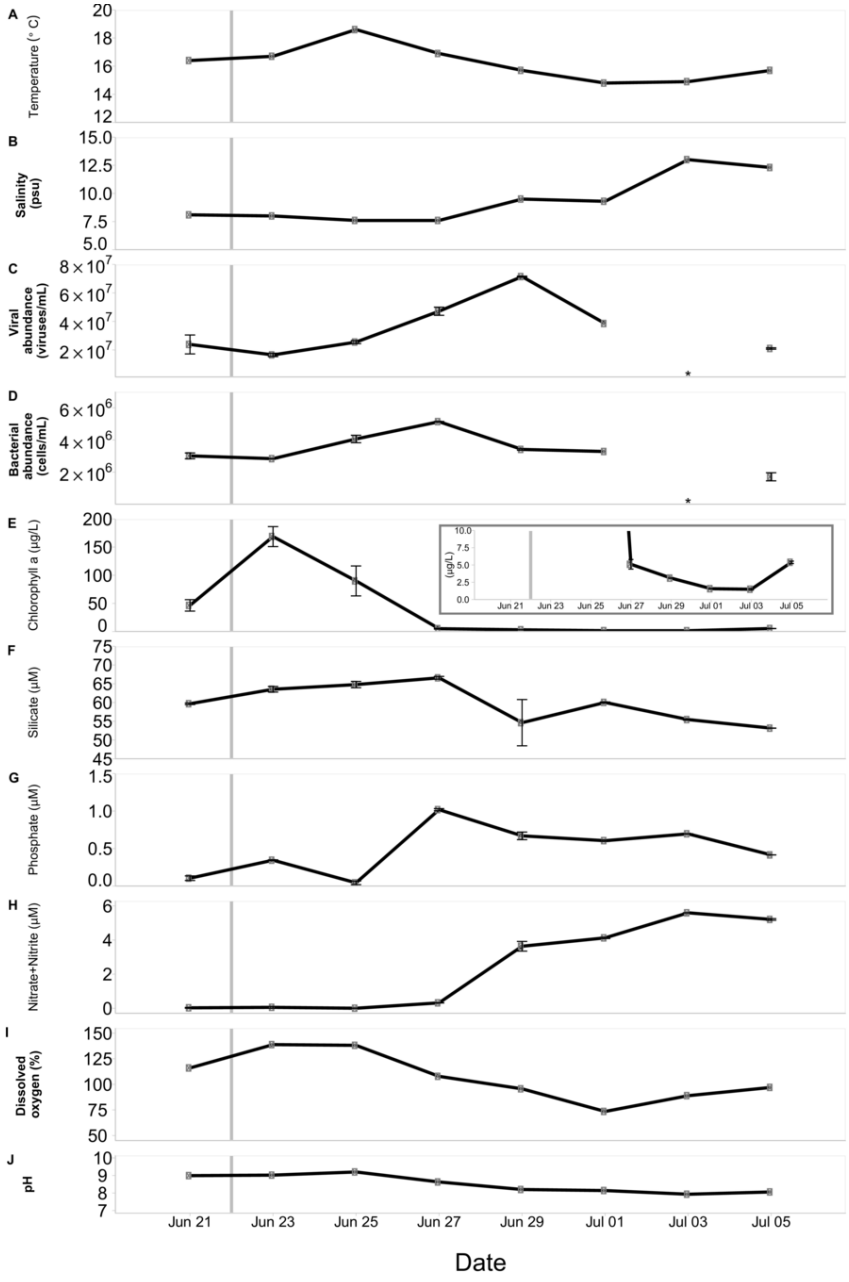

FIGURE 5.1: Environmental parameters during summer algal bloom 21 June -5 July 2011. Error bars represent standard error of the mean. A) temperature, B) salinity, C) viral abundance, D) bacterial abundance, E) chlorophyll *a*, inset is chlorophyll *a* with shorter y-axis, F) silicate, G) phosphate, H) nitrate+nitrite I) dissolved oxygen (percent saturation), J) pH. Stars (*) represent missing data. Grey vertical line demarcates the beginning of summer, 22 June 2011.

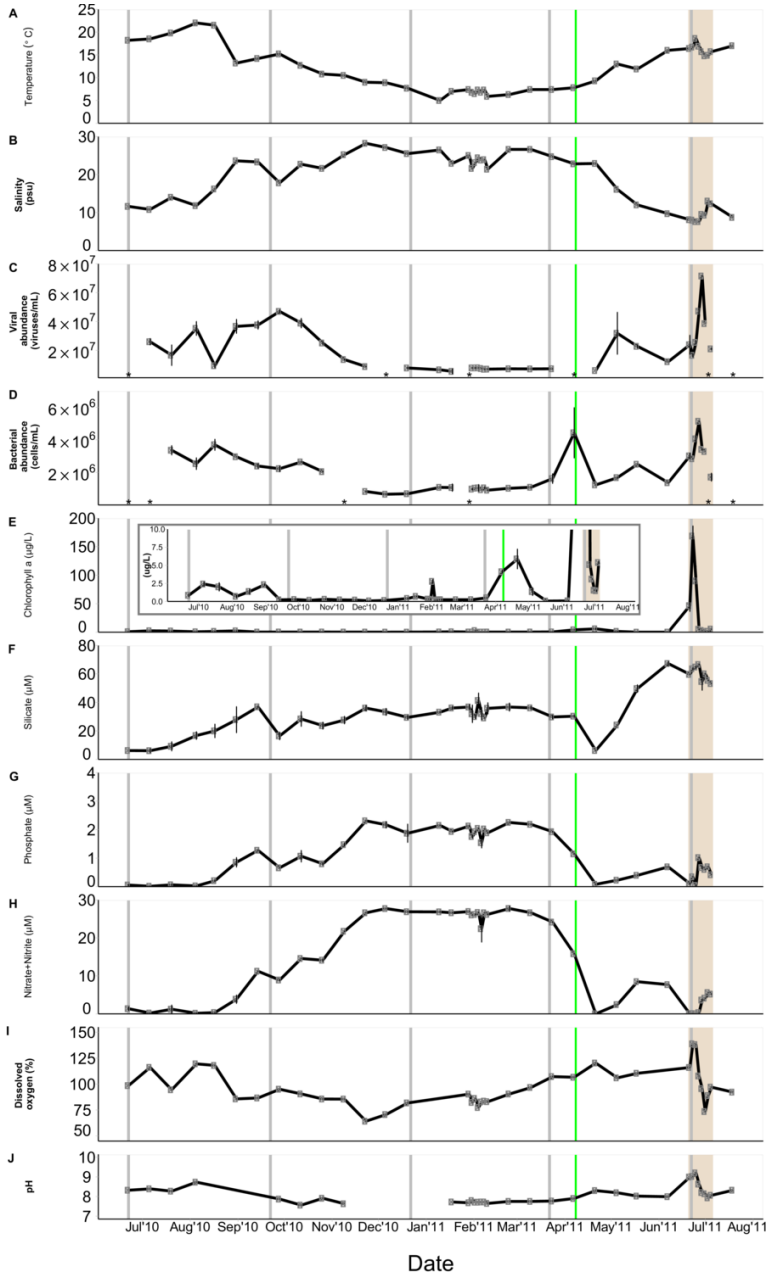

FIGURE 5.2: Environmental parameters during 1 year time series at Jericho Pier. Error bars represent standard error of the mean. A) temperature, B) salinity, C) viral abundance, D) bacterial abundance, E) chlorophyll *a*, inset is chlorophyll *a* with shorter y-axis, F) silicate, G) phosphate, H) nitrate+nitrite I) dissolved oxygen, and J) pH. Stars (*) represent missing data. The green vertical line indicates the spring diatom bloom. Grey vertical lines indicate seasons. Brown background indicates time of summer algal bloom.

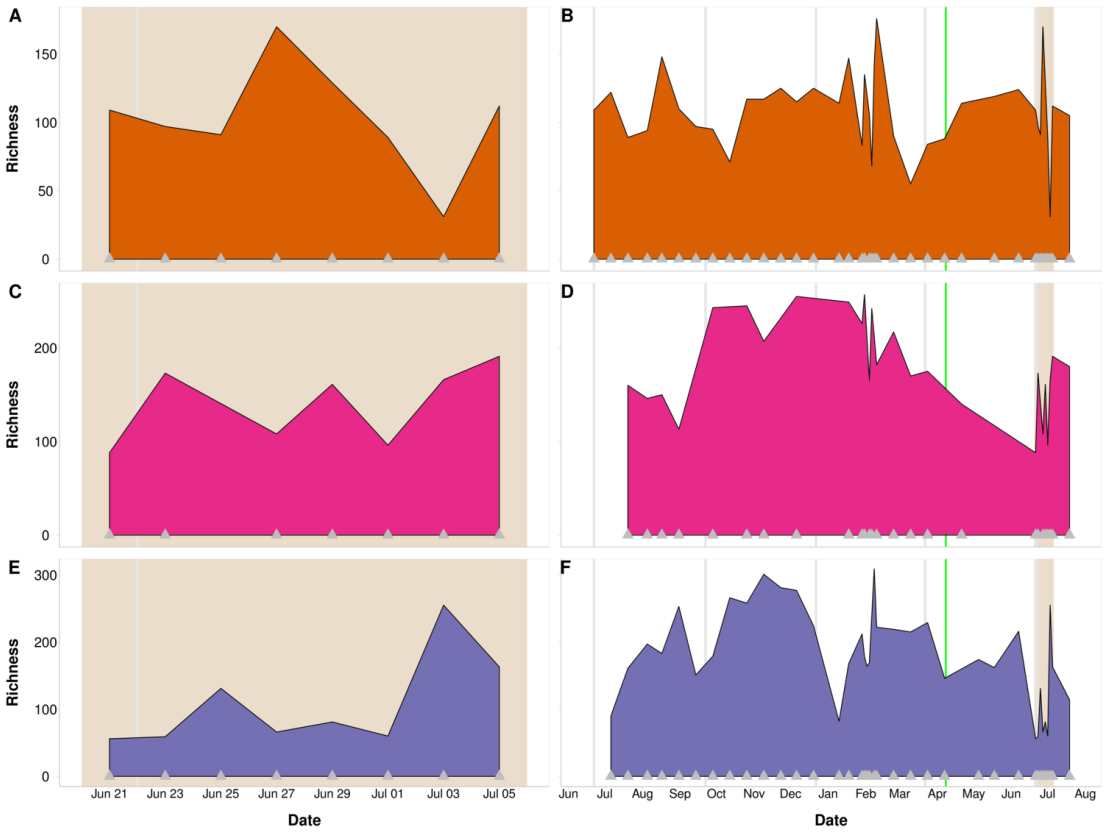

FIGURE 5.3: Richness of amplicons during the summer algal bloom (left panels) and overall 13 month time series (right panels). Richness calculated as number of observed OTUs after normalizing by sampling effort. A) and B) T4-like myoviruses C) and D) Bacteria E) and F) Eukaryotes. Grey arrows on x-axis indicate sampling time points for each amplicon. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom.

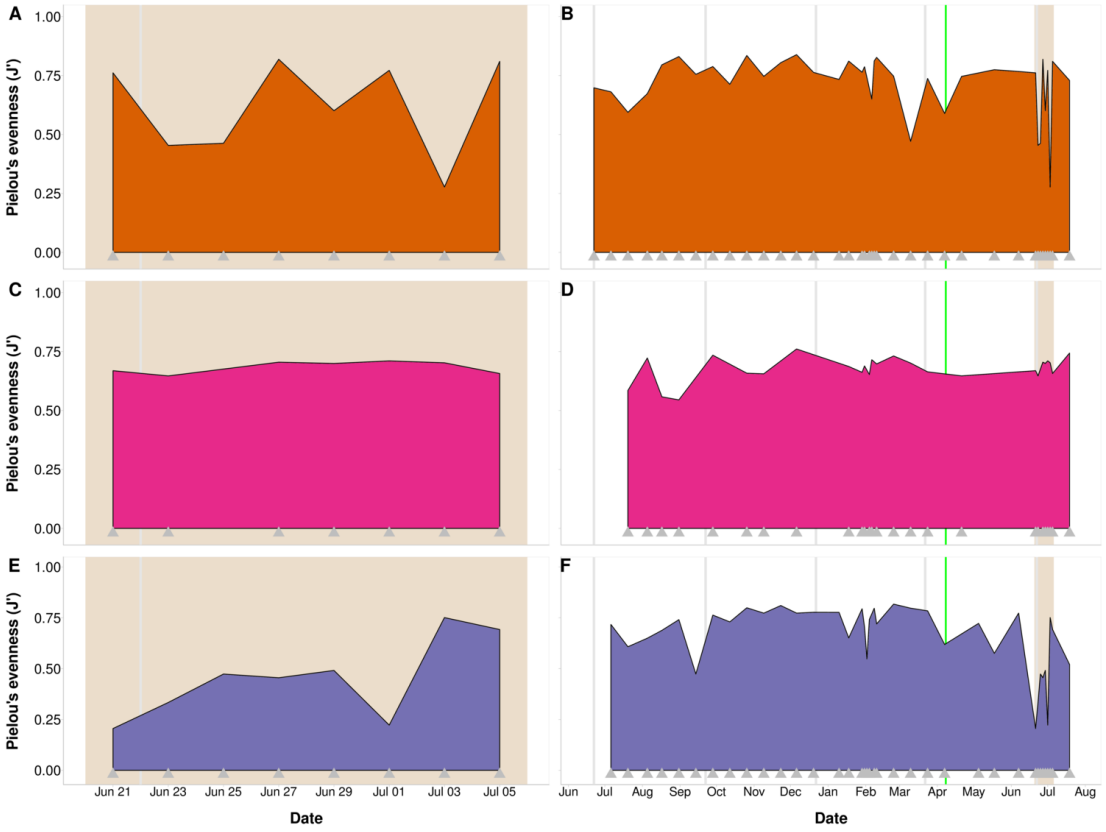

FIGURE 5.4: Evenness of amplicons during the summer algal bloom (left panels) and overall 13 month time series (right panels). Evenness calculated as Pielou's evenness (Pielou, 1966). A) and B) T4-like myoviruses C) and D) Bacteria E) and F) Eukaryotes. Grey arrows on x-axis indicate sampling time points. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom.

5.4.4 *Top 20 most abundant OTUs during the bloom*

The highest richness in the T4-like myoviral communities was observed on June 27 when the viral community composition was changed relative to the previous day (Figure 5.5A). The dominant OTUs from the one-year time series were in low abundance or undetectable during the bloom, and the bloom community was not made up of the top 20 most abundant OTUs found during the rest of the year (Figure 5.5A and B).

The relative abundances of the top 20 bacterial OTUs in the community was more stable during the bloom than observed for the T4-like myoviruses. At the peak of the first *Heterosigma* bloom on June 23, a Flavobacteriaceae OTU (bacterial OTU 6) became a dominant member of the bacterial community. In contrast, an alphabacterial OTU in the family Rhodobacteraceae (bacterial OTU 2) was persistent throughout the time series, but became dominant after July 1, and made up more than 35% of the community on July 5 (Figure 5.5C and D).

The eukaryotic community also had large temporal dynamics. Before the peak in chlorophyll *a* the community was dominated by *Heterosigma* (eukaryotic OTU 2) (Figure 5.5E). After chlorophyll peaked, the OTUs increased from 189 to 336 (Figure 5.3E and F), with different OTUs dominating the communities. On June 29 *Heterosigma* (eukaryotic OTU 2) dominated the community again before being replaced by a higher diversity of eukaryotic OTUs.

The marine picorna-like viral community, as seen in Chapter 3 (p. 68), is much less even than the other communities (Figure 5.5G and H). A higher percentage of the marine picorna-like virus communities was comprised of the top 20 OTUs than the T4-like myoviruses, bacteria and microeukaryotes. In the picorna-like viral community one OTU dominated the community during the summer algal bloom (OTU 288).

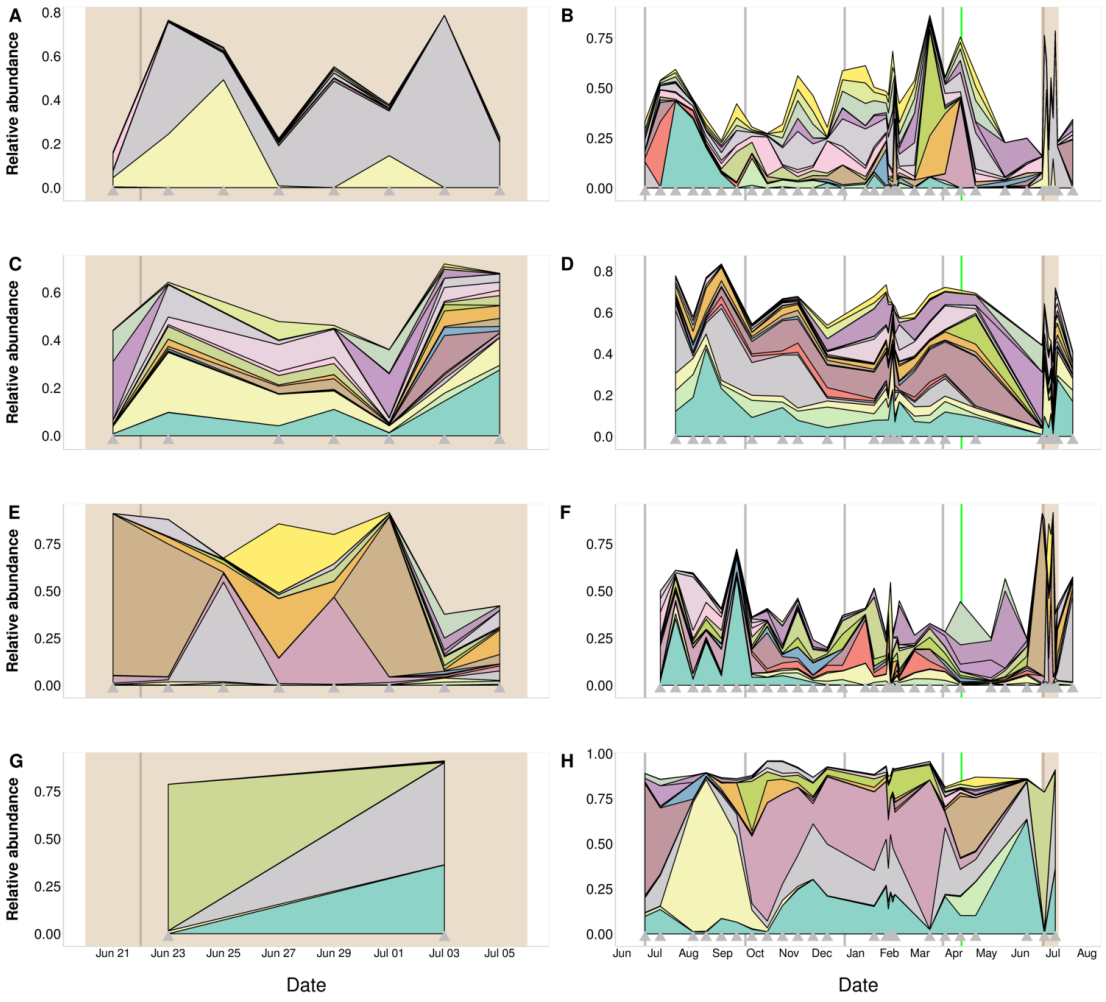

FIGURE 5.5: Top 20 most relatively abundant OTUs from each community during the summer algal bloom (left panels) and over the entire time series (right panels). A) and B) T4-like myoviruses, C) and D) Bacteria, E) and F) Eukaryotes, G) and H) Marine picorna-like viruses. Each contour represents a different OTU. Grey arrows on x-axis indicate sampling time points. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom.

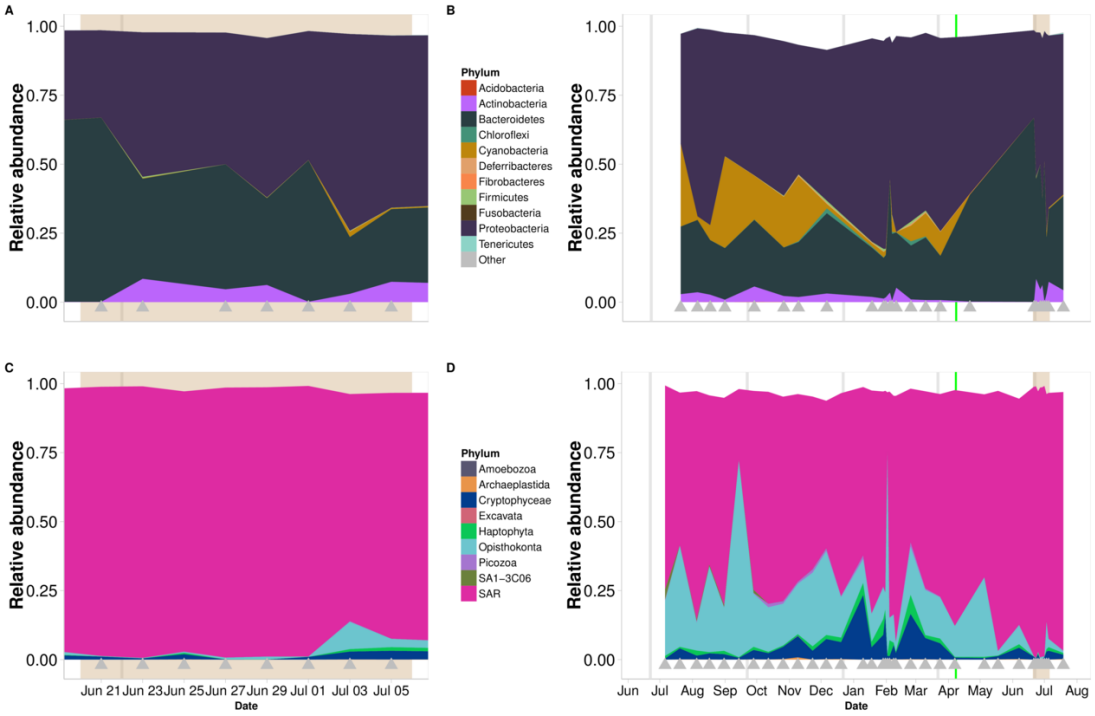

FIGURE 5.6: Relative abundance of bacterial (top panels) and eukaryotic (bottom panels) OTUs classified by phyla, during the summer algal bloom (left panels) and annually (right panels). Grey arrows on x-axis indicate sampling date. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom. In the bacterial community phyla representing less than 1% of relative abundance are grouped together into 'other' category. Classifications were done using the Wang algorithm as implemented in mothur (Schloss *et al.*, 2009) and using the Silva 119 database (Quast *et al.*, 2013)

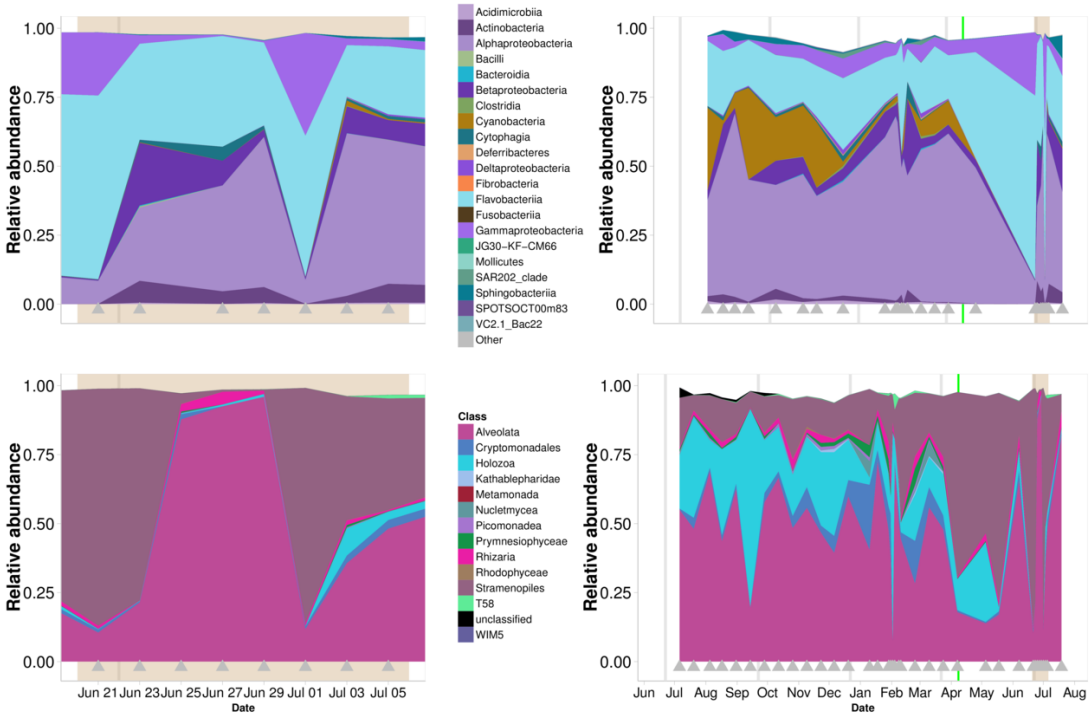

FIGURE 5.7: Relative abundance of bacterial (top panels) and eukaryotic (bottom panels) OTUs classified by class, during the summer algal bloom (left panels) and annually (right panels). Grey arrows on x-axis indicate sampling time points. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom. In the bacterial community classes representing less than 1% of relative abundance are grouped together into the 'other' category. Classifications were done using the Wang algorithm as implemented in mothur (Schloss *et al.*, 2009) and using the Silva 119 database (Quast *et al.*, 2013)

5.4.5 *Changes in OTUs or groups in bacterial and eukaryotic communities mirrored each other*

There were marked shifts in community composition during the study. For bacterial classes, the most dramatic shifts occurred in Gammaproteobacteria, which spiked in relative abundance during the bloom on June 21 and July 1 (Figure 5.7). At the peak of the bloom, the proportion of Betaproteobacteria and Flavobacteria was also higher, along with the rise and fall of a group of Cytophagia between the peak of the bloom and June 29. The eukaryotes were dominated by the SAR supergroup throughout the year (Figure 5.6) with Stramenopiles and Alveolates switching in relative dominance during the summer algal bloom (Figure 5.7).

5.4.6 *Individual eukaryotic OTUs*

The dominant eukaryotic OTU during the bloom was a raphidophyte (Figure 8A) that dominated between June 21 and June 23 and then again on July 1, while a ciliate (family Intramacronucleata) comprised >40% of the relative abundance on June 27 and 29 (Figure 5.8C). Dinoflagellates were at times dominant, and during the summer algal bloom three different OTUs dominated at different times (Figure 5.8G).

5.4.7 *Oligotypes of viruses, and bloom-forming microeukaryotes*

At the beginning of the summer algal bloom (June 21) two Raphidophyte oligotypes from OTU2 were detected (Figure 5.10 A); however, on the second bloom of this Raphidophyte (July 1), this OTU was dominated by one oligotype (Figure 5.10 A). The dinoflagellate OTU (classified as family Dinophyceae and genus *Gymnodiniphycidae*) that bloomed five days after the start of the *Heterosigma* bloom was primarily comprised of one oligotype (Figure 5.10 C); whereas, before the bloom more oligotypes were present (Figure 5.10 D). A second dinoflagellate OTU "bloomed" two days after *Heterosigma*,

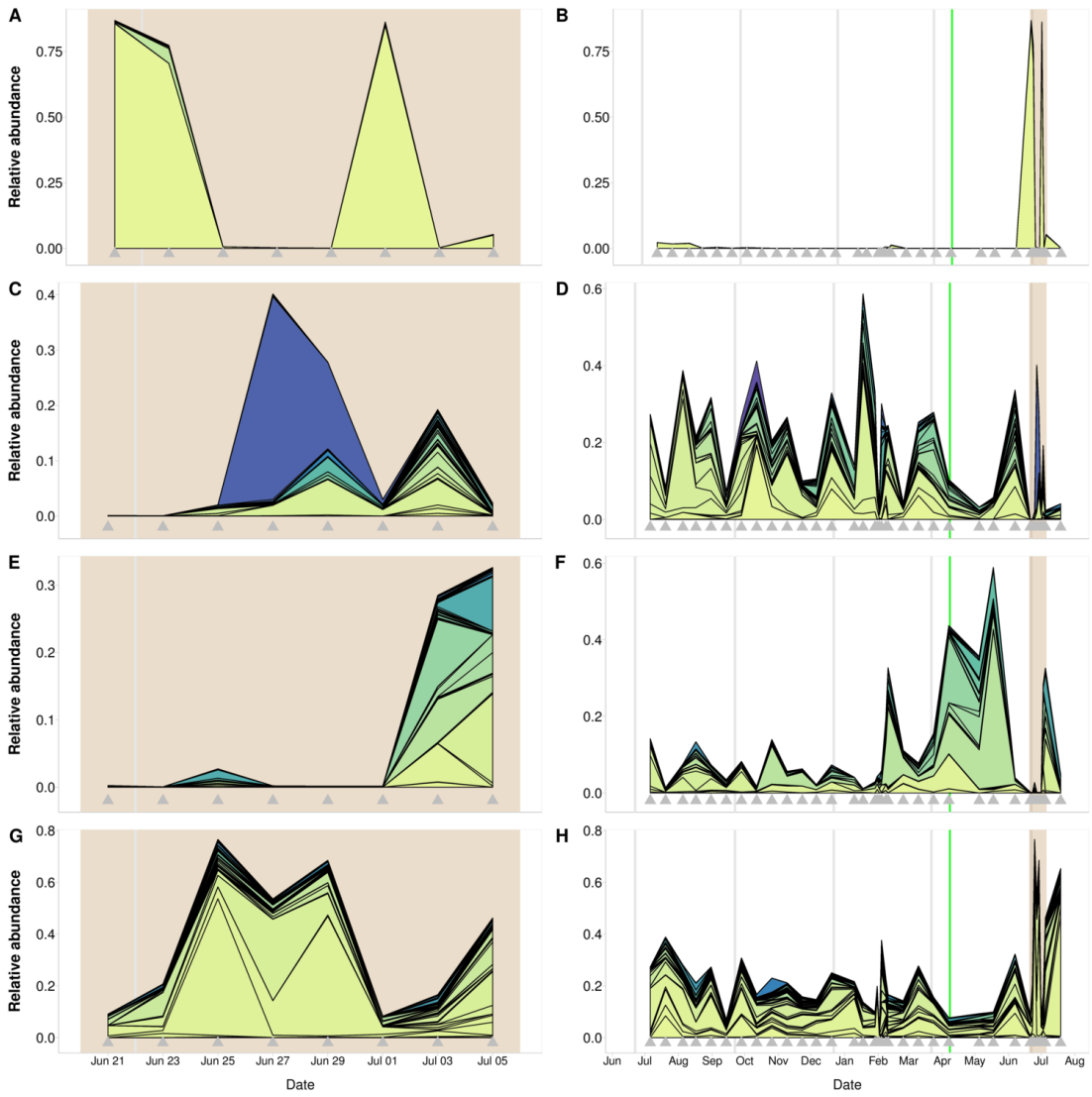

FIGURE 5.8: Change in relative abundance of OTUs for four eukaryotic orders during the summer algal bloom (left panels) and throughout the year (right panels). A) and B) Raphidophytes, C) and D) Ciliates, E) and F) Diatoms, G) and H) Dinoflagellates. Grey arrows on x-axis indicate sampling dates. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom. Classifications were done using the Wang algorithm as implemented in mothur (Schloss *et al.*, 2009) and using the Silva 119 database (Quast *et al.*, 2013)

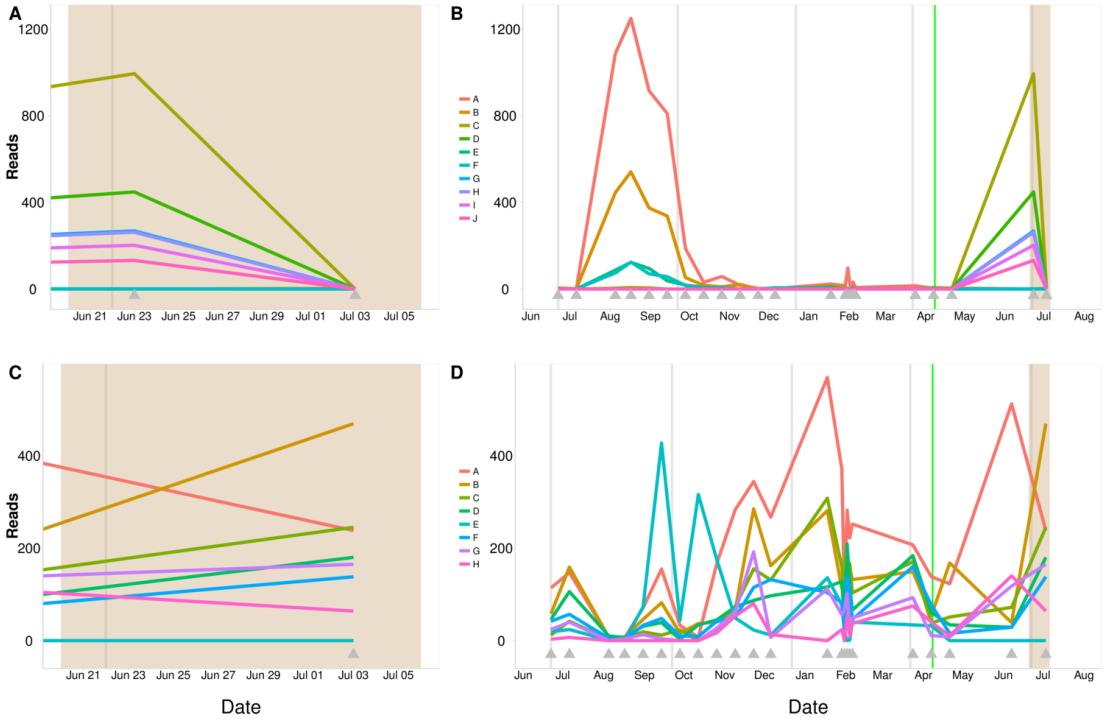

FIGURE 5.9: Oligotypes of specific OTUs pf marine picorna-like viruses during the summer algal bloom (left panels) and annually (right panels). A) and B) OTU 288, C) and D) OTU 3. Legend letters indicate different oligotypes, letters refer to different oligotypes in the top and bottom panels are not the same. Grey arrows on x-axis indicate sampling time points. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom. Oligotyping as in Eren *et al.* (2013).

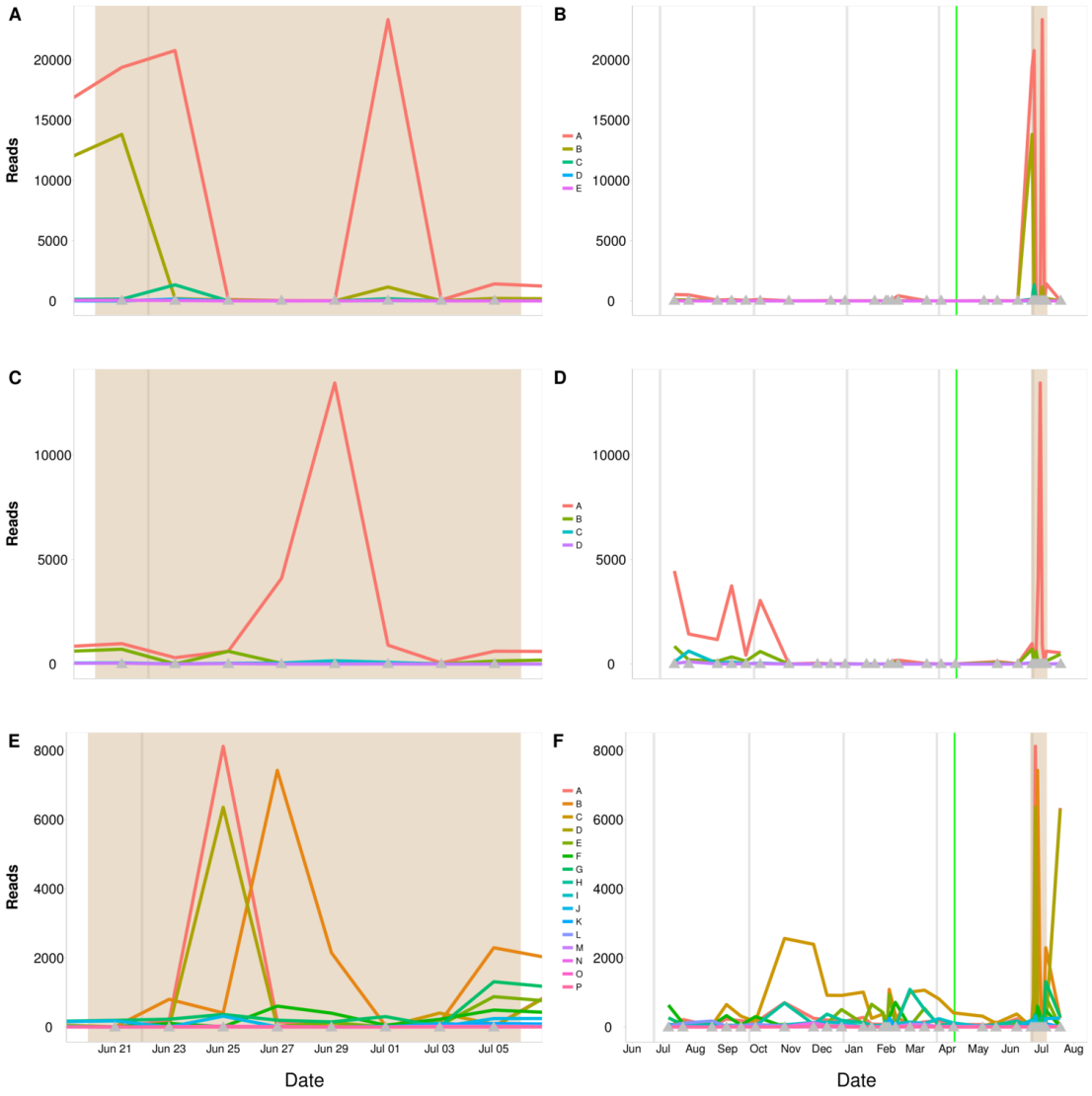

FIGURE 5.10: Oligotypes of OTUs during the bloom (left panels) and annually (right panel). A) and B) Raphidophytes, C) and D) Dinoflagellate OTU 1, E) and F) Dinoflagellate OTU 2. Legend letters indicate different oligotypes detected, letters between top and bottom panels are not the same. Grey arrows on x-axis indicate sampling time points. Grey vertical lines indicate season boundary and the green vertical line indicates the spring diatom bloom. Brown background indicates time of summer algal bloom. Classifications were done using the Wang algorithm as implemented in mothur (Schloss *et al.*, 2009) and using the Silva 119 database (Quast *et al.*, 2013) and oligotyping as in Eren *et al.* (2013).

and consisted of two oligotypes that were relatively abundant during the bloom and dominated thereafter (Figure 5.10 E and F).

5.5 DISCUSSION

High-throughput sequencing of samples taken every other day during a summer algal bloom revealed dynamics in the microbial communities, including a quick succession of several distinct dinoflagellate OTUs. The bacterial community mirrored the eukaryotic community at the order level, while marine picorna-like viruses showed shifts associated with the demise of the *Heterosigma*-dominated peak of bloom.

5.5.1 *Viral dynamics putatively explain host succession*

Viruses can affect the termination of phytoplankton blooms. In the bloom forming coccolithophore, *Emiliania huxleyi*, typically one genotype of large dsDNA viruses goes on to dominate the community as the bloom progresses (Martinez Martinez *et al.*, 2007; Sorensen *et al.*, 2009). Using several *Heterosigma akashiwo* strains and large DNA viruses (HAVs), Tarutani *et al.* (2000) found different patterns of resistance where a virus isolated at the end of the bloom would be most effective at lysing strains occurring during the bloom, but less effective at lysing host strains isolated after the bloom. Before the bloom, resistance to viral infection was low in *Heterosigma akashiwo* strains, while after the bloom it was high. Therefore, viruses may affect the clonal composition of populations, as previously demonstrated in cyanophage and *Synechococcus* in chemostat experiments (Marston *et al.*, 2012). These studies reinforce the hypothesis that there is a large cellular fitness cost to maintaining resistance to viruses. In bacteriophage, mutations in receptors either in structure, density, or access by virus show fitness costs (Lenski, 1988). Therefore before the bloom there are less resistant cells in the population.

In the samples from the one-year time series at Jericho Pier, there were high relative abundances of *Heterosigma* in the fall, as well as marine picorna-like virus OTU 1 (Figure

5.5). Based on Chapter 3 (p. 68) marine picorna-like virus OTU1 and OTU288 were closely related to *Heterosigma akashiwo* RNA virus (Tai *et al.*, 2003). During the summer algal bloom, the marine picorna-like viral OTU 288 fluctuated with the raphidophytes (Figure 5.5).

5.5.2 *Potential role of protists in termination of phytoplankton sub-blooms*

There was a second *Heterosigma* bloom on July 1, but none of the viral OTUs were associated with it, suggesting that the second bloom was terminated by either dinoflagellate competition (Figure 5.8D), ciliate grazing (Figure 5.8B), infection by different viruses, or nutrient limitation.

Dinoflagellates include microzooplankton grazers (e.g. *Ceratium* sp.) and phytoplankton which can be mixotrophic (e.g. *Dinophysis* sp.), and which are often a background species during blooms of other phytoplankton taxa (Anderson *et al.*, 2008). Similarly, the summer algal bloom at Jericho Pier was primarily *Heterosigma*, but there were also large signals from dinoflagellate OTUs (Figure 5.8G and H).

An increase in the relative abundance of ciliates, which was dominated by an OTU that was unique to the bloom and classified in the family Intramacronucleata, occurred two days after the rise of the dinoflagellate OTU 2. These ciliates could be consuming both *Heterosigma* and dinoflagellates that were in high abundance at this time and facilitated a different dinoflagellate to bloom two days later (Figure 5.8G and H). This agrees with field studies where tintinnids (a type of ciliate) showed opposite abundance dynamics to *Heterosigma* (Verity and Stoecker, 1982; Verity, 1987).

5.5.3 *Changes at the strain-level over time*

The main eukaryotic taxonomic groups during the bloom were generally also found in the one-year time series (Figure 5.5). Therefore, these OTUs were not ephemeral species. The raphidophyte, dinoflagellate and ciliate OTUs that showed dynamics during the bloom, had one or two main OTUs that fluctuated during the summer algal bloom.

To examine the dynamics within OTUs that are more than 97% similar, Shannon Entropy decomposition (oligotyping from Eren *et al*., 2013) was used to allow sub-OTU resolution. The oligotypes of raphidophyte OTUs decreased in diversity at the beginning of the bloom, and were quickly dominated by one oligotype (Figure 5.10). The oligotypes of the picorna-like viral OTU 288 (Figure 9), which was closely related to HaRNAV (shown in Chapter 3), followed similar dynamics as those of the raphidophytes throughout the time series, with peaks in abundance in August, September and February, but OTU 288 also had an oligotype that was not seen in the rest of the year, but dominated during the bloom. These dynamics are consistent with OTU 288 infecting one of the dominant bloom taxa, and potentially affecting the duration and composition of the bloom.

The dinoflagellate OTUs had their greatest relative abundance (0.13%) on June 25, the day after the first *Heterosigma* bloom. The most abundant dinoflagellate OTU-1 was classified to the family level as SCM15C8 and appeared to be a time generalist, as it was present throughout the time series. The oligotypes from the second dinoflagellate, OTU-2, showed high evenness the day after the first *Heterosigma* bloom, however, toward the end of the bloom (and after the second smaller bloom of *Heterosigma*) one oligotype dominated. The dinoflagellate OTU-1 was abundant after the second *Heterosigma* bloom, but had a much simpler population structure than OTU-2 with only one dominant oligotype observed during the bloom and in the overall time series. Thus, there was evidence for progression of strain-level taxa influenced by different ecological parameters during this bloom.

5.5.4 *Bacterial communities mirror shifts in eukaryotic communities*

During the bloom, the bacterial communities appeared to mirror the shifts in the eukaryotic communities (Figure 5.7, Mantel test on overall time series: 0.56, n=24). Members of the heterotrophic lineages Flavobacteriia, Rhodobacteracea (such as roseobacters), Alphaproteobacteria and Gammaproteobacteria (such as the Alteromonadaceae)

dominate the bacterial communities associated with phytoplankton blooms (Buchan *et al*., 2014). During the bloom, the Gammaproteobacteria increased with *Heterosigma*, while the Alphaproteobacteria, Actinobacteria and Flavobacteria decreased. Flavobacteria may have become most abundant after the peak of the *Heterosigma* bloom since they can convert high molecular weight (HMW) products of phytoplankton to low molecular weight (LMW) products (Buchan *et al*., 2014).

The bacterial communities had the highest evenness of all communities throughout two-week period examined (Figure 5.3). This could indicate functional stability, whereby different bacteria are using various phytoplankton products and thus the community is never completely dominated by one bacterial OTU. Similarly Delmont *et al*. (2014) found even, stable and distinct bacterial communities associated with *Phaeocystis* sp. blooms. Thus these blooms may support different niches for various bacteria, leading to evenness in the overall bacterial community.

5.5.5 *Community ecology*

Intense biological events may create unique niches over time that enable different organisms to exist in the communities, thus allowing distinct successional patterns. On June 27, four days after the peak in chl *a* (June 23rd) there was a spike in bacterial abundance, and two days later a peak in viral abundance (highest seen in this study) (Figure 5.1, Figure 5.2). In all the communities, the richness was greatest at the end of the two-week period, after the succession of phytoplankton blooms (Figure 5.3, Figure 5.4). This suggests that these communities may have responded to a disturbance event, in this case a phytoplankton bloom, by increasing in diversity as predicted by the intermediate disturbance hypothesis (Petraitis *et al*., 1989; Reynolds *et al*., 1993; Huston and DeAngelis, 1994). This contrasts previous studies in freshwater when during a *Prymnesium parvum* (haptophyte alga) bloom there were no changes in either eukaryotic or bacterial species richness, but community similarity was different between bloom and non-bloom times (Jones *et al*., 2013).

Communities can respond in different ways to disturbances: they can stay functionally the same (but composition will be different), they can return to the same composition after the disturbance event (resilience) or they can stay the same (resistance) (Allison and Martiny, 2008). A phytoplankton bloom, although a natural part of the community, can be considered a pulse disturbance (i.e. short duration, as opposed to a "press" disturbance which is longer-scale). For example, aquatic bacterial communities after a mechanical mixing event in a lake, returned to pre-disturbance composition seven days after the mixing event (Shade *et al.*, 2012b). Throughout the sampling, the evenness remained stable. Richness was low immediately following the disturbance, but then increased afterwards. They hypothesize that the biomass produced during the bloom was a source of nutrients and carbon for the bacterial communities, and provided distinct niches and food sources for many different organisms. Theoretically, intermediate levels of disturbance would generate the highest levels of species richness (Petraitis *et al.*, 1989). Moderate nutrient inputs can stimulate eukaryotic community diversity (Spatharis *et al.*, 2007) thus explaining the viral, bacterial and eukaryotic diversity after the bloom. Therefore, blooms of eukaryotic phytoplankton disturb the eukaryotic, bacterial, and subsets of the viral communities, provoking strain-level succession and stimulating an increase in richness in all communities following the bloom. This agrees with the Behrenfeld and Boss (2014) hypothesis that abiotic factors do not control bloom dynamics, but rather that biotic factors related to imbalances in predator-prey relationships could provoke a bloom.

5.6 CONCLUSIONS

Phytoplankton blooms have been well studied, but much less is known about bacterial and viral communities associated with blooms. There was succession of eukaryotic phytoplankton during a summer algal bloom. There were sub-blooms of *Heterosigma* and dinoflagellates throughout the sampling period, and evidence of viral and grazer control of

the summer algal bloom. Sub-strain progression within the Heterosigma bloom showed evidence for viral pressure on the genetic diversity of *Heterosigma* strains. The bacterial communities associated with the bloom, maintained evenness throughout the bloom even while mirroring shifts in the eukaryotic communities. The bacterial communities appeared taxonomically well-adapted to using the products of phytoplankton blooms. Overall this study showed that the natural disturbance of a bloom stimulated changes in succession and diversity in the viral, bacterial and microeukaryotic communities, and provided a glimpse into the complex dynamics within phytoplankton blooms.

CHAPTER 6

Conclusion and Future Directions

6.1 SUMMARY

To examine important ecological questions related to the diversity and community structure of marine viral communities, two ecologically important groups of viruses and their host communities were placed in a temporal context by using amplicon sequencing of a one-year time series. In addition, host datasets were examined for the potential influence of the viruses.

The identity of the members of the communities are not needed for measures of diversity and community structure, however, using high-throughput sequencing, the viral sequences were determined, thus, even though the viruses were mostly unknown, they were analysed in a phylogenetic context which provided more information about their relatedness and about the phylogenetic structure of viral communities. The sub-OTU dynamics, dynamics of reads within the chosen OTU percent similarity, of planktonic OTUs were determined during a eukaryotic phytoplankton bloom at a coastal site and showed successional shifts related to viral and heterotrophic protist hosts.

Considering the key findings of these studies, this dissertation has advanced the field of viral ecology in the following ways:

In Chapter 2 I examined the patchiness and continued production of marine picorna-like viruses using pyrosequencing of the RNA dependent RNA polymerase (RdRp) of viruses in the order *Picornavirales*. There was patchiness in the OTUs with low overlap between spatially proximate samples and high turnover in the mixed layer of the water column. These results implied continuous infection and lysis of eukaryotic phyto-

plankton by these viruses. This study also showed the potential and power of amplicon sequencing for addressing ecological questions related to temporal dynamics.

In Chapter 3 I examined the phylogenetic structure from two viral communities over time which revealed the phylogenetic structuring of these communities and the nature of ephemeral vs. persistent OTUs. The results implied that viruses influence the composition of the host communities, and that viral community structure is dependent on lifestyle (i.e. host range and burst size). These examinations provided refinements to the seed bank theory (including insight into phylogenetic dynamics of viral communities).

In Chapter 4 I constructed co-occurrence networks from the one-year time series of the eukaryotic, bacterial, T4-like myoviral and picorna-like communities. This analysis revealed the important role of environmental parameters in determining the co-occurrence of viruses and hosts. It also showed that eukaryotic and bacterial OTUs were more strongly correlated to environmental factors than the viral OTUs. Communities sampled in the fall were more strongly correlated to each other than any other season and the fall samples shared the greatest number of links with the winter timepoints. This demonstrated a time of stability for these communities. Based on the analysis of the environmental triplets, the environment plays a large role in filtering the host-virus pairs that occur in an environment and occur seasonally.

In Chapter 5 I looked at the dynamics of the eukaryotic, bacterial and viral communities during a eukaryotic phytoplankton bloom (peak of bloom composed mostly of *Heterosigma*). This project examined the effect of disturbances (ecological perturbations) on microbial communities. The observed succession of sub-OTUs (oligotypes) was linked to viral selective pressure early in the bloom and to protistan predation later in the bloom. Also, by examining the fine-scale dynamics it was observed, that eukaryotic phytoplankton blooms could be terminated by multiple different factors over a short time period.

This concluding chapter will discuss these advances in the context of the fields of viral and microbial ecology. Moreover, this chapter presents the next steps arising from this work to further advance these fields.

6.2 ADDITIONS TO THE "SEED BANK" THEORY

As seen in Chapters 3, 4, and 5, the "seed bank" model (or simply "Bank" model) describing community structure (Breitbart and Rohwer, 2005; Brum *et al.*, 2015) is a useful way to examine the work presented in this dissertation. The model posits that most members of the community are rare and there is a shuffling of the rare members that then become abundant based on the environment or on available hosts. This dissertation showed that there is a phylogenetic component to this shuffling in viral communities, meaning that closely related viruses seem to become abundant at the same time. Thus, there is an order to the "Bank" that was not previously incorporated whereby phylogenetically-related viruses show similar shifts in dominance of the community. The nature of viral replication, which is generally more error prone than host genome replication, could explain the structure where there is one relatively abundant virus and many related viruses since it could be the results of erroneous replication. Thus diversity is generated through errors and the shifts over time were based on the original dominant virus. The related viruses present that arise originally from the abundant dominant virus.

Studies have examined how the overall phylogenetic diversity changes in bacterial communities (Horner-Devine and Bohannan, 2006; Amend *et al.*, 2016), but phylogeny has neither been used to look at the shifts in the community nor to examine the relatedness of viral communities. In this dissertation I showed that the viral communities have large shifts in dominance over time and the data also suggest that this phylogenetic structure maintains the resilience and stability of these communities.

6.3 THE NATURE OF EPHEMERAL AND PERSISTENT OTUS OVER TIME

This dissertation examined which OTUs were persistent and which were ephemeral. This was mostly examined in Chapter 3, but Chapter 2 and Chapter 5 also examined which OTUs persisted. Most members of the community were ephemeral, and these members were usually present at low relative abundances. There was high turnover in the communities, especially in the viral communities. The OTUs in the viral communities that were persistent over time were often those that were the most relatively abundant over time. One hypothesis is that these viral taxa may be so abundant that their loss due to decay is slower over time (Wilhelm *et al.*, 1998) or it could be that they are being constantly produced (as suggested in Chapter 2) by continually infecting organisms, albeit those organisms might not always be at high abundance.

6.4 EFFECT OF ENVIRONMENTAL PARAMETERS ON DETERMINING THE CO-OCCURRENCE OF VIRUSES AND HOSTS

Focusing on the common, abundant OTUs (and setting aside the ephemeral OTUs), much can be learned about the overall ecosystem from the co-occurrence of these OTUs. Food-webs in marine systems are often very complex (Legendre and Rivkin, 2008; Weitz *et al.*, 2015). Microbial communities can have deeply interdependent relationships, however, these relationships can be hard to examine. In Chapter 4, based on network analysis of co-occurrence associations, the connections between viruses and hosts were mainly driven by nutrients, temperature, salinity, viral abundance, and bacterial abundance. This is complementary to what was observed using variation partitioning, where changes in different communities were driven by environmental parameters and by time. As discussed in the Introduction and in Chapter 4, positive co-occurrences can be attributed to host-virus pairs, mutualism, or to the same preferred niche. Viral predation could be represented by positive or negative links, positive links could represent symbiosis, or

shared niches, and negative links could represent predation, opposite niches, competitive exclusion.

6.5 EFFECT OF DISTURBANCES ON MICROBIAL COMMUNITIES

Microbes and viruses are thought to respond very quickly to pulses of nutrients (Buchan *et al.*, 2014), increases in temperature (Kendrick *et al.*, 2014), and other environmental changes. Thus it would be expected that if there was a disturbance in the ecosystem these organisms would be quickly affected. As described in the Introduction there are many different types of disturbances that can affect microbial communities. This dissertation focused on the effect of a summer algal bloom on microbial communities. Although the dominant phytoplankton in this bloom, *Heterosigma*, has been widely studied in the lab, and as part of compositional phytoplankton surveys, there have been few studies that have also looked at how the bacterial, viral, and overall eukaryotic communities change during a summer algal bloom.

6.5.1 *Resilience and stability*

Over the course of the eukaryotic phytoplankton bloom the richness and evenness of the bacterial community were stable, but the communities shifted in composition during that time. This has been documented for other phytoplankton using community fingerprinting and amplicon sequencing, but not yet for summer algal blooms dominated by *Heterosigma*. After the bloom all communities showed higher richness than immediately preceding the bloom. If systems with higher richness are considered to be more productive, there was an overall increase in biodiversity and productivity after this disturbance. This is in agreement with the intermediate disturbance hypothesis (Connell, 1978).

The eukaryotic community displayed interesting dynamics: first a *Heterosigma* OTU bloomed, then there were two separate sub-blooms of dinoflagellates, and finally a second bloom of the *Heterosigma* OTU (same oligotype) six days later. These sub-bloom

dynamics have, as of yet, received little attention. It was observed that these disturbances can affect the eukaryotic community differently than the bacterial community. For some bacterial populations, the disturbance acted like an input of nutrients or carbon, and thus there were shifts in the overall community structure, but the community still supported similar levels of richness. Whereas for the eukaryotic community it was a time of competitive exclusion with high overall unevenness.

6.6 IMPLICATIONS

- The phylogenetic relatedness plays an important role in the community assembly of viral and microbial communities and thus should be considered when examining community dynamics.
- Co-occurrence of hosts and viruses can be driven by environmental parameters such as nutrients, salinity and temperature and thus specific niches are important not just for the occurrence of certain organisms, but for their interactions.
- Phytoplankton blooms can be composed of many smaller sub-blooms of different organisms and these disturbances can generate diversity in all microbial and viral communities after the bloom.
- Overall microbial and viral diversity is driven by shifts in phylogenetically-related organisms over time and different environmental parameters.

When examining species on a rank abundance curve it is important to investigate the phylogeny to really understand how communities change over time and also could give insight into historical events of the community. Using only a one year time series it is hard to definitively identify hosts of viruses, but using high-throughput sequencing of longer data sets would enable deeper insights into these relationships. Nevertheless, this dissertation has shown putative host-virus interactions and the strong effect of the environment on these interactions.

6.7 FUTURE WORK

As with many studies performed, there are new or improved technologies that open up new or deeper avenues for the research. One of the relevant improvements for this study is in the area of metagenomics. Metagenomics has been used in the marine setting (Venter *et al.*, 2004) and for marine viruses (Breitbart *et al.*, 2002) since its inception, however, what has changed has been the amount of data retrieved from metagenomes and how many metagenomes can be compared. Early metagenomes used Sanger sequencing to examine genetic material, however, now with high-throughput sequencing, the number of reads retrieved from any metagenome has greatly increased. With this increase, a better representation of all of the communities and especially the rarer members can be achieved. Also, with the relative ease and low cost of processing metagenomes, it is now possible to analyse metagenomes of time series from viral and microbial communities.

Additionally, there have been large improvements in the functional annotation of metagenomes (i.e. the metabolic or enzymatic pathways present) through tools like Metacyc (Caspi *et al.*, 2014) and Metapathways (Konwar *et al.*, 2015). Therefore, more ecological questions can be examined using the functional roles and diversity of the communities. This leads to trait-based approaches as defined in the recent review focusing on microbial traits (Martiny *et al.*, 2015). Furthermore, applying the approaches in this dissertation to a longer time series or reanalyzing previous data sets to examine the temporal phylogenetic relatedness could provide deeper insights such as: 1) more opportunities to confirm how typical the dynamics are that were observed during the *Heterosigma* bloom 2) the ability to more closely examine virus-host co-occurrences enabling the identification of virus-host pairs and, 3) the ability to look at seasonality of composition of the communities and which members are driving this seasonality.

Finally, two relevant types of experiments with isolated viruses are prompted by this dissertation. First, is the continued need and importance of viral isolation from aquatic microbial hosts. The genetic diversity of both the T4-like myoviruses and the marine picorna-like viruses and lack of cultured representatives illustrated the impor-

tance of isolating viral-host systems (or of using techniques in which viral sequence can be specifically associated to host sequence, such as in single-cell sequencing of bacteria). Genomes can be sequenced from these viral isolates, which would be complementary approach to amplicon and metagenomic approaches. Second, is the examination of the dynamics between viruses infecting the same host but from different families or containing different genetic material (RNA vs DNA). An easily imaginable example would be for *Heterosigma akashiwo* where there is a DNA virus (HAV (Nagasaki *et al.*, 1994)) and a RNA virus (HaRNAV (Tai *et al.*, 2003)) that infect it. Considering that these viruses have different lifestyles and properties, what would be the conditions for one to be more successful than the other? Using two related DNA viruses Nissimov *et al.* (2016) determined that there was a "fight club" for the viruses, where one appeared to be better "competitor." A potential link illustrating that such dynamics could lead to co-infection is in the putative recombination of a ssRNA virus and ssDNA virus (Circovirus) discovered in metagenomes and then verified by long-range PCR from lakes in Yellowstone (Diemer and Stedman, 2012). Examining the dynamics of DNA and RNA viruses of one host could provide a way to test the hypothesis of the r- vs. K-selected viruses (Suttle, 2007) and whether there are environmental or other conditions that lead one to be that or another.

6.8 CONCLUSION

This dissertation examined the dynamics of the bacteria, eukaryotes and subsets of the viruses at the coastal site of Jericho Pier in Vancouver, British Columbia, Canada using high-throughput sequencing. By using an unprecedented combination of bacterial, eukaryotic and viral community data, this dissertation provided advances in the relatedness of viruses over time, the drivers of host-virus relationships, the dynamics and richness of coastal plankton communities during blooms, and updates to models of the community structure of viral communities. Altogether, this dissertation has advanced

knowledge about the community structure and dynamics of viral communities, and has identified important avenues to be explored in time series of microbes and viruses.

Bibliography

Acinas S.G., Sarma-Rupavtarm R., Klepac-Ceraj V., and Polz M.F. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16s rRNA clone libraries constructed from the same sample. Applied and Environmental Microbiology 71:8966–8969.

Adriaenssens E.M. and Cowan D.A. 2014. Using signature genes as tools to assess environmental viral ecology and diversity. Applied and Environmental Microbiology 80:4470–4480.

Alkire M.B., D'Asaro E., Lee C., Jane Perry M., Gray A., Cetinić I., Briggs N., Rehm E., Kallin E., Kaiser J., and González-Posada A. 2012. Estimates of net community production and export using high-resolution, Lagrangian measurements of O_2, NO_3^-, and POC through the evolution of a spring diatom bloom in the North Atlantic. Deep Sea Research Part I: Oceanographic Research Papers 64:157–174.

Allen S. and Wolfe M. 2013. Hindcast of the timing of the spring phytoplankton bloom in the Strait of Georgia, 1968-2010. Progress in Oceanography 115:6–13.

Allison S.D. and Martiny J.B. 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 105:11512–11519.

Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

Amend A.S., Martiny A.C., Allison S.D., Berlemont R., Goulden M.L., Lu Y., Treseder K.K., Weihe C., and Martiny J.B.H. 2016. Microbial response to simulated global change is phylogenetically conserved and linked with functional potential. The ISME Journal 10:109–118.

Amend A.S., Seifert K.A., and Bruns T.D. 2010. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Molecular Ecology 19:5555–5565.

Anderson C.R., Siegel D.A., Brzezinski M.A., and Guillocheau N. 2008. Controls on temporal patterns in phytoplankton community structure in the Santa Barbara Channel, California. Journal of Geophysical Research 113.

Andrews S. 2015. FastQC. http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ Version 0.11.4.

Angly F.E., Felts B., Breitbart M., Salamon P., Edwards R.A., Carlson C., Chan A.M., Haynes M., Kelley S., and Liu H. 2006. The marine viromes of four oceanic regions. PLoS biology 4:e368.

Apprill A., McNally S., Parsons R., and Weber L. 2015. Minor revision to V4 region SSU rRNA 806r gene primer greatly increases detection of SAR11 bacterioplankton. Aquatic Microbial Ecology 75:129–137.

Arslan D., Legendre M., Seltzer V., Abergel C., and Claverie J.M. 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proceedings of the National Academy of Sciences 108:17486–17491.

Azam F., Fenchel T., Field J.G., Gray J.S., Meyer-Reil L.A., and Thingstad F. 1983. The ecological role of water-column microbes in the sea. Marine ecology progress series. Oldendorf 10:257–263.

Baker G., Smith J., and Cowan D. 2003. Review and re-analysis of domain-specific 16**S** primers. Journal of Microbiological Methods 55:541–555.

Banks S.C., Cary G.J., Smith A.L., Davies I.D., Driscoll D.A., Gill A.M., Lindenmayer D.B., and Peakall R. 2013. How does ecological disturbance influence genetic diversity? Trends in Ecology & Evolution 28:670–679.

Barberán A., Bates S.T., Casamayor E.O., and Fierer N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. The ISME journal 6:343–351.

Behrenfeld M.J. and Boss E.S. 2014. Resurrecting the ecological underpinnings of ocean plankton blooms. Annual Review of Marine Science 6:167–194.

Bellas C.M. and Anesio A.M. 2013. High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. Extremophiles 17:861–870.

Bellec L., Clerissi C., Edern R., Foulon E., Simon N., Grimsley N., and Desdevises Y. 2014. Cophylogenetic interactions between marine viruses and eukaryotic picophytoplankton. BMC Evolutionary Biology 14:59.

Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., and Wheeler D.L. 2007. GenBank. Nucleic Acids Research 35:D21–D25.

Benton M.J. 2009. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. Science 323:728–732.

Berger S.A., Krompass D., and Stamatakis A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Systematic Biology 60:291–302.

Bergh O., Borsheim K.Y., Bratbak G., and Heldal M. 1989. High abundance of viruses found in aquatic environments. Nature 340:467–468.

Berry D. and Widder S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in Microbiology 5.

Bettarel Y., Bouvier T., and Bouvy M. 2009. Viral persistence in water as evaluated from a tropical/temperate cross-incubation. Journal of Plankton Research 31:909–916.

Blain S., Guieu C., Claustre H., Leblanc K., Moutin T., Quéguiner B., Ras J., and Sarthou G. 2004. Availability of iron and major nutrients for phytoplankton in the northeast Atlantic Ocean. Limnology and Oceanography 49:2095–2104.

Bokulich N.A., Subramanian S., Faith J.J., Gevers D., Gordon J.I., Knight R., Mills D.A., and Caporaso J.G. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods 10:57–59.

Bolger A.M., Lohse M., and Usadel B. 2014. Trimmomatic a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Borcard D., Legendre P., and Drapeau P. 1992. Partialling out the spatial component of ecological variation. Ecology 73:1045.

Bouvier T. and del Giorgio P.A. 2007. Key role of selective viral-induced mortality in determining marine bacterial community composition. Environmental Microbiology 9:287–297.

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

Bratbak G., Heldal M., Norland S., and Thingstad T.F. 1990. Viruses as partners in spring bloom microbial trophodynamics. Applied and Environmental Microbiology 56:1400–1405.

Bratbak G., Wilson W., and Heldal M. 1996. Viral control of *Emiliania huxleyi* blooms? Journal of Marine Systems 9:75–81.

Breitbart M. 2012. Marine viruses: truth or dare. Annual Review of Marine Science 4:425–448.

Breitbart M. and Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? Trends in Microbiology 13:278–284.

Breitbart M., Salamon P., Andresen B., Mahaffy J.M., Segall A.M., Mead D., Azam F., and Rohwer F. 2002. Genomic analysis of uncultured marine viral communities. Proceedings of the National Academy of Sciences 99:14250–14255.

Brown C.T., Hug L.A., Thomas B.C., Sharon I., Castelle C.J., Singh A., Wilkins M.J., Wrighton K.C., Williams K.H., and Banfield J.F. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523:208–211.

Brum J.R., Ignacio-Espinoza J.C., Roux S., Doulcier G., Acinas S.G., Alberti A., Chaffron S., Cruaud C., de Vargas C., Gasol J.M., and others. 2015. Patterns and ecological drivers of ocean viral communities. Science 348:1261498.

Brum J.R., Schenck R.O., and Sullivan M.B. 2013. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. The ISME journal 7:1738–1751.

Brum J.R. and Sullivan M.B. 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. Nature Reviews Microbiology 13:147–159.

Brussaard C.P. 2004a. Viral control of phytoplankton populations—a review. Journal of Eukaryotic Microbiology 51:125–138.

Brussaard C.P.D. 2004b. Optimization of procedures for counting viruses by flow cytometry. Applied and Environmental Microbiology 70:1506–1513.

Brussaard C.P.D., Noordeloos A.A.M., Sandaa R.A., Heldal M., and Bratbak G. 2004. Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. Virology 319:280–291.

Buchan A., LeCleir G.R., Gulvik C.A., and González J.M. 2014. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nature Reviews Microbiology 12:686–698.

Butina T.V., Belykh O.I., Potapov S.A., and Sorokovikova E.G. 2013. Diversity of the major capsid genes (g23) of T4-like bacteriophages in the eutrophic Lake Kotokel in East Siberia, Russia. Archives of Microbiology 195:513–520.

Capella-Gutierrez S., Silla-Martinez J.M., and Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Pena A.G., Goodrich J.K., and Gordon J.I. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:335–336.

Caporaso J.G., Lauber C.L., Walters W.A., Berg-Lyons D., Lozupone C.A., Turnbaugh P.J., Fierer N., and Knight R. 2011. Global patterns of 16**S** rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences 108:4516–4522.

Caspi R., Altman T., Billington R., Dreher K., Foerster H., Fulcher C.A., Holland T.A., Keseler I.M., Kothari A., Kubo A., Krummenacker M., Latendresse M., Mueller L.A., Ong Q., Paley S., Subhraveti P., Weaver D.S., Weerasinghe D., Zhang P., and Karp P.D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Research 42:D459–D471.

Chaffron S., Rehrauer H., Pernthaler J., and von Mering C. 2010. A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Research 20:947–959.

Chen F. and Suttle C.A. 1995. Amplification of DNA polymerase gene fragments from viruses infecting microalgae. Applied and Environmental Microbiology 61:1274–1278.

Chen F., Wang K., Huang S., Cai H., Zhao M., Jiao N., and Wommack K.E. 2009. Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. Environmental Microbiology 11:2884–2892.

Chesson P. 2000. Mechanisms of maintenance of species diversity. Annual review of Ecology and Systematics 31:343–366.

Chesson P. and Huntly N. 1997. The roles of harsh and fluctuating conditions in the dynamics of ecological communities. The American Naturalist 150:519–553.

Chiswell S.M., Calil P.H.R., and Boyd P.W. 2015. Spring blooms and annual cycles of phytoplankton: a unified perspective. Journal of Plankton Research 37:500–508.

Chow C.E.T. and Fuhrman J.A. 2012. Seasonality and monthly dynamics of marine myovirus communities: Marine myovirus community dynamics at SPOT. Environmental Microbiology 14:2171–2183.

Chow C.E.T., Kim D.Y., Sachdeva R., Caron D.A., and Fuhrman J.A. 2014. Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. The ISME journal 8:816–829.

Chow C.E.T., Sachdeva R., Cram J.A., Steele J.A., Needham D.M., Patel A., Parada A.E., and Fuhrman J.A. 2013. Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. The ISME Journal 7:2259–2273.

Chow C.E.T. and Suttle C.A. 2015. Biogeography of Viruses in the Sea. Annual Review of Virology 2:41–66.

Christen A. 2013. Shortwave Irradiance at Vancouver Sunset (Tower) 2011. http://ibis.geog.ubc.ca/~achristn/infrastructure/climate-data/STSDA2-01-2011-to-12-2011.html .

Clasen J., Hanson C., Ibrahim Y., Weihe C., Marston M., and Martiny J. 2013. Diversity and temporal dynamics of Southern California coastal marine cyanophage isolates. Aquatic Microbial Ecology 69:17–31.

Clasen J.L. and Suttle C.A. 2009. Identification of freshwater *Phycodnaviridae* and their potential phytoplankton hosts, using DNA *pol* sequence fragments and a genetic-distance analysis. Applied and Environmental Microbiology 75:991–997.

Cole J.J. 1982. Interactions between bacteria and algae in aquatic ecosystems. Annual Review of Ecology and Systematics 13:291–314.

Cole J.R., Wang Q., Fish J.A., Chai B., McGarrell D.M., Sun Y., Brown C.T., Porras-Alfaro A., Kuske C.R., and Tiedje J.M. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research 42:D633–D642.

Combe M. and Sanjuán R. 2014. Variation in RNA virus mutation rates across host cells. PLoS Pathogens 10:e1003855.

Comeau A.M., Arbiol C., and Krisch H.M. 2010. Gene network visualization and quantitative synteny analysis of more than 300 marine T4-like phage scaffolds from the GOS metagenome. Molecular Biology and Evolution 27:1935–1944.

Comte J., Lovejoy C., Crevecoeur S., and F. Vincent W. 2015. Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes. Biogeosciences Discussions 12:10233–10269.

Connell J.H. 1978. Diversity in tropical rain forests and coral reefs. Science 199:1302–1310.

Cram J., Sun F., and Fuhrman J.A. 2013. Marine bacterial, archaeal, and protistan association networks. *In* Encyclopedia of Metagenomics, pages 1–10, Springer New York, New York, NY.

Cram J.A., Xia L.C., Needham D.M., Sachdeva R., Sun F., and Fuhrman J.A. 2015. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. ISME J 9:2573–2586.

Croft M.T., Lawrence A.D., Raux-Deery E., Warren M.J., and Smith A.G. 2005. Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria. Nature 438:90–93.

Crump B.C., Hopkinson C.S., Sogin M.L., and Hobbie J.E. 2004. Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. Applied and Environmental Microbiology 70:1494–1505.

Csardi G. and Nepusz T. 2006. The igraph software package for complex network research. InterJournal Complex Systems:1695.

Cullen J.J. 1991. Hypotheses to explain high-nutrient conditions in the open sea. Limnology and Oceanography 36:1578–1599.

Culley A.I., Lang A., and Suttle C.A. 2003. High diversity of unknown picorna-like viruses in the sea. Nature 424:1054–1057.

Culley A.I., Lang A., and Suttle C.A. 2006. Metagenomic analysis of coastal RNA virus communities. Science 312:1795–1798.

Culley A.I., Mueller J.A., Belcaid M., Wood-Charlson E.M., Poisson G., and Steward G.F. 2014. The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. mBio 5:e01210–14–e01210–14.

Culley A.I. and Steward G.F. 2007. New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. Applied and Environmental Microbiology 73:5937–5944.

Dale T., Rey F., and Heimdal B.R. 1999. Seasonal development of phytoplankton at a high latitude oceanic site. Sarsia 84:419–435.

Darriba D., Taboada G.L., Doallo R., and Bangor D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Delmont T.O., Hammar K.M., Ducklow H.W., Yager P.L., and Post A.F. 2014. *Phaeocystis antarctica* blooms strongly influence bacterial community structures in the Amundsen Sea polynya. Frontiers in Microbiology 5.

Derelle E., Ferraz C., Escande M.L., Eychenié S., Cooke R., Piganeau G., Desdevises Y., Bellec L., Moreau H., and Grimsley N. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. PLoS ONE 3:e2250.

DeSantis T.Z., Hugenholtz P., Larsen N., Rojas M., Brodie E.L., Keller K., Huber T., Dalevi D., Hu P., and Andersen G.L. 2006. Greengenes, a chimera-checked 16s rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology 72:5069–5072.

Dia A., Guillou L., Mauger S., Bigeard E., Marie D., Valero M., and Destombe C. 2014. Spatiotemporal changes in the genetic diversity of harmful algal blooms caused by the toxic dinoflagellate *Alexandrium minutum*. Molecular Ecology 23:549–560.

Diemer G.S. and Stedman K.M. 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. Biology Direct 7:13.

Diez B., Pedros-Alio C., Marsh T.L., and Massana R. 2001. Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. Applied and Environmental Microbiology 67:2942–2951.

Djikeng A., Kuzmickas R., Anderson N.G., and Spiro D.J. 2009. Metagenomic analysis of RNA viruses in a fresh water lake. PLoS One 4:e7264.

Domingo E., Sheldon J., and Perales C. 2012. Viral quasispecies evolution. Microbiology and Molecular Biology Reviews 76:159–216.

Domingues R.B., Barbosa A., and Galvão H. 2005. Nutrients, light and phytoplankton succession in a temperate estuary (the Guadiana, south-western Iberia). Estuarine, Coastal and Shelf Science 64:249–260.

DuRand M.D., Olson R.J., and Chisholm S.W. 2001. Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. Deep Sea Research Part II: Topical Studies in Oceanography 48:1983–2003.

Dutilh B.E., Cassman N., McNair K., Sanchez S.E., Silva G.G.Z., Boling L., Barr J.J., Speth D.R., Seguritan V., Aziz R.K., Felts B., Dinsdale E.A., Mokili J.L., and Edwards R.A. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nature Communications 5.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

Edgar R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

Emerson J.B., Thomas B.C., Andrade K., Heidelberg K.B., and Banfield J.F. 2013. New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. Applied and Environmental Microbiology 79:6755–6764.

Eren A.M., Maignien L., Sul W.J., Murphy L.G., Grim S.L., Morrison H.G., and Sogin M.L. 2013. Oligotyping: differentiating between closely related microbial taxa using 16**S** rRNA gene data. Methods in Ecology and Evolution 4:1111–1119.

Faith D.P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61:1–10.

Faust K., Lahti L., Gonze D., de Vos W.M., and Raes J. 2015a. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology 25:56–66.

Faust K., Lima-Mendez G., Lerat J.S., Sathirapongsasuti J.F., Knight R., Huttenhower C., Lenaerts T., and Raes J. 2015b. Cross-biome comparison of microbial association networks. Frontiers in Microbiology 6.

Faust K. and Raes J. 2012. Microbial interactions: from networks to models. Nature Reviews Microbiology 10:538–550.

Field C.B., Behrenfeld M.J., Randerson J.T., and Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281:237–240.

Filée J., Tétart F., Suttle C.A., and Krisch H.M. 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proceedings of the National Academy of Sciences 102:12471–12476.

Frederickson C.M., Short S.M., and Suttle C.A. 2003. The physical environment affects cyanophage communities in British Columbia inlets. Microbial Ecology 46:348–357.

Fuhrman J.A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.

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

Fuhrman J.A., Cram J.A., and Needham D.M. 2015. Marine microbial community dynamics and their ecological interpretation. Nature Reviews Microbiology 13:133–146.

Fuhrman J.A., Hewson I., Schwalbach M.S., Steele J.A., Brown M.V., and Naeem S. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences 103:13104–13109.

Fujiki T., Matsumoto K., Mino Y., Sasaoka K., Wakita M., Kawakami H., Honda M.C., Watanabe S., and Saino T. 2014. Seasonal cycle of phytoplankton community structure and photophysiological state in the western subarctic gyre of the North Pacific. Limnology and Oceanography 59:887–900.

Gaston K.J. 2000. Global patterns in biodiversity. Nature 405:220–227.

Ghiglione J.F. and Murray A.E. 2012. Pronounced summer to winter differences and higher wintertime richness in coastal Antarctic marine bacterioplankton: Temporal variation in Southern Ocean coastal bacterioplankton. Environmental Microbiology 14:617–629.

Gibbons S.M., Caporaso J.G., Pirrung M., Field D., Knight R., and Gilbert J.A. 2013. Evidence for a persistent microbial seed bank throughout the global ocean. Proceedings of the National Academy of Sciences 110:4651–4655.

Gilbert J.A., Field D., Swift P., Newbold L., Oliver A., Smyth T., Somerfield P.J., Huse S., and Joint I. 2009. The seasonal structure of microbial communities in the Western English Channel. Environmental Microbiology 11:3132–3139.

Gilbert J.A., Steele J.A., Caporaso J.G., Steinbrück L., Reeder J., Temperton B., Huse S., McHardy A.C., Knight R., Joint I., Somerfield P., Fuhrman J.A., and Field D. 2011. Defining seasonal marine microbial community dynamics. The ISME Journal 6:298–308.

Giovannoni S., Temperton B., and Zhao Y. 2013. Giovannoni et al. reply. Nature 499:E4–E5.

Gobet A., Böer S.I., Huse S.M., van Beusekom J.E., Quince C., Sogin M.L., Boetius A., and Ramette A. 2012. Diversity and dynamics of rare and of resident bacterial populations in coastal sands. The ISME journal 6:542–553.

Goldsmith D.B., Parsons R.J., Beyene D., Salamon P., and Breitbart M. 2015. Deep sequencing of the viral *phoH* gene reveals temporal variation, depth-specific composition, and persistent dominance of the same viral *phoH* genes in the Sargasso Sea. PeerJ 3:e997.

Gorbalenya A.E., Enjuanes L., Ziebuhr J., and Snijder E.J. 2006. *Nidovirales*: Evolving the largest RNA virus genome. Virus Research 117:17–37.

Gordon A. 2014. fastx_toolkit. https://github.com/agordon/fastx_toolkit Version-0.0.14.

Guillou L., Bachar D., Audic S., Bass D., Berney C., Bittner L., Boutte C., Burgaud G., de Vargas C., Decelle J., del Campo J., Dolan J.R., Dunthorn M., Edvardsen B., Holzmann M., Kooistra W.H., Lara E., Le Bescot N., Logares R., Mahé F., Massana R., Montresor M., Morard R., Not F., Pawlowski J., Probert I., Sauvadet A.L., Siano R., Stoeck T., Vaulot D., Zimmermann P., and Christen R. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41:D597–D604.

Gustavsen J. 2015. RdRp 454 amplicons Jericho and SOG scripts. http://github.com/jooolia/RdRp_454_amplicons_Jericho_and_SOG Doi: 10.5281/zenodo.12509.

Gustavsen J. 2016. Temporal data scripts. https://github.com/jooolia/phylo_temporal_jericho DOI Zenodo: TBA.

Gustavsen J.A., Winget D.M., Tian X., and Suttle C.A. 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.

Haigh R., Taylor F., and Sutherland T. 1992. Phytoplankton ecology of Sechelt Inlet, a fjord system on the British Columbia coast. I. General features of the nano- and microplankton. Marine Ecology Progress Series 89:117–134.

Harrell F. 2014. Hmisc: A package of miscellaneous R functions. https://cran.r-project.org/web/packages/Hmisc/index.html Version 3.17-1.

Harrison P.J., Fulton J.D., Taylor F.J.R., and Parsons T.R. 1983. Review of the biological oceanography of the Strait of Georgia: pelagic environment. Canadian Journal of Fisheries and Aquatic Sciences 40:1064–1094.

Harvey E.L. and Menden-Deuer S. 2012. Predator-induced fleeing behaviors in phytoplankton: a new mechanism for harmful algal bloom formation? PLoS ONE 7:e46438.

Harvey P. and Purvis A. 1991. Comparative methods for explaining adaptations. Nature 351:619–624.

Heldal M. and Bratbak G. 1991. Production and decay of viruses in aquatic environments. Mar. Ecol. Prog. Ser 72:205–212.

Hewson I., Steele J.A., Capone D.G., and Fuhrman J.A. 2006a. Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. Marine Ecology Progress Series 311:67–77.

Hewson I., Vargo G.A., and Fuhrman J.A. 2003. Bacterial diversity in shallow oligotrophic marine benthos and overlying waters: effects of virus infection, containment, and nutrient enrichment. Microbial Ecology 46:322–336.

Hewson I., Winget D.M., Williamson K.E., Fuhrman J.A., and Wommack K.E. 2006b. Viral and bacterial assemblage covariance in oligotrophic waters of the West Florida Shelf (Gulf of Mexico). Journal of the Marine Biological Association of the United Kingdom 86:591–604.

Highfield A., Evans C., Walne A., Miller P.I., and Schroeder D.C. 2014. How many Coccolithovirus genotypes does it take to terminate an *Emiliania huxleyi* bloom? Virology 466-467:138–145.

Hijmans R.J. 2015. geosphere: Spherical Trigonometry. http://CRAN.R-project.org/package=geosphere R package version 1.5-1.

Holmes E. 2009. Evolution and emergence of RNA viruses. Oxford Series in Ecology and Evolution, Oxford University Press, USA.

Holmes E.C. 2010. The RNA Virus Quasispecies: Fact or Fiction? Journal of Molecular Biology 400:271–273.

Holt R.D. 2008. Theoretical perspectives on resource pulses. Ecology 89:671–681.

Holt R.D. and Pickering J. 1985. Infectious disease and species coexistence: a model of Lotka-Volterra form. The American Naturalist 126:196–211.

Horner R.A., Garrison D.L., and Plumley F.G. 1997. Harmful algal blooms and red tide problems on the US west coast. Limnology and Oceanography 42:1076–1088.

Horner-Devine M.C. and Bohannan B.J. 2006. Phylogenetic clustering and overdispersion in bacterial communities. Ecology 87:S100–S108.

Huang S., Wilhelm S.W., Jiao N., and Chen F. 2010. Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. The ISME Journal 4:1243–1251.

Hurwitz B.L. and Sullivan M.B. 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS ONE 8:e57355.

Huse S.M., Dethlefsen L., Huber J., Welch D.M., Relman D., and Sogin M.L. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genetics 4:e1000255.

Huston M.A. and DeAngelis D.L. 1994. Competition and coexistence: the effects of resource transport and supply rates. The American Naturalist 144:954–977.

Ibarbalz F.M., Pérez M.V., Figuerola E.L.M., and Erijman L. 2014. The bias associated with amplicon sequencing does not affect the quantitative assessment of bacterial community dynamics. PLoS ONE 9:e99722.

Irvine J. and Crawford W. 2012. State of physical, biological, and selected fishery resources of Pacific Canadian marine ecosystems in 2011. J.R. Irvine and W.R. Crawford.

Jameson E., Mann N.H., Joint I., Sambles C., and Mühling M. 2011. The diversity of cyanomyovirus populations along a North–South Atlantic Ocean transect. The ISME journal 5:1713–1721.

Jia Z., Ishihara R., Nakajima Y., Asakawa S., and Kimura M. 2007. Molecular characterization of T4-type bacteriophages in a rice field. Environmental Microbiology 9:1091–1096.

Jiang S.C. and Paul J.H. 1994. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. Marine Ecology-Progress Series 104:163–172.

Johnson L.S., Eddy S.R., and Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC bioinformatics 11:431.

Jones A.C., Liao T.S.V., Najar F.Z., Roe B.A., Hambright K.D., and Caron D.A. 2013. Seasonality and disturbance: annual pattern and response of the bacterial and microbial eukaryotic assemblages in a freshwater ecosystem: Freshwater microbial community structure. Environmental Microbiology 15:2557–2572.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., and Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D., Blomberg S.P., and Webb C.O. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kemp P.F. and Aller J.Y. 2004. Estimating prokaryotic diversity: when are 16**S** rDNA libraries large enough? Limnology and Oceanography: Methods 2:114–125.

Kendrick B.J., DiTullio G.R., Cyronak T.J., Fulton J.M., Van Mooy B.A.S., and Bidle K.D. 2014. Temperature-induced viral resistance in *Emiliania huxleyi* (Prymnesiophyceae). PLoS ONE 9:e112134.

Khan S., Arakawa O., and Onoue Y. 1997. Neurotoxins in a toxic red tide of *Heterosigma akashiwo* (Raphidophyceae) in Kagoshima Bay, Japan. Aquaculture Research 28:9–14.

Kimmance S., Allen M., Pagarete A., Martínez Martínez J., and Wilson W. 2014. Reduction in photosystem II efficiency during a virus-controlled *Emiliania huxleyi* bloom. Marine Ecology Progress Series 495:65–76.

Kirchman D.L., Cottrell M.T., and Lovejoy C. 2010. The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16**S** rRNA genes: Arctic Bacteria in Winter and Summer. Environmental Microbiology 12:1132–1143.

Kirchman D.L., Suzuki Y., Garside C., and Ducklow H.W. 1991. High turnover rates of dissolved organic carbon during a spring phytoplankton bloom. Nature 352:612–614.

Klopfer P. and MacArthur R. 1960. Niche size and faunal diversity. The American Naturalist 94:293–300.

Koeppel A.F. and Wu M. 2013a. Species matter: the role of competition in the assembly of congeneric bacteria. The ISME journal 8:531–540.

Koeppel A.F. and Wu M. 2013b. Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. Nucleic Acids Research 41:5175–5188.

Konwar K.M., Hanson N.W., Bhatia M.P., Kim D., Wu S.J., Hahn A.S., Morgan-Lang C., Cheung H.K., and Hallam S.J. 2015. MetaPathways v2.5: quantitative functional, taxonomic and usability improvements. Bioinformatics 31:3345–3347.

Koonin E.V. 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. Journal of General Virology 72:2197–2206.

Koonin E.V., Wolf Y.I., Nagasaki K., and Dolja V.V. 2008. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. Nature Reviews Microbiology 6:925–939.

Kunin V., Engelbrektson A., Ochman H., and Hugenholtz P. 2009. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental Microbiology 12:118–123.

Kyle J.E., Eydal H.S., Ferris F.G., and Pedersen K. 2008. Viruses in granitic groundwater from 69 to 450 m depth of the Äspö hard rock laboratory, Sweden. The ISME Journal 2:571–574.

Ladau J., Sharpton T.J., Finucane M.M., Jospin G., Kembel S.W., O'Dwyer J., Koeppel A.F., Green J.L., and Pollard K.S. 2013. Global marine bacterial diversity peaks at high latitudes in winter. The ISME journal 7:1669–1677.

Landa M., Blain S., Christaki U., Monchy S., and Obernosterer I. 2016. Shifts in bacterial community composition associated with increased carbon cycling in a mosaic of phytoplankton blooms. ISME J 10:39–50.

Lang A.S., Culley A.I., and Suttle C.A. 2004. Genome sequence and characterization of a virus (HaRNAV) related to picorna-like viruses that infects the marine toxic bloom-forming alga *Heterosigma akashiwo*. Virology 320:206–217.

Lang A.S., Rise M.L., Culley A.I., and Steward G.F. 2009. RNA viruses in the sea. FEMS Microbiology Reviews 33:295–323.

Larsen A., Flaten G.A.F., Sandaa R.A., Castberg T., Thyrhaug R., Erga S.R., Jacquet S., and Bratbak G. 2004. Spring phytoplankton bloom dynamics in Norwegian coastal waters: Microbial community succession and diversity. Limnology and Oceanography 49:180–190.

Larsen J.B., Larsen A., Thyrhaug R., Bratbak G., and Sandaa R.A. 2008. Response of marine viral populations to a nutrient induced phytoplankton bloom at different pCO_2 levels. Biogeosciences 5:523–533.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278.

Lawrence J.E. and Suttle C.A. 2004. Effect of viral infection on sinking rates of *Heterosigma akashiwo* and its implications for bloom termination. Aquatic Microbial Ecology 37:1–7.

Le Gall O., Christian P., Fauquet C.M., King A.M.Q., Knowles N.J., Nakashima N., Stanway G., and Gorbalenya A.E. 2008. *Picornavirales*, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. Archives of Virology 153:715–727.

Lee C.K., Herbold C.W., Polson S.W., Wommack K.E., Williamson S.J., McDonald I.R., and Cary S.C. 2012. Groundtruthing next-gen sequencing for microbial ecology–biases and errors in community structure estimates from PCR amplicon pyrosequencing. PLoS ONE 7:e44224.

Legendre L. and Rivkin R. 2008. Planktonic food webs: microbial hub approach. Marine Ecology Progress Series 365:289–309.

Legendre P. and Legendre L. 1998. Numerical Ecology.

Lennon J.T., Aanderud Z.T., Lehmkuhl B.K., and Schoolmaster Jr D.R. 2012. Mapping the niche space of soil microorganisms using taxonomy and traits. Ecology 93:1867–1879.

Lenski R.E. 1988. Dynamics of interactions between bacteria and virulent bacteriophage. *In* Advances in Microbial Ecology, pages 1–44, Springer.

Levin S.A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. Ecology 73:1943–1967.

Lewitus A.J., Horner R.A., Caron D.A., Garcia-Mendoza E., Hickey B.M., Hunter M., Huppert D.D., Kudela R.M., Langlois G.W., Largier J.L., Lessard E.J., RaLonde R., Jack Rensel J., Strutton P.G., Trainer V.L., and Tweddle J.F. 2012. Harmful algal blooms along the North American west coast region: History, trends, causes, and impacts. Harmful Algae 19:133–159.

Liang Y., Li L., Luo T., Zhang Y., Zhang R., and Jiao N. 2014. Horizontal and vertical distribution of marine virioplankton: a basin scale investigation based on a global cruise. PLoS ONE 9:e111634.

Lima-Mendez G., Faust K., Henry N., Decelle J., Colin S., Carcillo F., Chaffron S., Ignacio-Espinosa J.C., Roux S., Vincent F., and others. 2015. Determinants of community structure in the global plankton interactome. Science 348:1262073.

Logares R., Audic S., Bass D., Bittner L., Boutte C., Christen R., Claverie J.M., Decelle J., Dolan J.R., Dunthorn M., Edvardsen B., Gobet A., Kooistra W.H., Mahe F., Not F., Ogata H., Pawlowski J., Pernice M.C., Romac S., Shalchian-Tabrizi K., Simon N., Stoeck T., Santini S., Siano R., Wincker P., Zingone A., Richards T.A., de Vargas C., and Massana R. 2014. Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24:813–821.

Loman N.J., Misra R.V., Dallman T.J., Constantinidou C., Gharbia S.E., Wain J., and Pallen M.J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnology 30:434–439.

Long J.R., Pittet V., Trost B., Yan Q., Vickers D., Haakensen M., and Kusalik A. 2014. Equivalent input produces different output in the UniFrac significance test. BMC bioinformatics 15:1.

Lozupone C. and Knight R. 2005. Unifrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71:8228–8235.

Lozupone C.A. and Knight R. 2007. Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences 104:11436–11440.

Lozupone C.A. and Knight R. 2015. The unifrac significance test is sensitive to tree topology. BMC Bioinformatics 16:1–3.

López-Bueno A., Rastrojo A., Peiró R., Arenas M., and Alcamí A. 2015. Ecological connectivity shapes quasispecies structure of RNA viruses in an Antarctic lake. Molecular Ecology 24:4812–4825.

López-Bueno A., Tamames J., Velázquez D., Moya A., Quesada A., and Alcamí A. 2009. High diversity of the viral community from an Antarctic lake. Science 326:858–861.

MacArthur R.H. and Wilson E.O. 1967. The theory of island biogeography, vol. 1. Princeton University Press.

Mahadevan A., D'Asaro E., Lee C., and Perry M.J. 2012. Eddy-driven stratification initiates north atlantic spring phytoplankton blooms. Science 337:54–58.

Marchler-Bauer A., Lu S., Anderson J.B., Chitsaz F., Derbyshire M.K., DeWeese-Scott C., Fong J.H., Geer L.Y., Geer R.C., Gonzales N.R., Gwadz M., Hurwitz D.I., Jackson J.D., Ke Z., Lanczycki C.J., Lu F., Marchler G.H., Mullokandov M., Omelchenko M.V., Robertson C.L., Song J.S., Thanki N., Yamashita R.A., Zhang D., Zhang N., Zheng C., and Bryant S.H. 2010. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research 39:D225–D229.

Marston M.F., Pierciey F.J., Shepard A., Gearin G., Qi J., Yandava C., Schuster S.C., Henn M.R., and Martiny J.B.H. 2012. Rapid diversification of coevolving marine *Synechococcus* and a virus. Proceedings of the National Academy of Sciences 109:4544–4549.

Marston M.F. and Sallee J.L. 2013. Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. Applied and Environmental Microbiology 69:4639–4647.

Marston M.F., Taylor S., Sme N., Parsons R.J., Noyes T.J.E., and Martiny J.B.H. 2013. Marine cyanophages exhibit local and regional biogeography. Environmental Microbiology 15:1452–1463.

Martinez Martinez J., Schroeder D.C., Larsen A., Bratbak G., and Wilson W.H. 2007. Molecular dynamics of *Emiliania huxleyi* and cooccurring viruses during two separate mesocosm studies. Applied and Environmental Microbiology 73:554–562.

Martiny J.B.H., Jones S.E., Lennon J.T., and Martiny A.C. 2015. Microbiomes in light of traits: A phylogenetic perspective. Science 350:aac9323–aac9323.

Massana R., Gobet A., Audic S., Bass D., Bittner L., Boutte C., Chambouvet A., Christen R., Claverie J.M., Decelle J., Dolan J.R., Dunthorn M., Edvardsen B., Forn I., Forster D., Guillou L., Jaillon O., Kooistra W.H.C.F., Logares R., Mahé F., Not F., Ogata H., Pawlowski J., Pernice M.C., Probert I., Romac S., Richards T., Santini S., Shalchian-Tabrizi K., Siano R., Simon N., Stoeck T., Vaulot D., Zingone A., and de Vargas C. 2015. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environmental Microbiology 17:4035–4049.

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

Matteson A.R., Loar S.N., Pickmere S., DeBruyn J.M., Ellwood M.J., Boyd P.W., Hutchins D.A., and Wilhelm S.W. 2012. Production of viruses during a spring phytoplankton bloom in the South Pacific Ocean near of New Zealand. FEMS Microbiology Ecology 79:709–719.

Mayer J.A. and Taylor F.R. 1979. A virus which lyses the marine nanoflagellate *Micromonas pusilla*. Nature 281:299–301.

Middelboe M. 2000. Bacterial growth rate and marine virus–host dynamics. Microbial Ecology 40:114–124.

Middelboe M., Hagström A., Blackburn N., Sinn B., Fischer U., Borch N., Pinhassi J., Simu K., and Lorenz M. 2001. Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria. Microbial Ecology 42:395–406.

Middelboe M., Holmfeldt K., Riemann L., Nybroe O., and Haaber J. 2009. Bacteriophages drive strain diversification in a marine *Flavobacterium* : implications for phage resistance and physiological properties. Environmental Microbiology 11:1971–1982.

Miki T. and Jacquet S. 2010. Indirect interactions in the microbial world: specificities and similarities to plant–insect systems. Population Ecology 52:475–483.

Miller M., Pfeiffer W., Schwartz T., and others. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *In* Gateway Computing Environments Workshop (GCE), 2010, pages 1–8, IEEE.

Milo R., Jorgensen P., Moran U., Weber G., and Springer M. 2010. BioNumbers the database of key numbers in molecular and cell biology. Nucleic Acids Research 38:D750–D753.

Mizumoto H., Tomaru Y., Takao Y., Shirai Y., and Nagasaki K. 2006. Intraspecies host specificity of a single-stranded RNA virus infecting a marine photosynthetic protist is determined at the early steps of infection. Journal of Virology 81:1372–1378.

Mojica K.D. and Brussaard C.P. 2014. Factors affecting virus dynamics and microbial host-virus interactions in marine environments. FEMS Microbiology Ecology 89:495–515.

Monchy S., Grattepanche J.D., Breton E., Meloni D., Sanciu G., Chabé M., Delhaes L., Viscogliosi E., Sime-Ngando T., and Christaki U. 2012. Microplanktonic community structure in a coastal system relative to a *Phaeocystis* bloom inferred from morphological and tag pyrosequencing methods. PLoS ONE 7:e39924.

Moreau H., Piganeau G., Desdevises Y., Cooke R., Derelle E., and Grimsley N. 2010. Marine Prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. Journal of Virology 84:12555–12563.

Morris R.M., Vergin K.L., Cho J.C., Rappé M.S., Carlson C.A., and Giovannoni S.J. 2005. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. Limnology and Oceanography 50:1687–1696.

Murray A.G. and Jackson G.A. 1992. Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. Marine Ecology Progress Series 89:103–116.

Muyzer G., Teske A., Wirsen C.O., and Jannasch H.W. 1995. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16**S** rDNA fragments. Archives of Microbiology 164:165–172.

Mühling M., Fuller N.J., Millard A., Somerfield P.J., Marie D., Wilson W.H., Scanlan D.J., Post A.F., Joint I., and Mann N.H. 2005. Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. Environmental Microbiology 7:499–508.

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

Nagasaki K., Shirai Y., Takao Y., Mizumoto H., Nishida K., and Tomaru Y. 2005. Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate *Heterocapsa circularisquama*. Applied and Environmental Microbiology 71:8888–8894.

Nagasaki K., Tomaru Y., Katanozaka N., Shirai Y., Nishida K., Itakura S., and Yamaguchi M. 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.

Nagasaki K. and Yamaguchi M. 1997. Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (Raphidophyceae). Aquatic Microbial Ecology 13:135–140.

Nakamura A., Okamoto T., Komatsu N., Ooka S., Oda T., Ishimatsu A., and Muramatsu T. 1998. Fish mucus stimulates the generation of superoxide anion by *Chattonella marina* and *Heterosigma akashiwo*. Fisheries Science 64:866–869.

Needham D.M., Chow C.E.T., Cram J.A., Sachdeva R., Parada A., and Fuhrman J.A. 2013. Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. The ISME Journal 7:1274–1285.

Nemergut D.R., Schmidt S.K., Fukami T., O'Neill S.P., Bilinski T.M., Stanish L.F., Knelman J.E., Darcy J.L., Lynch R.C., Wickey P., and Ferrenberg S. 2013. Patterns and processes of microbial community assembly. Microbiology and Molecular Biology Reviews 77:342–356.

Ng T.F.F., Marine R., Wang C., Simmonds P., Kapusinszky B., Bodhidatta L., Oderinde B.S., Wommack K.E., and Delwart E. 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. Journal of Virology 86:12161–12175.

Nissimov J.I., Napier J.A., Allen M.J., and Kimmance S.A. 2016. Intragenus competition between coccolithoviruses: an insight on how a select few can come to dominate many. Environmental Microbiology 18:133–145.

Noble R.T. and Fuhrman J.A. 1997. Virus decay and its causes in coastal waters. Applied and Environmental Microbiology 63:77–83.

Oda T., Nakamura A., Okamoto T., Ishimatsu A., and Muramatsu T. 1998. Lectin-induced enhancement of superoxide anion production by red tide phytoplankton. Marine Biology 131:383–390.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., and Wagner H. 2013. vegan: Community Ecology Package. R package version 2.0-8.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., and Wagner H. 2015. vegan: Community Ecology Package. R package version 2.3-0.

Ono K., Khan S., and Onoue Y. 2000. Effects of temperature and light intensity on the growth and toxicity of *Heterosigma akashiwo* (Raphidophyceae). Aquaculture Research 31:427–433.

Ono K., Muetze T., Kolishovski G., Shannon P., and Demchak B. 2015. CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API [version 1; referees: 2 approved]. F1000Research 4.

Pagarete A., Chow C.E.T., Johannessen T., Fuhrman J.A., Thingstad T.F., and Sandaa R.A. 2013. Strong seasonality and interannual recurrence in marine myovirus communities. Applied and Environmental Microbiology 79:6253–6259.

Parada A., Needham D.M., and Fuhrman J.A. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. Environmental Microbiology 18:1403–1414.

Paradis E. 2012. Analysis of Phylogenetics and Evolution with R. Use R!, Springer, New York.

Parsons R.J., Breitbart M., Lomas M.W., and Carlson C.A. 2012. Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. The ISME journal 6:273–284.

Parsons T., Maita Y., and Lalli C. 1984. A manual of chemical and biological methods for seawater analysis. Pergamon international library of science, technology, engineering, and social studies, Pergamon Press, New York.

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

Payet J. and Suttle C. 2014. Viral infection of bacteria and phytoplankton in the Arctic Ocean as viewed through the lens of fingerprint analysis. Aquatic Microbial Ecology 72:47–61.

Pedrós-Alió C. 2012. The rare bacterial biosphere. Annual Review of Marine Science 4:449–466.

Petraitis P.S., Latham R.E., and Niesenbaum R.A. 1989. The maintenance of species diversity by disturbance. The Quarterly Review of Biology 64:393–418.

Peura S., Bertilsson S., Jones R.I., and Eiler A. 2015. Resistant microbial cooccurrence patterns inferred by network topology. Applied and Environmental Microbiology 81:2090–2097.

Philippe N., Legendre M., Doutre G., Couté Y., Poirot O., Lescot M., Arslan D., Seltzer V., Bertaux L., Bruley C., Garin J., Claverie J.M., and Abergel C. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341:281–286.

Pielou E.C. 1966. The measurement of diversity in different types of biological collections. Journal of Theoretical Biology 13:131 – 144.

Pinhassi J., Sala M.M., Havskum H., Peters F., Guadayol O.S., Malits A., and Marrase C. 2004. Changes in bacterioplankton composition under different phytoplankton regimens. Applied and Environmental Microbiology 70:6753–6766.

Pinto A.J. and Raskin L. 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS ONE 7:e43093.

Powers L., Creed I.F., and Trick C.G. 2012. Sinking of *Heterosigma akashiwo* results in increased toxicity of this harmful algal bloom species. Harmful Algae 13:95–104.

Price M.N., Dehal P.S., and Arkin A.P. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one 5:e9490.

Proctor L.M. and Fuhrman J.A. 1990. Viral mortality of marine bacteria and cyanobacteria. Nature 343:60–62.

Qiu X., Wu L., Huang H., McDonel P.E., Palumbo A.V., Tiedje J.M., and Zhou J. 2001. Evaluation of PCR-Generated chimeras, mutations, and heteroduplexes with 16**S** rRNA gene-based cloning. Applied and Environmental Microbiology 67:880–887.

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., and Glockner F.O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41:D590–D596.

Quince C., Lanzén A., Curtis T.P., Davenport R.J., Hall N., Head I.M., Read L.F., and Sloan W.T. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods 6:639–641.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A. 2014. FigTree. http://tree.bio.ed.ac.uk/software/figtree/ Version 1.4.2.

Reeder J. and Knight R. 2010. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nature Methods 7:668–669.

Renzette N., Pokalyuk C., Gibson L., Bhattacharjee B., Schleiss M.R., Hamprecht K., Yamamoto A.Y., Mussi-Pinhata M.M., Britt W.J., Jensen J.D., and Kowalik T.F. 2015. Limits and patterns of cytomegalovirus genomic diversity in humans. Proceedings of the National Academy of Sciences 112:E4120–E4128.

Reynolds C.S., Padisák J., and Sommer U. 1993. Intermediate disturbance in the ecology of phytoplankton and the maintenance of species diversity: a synthesis. Hydrobiologia 249:183–188.

Rho M., Tang H., and Ye Y. 2010. FragGeneScan predicting genes in short and error-prone reads. Nucleic Acids Research 38:e191–e191.

Rodriguez-Brito B., Li L., Wegley L., Furlan M., Angly F., Breitbart M., Buchanan J., Desnues C., Dinsdale E., Edwards R., Felts B., Haynes M., Liu H., Lipson D., Mahaffy J., Martin-Cuadrado A.B., Mira A., Nulton J., Pašić L., Rayhawk S., Rodriguez-Mueller J., Rodriguez-Valera F., Salamon P., Srinagesh S., Thingstad T.F., Tran T., Thurber R.V., Willner D., Youle M., and Rohwer F. 2010. Viral and microbial community dynamics in four aquatic environments. The ISME Journal 4:739–751.

Rodriguez-Valera F., Martin-Cuadrado A.B., Rodriguez-Brito B., Pašić L., Thingstad T.F., Rohwer F., and Mira A. 2009. Explaining microbial population genomics through phage predation. Nature Reviews Microbiology 7:828–836.

Romano C.M., Lauck M., Salvador F.S., Lima C.R., Villas-Boas L.S., Araújo E.S.A., Levi J.E., Pannuti C.S., O'Connor D., and Kallas E.G. 2013. Inter- and intra-host viral diversity in a large seasonal DENV2 outbreak. PLoS ONE 8:e70318.

Rosenzweig M.L. 1995. Species diversity in space and time. Cambridge University Press.

Rosetta C.H. and McManus G.B. 2003. Feeding by ciliates on two harmful algal bloom species, *Prymnesium parvum* and *Prorocentrum minimum*. Harmful Algae 2:109–126.

Rozon R.M. and Short S.M. 2013. Complex seasonality observed amongst diverse phytoplankton viruses in the Bay of Quinte, an embayment of Lake Ontario. Freshwater Biology 58:2648–2663.

Ruan Q., Dutta D., Schwalbach M.S., Steele J.A., Fuhrman J.A., and Sun F. 2006. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22:2532–2538.

Sandaa R.A., Gómez-Consarnau L., Pinhassi J., Riemann L., Malits A., Weinbauer M.G., Gasol J.M., and Thingstad T.F. 2009. Viral control of bacterial biodiversity - evidence from a nutrient-enriched marine mesocosm experiment. Environmental Microbiology 11:2585–2597.

Sanfaçon H., Wellink J., Gall O., Karasev A., Vlugt R., and Wetzel T. 2009. Secoviridae: a proposed family of plant viruses within the order *Picornavirales* that combines the families *Sequiviridae* and *Comoviridae*, the unassigned genera *Cheravirus* and *Sadwavirus*, and the proposed genus Torradovirus. Archives of Virology 154:899–907.

Sanjuan R., Nebot M.R., Chirico N., Mansky L.M., and Belshaw R. 2010. Viral mutation rates. Journal of Virology 84:9733–9748.

Schirmer M., Ijaz U.Z., D'Amore R., Hall N., Sloan W.T., and Quince C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Research 43:e37.

Schloss P.D., Gevers D., and Westcott S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16s rRNA-based studies. PLoS ONE 6:e27310.

Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Lesniewski R.A., Oakley B.B., Parks D.H., Robinson C.J., Sahl J.W., Stres B., Thallinger G.G., Van Horn D.J., and Weber C.F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75:7537–7541.

Schluter D. 1984. A variance test for detecting species associations, with some example applications. Ecology 65:998–1005.

Schmidt T.S.B., Matias Rodrigues J.F., and von Mering C. 2015. Limits to robustness and reproducibility in the demarcation of operational taxonomic units: Robustness and reproducibility in OTU demarcation. Environmental Microbiology 17:1689–1706.

Schroeder D.C., Oke J., Hall M., Malin G., and Wilson W.H. 2003. Virus succession observed during an *Emiliania huxleyi* bloom. Applied and Environmental Microbiology 69:2484–2490.

Schwalbach M.S., Hewson I., and Fuhrman J.A. 2004. Viral effects on bacterial community composition in marine plankton microcosms. Aquatic Microbial Ecology 34:117–127.

Seymour J.R., Seuront L., Doubell M., Waters R.L., and Mitchell J.G. 2006. Microscale patchiness of virioplankton. Journal of the Marine Biological Association of the United Kingdom 86:551–562.

Shade A., Jones S.E., Caporaso J.G., Handelsman J., Knight R., Fierer N., and Gilbert J.A. 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio 5:e01371–14–e01371–14.

Shade A., Peter H., Allison S.D., Baho D.L., Berga M., Bürgmann H., Huber D.H., Langenheder S., Lennon J.T., Martiny J.B.H., Matulich K.L., Schmidt T.M., and Handelsman J. 2012a. Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology 3.

Shade A., Read J.S., Youngblut N.D., Fierer N., Knight R., Kratz T.K., Lottig N.R., Roden E.E., Stanley E.H., and Stombaugh J. 2012b. Lake microbial communities are resilient after a whole-ecosystem disturbance. The ISME journal 6:2153–2167.

Shannon C. 1948. A mathematical theory of distribution. The Bell System Technical Journal 27:623.

Sher D., Thompson J.W., Kashtan N., Croal L., and Chisholm S.W. 2011. Response of *Prochlorococcus* ecotypes to co-culture with diverse marine bacteria. The ISME journal 5:1125–1132.

Shirai Y., Tomaru Y., Takao Y., Suzuki H., Nagumo T., and 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.

Short C.M., Rusanova O., and Short S.M. 2010. Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. The ISME Journal 5:810–821.

Short C.M. and Suttle C.A. 2005. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Applied and Environmental Microbiology 71:480–486.

Short S.M. 2012. The ecology of viruses that infect eukaryotic algae: Algal virus ecology. Environmental Microbiology 14:2253–2271.

Short S.M. and Short C.M. 2009. Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. Environmental Microbiology 11:2639–2648.

Short S.M. and Suttle C.A. 2002. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. Applied and Environmental Microbiology 68:1290–1296.

Short S.M. and Suttle C.A. 2003. Temporal dynamics of natural communities of marine algal viruses and eukaryotes. Aquatic Microbial Ecology 32:107–119.

Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Soding J., Thompson J.D., and Higgins D.G. 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539–539.

Simon M., López-García P., Deschamps P., Moreira D., Restoux G., Bertolino P., and Jardillier L. 2015. Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. The ISME Journal 9:1941–1953.

Simpson E. 1949. Measurement of diversity. Nature 163:688–688.

Sipos R., Székely A.J., Palatinszky M., Révész S., Márialigeti K., and Nikolausz M. 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16s rRNA gene-targetting bacterial community analysis. FEMS Microbiology Ecology 60:341–350.

Smoot M.E., Ono K., Ruscheinski J., Wang P.L., and Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432.

Sogin M.L., Morrison H.G., Huber J.A., Welch D.M., Huse S.M., Neal P.R., Arrieta J.M., and Herndl G.J. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences 103:12115–12120.

Sorensen G., Baker A.C., Hall M.J., Munn C.B., and Schroeder D.C. 2009. Novel virus dynamics in an *Emiliania huxleyi* bloom. Journal of Plankton Research 31:787–791.

Spatharis S., Tsirtsis G., Danielidis D.B., Chi T.D., and Mouillot D. 2007. Effects of pulsed nutrient inputs on phytoplankton assemblage structure and blooms in an enclosed coastal area. Estuarine, Coastal and Shelf Science 73:807–815.

Spencer R. 1955. A marine bacteriophage. Nature 175:690–691.

Spencer R. 1960. Indigenous marine bacteriophages. Journal of Bacteriology 79:614.

Srivastava D.S., Cadotte M.W., MacDonald A.A.M., Marushia R.G., and Mirotchnick N. 2012. Phylogenetic diversity and the functioning of ecosystems. Ecology Letters 15:637–648.

Stackebrandt E. and Goebel B.M. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16s rRNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic Bacteriology 44:846–849.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stamatakis A., Hoover P., and Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Systematic Biology 57:758–771.

Steele J.A., Countway P.D., Xia L., Vigil P.D., Beman J.M., Kim D.Y., Chow C.E.T., Sachdeva R., Jones A.C., and Schwalbach M.S. 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. The ISME journal 5:1414–1425.

Steward G.F., Culley A.I., Mueller J.A., Wood-Charlson E.M., Belcaid M., and Poisson G. 2012. Are we missing half of the viruses in the ocean? The ISME Journal 7:672–679.

Steward G.F., Montiel J.L., and Azam F. 2000. Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. Limnology and Oceanography 45:1697–1706.

Storesund J.E., Erga S.R., Ray J.L., Thingstad T.F., and Sandaa R.A. 2015. Top-down and bottom-up control on bacterial diversity in a western Norwegian deep-silled fjord. FEMS Microbiology Ecology 91:fiv076.

Storey J.D., Xiao W., Leek J.T., Tompkins R.G., and Davis R.W. 2005. Significance analysis of time course microarray experiments. Proceedings of the National Academy of Sciences of the United States of America 102:12837–12842.

Sullivan M.B. 2015. Viromes, not gene markers, for studying double-stranded DNA virus communities. Journal of Virology 89:2459–2461.

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

Suttle C.A. 2005. Viruses in the sea. Nature 437:356–361.

Suttle C.A. 2007. Marine viruses - major players in the global ecosystem. Nat Rev Micro 5:801–812.

Suttle C.A. and Chan A.M. 1994. Dynamics and distribution of cyanophages and their effect on marine Synechococcus spp. Applied and Environmental Microbiology 60:3167–3174.

Suttle C.A., Chan A.M., and Cottrell M.T. 1990. Infection of phytoplankton by viruses and reduction of primary productivity. Nature 347:467–469.

Suttle C.A., Chan A.M., and Cottrell M.T. 1991. Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. Applied and Environmental Microbiology 57:721–726.

Suttle C.A. and Chen F. 1992. Mechanisms and rates of decay of marine viruses in seawater. Applied and Environmental Microbiology 58:3721–3729.

Tai V., Lawrence J.E., Lang A.S., Chan A.M., Culley A.I., and Suttle C.A. 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., and 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.

Tan J., Kelly C.K., and Jiang L. 2013. Temporal niche promotes biodiversity during adaptive radiation. Nature Communications 4.

Tarutani K., Nagasaki K., and Yamaguchi M. 2000. Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. Applied and Environmental Microbiology 66:4916–4920.

Taylor F. and Haigh R. 1993. The ecology of fish-killing blooms of the chloromonad flagellate *Heterosigma* in the Strait of Georgia and adjacent waters. *In* Toxic Phytoplankton Blooms in the Sea, pages 705–710.

Taylor F., Haigh R., and Sutherland T. 1994. Phytoplankton ecology of Sechelt Inlet, a fjord system on the British Columbia coast. II. Potentially harmful species. Marine Ecology Progress Series 104:151–164.

te Velthuis A.J.W. 2014. Common and unique features of viral RNA-dependent polymerases. Cellular and Molecular Life Sciences 71:4403–4420.

Teeling H., Fuchs B.M., Becher D., Klockow C., Gardebrecht A., Bennke C.M., Kassabgy M., Huang S., Mann A.J., Waldmann J., Weber M., Klindworth A., Otto A., Lange J., Bernhardt J., Reinsch C., Hecker M., Peplies J., Bockelmann F.D., Callies U., Gerdts G., Wichels A., Wiltshire K.H., Glöckner F.O., Schweder T., and Amann R. 2012. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science 336:608–611.

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

Thingstad T.F., Pree B., Giske J., and Våge S. 2015. What difference does it make if viruses are strain-, rather than species-specific? Frontiers in Microbiology 6:320.

Thingstad T.F., Våge S., Storesund J.E., Sandaa R.A., and Giske J. 2014. A theoretical analysis of how strain-specific viruses can control microbial species diversity. Proceedings of the National Academy of Sciences 111:7813–7818.

Tomaru Y. 2015. Marine diatom viruses and their hosts: Resistance mechanisms and population dynamics. Perspectives in Phycology 2:69–81.

Tomaru Y., Katanozaka N., Nishida K., Shirai Y., Tarutani K., Yamaguchi M., and Nagasaki K. 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., Kimura K., and Nagasaki K. 2015. Marine protist viruses. *In* Marine Protists (S. Ohtsuka, T. Suzaki, T. Horiguchi, N. Suzuki, and F. Not, eds.), pages 501–517, Springer Japan, Tokyo.

Tomaru Y., Takao Y., Suzuki H., Nagumo T., and 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., and Nagasaki K. 2012. First evidence for the existence of pennate diatom viruses. The ISME Journal 6:1445–1448.

Tucker C.M. and Cadotte M.W. 2013. Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. Diversity and Distributions 19:845–854.

Van Etten J.L., Meints R.H., Burbank D.E., Kuczmarski D., Cuppels D.A., and Lane L.C. 1981. Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis*. Virology 113:704–711.

Van Valen L. 1973. A new evolutionary law. Evolutionary theory 1:1–30.

Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A., Wu D., Paulsen I., Nelson K.E., Nelson W., Fouts D.E., Levy S., Knap A.H., Lomas M.W., Nealson K., White O., Peterson J., Hoffman J., Parsons R., Baden-Tillson H., Pfannkoch C., Rogers Y.H., and Smith H.O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74.

Verity P.G. 1987. Abundance, community composition, size distribution, and production rates of tintinnids in Narragansett Bay, Rhode Island. Estuarine, Coastal and Shelf Science 24:671–690.

Verity P.G. and Stoecker D. 1982. Effects of *Olisthodiscus luteus* on the growth and abundance of tintinnids. Marine Biology 72:79–87.

Våge S., Storesund J.E., and Thingstad T.F. 2013. SAR11 viruses and defensive host strains. Nature 499:E3–E4.

Wang K., Wommack K.E., and Chen F. 2011. Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. Applied and Environmental Microbiology 77:7459–7468.

Wang Q., Garrity G.M., Tiedje J.M., and Cole J.R. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73:5261–5267.

Waterbury J.B. and Valois F.W. 1993. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. Applied and Environmental Microbiology 59:3393–3399.

Watson S.J., Welkers M.R.A., Depledge D.P., Coulter E., Breuer J.M., de Jong M.D., and Kellam P. 2013. Viral population analysis and minority-variant detection using short read next-generation sequencing. Philosophical Transactions of the Royal Society B: Biological Sciences 368:20120205.

Weinbauer M.G. 2004. Ecology of prokaryotic viruses. FEMS Microbiology Reviews 28:127–181.

Weitz J.S., Hartman H., and Levin S.A. 2005. Coevolutionary arms races between bacteria and bacteriophage. Proceedings of the National Academy of Sciences 102:9535–9540.

Weitz J.S., Stock C.A., Wilhelm S.W., Bourouiba L., Coleman M.L., Buchan A., Follows M.J., Fuhrman J.A., Jover L.F., Lennon J.T., Middelboe M., Sonderegger D.L., Suttle C.A., Taylor B.P., Frede Thingstad T., Wilson W.H., and Eric Wommack K. 2015. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. ISME J 9:1352–1364.

Weitz J.S. and Wilhelm S.W. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol Rep 4:17.

Wickham H. 2009. ggplot2. Springer New York, New York, NY.

Wilhelm S.W. and Suttle C.A. 1999. Viruses and nutrient cycles in the sea. BioScience 49:781–788.

Wilhelm S.W., Weinbauer M.G., Suttle C.A., and Jeffrey W.H. 1998. The role of sunlight in the removal and repair of viruses in the sea. Limnology and Oceanography 43:586–592.

Wilkinson L. 2011. venneuler: Venn and Euler Diagrams. R package version 1.1-0.

Williams R.J., Howe A., and Hofmockel K.S. 2014. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. Frontiers in Microbiology 5.

Williamson S.J., Rusch D.B., Yooseph S., Halpern A.L., Heidelberg K.B., Glass J.I., Andrews-Pfannkoch C., Fadrosh D., Miller C.S., Sutton G., Frazier M., and Venter J.C. 2008. The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. PLoS ONE 3:e1456.

Wilson W., Tarran G., Schroeder D., Cox M., Oke J., and Malin G. 2002a. Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. Journal of the Marine Biological Association of the UK 82:369–377.

Wilson W., Tarran G., and Zubkov M.V. 2002b. Virus dynamics in a coccolithophore-dominated bloom in the North Sea. Deep Sea Research Part II: Topical Studies in Oceanography 49:2951–2963.

Winget D.M., Helton R.R., Williamson K.E., Bench S.R., Williamson S.J., and Wommack K.E. 2011. Repeating patterns of virioplankton production within an estuarine ecosystem. Proceedings of the National Academy of Sciences 108:11506–11511.

Winget D.M. and Wommack K.E. 2008. Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. Applied and Environmental Microbiology 74:2612–2618.

Winter C., Bouvier T., Weinbauer M.G., and Thingstad T.F. 2010. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "Killing the Winner" hypothesis revisited. Microbiology and Molecular Biology Reviews 74:42–57.

Woese C.R. and Fox G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences 74:5088–5090.

Woese C.R., Kandler O., and Wheelis M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences 87:4576–4579.

Worden A.Z., Follows M.J., Giovannoni S.J., Wilken S., Zimmerman A.E., and Keeling P.J. 2015. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. Science 347:1257594–1257594.

Xia L., Steele J., Cram J., Cardon Z., Simmons S., Vallino J., Fuhrman J., and Sun F. 2011. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. BMC Systems Biology 5:S15.

Yang Y., Motegi C., Yokokawa T., and Nagata T. 2010. Large-scale distribution patterns of virioplankton in the upper ocean. Aquatic Microbial Ecology 60:233–246.

Yokoyama R. and Honda D. 2007. Taxonomic rearrangement of the genus *Schizochytrium sensu lato* based on morphology, chemotaxonomic characteristics, and 18**S** rRNA gene phylogeny (Thraustochytriaceae, Labyrinthulomycetes): emendation for *Schizochytrium* and erection of *Aurantiochytrium* and *Oblongichytrium* gen. nov. Mycoscience 48:199–211.

Yu G., Smith D., Zhu H., Guan Y., and Lam T.T.Y. 2016. ggtree: an R package for visualization and annotation of phylogenetic tree with different types of meta-data, submitted to Methods in Ecology and Evolution.

Zhang J. and Byrne C. 1999. Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. Biochem. J 337:231–241.

Zhang J., Kobert K., Flouri T., and Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620.

Zhao Y., Temperton B., Thrash J.C., Schwalbach M.S., Vergin K.L., Landry Z.C., Ellisman M., Deerinck T., Sullivan M.B., and Giovannoni S.J. 2013. Abundant SAR11 viruses in the ocean. Nature 494:357–360.

Zhong X. and Jacquet S. 2014. Differing assemblage composition and dynamics in T4-like myophages of two neighbouring sub-alpine lakes. Freshwater Biology 59:1577–1595.

Zhou J., Wu L., Deng Y., Zhi X., Jiang Y.H., Tu Q., Xie J., Van Nostrand J.D., He Z., and Yang Y. 2011. Reproducibility and quantitation of amplicon sequencing-based detection. The ISME journal 5:1303–1313.

Zingone A., Dubroca L., Iudicone D., Margiotta F., Corato F., Ribera d'Alcalà M., Saggiomo V., and Sarno D. 2009. Coastal phytoplankton do not rest in winter. Estuaries and Coasts 33:342–361.

APPENDIX A

Supplementary Information to Chapter 2

A.1 SUPPLEMENTARY DATA AND FIGURES

A.1.1 *Control libraries:*

Methods: — Two control libraries were prepared in addition to the 5 sample libraries prepared. To get the Sanger sequence for 1 amplicon, from one PCR amplification, a single sequence was cloned into TOPO TA vector (Invitrogen) and transformed into E. coli grown in LB + amp at 37°C overnight. Colony PCR was performed on several clones. Positive PCR products were cleaned using the Qiagen Minelute PCR cleanup and sequenced using the M13 forward primer at NAPS (UBC) on an ABI sequencer using Big Dye Chemistry. To make the unamplified control sequence the plasmid containing the clone was grown in large quantities overnight (6x75ml cultures). Cells were harvested by centrifugation at 3200 *g* at 4° for 20 min. The plasmids were extracted using the Qiagen mini-prep plasmid kit and digested with Ecor I for 2 h (600 μl DNA, 60 μl React 3 buffer, 300 μl Ecor I, 38.5 μl H20) to cut out the product from the vector. The digested extracts were run on 1.5% agarose gels and the Qiagen Minelute gel extraction kit was to purify the desired cut product. The purified cut product was processed in library preparation like the other samples (see Chapter 2 Materials and Methods). To make the amplified control sequences, the purified digested material was used as template in a PCR reaction as detailed above and the product was used in library preparation. Alignments of the control sequences were visualized in Geneious (v.6.1.6)(Kearse *et al.*, 2012).

RESULTS — The control libraries contained a total of 94 reads. Three reads were recovered from the non-amplified control sequence and 91 reads from the amplified cloned sequence. The non-amplified control sequences had no errors. However, there were only 3 reads recovered from that library. Therefore, it is difficult to compare to the amplified library. The 91 reads from the amplified library contained some sequences with insertions and some with erroneous base-calls (Figure A.2).

These control libraries enabled confident testing of the error-correction algorithm (Reeder and Knight, 2010). There were errors such as homopolymers and insertions attributable to PCR amplification and 454 pyrosequencing. However, the denoiser algorithm adequately corrected the viral OTU reads.

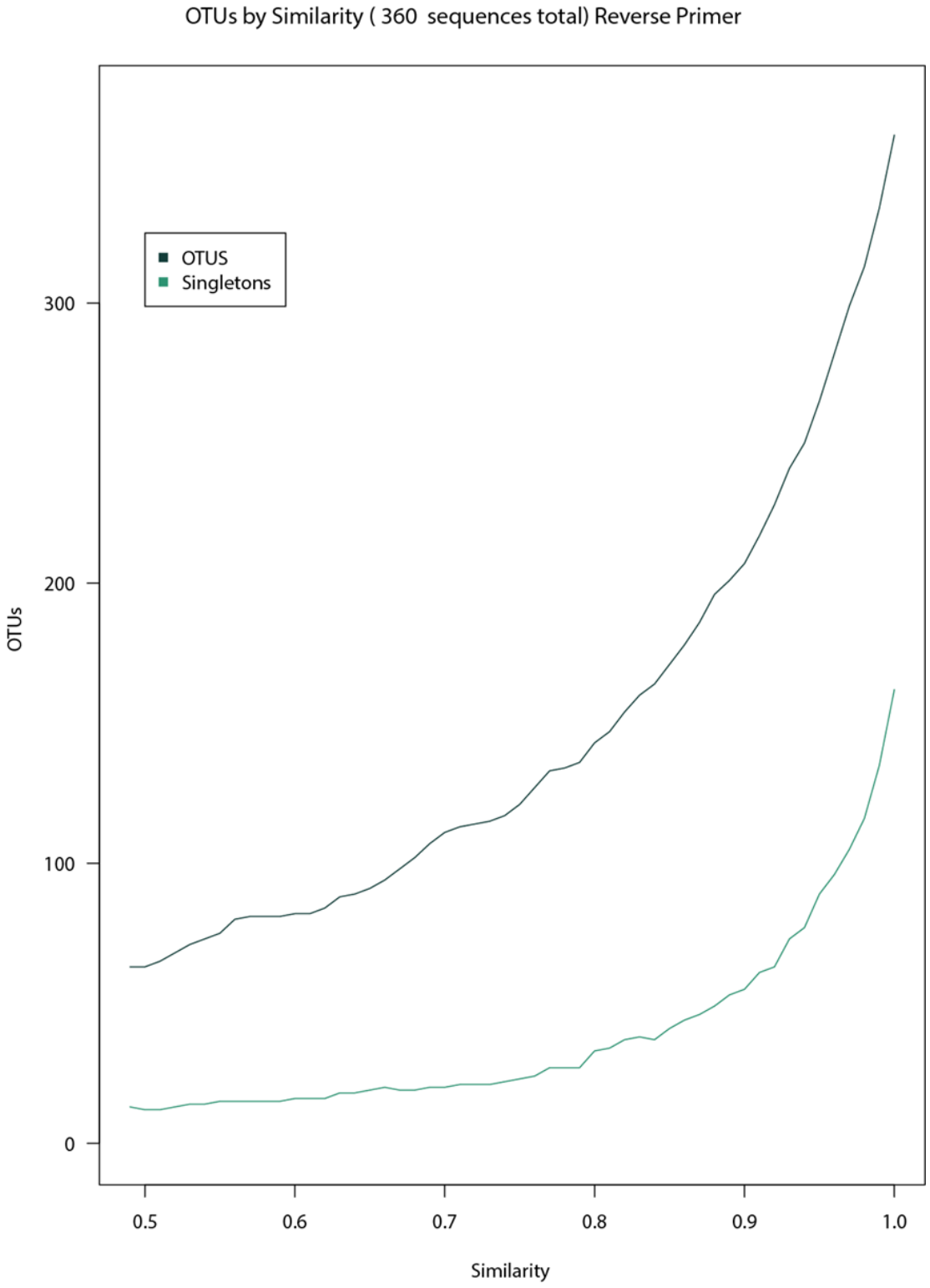

FIGURE A.1: Percent similarity vs. number of OTUs. All sequences were translated to amino acids using FragGeneScan with the 454_10 training option (Rho *et al.*, 2010) and were clustered with uclust (Edgar, 2010).

FIGURE A.2: Control sequence reads clustered at 95% similarity. All sequences were translated to amino acids using FragGeneScan with the 454_10 training option (Rho *et al.*, 2010) and clustered with uclust at 95% similarity using centroids as the output (Edgar, 2010). Sequences were aligned using default parameters for MUSCLE(Edgar, 2004). Mismatches in clustered sequences are highlighted in colours. Screenshot was taken from Geneious (v.6.1.6)(Kearse *et al.*, 2012).

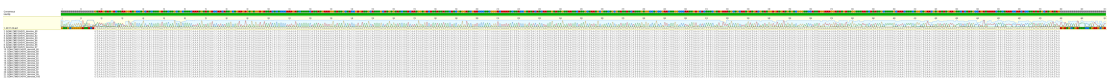

FIGURE A.3: Control sequence and PCR amplified reads denoised at different percentages using the QIIME denoiser Titanium settings (Reeder and Knight, 2010). All sequences were translated to amino acids using FragGeneScan with the 454_10 training option (Rho *et al.*, 2010) and clustered with uclust at 95% similarity using centroids as the output (Edgar, 2010). Sequences were aligned using default parameters for MUSCLE (Edgar, 2004). Screenshot was taken from Geneious (v.6.1.6) (Kearse *et al.*, 2012).

APPENDIX B

Supplementary Information to Chapter 3

B.1 SUPPLEMENTARY FIGURES

B.1.1 *Detailed phylogenetic trees*

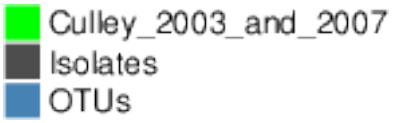

FIGURE B.1: Legend for tip colours for Maximum likelihood trees for marine picorna-like viruses.

MARINE PICORNA-LIKE VIRUSES —

T4-LIKE MYOVIRUSES —

B.1.2 *Heatmaps over time*

EUKARYOTES — There are no large scale fluctuations over time in the eukaryotic OTUs (Figure B.18). The OTUs that dominated the community fluctuated and were generally present, but some groups come and go a bit. Most notably the 1st and 2nd samples in January seem to have the most shifts between them.

BACTERIA — In the bacterial OTUs many OTUs are persistent and present except in 1-2 samples (Figure B.19). One sample in late June shows that there is a large turnover in the whole community and the community is then dominated by one group of related bacteria.

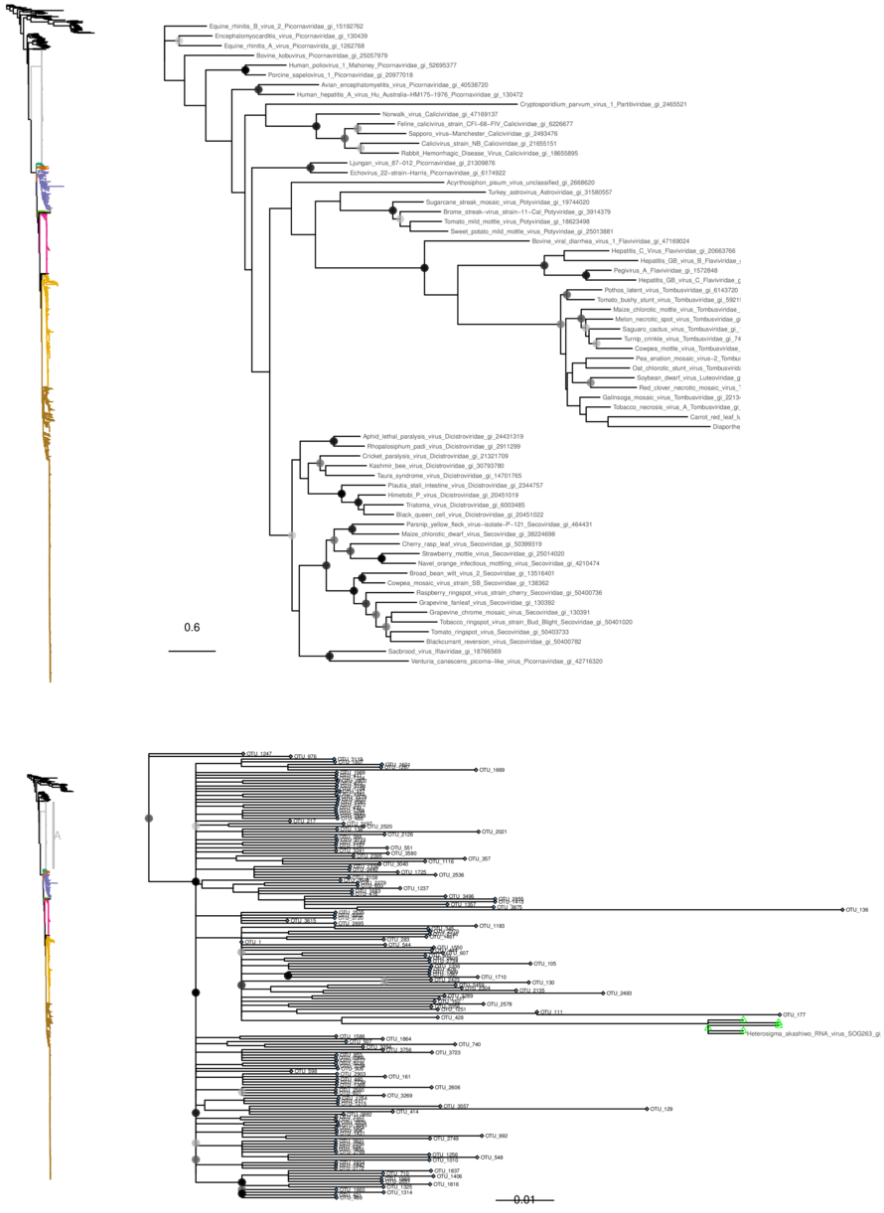

FIGURE B.2: Maximum likelihood phylogenetic tree (RAxML) of subsection A of RdRp including reference sequences and OTUs generated in this study. Subsection views are for the reference isolates (in black) and Group A (in grey). Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). OTUs at 95% similarity at the amino-acid level.

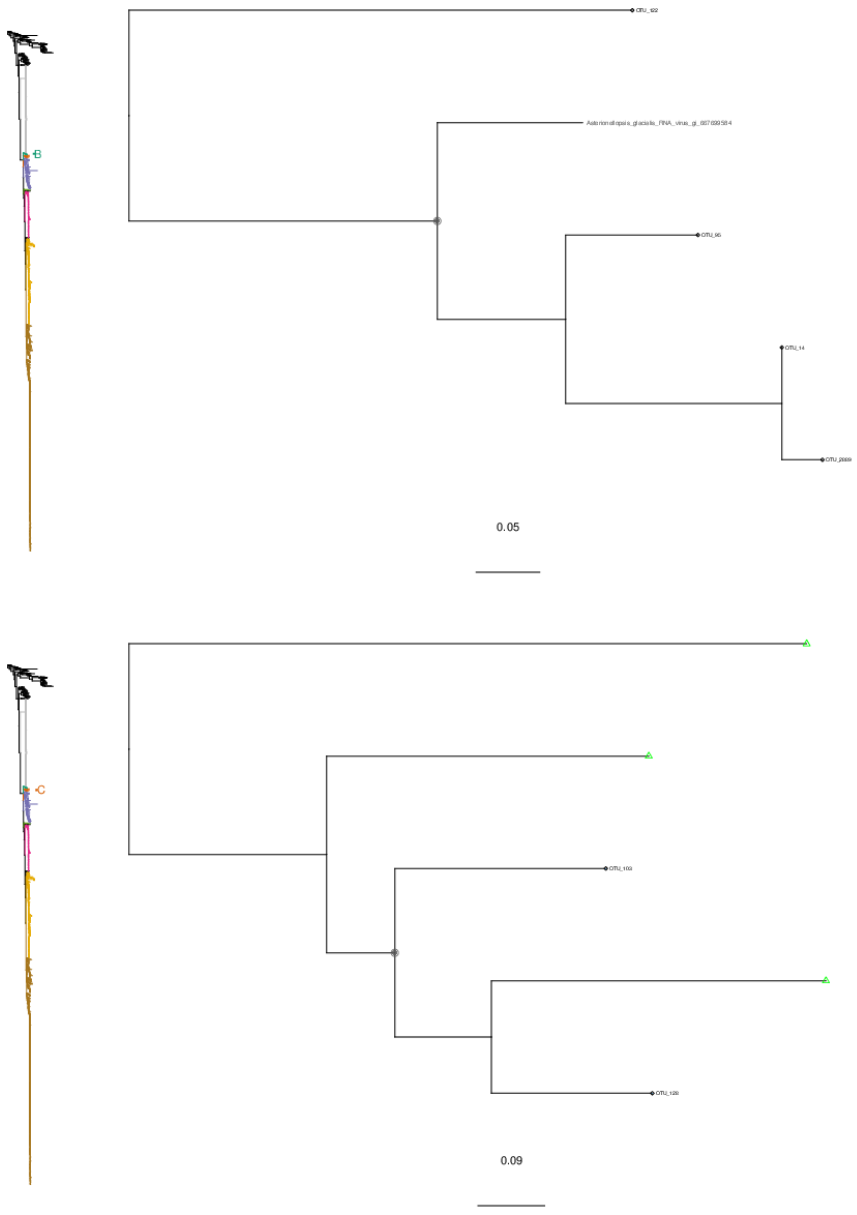

FIGURE B.3: Maximum likelihood phylogenetic tree (RAxML) of subsections B and C of RdRp including reference sequences and OTUs generated in this study. Subsection views are for the Group B (in blue) and Group C (in orange). Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). OTUs at 95% similarity at the amino-acid level.

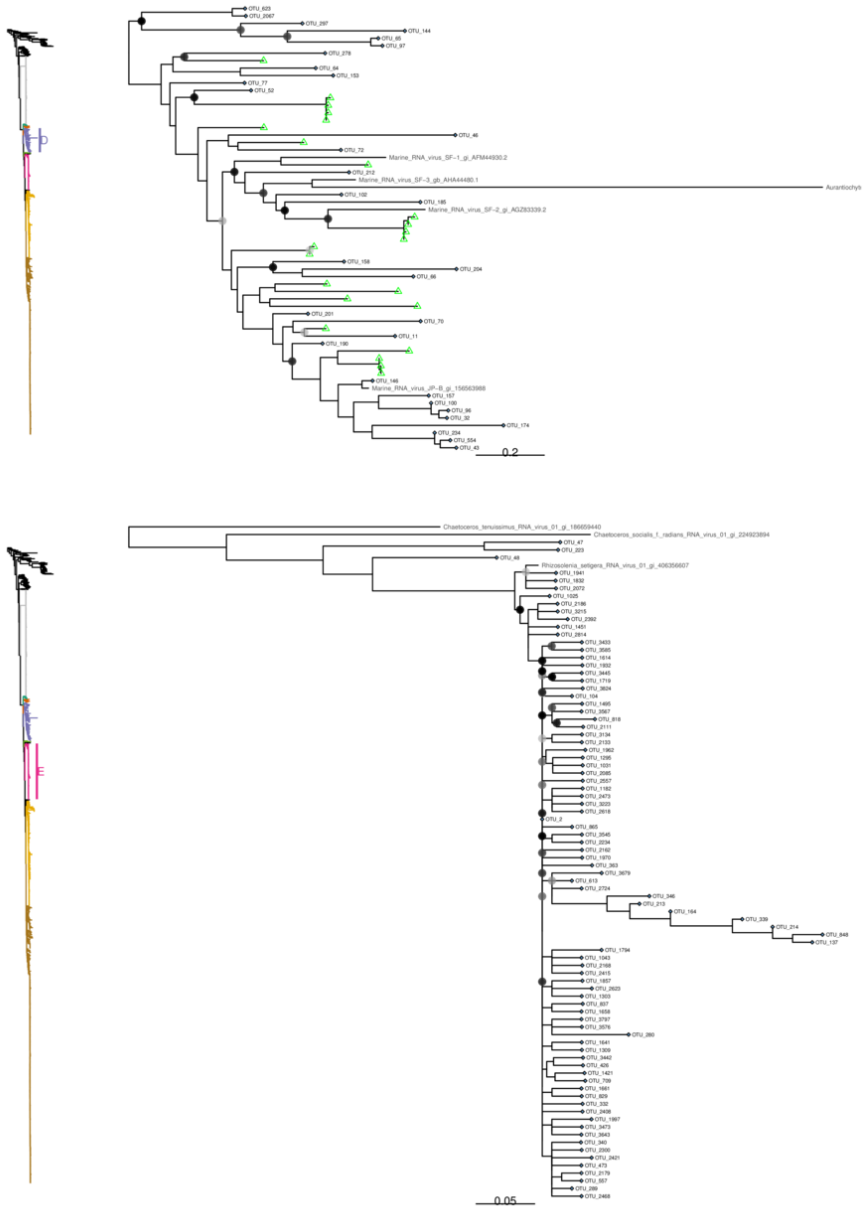

FIGURE B.4: Maximum likelihood phylogenetic tree (RAxML) of subsections D and E of RdRp including reference sequences and OTUs generated in this study. Subsection views are for the Group D (in purple) and Group E (in fushia). Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). OTUs at 95% similarity at the amino-acid level.

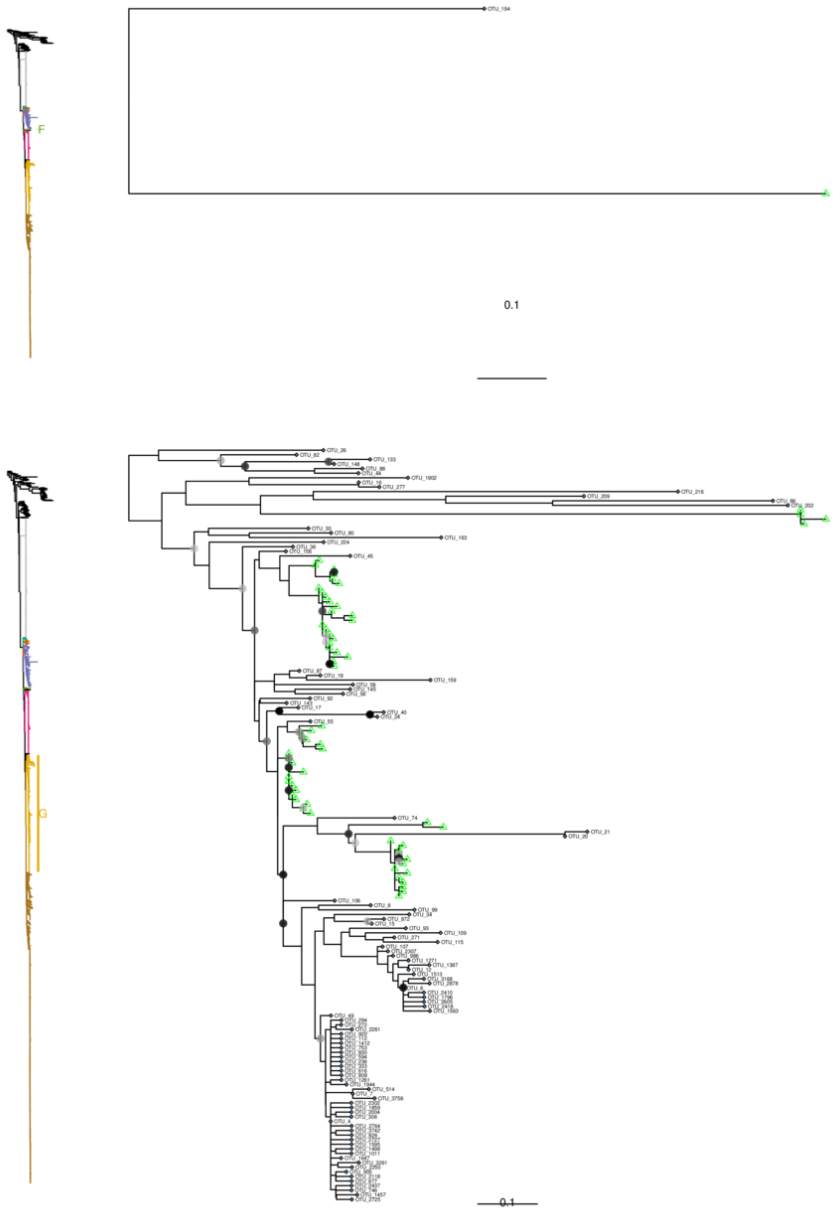

FIGURE B.5: Maximum likelihood phylogenetic tree (RAxML) of subsection F and G of RdRp including reference sequences and OTUs generated in this study. Subsection views are for the Group F (in green) and Group G (in yellow). Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). OTUs at 95% similarity at the amino-acid level.

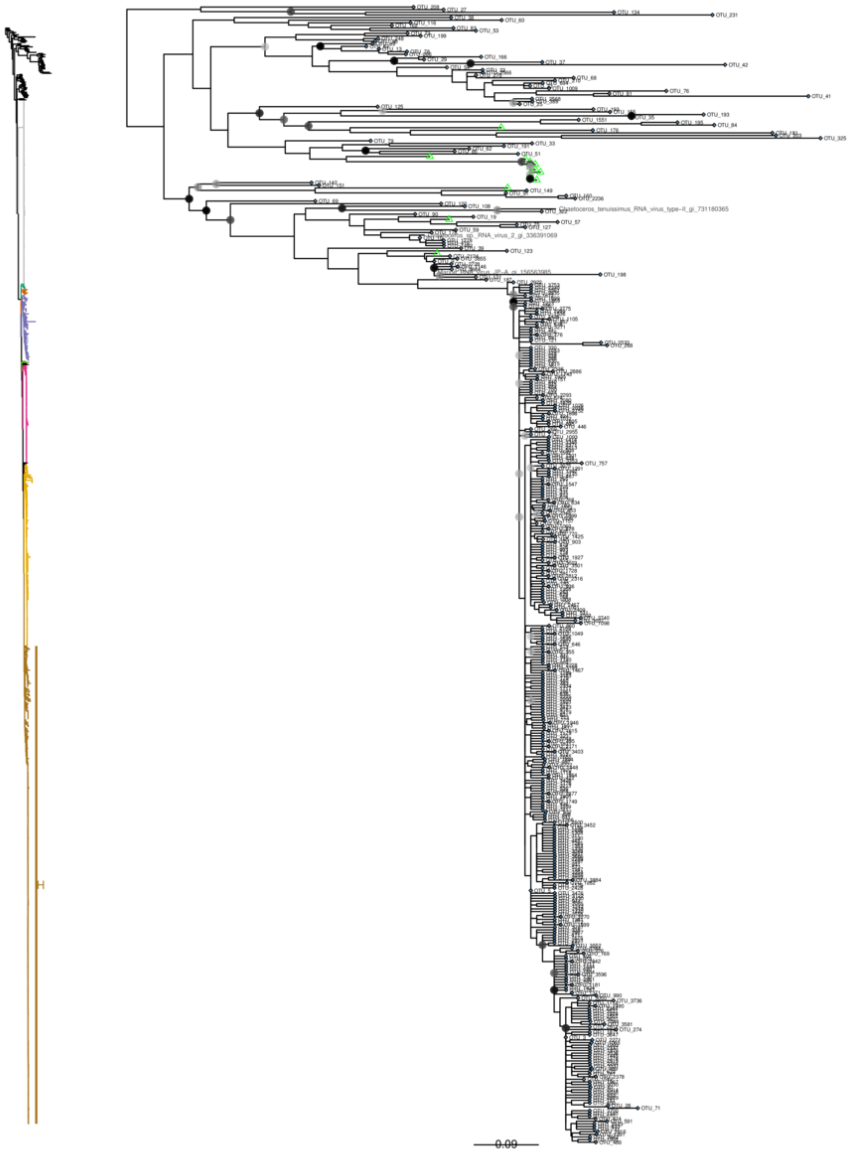

FIGURE B.6: Maximum likelihood phylogenetic tree (RAxML) of subsection H of RdRp including reference sequences and OTUs generated in this study. Subsection views are for the Group H (in brown). Outgroup is virus Equine rhinitis B virus (*Picornaviridae*). OTUs at 95% similarity at the amino-acid level.

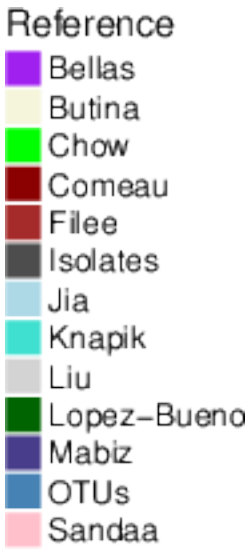

FIGURE B.7: Legend for tip colours for the Maximum likelihood tree of T4-like myovirises.

FIGURE B.8: Maximum likelihood phylogenetic tree (RAxML) of subsection A of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for the reference isolates (in black) and Group A (in grey). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

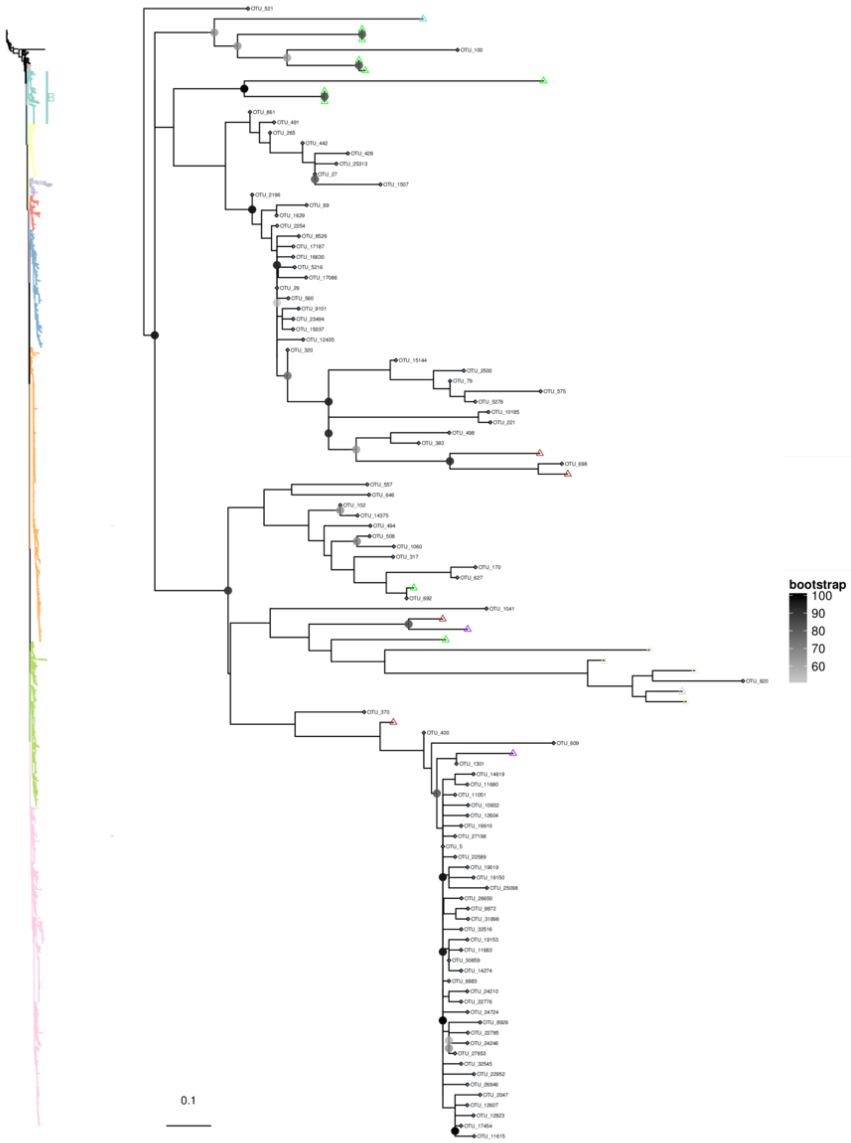

FIGURE B.9: Maximum likelihood phylogenetic tree (RAxML) of subsection B of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for Group B (in turquoise). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

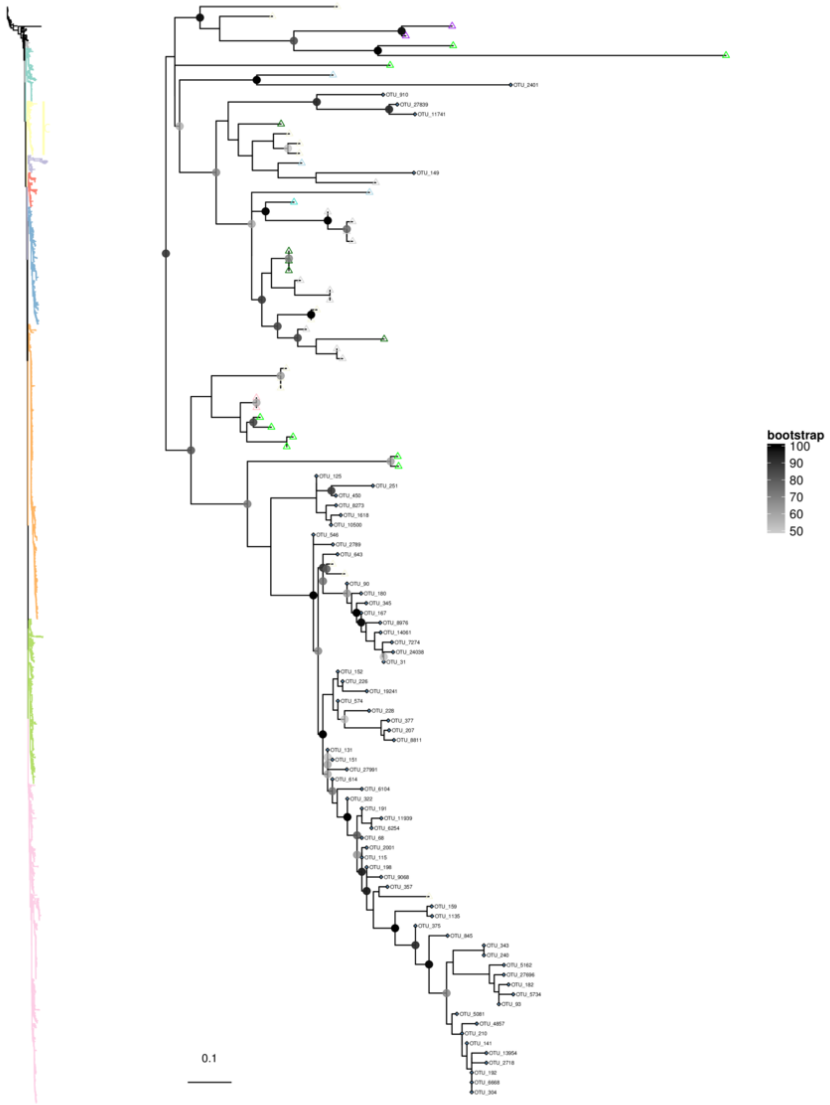

FIGURE B.10: Maximum likelihood phylogenetic tree (RAxML) of subsection C of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for Group C (in yellow). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

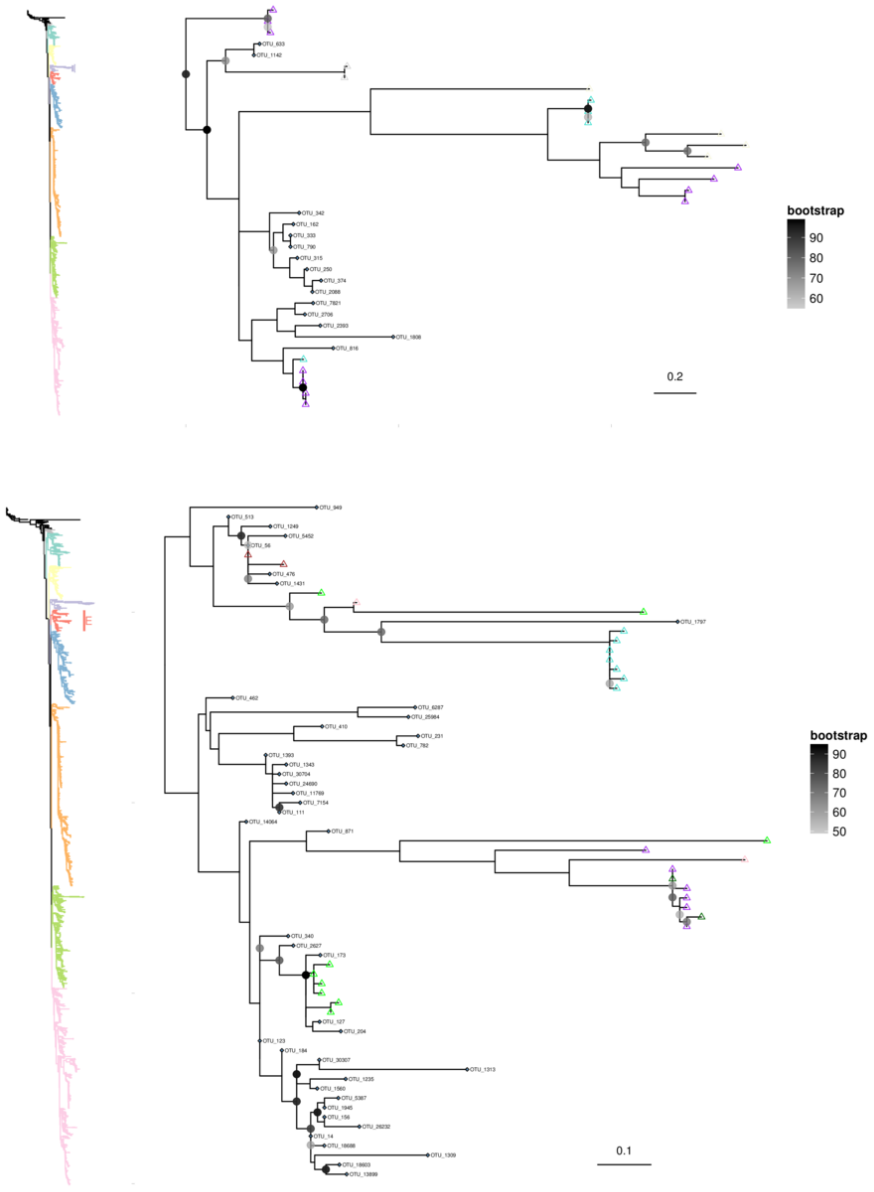

FIGURE B.11: Maximum likelihood phylogenetic tree (RAxML) of subsection D and E of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for Group D (in purple) and Group E (in red). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

FIGURE B.12: Maximum likelihood phylogenetic tree (RAxML) of subsection F of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for Group F (in blue). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

FIGURE B.13: Maximum likelihood phylogenetic tree (RAxML) of subsection G (top) of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for the top portion of Group G (in orange). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

FIGURE B.14: Maximum likelihood phylogenetic tree (RAxML) of subsection G (bottom) of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for the bottom portion of Group G (see previous sub tree for the top portion). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

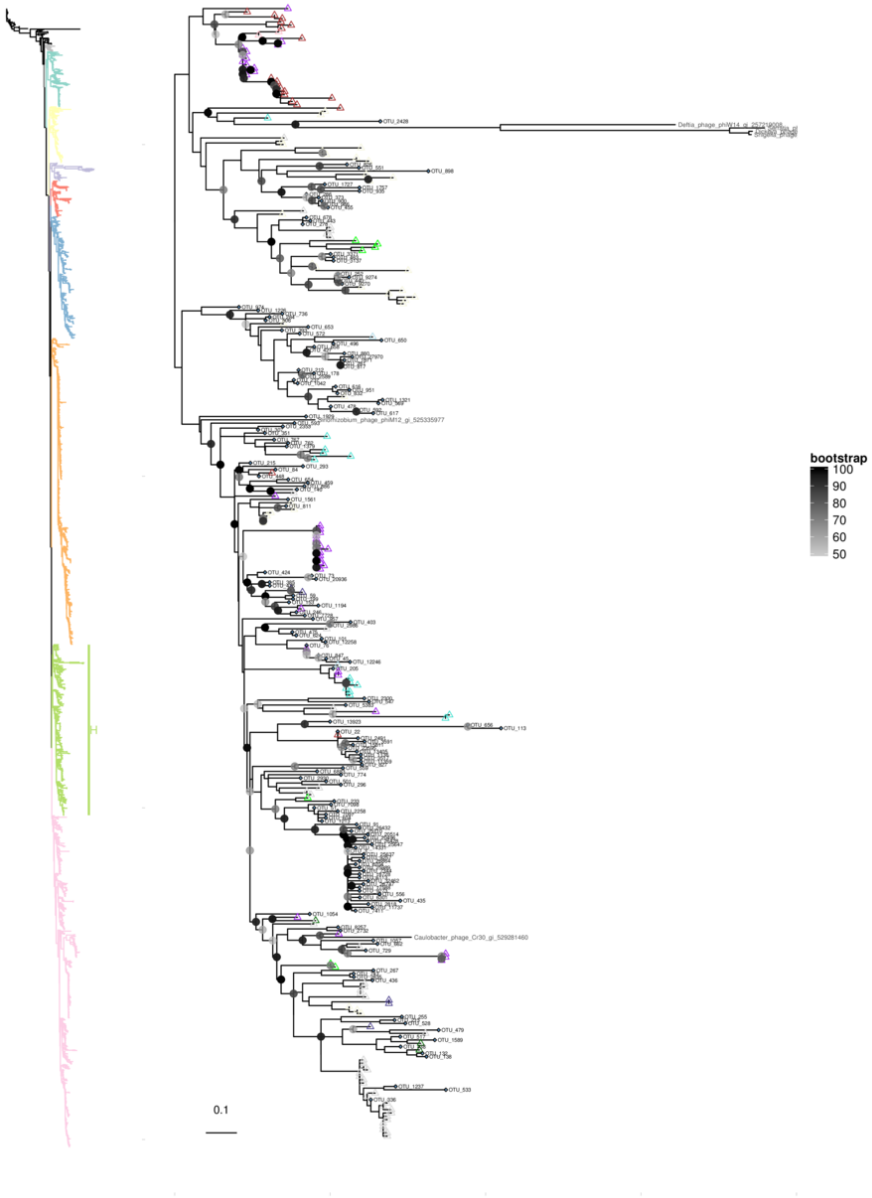

FIGURE B.15: Maximum likelihood phylogenetic tree (RAxML) of subsection H of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for Group H (in green). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

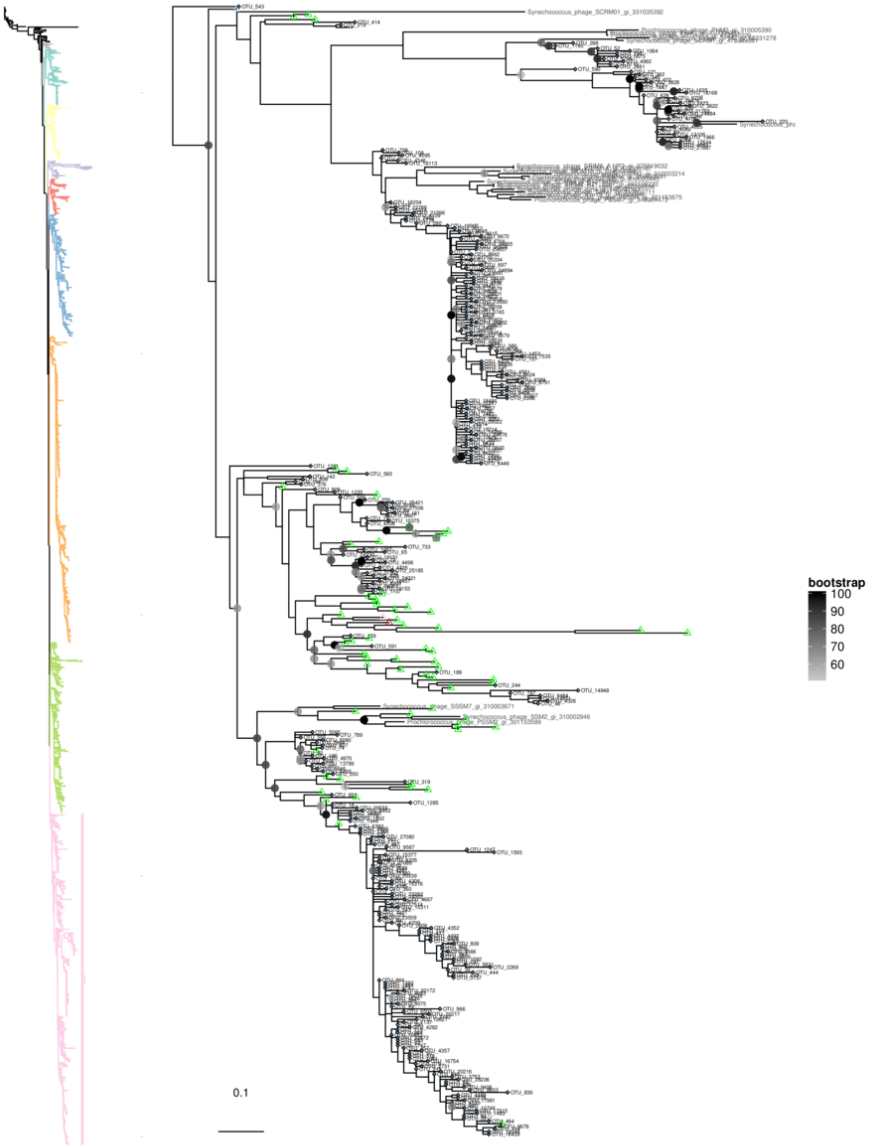

FIGURE B.16: Maximum likelihood phylogenetic tree (RAxML) of subsection I (top) of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for the top portion of Group I (in pink). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

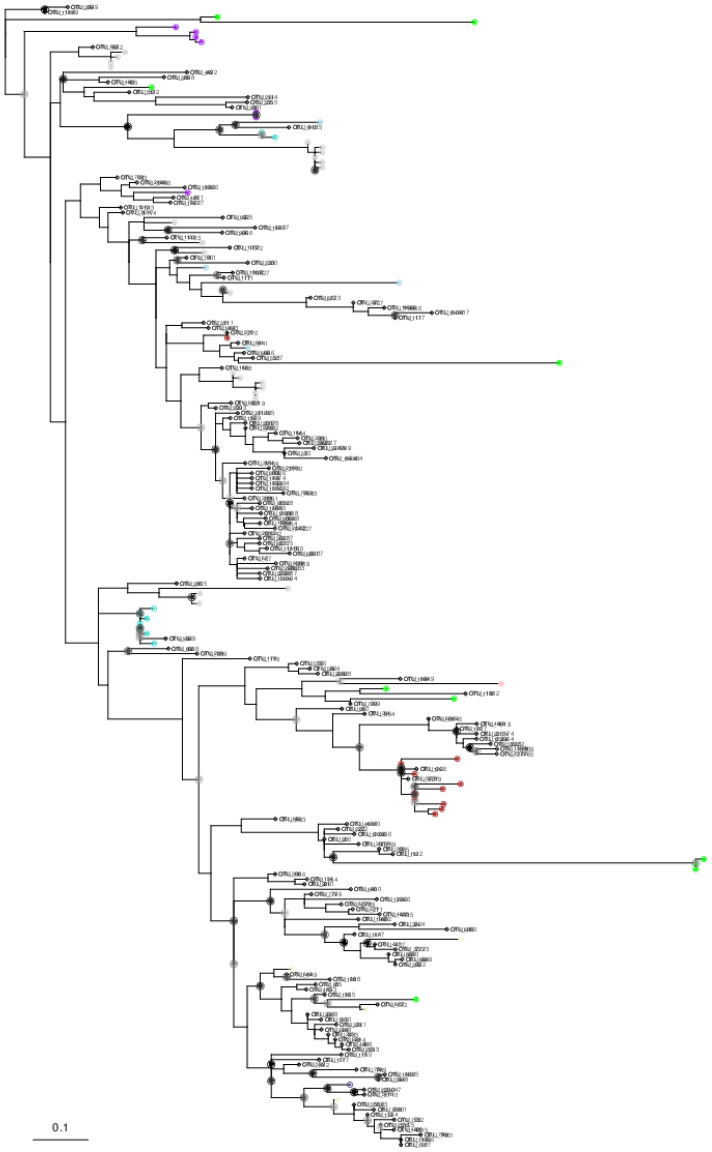

FIGURE B.17: Maximum likelihood phylogenetic tree (RAxML) of subsection I (bottom) of gp23 (marker for T4-like myoviruses) including reference sequences and OTUs generated in this study. Subsection views are for the bottom portion of Group I (see previous sub tree for the top portion). Outgroup is Enterobacteria phage T4. OTUs at 95% similarity at the amino acid level.

FIGURE B.18: Heatmap of relative abundance of eukaryotic OTUs (97% similarity) over time. Each column is a time point

FIGURE B.19: Heatmap of relative abundance of bacterial OTUs (97% similarity) over time. Each column is a time point

FIGURE B.20: Heatmap of relative abundance of T4-like myoviral OTUs (95% similarity amino acid) over time ordered by phylogenetic tree tree (Figure B.3.7). Each column is a time point

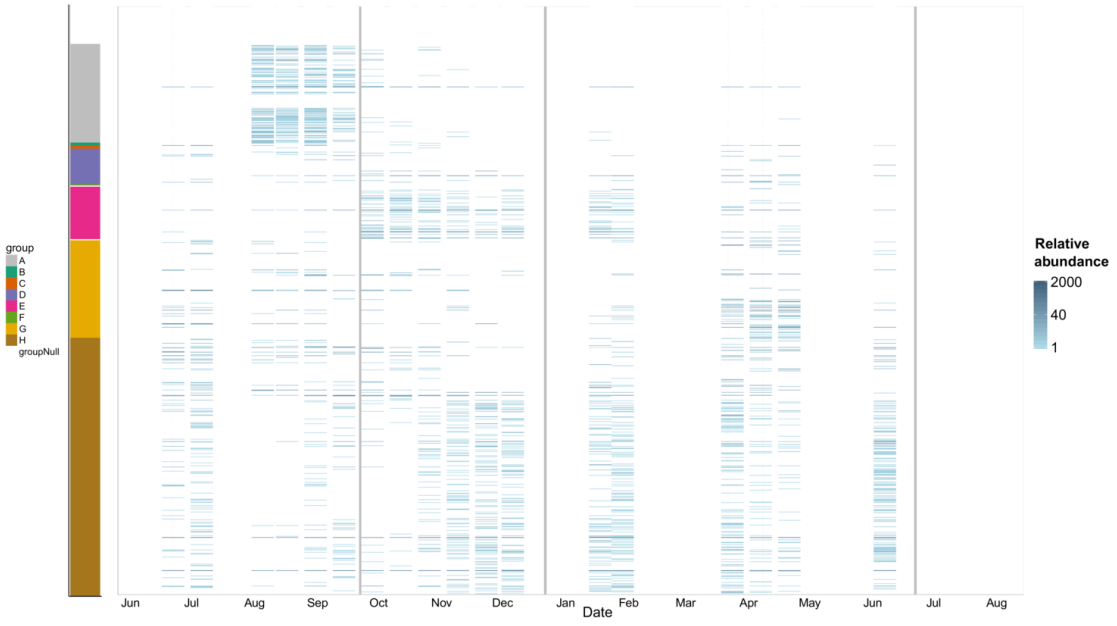

FIGURE B.21: Heatmap of relative abundance of marine picorna-like OTUs (95% similarity amino acid) over time ordered by phylogenetic tree tree (Figure B.3.6). Each column is a time point

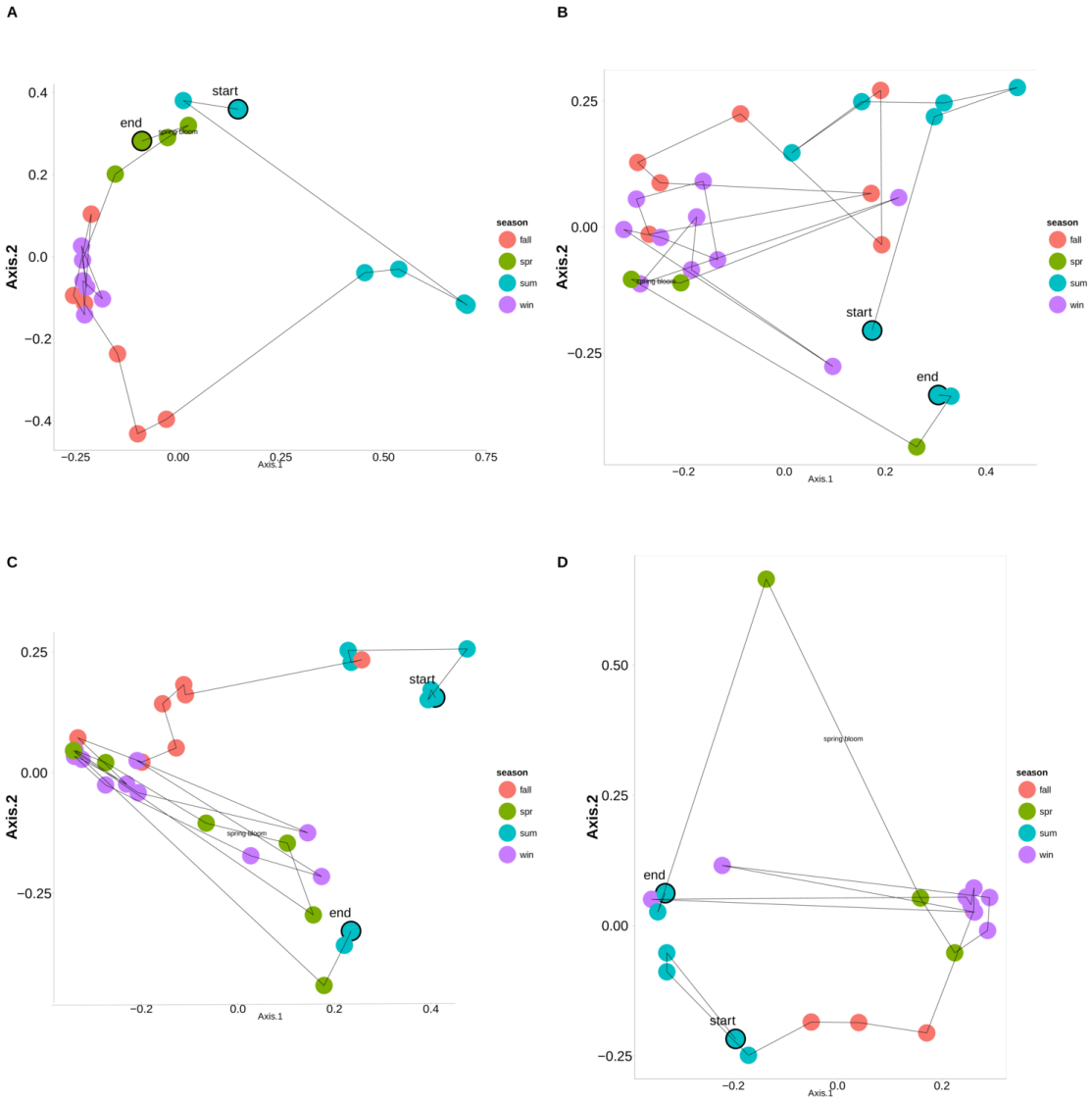

FIGURE B.22: Non-metric dimensional scaling (NMDS) plot of the microbial communities coloured by season. A) NMDS of eukaryotic communities using Hellinger's distance. Seasons are defined according to text. B) Bacterial(16S) communitiy NMDS. C) Marine picorna-like viral communitiy NMDS. D) T4-like myoviral community NMDS. Colours by season. Lines connect sequential sampling times.

APPENDIX C

Supplementary Information to Chapter 4

C.1 SUPPLEMENTARY FIGURES

C.1.1 *Network simulations over time*

Simulations were performed with the same number of nodes and edges as from the observed networks.

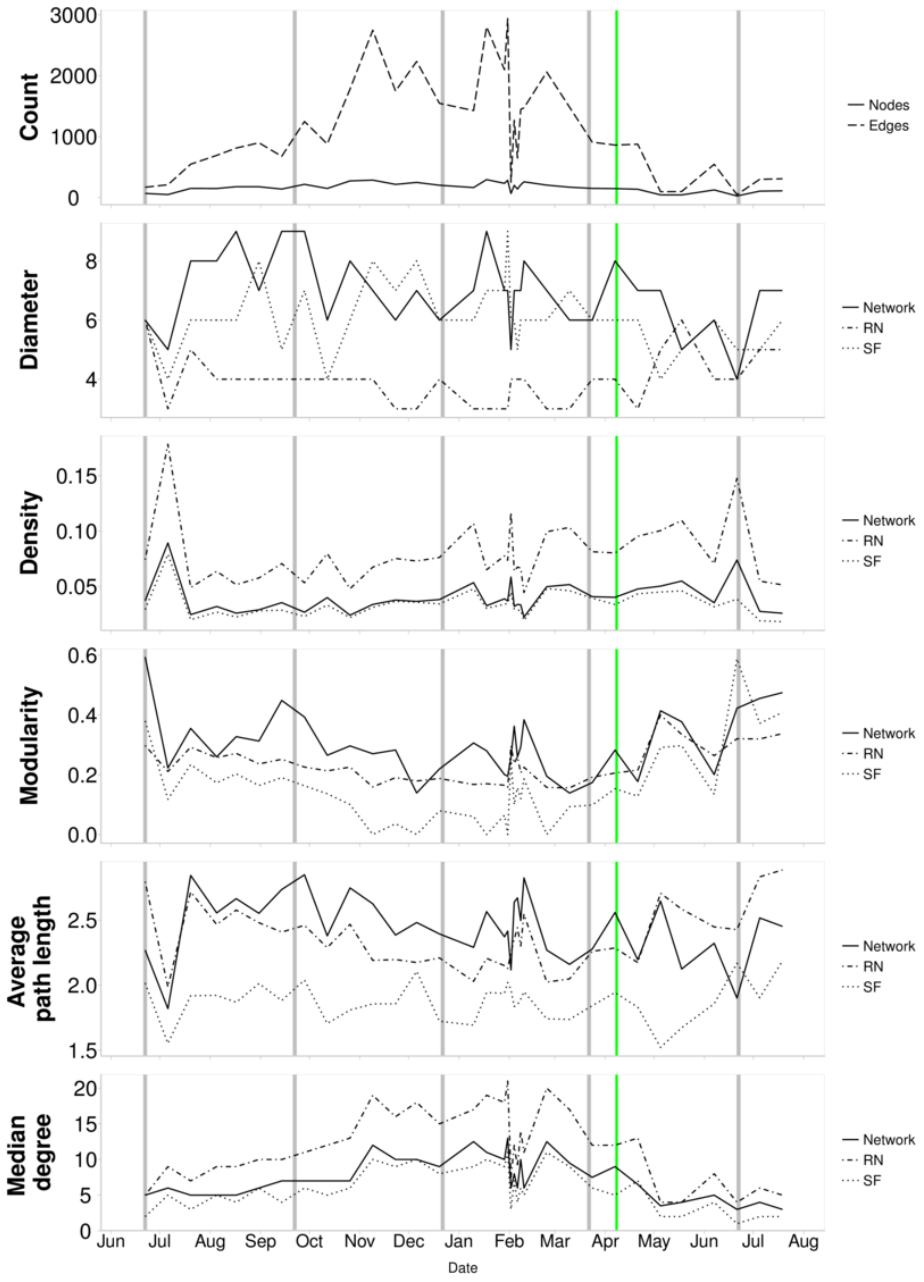

FIGURE C.1: Overall network (includes OTUs from eukaryotic, bacterial, T4-like myoviruses and environmental parameters) compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.

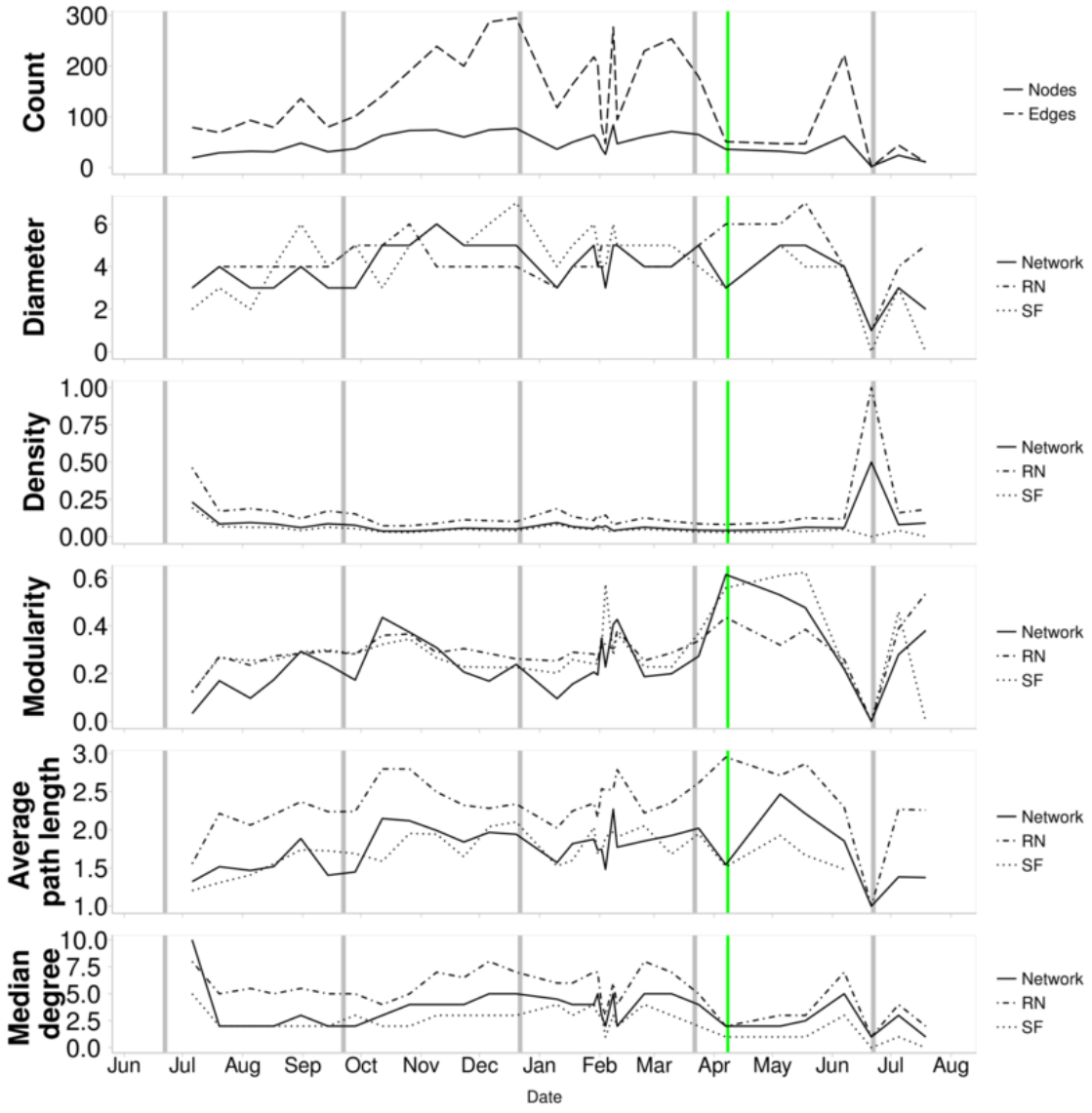

FIGURE C.2: Network of eukaryotic OTUs compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.

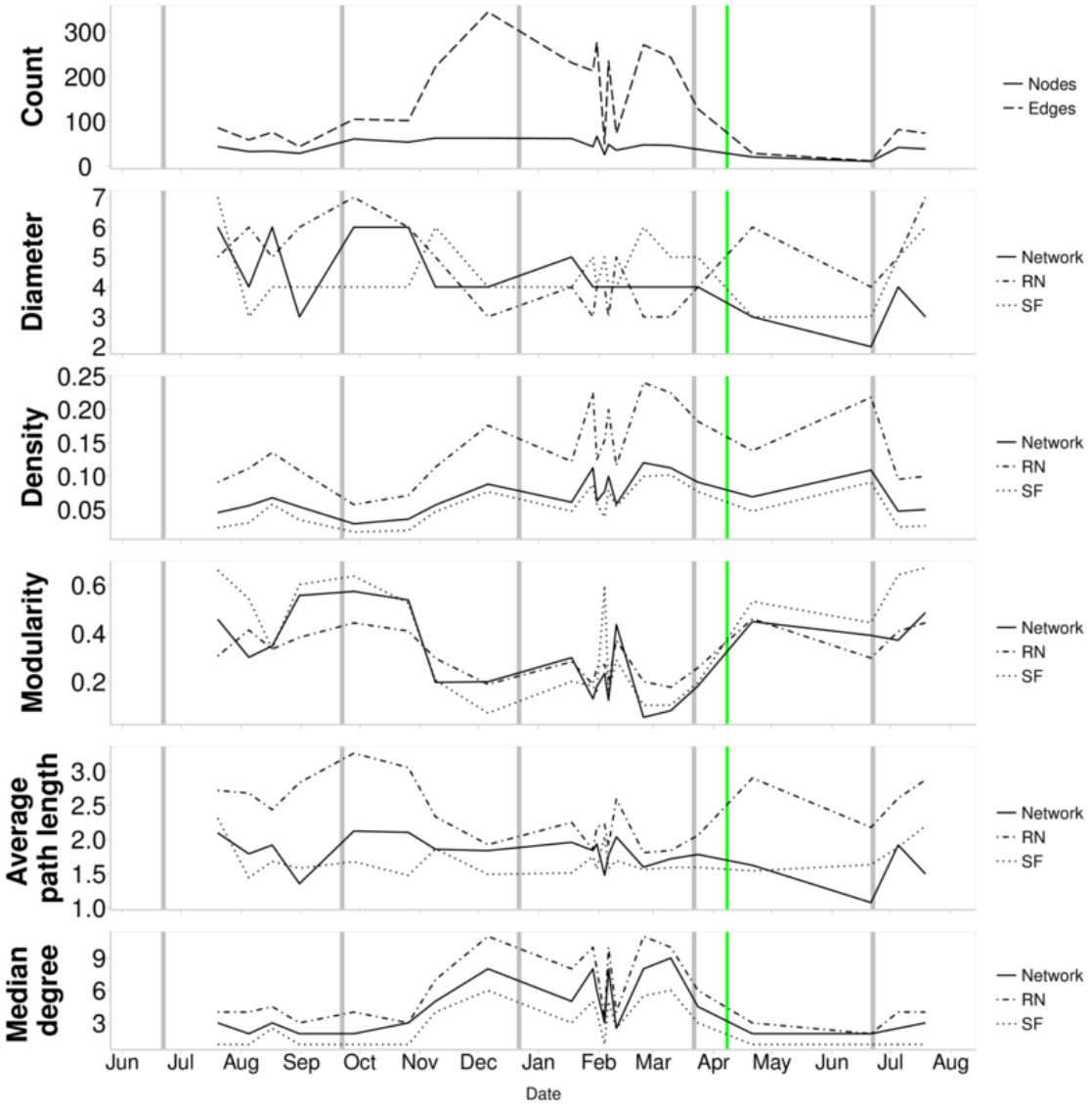

FIGURE C.3: Network of bacterial OTUs compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.

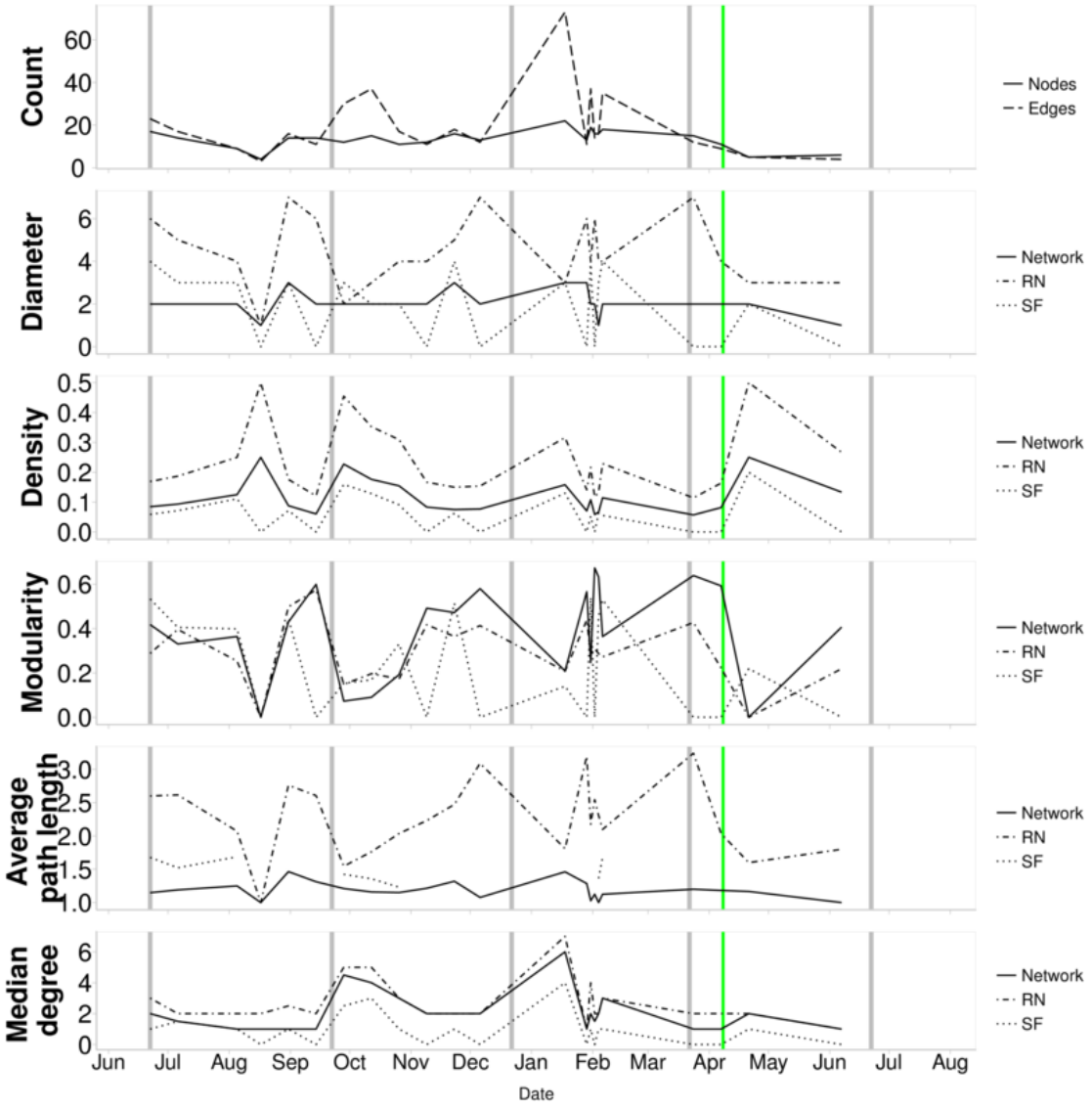

FIGURE C.4: Network of marine picorna-like viral OTUs compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.

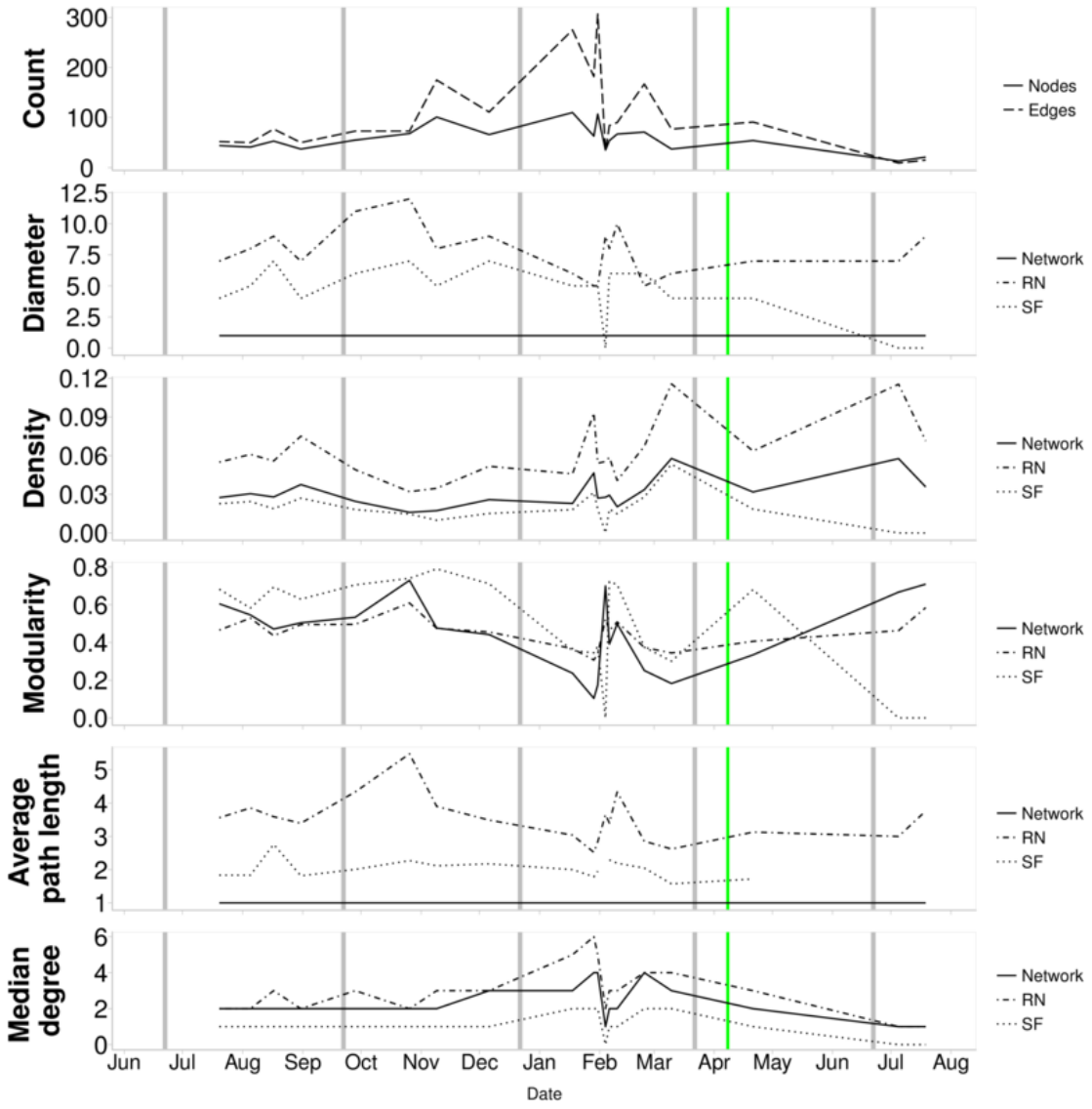

FIGURE C.5: Network of bacterial and T4-like myoviral OTUs compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.

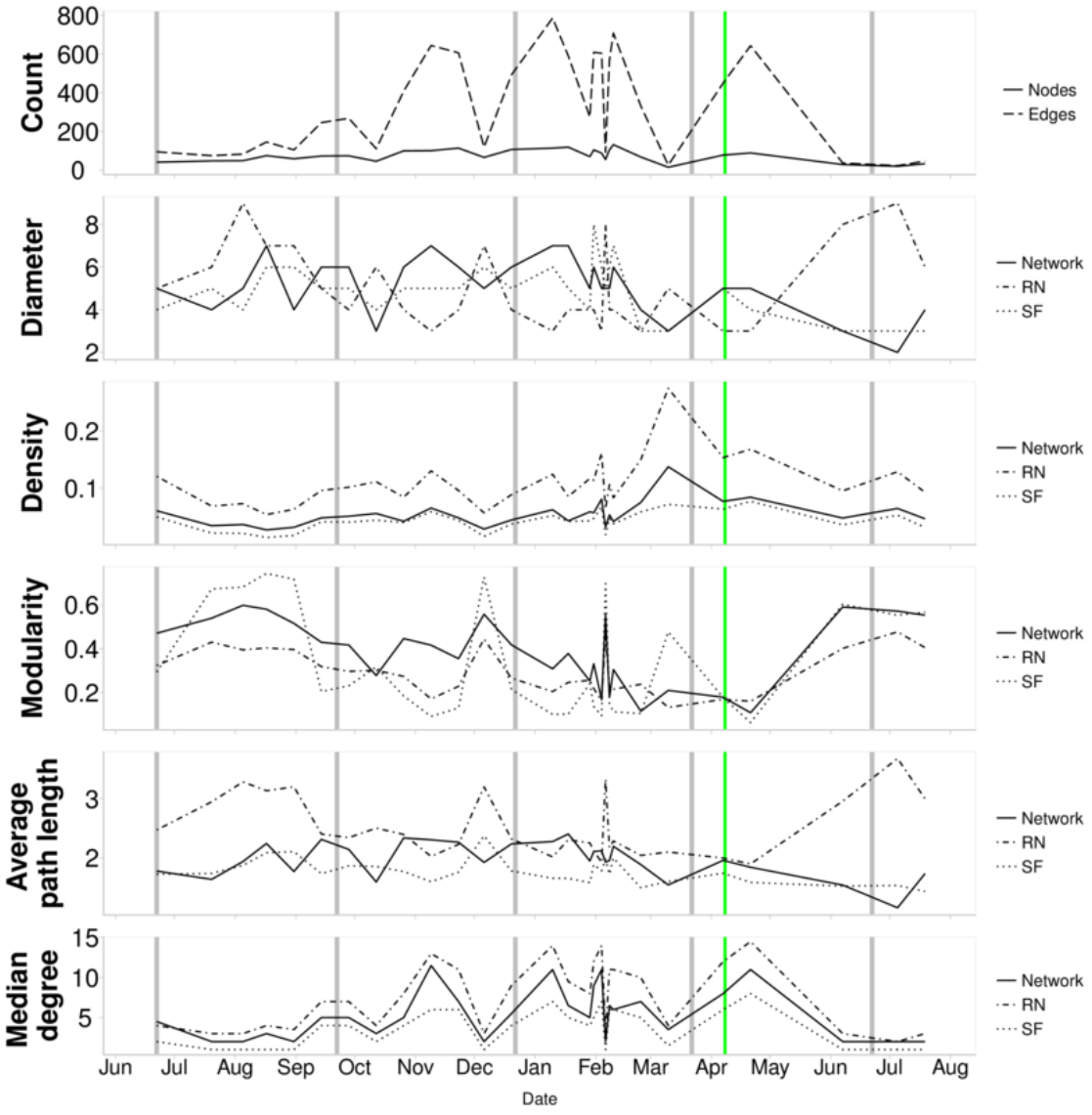

FIGURE C.6: Network of T4-like myoviral OTUs compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.

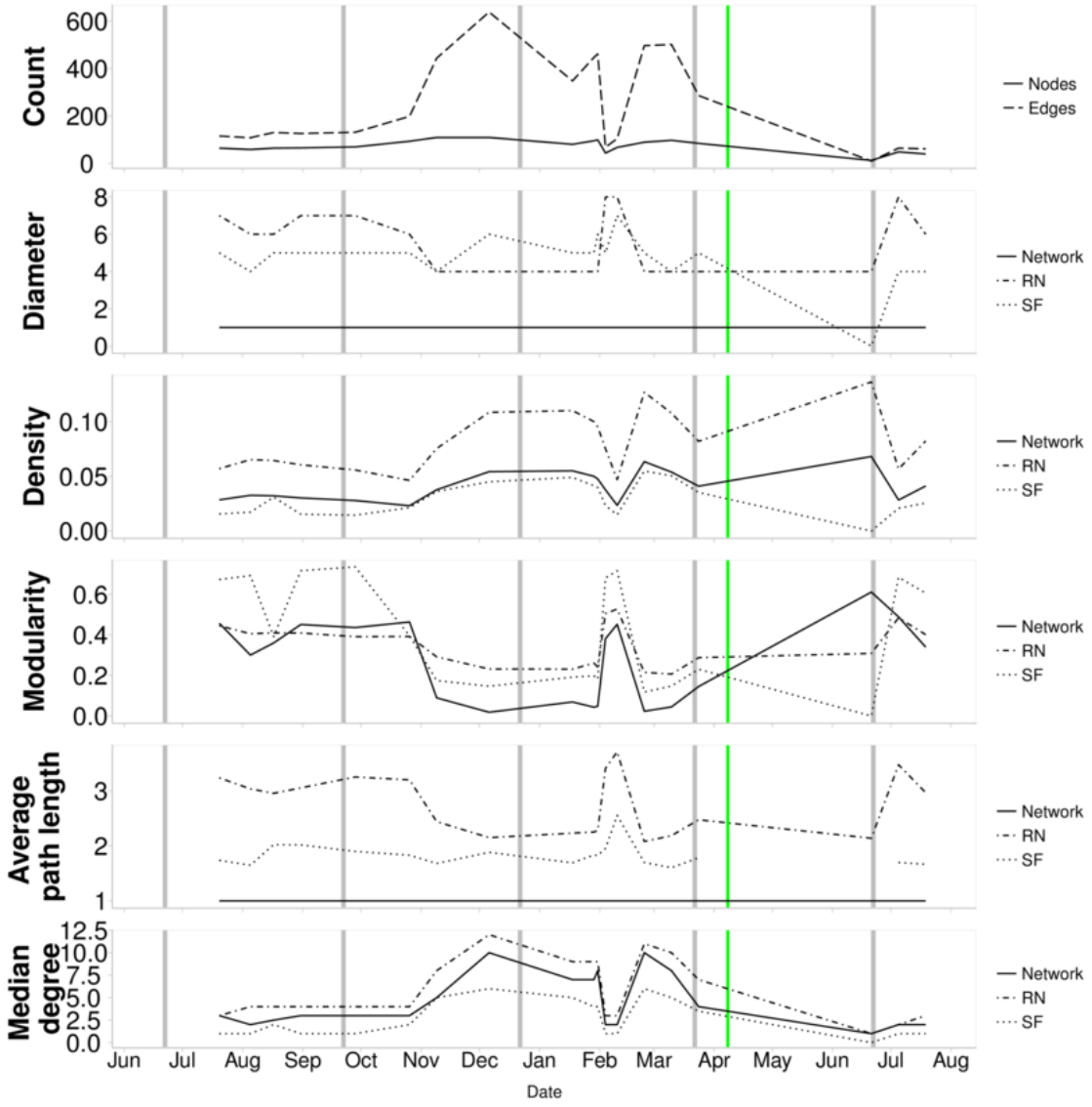

FIGURE C.7: Network of eukaryotic and bacterial OTUs compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.

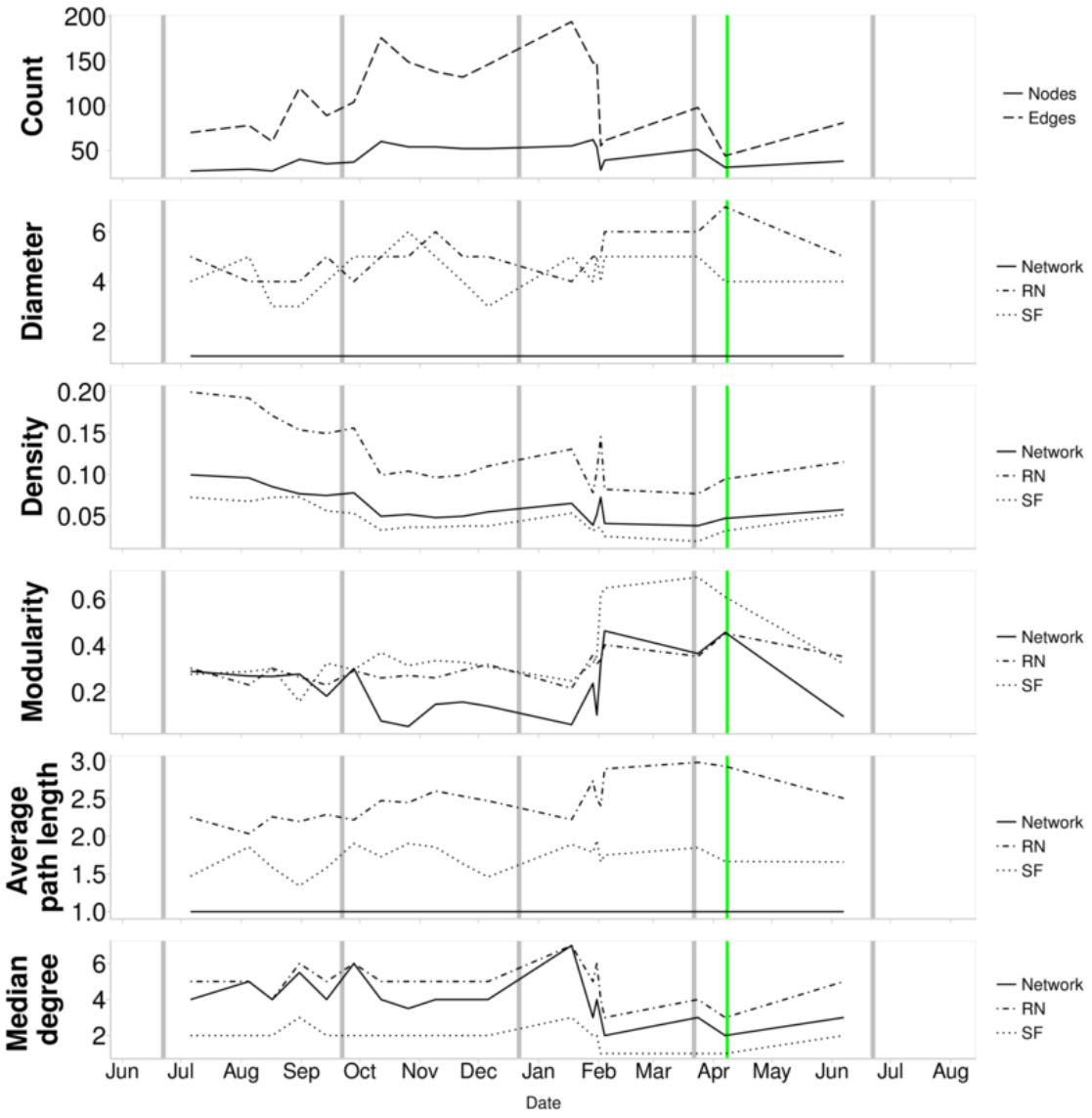

FIGURE C.8: Network of eukaryotic and marine picorna-like viral OTUs compared to simulated networks created from same number of nodes and edges. Random Network (RN) was generated according to the Erdos-Renyi model and the Scale-Free (SF) were generated according to the Barabasi-Albert model. The green vertical line corresponds to the annual spring bloom and grey lines correspond to divisions between seasons.